ASPECTS OF SYMPLECTIC DYNAMICS AND TOPOLOGY: GAUGE ANOMALIES, CHIRAL KINETIC THEORY AND TRANSFER MATRICES

BY

VATSAL DWIVEDI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Associate Professor Shinsei Ryu, Chair
Professor Michael Stone, Director of Research
Professor James N. Eckstein
Professor Jared Bronski

Abstract

This thesis presents some work on two quite disparate kinds of dynamical systems described by Hamiltonian dynamics.

The first part describes a computation of gauge anomalies and their macroscopic effects in a semiclassical picture. The geometric (symplectic) formulation of classical mechanics is used to describe the dynamics of Weyl fermions in even spacetime dimensions, the only quantum input to the symplectic form being the Berry curvature that encodes the spin-momentum locking. The (semi-)classical equations of motion are used in a kinetic theory setup to compute the gauge and singlet currents, whose conservation laws reproduce the nonabelian gauge and singlet anomalies. Anomalous contributions to the hydrodynamic currents for a gas of Weyl fermions at a finite temperature and chemical potential are also calculated, and are in agreement with similar results in literature which were obtained using thermodynamic and/or quantum field theoretical arguments.

The second part describes a generalized transfer matrix formalism for noninteracting tight-binding models. The formalism is used to study the bulk and edge spectra, both of which are encoded in the spectrum of the transfer matrices, for some of the common tight-binding models for noninteracting electronic topological phases of matter. The topological invariants associated with the boundary states are interpreted as winding numbers for windings around noncontractible loops on a Riemann sheet constructed using the algebraic structure of the transfer matrices, as well as with a Maslov index on a symplectic group manifold, which is the space of transfer matrices.

Acknowledgments

At this end of the first step of my scientific career, I would like to express my gratitude to all who have facilitated the work that went into this thesis.

First and foremost must be my adviser, Prof Michael Stone, who, once past his initial reluctance in accepting new students, has been an exceptional guide over the past five years. He has always been free and willing to talk and entertain my many, and in retrospect, occasionally silly, questions, as long as I was willing to be up early enough to catch him in his office. For his support, and well as for his willingness to let me pursue my own audacious *adventures*(a privilege that only a few have in grad school, as I have come to realize), I am eternally grateful to him.

Among other people, the *primus inter pares* must be Dr Victor Chua, the colleague, the quintessential collaborator, the co-adventurer and the co-conspirator. I am grateful for all the physics that I have learnt from him, as well as for all the crazy physico-mathematical *rabbit holes* that, should I choose to indulge in, would be enough to keep me intellectually occupied for at least a few years.

I am grateful to all the teachers who have, over the years, taught me physics, and more importantly, taught me how to learn physics. Named must be Prof A Taraphder and Prof S P Khastgir at IIT Kharagpur and Prof Jared Bronski, Prof Nigel Goldenfeld and Prof Rob Leigh at UIUC. I would like to thank Prof Lance Cooper, the associate head of the UIUC physics graduate program, for having answers to all questions to do with *grad school logistics*, as well as for promptly responding to emails, even at the most unusual of hours.

Among the "deities", I must extend my gratitude to Stephen Wolfram for Mathematica and Donald Knuth and Leslie Lamport for LaTeX! I am grateful to the Engineering and Mathematics libraries at UIUC, as well as the I-Share system. Last but not the least are the coffee shops of Urbana, for all the caffeine-fueled hours in the past five years, and especially the past two months, that have made this work possible.

For part of my graduate school, I was supported by the National Science Foundation through Grant NSF DMR 1306011 as a research assistant under Prof Michael Stone. For part of the work, my collaborator Dr Victor Chua was also supported by the Gordon and Betty Moore Foundation's EPiQS Initiative through Grant GBMF4305.

Table of Contents

Appendix

Notation

Spaces: The n dimensional real and complex Euclidean spaces are denoted by $\mathbb{R}^n$ and $\mathbb{C}^n$. The tangent and cotangent spaces of a manifold M are denoted by TM and T^*M, respectively. The set of vector fields and 1-forms on M are denoted by $\mathrm{Vect}(M)$ and $\Omega^1(M)$, respectively. We denote the Hodge dual of a differential form X as either $\star X$ or $\bar{X}$.

The standard basis of $\mathbb{R}^n$ or $\mathbb{C}^n$ is denoted by with $\mathbf{e}_i, i = 1, \dots n$, where $(\mathbf{e}_i)_j = \delta_{ij}$. Given a set of oriented orthonormal axes x^i on a n-dimensional manifold M, we have defined $d^m x \equiv dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n$.

Indices: The Greek indices (μ, ν) run over all the spacetime coordinates and the Latin indices from the middle of the alphabet (i, j, k) run over only the space coordinates. The Latin indices from the beginning of the alphabet are used for Lie algebras. Einstein summation for repeated indices is always assumed unless stated otherwise.

Brackets: The angled brackets $\langle\, , \rangle$ have been used to denote inner products. The braces $\{\, , \}$ have been used for Poisson brackets. The commutators are denoted by $[\, , \,]_-$ or simply $[\, , \,]$, while the anticommutators are denoted by $[\, , \,]_+$.

Matrices: The space of all $n \times n$ real and complex matrices is denoted by $\mathrm{Mat}(n, \mathbb{R})$ and $\mathrm{Mat}(n, \mathbb{C})$, respectively, while $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{GL}(n, \mathbb{C})$ denote the corresponding subspaces of nonsingular matrices. The Pauli matrices are defined as

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Physics: The Minkowski metric is defined as $\eta^{\mu\nu} = \mathrm{diag}\{-1, 1, \dots 1\}$ on $\mathbb{R}^{2N+1,1}$, following the general relativity convention. We set $\hbar = c = e = k_B = 1$.

Misc: We use the notation $\{X_i\}_{i=1}^n$ to denote sets of the form $\{X_1, X_2, \dots X_n\}$.

1 Introduction

This thesis consists of two largely disjoint parts. The first part is based on a series of papers[1, 2, 3, 4] with Prof Michael Stone, on the connection between semiclassical dynamics and anomalies in quantum field theories(QFTs), derived using the *chiral kinetic theory*. The second part is based on a long paper[5] in collaboration with Dr Victor Chua, on a general construction of transfer matrices for tight binding models in condensed matter physics. In the rest of this chapter, I introduce the two sections separately and conclude with a briefly discussion of certain similar themes in the underlying mathematical structures.

1.1 Part I: Anomalies and Semiclassics

In quantum field theory, an *anomaly*[6] is a breakdown of a classical symmetry when the theory is quantized, manifesting in the breakdown of the conservation law for the corresponding Noether current for continuous symmetries. Since the discovery of the U(1) anomaly by Adler[7], and independently Bell and Jackiw[8] to explain the anomalous pion decay, they have been an integral part of the quantum field theory of chiral fermions coupled to gauge degrees of freedom. At first sight, the anomalies seem to be quite undesirable! Indeed, if a classical gauge symmetry becomes anomalous on quantization, the theory is rendered inconsistent, since the gauge symmetry is simply a redundancy in description. This can sometimes be mended by a precise anomaly cancellation, as is the case with the standard model of particle physics[9].

However, an anomaly in a global symmetry can be a powerful theoretical probe into the behavior of the theory. One primary reason is their nonperturbative nature; anomalies computed at one loop level do not get renormalized at higher orders. Furthermore, they are related to certain topological data associated with the theory, e.g, to the index of the Dirac operator, via *index theorems*[6]. The topological nature of anomalies implies that they must appear, in various disguises, in various effective descriptions of the theory at different energy/length scales. The conventional examples are the Chern-Simons and the Wess-Zumino-Witten theories, which have found various applications in condensed matter physics lately.

A particular low energy description of QFTs is relativistic hydrodynamics[10, 11], which provides an ideal setup to study the finite temperature/chemical potential behavior of the theory. A hydrodynamic

description of a QFT can often be constructed systematically from the symmetries of the QFT, which has proved useful in the study of *gauge-gravity duality*[12]. The manifestation of anomalies in hydrodynamics as anomalous contributions to various transport coefficients have been investigated over the past decade[13, 14, 15]. A particularly striking result, conjectured in Ref [16] and further studied in Refs [17, 18, 19, 20, 21], relates the anomalous contributions in the hydrodynamic currents to the mixed anomaly polynomials in 2 higher dimensions by a simple substitution of the gauge curvatures in the anomaly polynomials with the thermodynamic variables, termed *replacement rules*.

Another low energy description for general quantum theory is a semiclassical description, which is essentially classical Hamiltonian dynamics with additional "features" added to include some quantum effects. A particular example is wavepackets[22], which can be treated as classical (point) particles with an additional Berry curvature term. Such descriptions have been used since the beginnings of quantum mechanics to study transport in metals and semiconductors[23, 24, 25].

It is somewhat surprising that anomalies, whose computation generally requires sophisticated quantum mechanical computations, can be derived in a classical setup, with the only quantum inputs being the phase space volume, as is conventionally done in classical kinetic theory, and the Berry curvature. The first step in this direction was by Stephanov and Yin[26], who derived a semiclassical action for positive energy, positive chirality Weyl fermions and showed that a computation of the gradient of the particle current using a *(chiral) kinetic theory* reproduces the expression for the U(1) (Adler-Bell-Jackiw) anomaly. The intuitive picture of their argument hinges on the fact that the U(1) anomaly manifests itself as a breakdown of the conservation of particle number for chiral (Weyl) fermions, and near the Fermi surface, and well away from the Dirac point, a semiclassical effect is sufficiently accurate that the influx of extra particles can be counted reliably in a kinetic theory setup.

In Refs [1, 2], we generalized the semiclassical calculation to compute nonabelian gauge and singlet anomalies in arbitrary even spacetime dimensions. The central mathematical tool for this generalization was the geometric formulation of classical mechanics[27, 28], and the anomaly manifests itself as a symplectic form that fails to be closed at the diabolical (Weyl) point, thereby violating Liouville's theorem. Thus, the anomalies are, in fact, encoded in the phase space structure.

An increasing application of quantum field theory techniques to study condensed matter systems has made the study of macroscopic effects of anomalies particularly relevant. In particular, Weyl semimetals[29, 30, 31] provide a condensed matter realization of Weyl fermions, and hence the Adler-Bell-Jackiw anomaly[32]. The effect of anomaly on electronic transport[33] has been widely studied in recent years using various techniques, including the kinetic theory approach[34] discussed above. Transport phenomenon associated with gauge

anomalies have also been studied in quark-gluon plasma formed in heavy-ion collisions. In this case, the anomaly manifests as the *chiral magnetic effect*(CME)[35, 36] and *chiral vortical effect*(CVE)[37], and the computation of relevant coefficients have been approached from many direction, including hydrodynamics[13, 14], kinetic theory[26] and holography[38].

On the theoretical side, the chiral kinetic theory provides an interesting perspective on gauge anomaly, and it would be interesting to generalize it to curved spacetime backgrounds so as to derive the gravitational contribution to gauge anomaly. However, this approach, being based on Hamiltonian dynamics, treats space and time differently, and is thus not manifestly Lorentz invariant. The Lorentz symmetry is implemented on the position and momentum coordinates in an unusual representation of the Lorentz group[3, 39]. A manifestly Lorentz invariant formalism would be the first step in studying chiral kinetic theory on curved spacetime background, required to compute the most general mixed anomaly polynomials for Weyl fermions. Our first steps in this direction are discussed in Ref [3].

1.2 Part II: Transfer matrices

The topological phases of matter[40, 41, 42] have been a subject of considerable interest in recent years. In the most general (and vague) sense, they are *ordered* phases of matter, whose order cannot be characterized by a local order parameter *á la* Landau theory of phase transitions, or, equivalently, cannot be understood as a "symmetry breaking". Instead, they are characterized by a global invariant, which is "topological" in the sense that it is invariant under "continuous deformations" of the system.

We unpack this vague terminology for a quite *plebeian* system: one consisting of noninteracting fermions on a lattice, described by a Bloch Hamiltonian over the Brillouin zone (i.e, the reciprocal lattice). The eigenstates of the Bloch Hamiltonian corresponding to a given energy eigenvalue then form a vector bundle over the Brillouin zone. We deem two states, i.e, two Bloch Hamiltonians, to be "topologically equivalent" if they can be deformed into each other without closing the gap, i.e, without changing the topology of the corresponding vector bundles. The topological invariant is the Chern number of the vector bundle[43, 44].

Since tuning such a *topological insulator* (nonzero Chern number) to a *trivial insulator* must involve closing the bulk gap, it follows that whenever the topological insulator is placed next to vacuum (a trivial insulator) or another topological insulator with a different Chern number, the bulk gap must close at the interface, leading to gapless *edge states*. These edge states also carry topological information: for instance, the (signed) intersections of the edge spectrum with a given energy level is a topological invariant determined by the bulk, and cannot be changed by local perturbations at the edge.

The bulk invariants of the system can be computed, in principle, by using the momentum state representation of the system. However, in order to investigate the edge behavior of the system, one needs to open an edge along at least one space dimension, thereby breaking the lattice translation invariance in that direction and rendering the momentum space picture inapplicable. The edge spectrum is then usually calculated either using exact diagonalization, which can be computationally expensive for large systems (esp for 3d systems), or using a decaying ansatz[45, 46], which usually needs additional input regarding the existence of edge modes, for instance, from symmetry considerations. However, another way of studying these systems is using transfer matrices. The central idea is that given a quasi-1D system (obtained by Fourier transforming a given system in d space dimensions along the $d-1$ periodic directions), once we obtain the transfer matrix for translation by a period, we can obtain both bulk and edge characteristics by looking at the eigenvalues of the transfer matrix.

Transfer matrices have been studied in diverse contexts. They have been used to study electronic band structure[47, 48, 49, 50], conductivity[51], Majorana fermions[52] and wave motion in electromechanical systems[53]. They have also been studied by mathematicians under the banner of Floquet theory[54, 55, 56], known to physicists as Bloch theory. The monodromy matrix[54, 57] in Floquet theory, which translates the solution by one period, is the direct analogue of the transfer matrix in condensed matter systems. However, often there is additional structure associated with the monodromy matrix. For instance, if the system is Hamiltonian, the transfer matrix turns out to be symplectic[58], in which case the eigenvalues of the transfer matrix can be computed and hence its eigenstates classified as growing/decaying or propagating, using a set of quantities called Floquet discriminants[59], which are derived directly from the traces of powers of the transfer matrix.

The properties of a transfer matrix associated with a topological state was studied, to interesting results, by Y Hatsugai[47, 60]. He computed the transfer matrix for the Hofstadter model with flux $\phi = p/q$ per plaquette on a cylinder, finite along x and periodic along y, so that k_y is a good quantum number, and showed that one can obtain the bulk bands as well as the edge states using the transfer matrix. Furthermore, the eigenvalues of his transfer matrix are given by

$$\rho_{\pm} = \frac{1}{2}\left[\Delta(\varepsilon, k_y) \pm \sqrt{\Delta^2(\varepsilon, k_y) - 4}\right], \tag{1.1}$$

where $\Delta(\varepsilon, k_y)$, the trace of the transfer matrix, is a polynomial in ε. Given the square root, it is natural to allow ε to be a variable on a two-sheeted Riemann surface, glued along branch cuts corresponding to the bulk bands. This ε-Riemann surface is a two-dimensional complex manifold with genus $g = q-1$, where q is

the number of bulk bands. Hatsugai shows that in such a picture, the topological nature of the edge states manifests as windings around the holes of the ε-Riemann surface, the winding number being equal to the Chern number of the filled bulk bands.

This equality is an instance of a more general issue of fundamental importance, *viz*, the connection (usually equality) between the bulk topological invariant and the number of edge modes in a gap, usually referred to as the *bulk-boundary correspondence*. Various proofs of this correspondence have been discussed in literature for different classes of topological phases. In particular, using the methods of non-commutative geometry, the equality of the Chern number and the Hall conductivity at finite disorder have been rigorously addressed within the mathematical physics literature[61, 62], and with generalizations to time-reversal invariant topological insulators[63]. Complementary to this are approaches based on Green's functions[64, 65], which have also demonstrated the bulk-boundary correspondence from a field theoretic perspective. In addition, aspects of this correspondence have also been discussed from the viewpoint of quantum transport using S-matrices[66, 67]. Other notable citations are Refs [68, 45, 60], however, none of them, to our knowledge, are entirely general.

In this work, we generalize the transfer matrix construction to certain systems where the hopping matrix turns out to be singular, so that the transfer matrices cannot be constructed using the conventional techniques. The generalization also provides us insight into the algebraic and geometric structure of the problem, as well as new ways to characterize the topological invariant(s) associated with the nontrivial boundary states for the topological phases of matter. This approach can potentially lead to a purely algebraic proof of the bulk-boundary correspondence.

1.3 Outlook and Outline

A common thread running through the two seemingly disparate parts of this dissertation is the general idea of *dynamical systems*, and in particular, those described by Hamiltonian dynamics. We have used many results commonly studied in the mathematics literature under that banner, and applied them to physically relevant systems.

A related idea, which makes its appearance in both sections, is that of a symplectic structure. In Part I, the classical Hamiltonian dynamics can be recast as flows on a symplectic manifold, and the Hamiltonian flows are those which generate symplectomorphisms, which can be crudely thought of as elements of an (infinite-dimensional) symplectic group. In the second part, for all of the tight-binding models studied, the generalized transfer matrix turns out to be a symplectic matrix, i.e, an element of a (finite-dimensional)

symplectic group, which leads to various simplifications in computation.

The rest of this thesis is organized as follows:

Part I

In Ch 2, we describe the symplectic formulation of classical mechanics, and a generalization to contact manifolds to for time-dependent systems, and reformulate the conventional kinetic theory in geometric terms. In Ch 3, we discuss the basis of gauge theories and anomalies, and use the anomalous conservation laws and thermodynamic constraints to compute the anomalous contributions to the hydrodynamic currents in arbitrary spacetime dimensions. In Ch 4, we introduce (abelian and nonabelian) Berry phases and their various topological aspects, and compute the Chern numbers for various systems. In Ch 5, we describe the chiral kinetic theory computation of the U(1) anomaly and assemble the ingredients from Ch 2, 3 and 4 to generalize it to Weyl fermions coupled to nonabelian gauge fields in arbitrary even dimensions. We also use the generalization to compute the anomalous contribution to the hydrodynamic currents, and compare with the replacement rules discussed in Ch 3.

Part II

In Ch 6, we start from general noninteracting tight-binding models and construct generalized transfer matrices. We also discuss the computation of bulk/edge spectra using the transfer matrices. In Ch 7, we apply the generalized transfer matrices to a wide variety of commonly studied topological phases, including disordered Hamiltonians, and study windings associated with edge states for certain cases.

We present our general conclusions, as well as outlook, in Ch 8. Formal definitions of certain mathematical objects used in the thesis are collected in Appendix A, while Appendices B and C contain a mixed bag of general results that turned out to be useful for various parts of the thesis, as well as details of certain tedious or not-particularly-informative calculations.

Part I

Anomalies and Semiclassical Dynamics

2 Classical mechanics

Classical mechanics is the vast body of analytical mechanics developed before the 20th century to describe the mechanics of material bodies. In recent times, it has been described more accurately only in contradistinction to the newer physical theories, particularly quantum mechanics and special theory of relativity. Tracing its origins in the work of Galileo and Newton in the 17th century, classical mechanics has acquired many analytical tools and reformulations, the Hamiltonian and the Lagrangian mechanics being the major examples. These formulations, besides simplifying the analysis of complicated systems, also provide useful insights into the structure of classical mechanics.

The latest addition to this repertoire is the geometric formulation of classical mechanics in terms of flows on symplectic manifolds, developed in the 20th century[28, 27]. Since smooth manifolds generically admit coordinate systems only locally, it is more efficient to use Cartan's intrinsic calculus instead of the traditional Newtonian calculus, thereby replacing analytical methods with those of differential geometry and topology. This perspective also lets us generalize the conventional classical mechanics to include arbitrary symplectic manifolds, and thus make use of many ideas that historically emerged from the study of dynamical systems.

In the first part of this dissertation, we shall often use an extension of the symplectic formulation of classical mechanics, usually termed *extended phase space* or *contact structure*, necessary to deal with time-dependent systems. Thus, in this chapter, we describe the symplectic formulation of classical mechanics and a reformulation of kinetic theory in geometric terms. Besides being a primer on these topics, this chapter would set up the notations and basic results used in rest of Part I.

2.1 Newton, Lagrange and Hamilton

The foundation of classical mechanics lies in the Newton's laws of mechanics, esp the second law, which describes the mechanics of a *point particle* as a function of time $t \in \mathbb{R}$ in presence of external "forces"[69]. Mathematically speaking, associated with a point particle in Newtonian mechanics is its mass, $m \in (0, \infty)$, and its position, $\mathbf{x} \in \mathbb{R}^d$, in a fixed coordinate system. In an inertial frame, given an external force $\mathbf{F}(\mathbf{x}, t)$ and some data at a the time origin ($t = 0$), one is interested in computing the trajectory, i.e, a continuous

curve $\mathbf{r}\colon \mathbb{R} \to \mathbb{R}^d$ so that the position of the particle at a given time t is $\mathbf{x} = \mathbf{r}(t)$. Note that conventionally, both the coordinates and the trajectories are notated by the same letter; however, for this section we shall keep that distinction explicit.

The linear momentum of the point particle is defined as $\mathbf{p} = m\dot{\mathbf{r}}$. Newtonian mechanics posits that the state of a point particle is completely determined by the knowledge of its position $\mathbf{x}$ and momentum $\mathbf{p}$, so that one can define a *state space* $\mathbb{R}^{2d} \ni (\mathbf{x}, \mathbf{p})$. Furthermore, Newton's second law states that the rate of change of momentum is equal to the force acting on the body, i.e, $\dot{\mathbf{p}} = \mathbf{F}$. Assuming m to be independent of time, the trajectory is then obtained as a solution of a second order initial value problem(IVP)

$$m\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r}(t), t); \qquad \mathbf{r}(t_z) = \mathbf{r}_0, \quad \dot{\mathbf{r}}(t_z) = \mathbf{v}_0. \tag{2.1}$$

A force $\mathbf{F}$ is termed *conservative* if it is time-independent and can be written as a gradient of a scalar function, i.e, $\mathbf{F}_{\mathbf{x}}(\mathbf{x}) = -\nabla\phi(\mathbf{x})$. For such a force, we can integrate the IVP by multiplying with $\dot{\mathbf{r}}$:

$$0 = \dot{\mathbf{r}} \cdot (m\ddot{\mathbf{r}} - \mathbf{F}) = m\dot{\mathbf{r}} \cdot \ddot{\mathbf{r}} + \dot{\mathbf{r}} \cdot \nabla\phi = \frac{d}{dt}\left(\frac{1}{2}m|\dot{\mathbf{r}}|^2 + \phi\right).$$

Thus, for a given trajectory $\mathbf{r}(t)$, the quantity in the bracket is conserved (i.e, independent of time), which we define as the *total mechanical energy* of the system:

$$\varepsilon = \frac{1}{2}m|\dot{\mathbf{r}}|^2 + \phi(\mathbf{r}) = \frac{|\mathbf{p}|^2}{2m} + \phi(\mathbf{r}). \tag{2.2}$$

In the subsequent discussions, we shall only consider conservative forces.

Thus, the mechanics of a point particle is described by a second order IVP in d variables r_i. These can readily be reduced to a first order IVP in $2d$ variables, which define flows on $\mathbb{R}^{2d}$, the 2d-dimensional state space. In this formulation, the $2d$ variables can be treated as independent. There are two distinct choices of independent variables, viz, $(\mathbf{r}, \dot{\mathbf{r}})$ and $(\mathbf{r}, \mathbf{p})$, which correspond to two alternative formulations of classical mechanics, viz, the Lagrangian and the Hamiltonian formalisms, respectively.

We begin with Hamiltonian mechanics, which takes place on a *phase space* $\mathcal{M}$, which is the space of all possible positions and momenta. For a point particle on $\mathbb{R}^d$, the (classical) phase space is simply $\mathbb{R}^d \times \mathbb{R}^d$, with coordinates $\boldsymbol{\zeta} = (\mathbf{p}, \mathbf{x})$. Then, we seek to compute the trajectory $\boldsymbol{\gamma} : \mathbb{R} \to \mathcal{M}$, where the state of the particle at time t is $\boldsymbol{\gamma}(t) = (\boldsymbol{\gamma}_{\mathbf{p}}(t), \boldsymbol{\gamma}_{\mathbf{x}}(t))$. The Newton's second law reduces to

$$\dot{\boldsymbol{\gamma}}_{\mathbf{r}}(t) = \frac{\boldsymbol{\gamma}_{\mathbf{p}}}{m}, \quad \dot{\boldsymbol{\gamma}}_{\mathbf{p}}(t) = \mathbf{F}\left(\boldsymbol{\gamma}_{\mathbf{r}}(t)\right) = -\nabla\phi(\mathbf{x})\Big|_{\mathbf{x}=\boldsymbol{\gamma}_{\mathbf{r}}(t)}. \tag{2.3}$$

Furthermore, taking a cue from eq. (2.2), one defines a "Hamiltonian" $H\colon \mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ as

$$H(t,\mathbf{x},\mathbf{p}) = \frac{|\mathbf{p}|^2}{2m} + \phi(\mathbf{x},t), \tag{2.4}$$

Note that this quantity is defined in general for any system where the force can be written as a gradient, i.e, $\mathbf{F}(\mathbf{x},t) = -\nabla_{\mathbf{x}}\phi(\mathbf{x},t)$. However, it is a constant along the trajectories if ϕ is independent of t, i.e, if the system has time translation symmetry.

In terms of the Hamiltonian, the equations of motion become

$$\dot{\boldsymbol{\gamma}}_{\mathbf{r}}(t) = \nabla_{\mathbf{p}}H(t,\mathbf{x},\mathbf{p})\Big|_{(\mathbf{x},\mathbf{p})=\boldsymbol{\gamma}(t)}, \quad \dot{\boldsymbol{\gamma}}_{\mathbf{p}}(t) = -\nabla_{\mathbf{x}}H(t,\mathbf{x},\mathbf{p})\Big|_{(\mathbf{x},\mathbf{p})=\boldsymbol{\gamma}(t)}. \tag{2.5}$$

Conventionally, we 'forget' the distinction between $\mathbf{x}$ and $\boldsymbol{\gamma}_{\mathbf{x}}$, and between $\mathbf{p}$ and $\boldsymbol{\gamma}_{\mathbf{p}}$. Thus, the Hamilton's equations are simply written as

$$\dot{x}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial x_i}, \tag{2.6}$$

which can be combined as

$$\dot{\zeta}^i = \mathcal{J}^{ij}\frac{\partial H}{\partial \zeta^j}, \qquad \mathcal{J} = \begin{pmatrix} 0 & -\mathbb{1}_d \\ \mathbb{1}_d & 0 \end{pmatrix}, \quad i,j = 1,\ldots,2d. \tag{2.7}$$

The structure of this differential equation is termed *symplectic*. This equation would be the starting point for a geometric formulation of classical mechanics, as discussed in Sec 2.2.

Next, we discuss Lagrangian mechanics, which reduces the IVP to a boundary value problem (BVP). Given all possible (smooth) trajectories $\mathbf{r}(t)$ connecting the initial position $\mathbf{r}(t_0) = \mathbf{r}_0$ and the final position $\mathbf{r}(t_1) = \mathbf{r}_1$, one defines an action functional $S\colon \mathbf{r}(t) \mapsto S[\mathbf{r}(t)] \in \mathbb{R}$. The actual trajectory of the particle is then the trajectory $\mathbf{r}(t)$ that extremizes the action, i.e, for which the first order variation $\delta S[\mathbf{r}(t)] = 0$. The idea of writing dynamical equation of motion as extremization of a functional, commonly referred to as *Hamilton's principle* or the *principle of least action*, has wide applicability beyond classical mechanics, for instance, in classical field theories or general relativity.

The action can be written as an integral

$$S[\mathbf{r}(t)] = \int_{t_z}^{t_1} dt\, L\left(\mathbf{r}(t), \dot{\mathbf{r}}(t), t\right), \tag{2.8}$$

where $L(\mathbf{x},\dot{\mathbf{x}},t)$ is the *Lagrangian*, and $\mathbf{r}$ and $\dot{\mathbf{r}}$, being the coordinates on the state space, are to be treated

as independent variables. Thus, a variation of action can then be written as

$$\begin{aligned}
\delta S[\mathbf{r}] &= S[\mathbf{r}+\delta\mathbf{r}] - S[\mathbf{r}] \\
&= \int_{t_0}^{t_1} dt \left[L\left(\mathbf{r}(t)+\delta\mathbf{r}(t), \dot{\mathbf{r}}(t)+\delta\dot{\mathbf{r}}(t), t\right) - L\left(\mathbf{r}(t), \dot{\mathbf{r}}(t), t\right)\right] \\
&= \int_{t_0}^{t_1} dt \left[\delta r_i(t) \frac{\partial L}{\partial r_i} + \delta \dot{r}_i(t) \frac{\partial L}{\partial \dot{r}_i} \right] \\
&= \delta r_i(t) \left. \frac{\partial L}{\partial \dot{r}_i} \right|_{t_0}^{t_1} + \int_{t_0}^{t_1} dt\, \delta r_i(t) \left[\frac{\partial L}{\partial r_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}_i} \right) \right].
\end{aligned} \tag{2.9}$$

The boundary term vanishes as we demand that each trajectory satisfies the boundary conditions, so that $\delta\mathbf{r}(t_0) = \delta\mathbf{r}(t_1) = 0$. Then, the condition on $\mathbf{r}(t)$ for $S[\mathbf{r}]$ to achieve an extremum becomes

$$\frac{\partial L}{\partial r_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}_i} \right) = 0. \tag{2.10}$$

These are the *Euler-Lagrange* equations of motion. For instance, for a single particle moving under a conservative force $\mathbf{F}(\mathbf{x}) = -\nabla\phi(\mathbf{x})$, we have

$$L = \frac{1}{2} m |\dot{\mathbf{x}}|^2 - \phi(\mathbf{x}, t) \implies 0 = \frac{\partial \phi}{\partial r_i} + \frac{d}{dt}\left(m \dot{r}_i \right) = m \ddot{r}_i - F_i, \tag{2.11}$$

which precisely reproduces the Newton's second law.

We note that there is a subtlety in going from Newtonian to Lagrangian formulation. If the forces are smooth (more precisely, Lifshitz) functions of coordinates, then the IVP has a unique solution for all initial conditions. However, for BVP, this is not the case in general. A typical example would be motion of a point particle on a circle $S^1 = [-\pi, \pi)$, with the boundary conditions $x(0) = x(T) = 0$. The corresponding action does not have a unique extremum; instead, it is extremized by any trajectory that winds around S^1 an integer number of times[1]

The Hamiltonian and Lagrangian for a given system are related by a *Legendre transform*. Explicitly,

$$L = \mathbf{p} \cdot \dot{\mathbf{x}} - H. \tag{2.12}$$

The existence of requires that $H(\mathbf{x}, \mathbf{p})$ be a convex function of $\mathbf{p}$. We shall use this to write an action functional for a given Hamiltonian.

The original Newtonian approach can often be quite complicated for many practical systems. A particular

[1] These different winding numbers are just the *topological sectors*! When summing over "all paths" to quantize the theory using Feynman's path integral, one needs to sum over this sectors explicitly.

example is systems with holonomic constraints, i.e, we require the trajectory to lie on the level set $\mathbf{f}(\mathbf{x}) = 0$, where $f\colon \mathbb{R}^d \to \mathbb{R}^{d'}$. The resulting space $\mathcal{R} \subset \mathbb{R}^d$ is generically $d-d'$ dimensional, and usually has a manifold structure for systems of interest in classical mechanics. The Lagrangian and Hamiltonian formalisms are convenient to deal with such cases, as one can define a set of coordinates on $\mathcal{R}$, conventionally termed "generalized coordinates", and project the Hamiltonian or the action down to these coordinates. Since both of these approaches describe flows on some $2d$-dimensional manifolds, they can readily be generalized to Hamiltonian flows on more general smooth manifolds. The symplectic formalism is just such a generalization of the Hamiltonian mechanics.

2.2 Symplectic formulation of classical mechanics

Consider the Hamiltonian formulation of classical mechanics, where the underlying space is a even-dimensional smooth manifold, *viz*, the phase space $\mathcal{M} \cong \mathbb{R}^{2m}$, and given a time-independent Hamiltonian $H\colon \mathcal{M} \to \mathbb{R}$ and a point $\boldsymbol{\zeta}_0 \in \mathcal{M}$, the Hamilton's equations of motion associate with it a trajectory passing through $\boldsymbol{\zeta}_0$. Geometrically, these trajectories are simply smooth 1-parameter family of curves, generated by a vector field $\xi_H \in \mathrm{Vect}(\mathcal{M})$, which can be explicitly written as

$$\xi_H = \xi_H^i \frac{\partial}{\partial \zeta^i}, \qquad \xi_H^i = \dot{\zeta}^i = \mathcal{J}^{ij} \frac{\partial H}{\partial \zeta^j}, \tag{2.13}$$

with $\mathcal{J}$ as defined in eq. (2.7), and $\frac{\partial H}{\partial \zeta^j}$ are simply the components of dH. We can potential generalize this to a larger class of even-dimensional smooth manifolds $\mathcal{M}$ on which have a "sensible" antisymmetric matrix analogous to $\mathcal{J}$. Such a structure is termed a symplectic structure, and such manifolds are termed symplectic manifolds.

Geometrically speaking, Hamiltonian mechanics associates vector fields to the differentials of functions (Hamiltonians) $H\colon \mathcal{M} \to \mathbb{R}$. This amounts to a linear map $\Omega^1(\mathcal{M}) \to \mathrm{Vect}(\mathcal{M})$, which can be defined using a map $T^*\mathcal{M} \to T\mathcal{M}$. Such a map is induced by a symplectic structure on the manifold $\mathcal{M}$.

Symplectic manifolds

A *symplectic manifold* $(\mathcal{M}, \rho)$ is an even-dimensional smooth manifold M equipped with a closed, nondegenerate 2-form ρ, termed the symplectic form. More explicitly, $\rho \in \Omega^2(\mathcal{M})$ such that $d\rho = 0$, and the nondegeneracy of ρ means that if for some $\xi \in \mathrm{Vect}(M)$, the symplectic form satisfies $\rho(\xi, \xi') = 0 \ \ \forall\, \xi' \in \mathrm{Vect}(M)$, then $\xi = 0$.

The symplectic form induces a canonical isomorphism, $\widetilde{\rho}\colon \mathrm{Vect}(\mathcal{M}) \to \Omega^1(\mathcal{M})$, which maps $\widetilde{\rho}\colon \xi \mapsto i_\xi \rho$, where ξ is a vector field on $\mathcal{M}$. The nondegeneracy of the symplectic form implies that the map $\widetilde{\rho}$ is invertible (see eq. (2.17)). Using its inverse, we define the *symplectic gradient* $\mathfrak{d}\colon C^\infty(\mathcal{M}, \mathbb{R}) \to \mathrm{Vect}(\mathcal{M})$, which acts on a smooth function $f\colon \mathcal{M} \to \mathbb{R}$ as

$$\mathfrak{d}f = -\left(\widetilde{\rho}\right)^{-1} df \iff i_{\mathfrak{d}f}\rho = -df. \tag{2.14}$$

Thus, given a Hamiltonian H, the symplectic derivate associates to it the vector field $\mathfrak{d}H$, which defines the trajectories for given initial conditions. The Hamilton's equation of motions can be written concisely in a coordinate-independent form as

$$i_{\mathfrak{d}H}\rho = -dH. \tag{2.15}$$

Vector fields of the form $\xi = \mathfrak{d}f$ for some $f \in C^\infty(\mathcal{M})$ are hereafter termed *Hamiltonian vector fields*.

More explicitly, consider a set of local coordinates $\boldsymbol{\zeta} = \{\zeta^i\}$ on some open set of $\mathcal{M}$. In these coordinates, the symplectic form can be written as

$$\rho = \frac{1}{2}\rho_{ij}(\boldsymbol{\zeta})\, d\zeta^i \wedge d\zeta^j, \tag{2.16}$$

where $\boldsymbol{\rho}$ is an antisymmetric matrix, and non-degeneracy of the 2-form ρ implies that $\boldsymbol{\rho}$ is invertible. Given a vector field $\xi \in \mathrm{Vect}(\mathcal{M})$, the map $\widetilde{\rho}$ acts as

$$\widetilde{\rho}\colon \xi = \xi^i \frac{\partial}{\partial \zeta^i} \mapsto i_\xi \rho = \frac{1}{2}\rho_{ij}\left(\xi^i d\zeta^j - \xi^j d\zeta^i\right) = -\rho_{ij}\,\xi^j d\zeta^i. \tag{2.17}$$

Thus, given $\theta \in \Omega^1(\mathcal{M})$, we have $\widetilde{\rho}^{-1}\colon \theta_i \mapsto -\left(\boldsymbol{\rho}^{-1}\right)^{ij}\theta_j$. Given a Hamiltonian H, the components of the associated Hamiltonian vector field $\xi_H = \mathfrak{d}H$ are then $\xi_H^i = \left(\boldsymbol{\rho}^{-1}\right)^{ij}\frac{\partial H}{\partial \zeta^j}$. Thus, the phase-space trajectory passing through $\boldsymbol{\zeta}_0 \in \mathcal{M}$ is given by the IVP

$$\rho_{ij}\dot{\zeta}^j(t) = \frac{\partial H}{\partial \zeta^i} \iff \dot{\zeta}^i(t) = \left(\boldsymbol{\rho}^{-1}\right)^{ij}\frac{\partial H}{\partial \zeta^j}, \tag{2.18}$$

with the initial value $\boldsymbol{\zeta}(t_0) = \boldsymbol{\zeta}_0$. In practice, one 'defines' the Hamiltonian vector field $\xi_H = \dot{\zeta}^i \frac{\partial}{\partial \zeta^i}$, and substitutes it in eq. (2.15) to derive the equations of motion[70].

We note that eq. (2.7) is now simply a special case of eq. (2.18), where $\boldsymbol{\rho} = \mathcal{J}^{-1}$ is a constant. For any symplectic manifold $(\mathcal{M}, \rho)$, one can locally set ρ to be constant, following Darboux theorem[71, 27]. More precisely, Darboux's theorem states that any symplectic manifold $(\mathcal{M}, \rho)$ is locally isomorphic to $(\mathbb{R}^{2m}, \rho_0)$, with $\rho_0 = dP_i \wedge dX^i$, where $(X^1, \ldots X^m, P_1, \ldots P_m)$ are coordinates on $\mathbb{R}^{2m}$. These are termed

the "canonical" coordinates on $\mathcal{M}$. The notations P and X are intended to be suggestive of momenta and positions. However, it must be kept in mind that these are only local coordinates on $\mathcal{M}$, which may not, in general, be globally well-defined.

Structure preserving maps

We begin with the transformations of symplectic manifolds induced by Hamiltonian dynamics. Recall that a vector field $\xi \in \mathrm{Vect}(\mathcal{M})$ can be defined as a generator of a 1-parameter family of curves. Explicitly, given coordinates $\boldsymbol{\zeta}$ and a point $\boldsymbol{\zeta}_0 \in \mathcal{M}$, the vector field ξ defines a curve $\boldsymbol{\zeta}(\tau)$ on $\mathcal{M}$ with $\boldsymbol{\zeta}(0) = \boldsymbol{\zeta}_0$. Given this Hamiltonian evolution, define a family of maps $\Xi_\tau \colon \mathcal{M} \to \mathcal{M},\ \tau \in \mathbb{R}$ as

$$\Xi_\tau \colon \boldsymbol{\zeta}_0 \mapsto \boldsymbol{\zeta}(t) \implies \frac{d}{d\tau}\left(\Xi_\tau \zeta_0^i\right)\Big|_{\tau=0} = \dot{\zeta}^i(0) = \xi^i(\boldsymbol{\zeta}_0). \tag{2.19}$$

These, by definition, follow the composition $\Xi_t \circ \Xi_{t'} = \Xi_{t+t'} = \Xi_{t'} \circ \Xi_t$. The variation of an arbitrary r-form θ under these maps can be defined using the Lie derivative. Recall that the Lie derivative of θ along ξ is defined as

$$\mathcal{L}_\xi \theta = \frac{d}{d\tau}\left(\Xi_\tau^* \theta\right)\Big|_{\tau=0}, \tag{2.20}$$

where Ξ_τ^* is the pullback of Ξ_τ. Equivalently, given a p-dimensional submanifold $\mathcal{S} \subseteq \mathcal{M}$,

$$\frac{d}{d\tau}\left[\int_{\Xi'_\tau \mathcal{S}} \theta\right]_{\tau=0} = \int_{\mathcal{S}} \frac{d}{d\tau} \Xi'^* \theta\Big|_{\tau=0} = \int_{\mathcal{S}} \mathcal{L}_{\xi'} \theta. \tag{2.21}$$

Furthermore,

$$\frac{d}{d\tau}\left(\Xi_\tau^* \theta\right)\Big|_{\tau=t} = \frac{d}{d\tau'}\left(\Xi_t^* \circ \Xi_{\tau'}^* \theta\right)\Big|_{\tau'=0} = \Xi_t^* \frac{d}{d\tau'}\left(\Xi_{\tau'}^* \theta\right)\Big|_{\tau'=0} = \Xi_t^* \left(\mathcal{L}_\xi \theta\right), \tag{2.22}$$

where we have set $\tau = t + \tau'$. An immediate corollary is

$$\mathcal{L}_\xi \theta = 0 \implies \theta \text{ is invariant under the flows induced by } \xi. \tag{2.23}$$

This follows from the fact that the vanishing of the Lie derivative implies that $\Xi_\tau^* \theta$ is a constant. However, Ξ_0 is the identity operator, so that $\Xi_\tau^* \theta = \Xi_0^* \theta = \theta$.

Given a symplectic manifold $(\mathcal{M}, \rho)$, consider the Hamiltonian vector field $\mathfrak{d}H$ for some Hamiltonian $H \in C^\infty(\mathcal{M})$. Using the fact that $\mathcal{L}_\xi = d i_\xi + i_\xi d$, we compute

$$\frac{d}{dt} \Xi_t^* \rho = \Xi_t^* \left[\left(d\, i_{\mathfrak{d}H} + i_{\mathfrak{d}H} d\right) \rho\right] = -\Xi_t^* \left[d(dH)\right] = 0. \tag{2.24}$$

Thus ρ is invariant under the Hamiltonian flows.

The structure preserving maps for symplectic manifolds are termed *symplectomorphisms*[71, 27]. Given symplectic manifolds $(\mathcal{M}, \rho)$ and $(\mathcal{M}', \rho')$, a map $f\colon \mathcal{M} \to \mathcal{M}'$ is a symplectomorphisms if $f^*\rho' = \rho$, where $f^*\colon T^*\mathcal{M}' \to T^*\mathcal{M}$ is the pullback of f. Thus, we have shown that Hamiltonian flows on $(\mathcal{M}, \rho)$ define a family of symplectomorphisms $(\mathcal{M}, \rho) \to (\mathcal{M}, \rho)$.

Action principle

We now reformulate the geometric form of the Hamilton's equations in the form of an action principle. Recall that since ρ is closed, by Poincaré's lemma, it is exact on any contractible subspace $\mathcal{U}_0 \subseteq \mathcal{M}$, i.e, $\exists\, \eta \in \Omega^1(\mathcal{U})$ with $t_1 > t_0$ such that $\rho = d\eta$. The 1-form η is referred to as the *symplectic potential* or the *Liouville 1-form*. Given the Liouville 1-form η and a Hamiltonian H, we define the action functional as

$$S[\mathcal{C}] = \int_{\mathcal{C}} \eta - \int_{t_z}^{t_1} H(\boldsymbol{\zeta}(t))\, dt, \tag{2.25}$$

where $\mathcal{C} \equiv \{\boldsymbol{\zeta}\colon [t_0, t_1] \to \mathcal{M} \mid \boldsymbol{\zeta}(t_0) = \boldsymbol{\zeta}_0,\ \boldsymbol{\zeta}(t_1) = \boldsymbol{\zeta}_1\}$ defines a curve on $\mathcal{M}$ with fixed end points $\boldsymbol{\zeta}_0$ and $\boldsymbol{\zeta}_1$.

Given a set of local coordinates $\boldsymbol{\zeta}$ on $\mathcal{M}$, the variation of action becomes

$$\begin{aligned} 0 = \delta \int_{t_0}^{t_1} \Big(\eta_i(\boldsymbol{\zeta})\dot{\zeta}^i - H\Big)\, dt &= \int_{t_0}^{t_1} \left(\frac{\partial \eta_i}{\partial \zeta^j}\delta\zeta^j\dot{\zeta}^i + \eta_i(\boldsymbol{\zeta})\delta\dot{\zeta}^i - \frac{\partial H}{\partial \zeta^i}\delta\zeta^i\right) dt \\ = \int_{t_0}^{t_1} \delta\zeta^i \left[\left(\frac{\partial \eta_j}{\partial \zeta^i}\dot{\zeta}^j - \frac{d\eta_i}{dt}\right) - \frac{\partial H}{\partial \zeta^i}\right] dt &= \int_{t_0}^{t_1} \delta\zeta^i \left[\left(\frac{\partial \eta_j}{\partial \zeta^i} - \frac{d\eta_i}{d\zeta^j}\right)\dot{\zeta}^j - \frac{\partial H}{\partial \zeta^i}\right] dt, \end{aligned} \tag{2.26}$$

which needs to be true for any variation $\delta\boldsymbol{\zeta}(t)$. Thus, we demand that

$$\rho_{ij}\dot{\zeta}^j(t) = \frac{\partial H}{\partial \zeta^i}, \qquad \rho_{ij} = \frac{\partial \eta_j}{\partial \zeta^i} - \frac{d\eta_i}{d\zeta^j}, \tag{2.27}$$

which is precisely eq. (2.18), the equation of motion in our given coordinates.

Poisson brackets and Poisson manifolds

As an aside, we now discuss an algebraic (as opposed to geometric) description of the symplectic manifolds, in terms of Poisson brackets. Recall that in conventional Hamiltonian mechanics, the Poisson bracket of two functions $f, g \in C^\infty(\mathbb{R}^{2m})$ is an antisymmetric bilinear product on $C^\infty(\mathcal{M})$, defined as

$$\{f, g\} = \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial x^i} = -\mathcal{J}^{ij}\frac{\partial f}{\partial \zeta^i}\frac{\partial g}{\partial \zeta^j}, \tag{2.28}$$

where $(\mathbf{p}, \mathbf{x})$ are the canonical coordinates on $\mathbb{R}^{2n}$. For a symplectic manifold $(\mathcal{M}, \rho)$ with coordinates $\boldsymbol{\zeta}$, this readily generalizes to

$$\{f, g\} = -\left(\boldsymbol{\rho}^{-1}\right)^{ij} \frac{\partial f}{\partial \zeta^i} \frac{\partial g}{\partial \zeta^j}, \tag{2.29}$$

which, using eq. (2.18), leads to a purely geometric expression for the Poisson bracket as

$$\{f, g\} = -\left(\rho^{-1}\right)^{ij} \rho_{ik} (\mathfrak{d} f)^k \rho_{j\ell} (\mathfrak{d} g)^\ell = (\mathfrak{d} f)^k \rho_{k\ell} (\mathfrak{d} g)^\ell = \rho(\mathfrak{d} f, \mathfrak{d} g). \tag{2.30}$$

Thus, we also have

$$\{f, g\} = i_{\mathfrak{d} f} i_{\mathfrak{d} g} \rho = -i_{\mathfrak{d} f} dg = -(\mathfrak{d} f)^\mu \partial_\mu g, \quad \text{also,} \quad \{f, g\} = (\mathfrak{d} g)^\mu \partial_\mu f, \tag{2.31}$$

so that geometrically, the Poisson bracket is simply the directional derivative of g along $-\mathfrak{d} f$, or the directional derivative of f along $\mathfrak{d} g$.

Given any time-dependent quantity f and a Hamiltonian H,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \dot{\zeta}^i \frac{\partial f}{\partial \zeta^i} = \frac{\partial f}{\partial t} + \left(\rho^{-1}\right)^{ij} \frac{\partial H}{\partial \zeta^j} \frac{\partial f}{\partial \zeta^i} = \frac{\partial f}{\partial t} + \{H, f\}. \tag{2.32}$$

If f does not depend explicitly on time, then $\{H, t\}$ is simply the time-derivative of f. Thus, H can be thought of as the generator of time translations.

Abstractly, a Poisson bracket is a map $\{\,,\,\}\colon C^\infty(\mathcal{M}) \times C^\infty(\mathcal{M}) \to C^\infty(\mathcal{M})$, which satisfies

1. Antisymmetry: $\{f, g\} = -\{g, f\}$,
2. Bilinearity: $\{\alpha f + \beta g, h\} = \alpha\{f, h\} + \beta\{g, h\},\ \alpha, \beta \in \mathbb{R}$, and
3. Jacobi identity: $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$.

Thus, the Poisson bracket defines an infinite-dimensional Lie algebra on $C^\infty(\mathcal{M})$. A manifold $\mathcal{M}$ for which $C^\infty(\mathcal{M})$ is equipped with a Poisson bracket is termed a *Poisson manifold*. More precisely, a smooth manifold $\left(\widetilde{\mathcal{M}}, c\right)$ is termed a Poisson manifold if it is endowed with a bivector field $c = c^{ij}\partial_i \partial_j$ which defines the Poisson bracket $\{f, g\} = c^{ij}\partial_i f \partial_j g$. Thus, symplectic manifolds are a special case[2] of Poisson manifolds, for which c^{ij} is invertible. Furthermore, when c^{ij} is noninvertible, the Poisson manifolds $\left(\widetilde{\mathcal{M}}, c\right)$ can be foliated by (i.e, represented as the disjoint union of) *symplectic leaves* $\mathcal{M}_\alpha$ such that $\left(\mathcal{M}_\alpha, c_\alpha^{-1}\right)$ forms a symplectic manifold, where c_α is the restriction of the bivector c to $\mathcal{M}_\alpha$.

[2]For symplectic manifolds, $\mathfrak{d}\colon C^\infty(\mathcal{M}) \to \mathrm{Vect}(\mathcal{M})$ becomes a Lie algebra homomorphism, i.e, $\mathfrak{d}\{f, g\} = [\mathfrak{d} f, \mathfrak{d} g]$, where $[\,,\,]$ denotes the Lie algebra of vector fields.

2.3 Examples of symplectic manifolds

There are three mathematical contexts that naturally give rise to a symplectic manifolds: contangent bundles of smooth manifolds, co-adjoint orbits of Lie groups and complex algebraic manifolds[72]. In this section, we discuss the first two cases.

2.3.1 Cotangent bundles

In the traditional classical mechanics, the phase space is defined as the cotangent bundle of the d-dimensional *configuration space*[3] $\mathcal{X}$, which, in conventional terminology, is simply the space of *positions*. More formally, let $\mathcal{M} \equiv T^*\mathcal{X} \xrightarrow{\pi} \mathcal{X}$ be the cotangent bundle of a smooth manifold $\mathcal{X}$, with π being the canonical projection. A point $\boldsymbol{\zeta} \in T^*_{\mathbf{x}}\mathcal{X}$ is actually a covector on $\mathcal{X}$, which takes vectors to c-numbers, i.e, $\boldsymbol{\zeta}\colon T_{\mathbf{x}}\mathcal{X} \to \mathbb{R}$. Given $\xi \in T_{\boldsymbol{\zeta}}\mathcal{M}$ with $\pi(\boldsymbol{\zeta}) = \mathbf{x}$, using $\pi_*\colon T_{\boldsymbol{\zeta}}\mathcal{M} \to T_{\pi(\boldsymbol{\zeta})}\mathcal{X}$, we can push it forward to $\pi_*(\xi) \in T_{\mathbf{x}}\mathcal{X}$. Thus, define a 1-form $\eta \in T^*_{\boldsymbol{\zeta}}\mathcal{M}$ by its action on $\xi \in T_{\boldsymbol{\zeta}}\mathcal{M}$ as

$$i_\xi \eta = i_{\pi_*(\xi)}\boldsymbol{\zeta}. \tag{2.33}$$

This defines the presymplectic form, and the symplectic form is simply $\rho = d\eta$, which is closed by definition. This is a geometric formulation of all of the conventional classical mechanics.

Explicitly, given local coordinates $\mathbf{x}$ on $\mathcal{X}$, the covectors are given by $p_i dx^i$, so that $\boldsymbol{\zeta} = (\mathbf{x}, \mathbf{p})$ are coordinates on $\mathcal{M}$, and the canonical projector is $\pi\colon (\mathbf{x}, \mathbf{p}) \mapsto \mathbf{x}$. Then, a vector $\xi \in T_{\boldsymbol{\zeta}}\mathcal{M}$ can be explicitly written as

$$\xi = \xi^i \frac{\partial}{\partial \zeta^i} = \xi^i_{\mathbf{x}} \frac{\partial}{\partial x^i} + \xi^i_{\mathbf{p}} \frac{\partial}{\partial p^i} \implies \pi_*(\xi) = \xi^i_{\mathbf{x}} \frac{\partial}{\partial x^i}, \tag{2.34}$$

where $\pi_*(\xi) \in T_{\mathbf{x}}\mathcal{X}$ and $\xi^i \in \mathbb{R}$. Then, using eq. (2.33)

$$\eta\left(\xi^i_{\mathbf{x}} \frac{\partial}{\partial x^i} + \xi^i_{\mathbf{p}} \frac{\partial}{\partial p^i}\right) = p_i dx^i \left(\xi^i_{\mathbf{x}} \frac{\partial}{\partial x^i}\right) = p_i \xi^i_{\mathbf{x}}, \tag{2.35}$$

which implies that $\eta = p_i dx^i$, and hence $\rho = dp_i \wedge dx^i$. These turn out to be the canonical (Darboux) coordinates on the phase space.

By definition, the phase space $\mathcal{M} = T^*\mathcal{X}$ is noncompact, since it associates to each point in $\mathcal{X}$ a (non-compact) vector space $T_{\mathbf{x}}\mathcal{X}$. Finally, if $\mathcal{X} = \mathbb{R}^d$, then $T^*\mathcal{X} = \mathbb{R}^{2d}$, on which the canonical coordinates $(\mathbf{x}, \mathbf{p})$ are defined globally, which is the setup for the conventional classical mechanics.

[3] For free point particles, $\mathcal{X} = \mathbb{R}^d$. However, this is not the case, for instance, in presence of holonomic constraints.

2.3.2 Coadjoint orbits

The second example of a symplectic manifold comes from the theory of Lie groups, and unlike the case of cotangent bundles, often leads to compact symplectic manifolds. Given a Lie group G with its Lie algebra $\mathfrak{g}$, the coadjoint orbits are defined as the orbits of elements of $\mathfrak{g}^*$, the dual of the Lie algebra, under the *coadjoint action* of the group. In this section, we shall only describe the co-adjoint orbits for matrix groups, for which they are identical to the *adjoint orbits*. For a more general and rigorous description, we refer the reader to Appendix A.

Adjoint and coadjoint orbits

Consider a matrix group $G \subseteq GL(N, \mathbb{C})$, its Lie algebra being $\mathfrak{g} \subset \mathrm{Mat}(N, \mathbb{C})$, and let $n = \dim(G)$ be the dimensionality of the Lie group manifold. As vector spaces, $\mathfrak{g} \cong \mathbb{R}^n$, and the group has a representation on $\mathfrak{g}$, the so called *adjoint representation*, defined as $\mathrm{Ad}(g)\colon X \mapsto g \cdot X \cdot g^{-1}$, where $g \in G$, $X \in \mathfrak{g}$ and '$\cdot$' denotes the matrix multiplication, which we shall hereafter omit. Furthermore, being a subset of $\mathrm{Mat}(N, \mathbb{R})$, $\mathfrak{g}$ inherits an inner product, *viz*, the *Killing form*, defined as $\langle X, Y \rangle = \mathrm{tr}\{XY\}$, which identifies $\mathfrak{g}$ with its dual vector space, $\mathfrak{g}^*$, so that we can define a canonical isomorphism $\varphi\colon \mathfrak{g}^* \to \mathfrak{g}$. The co-adjoint representation is then given by $\mathrm{K}(g)\colon X \mapsto g^{-1}Xg$. For a matrix Lie group, the adjoint and coadjoint representations are isomorphic, the isomorphism being $\mathrm{Ad}(g) = \mathrm{K}(g^{-1})$.

Given $X \in \mathfrak{g}^*$, its adjoint orbit is defined as

$$\mathcal{O}_X = \left\{ gXg^{-1} \mid g \in G \right\} \subseteq \mathfrak{g}^*, \tag{2.36}$$

which coincides with its coadjoint orbit[4]. Furthermore, note that $gXg^{-1} = X$ if, as elements of $Mat(N, \mathbb{C})$, g and X commute. Given X, define the set of g that commute with X as $\mathrm{Stab}(X) \subseteq G$, the stabilizer of G, which forms a subgroup of G. Then, $\mathcal{O}_X \cong G/\mathrm{Stab}(X)$. A trivial example is $X = 0 \in \mathrm{Mat}(n, \mathbb{R})$, where $\mathrm{Stab}(X) = G$, so that $\mathcal{O}_X$ can be identified with a single point. We are usually interested in the co-adjoint orbits of highest possible dimensions (termed *regular* orbits), corresponding to smallest possible $\mathrm{Stab}(X)$.

The regular coadjoint orbits can be defined using the maximal tori (i.e, the maximal connected commuting subgroup) T of the group. A maximal torus of G is generated by a Cartan subalgebra $\mathfrak{t} \subset \mathfrak{g}$, which can be crudely[5] defined as the vector space spanned by a subset of the Lie algebra whose elements commute with each other. The choice of a Cartan subalgebra and hence of a maximal torus is nonunique; however,

[4] This is not true in general, for instance, in infinite-dimensional cases, where it is possible that $\mathfrak{g} \ncong \mathfrak{g}^*$.

[5] More formally, a Cartan subalgebra $\mathfrak{t} \subseteq \mathfrak{g}$ is defined as a *nilpotent, self-normalizing* Lie algebra, i.e, $[\mathfrak{t}, [\mathfrak{t}, ..[\mathfrak{t}, \mathfrak{t}]..]] = 0$ for a finite number of nested commutators, and if $\exists\, Y$ such that $[X, Y] \in \mathfrak{t}\, \forall\, X \in \mathfrak{g}$, then $Y \in \mathfrak{t}$.

all maximal tori are isomorphic and hence of the same dimensionality. Given $X \in \mathfrak{g}^*$, choose a Cartan subalgebra of $\mathfrak{g}$ that commutes with $\varphi(X) \in \mathfrak{g}$, so that $T \subseteq \mathrm{Stab}(X)$. Thus, regular coadjoint orbits are generated by X for which $\mathrm{Stab}(X) = T$, so that $\mathcal{O}_X \cong G/T$, any point on which can be written as $Q = g_{-1} X g,\ g \in G/T$. Since $Q \in \mathfrak{g}^*$, this connects the two definitions of $\mathcal{O}_X$, viz, as a submanifold of $\mathfrak{g} \cong \mathbb{R}^n$ and as a quotient of G by T.

Vector fields and symplectic form

We next define the symplectic (Kirillov-Kostant) form[72] on $\mathcal{O}_X$ using intrinsic (Cartan) calculus, i.e, as a map $\mathrm{Vect}(\mathcal{O}_X) \times \mathrm{Vect}(\mathcal{O}_X) \to \mathbb{R}$. Since vector fields can be geometrically thought of as generators of curves[73, 70], for $\mathcal{O}_X$, every $Y \in \mathfrak{g}$ induces a G-invariant curve $\widetilde{X}(t)$ via the adjoint action, explicitly written as $\widetilde{X}(t) = e^{tY} X e^{-tY}$. Thus, we can define a vector $\xi_Y \in T_X \mathcal{O}_X$ as

$$\xi_Y(X) = \frac{d}{dt}\left(e^{tY} X e^{-tY}\right)\Big|_{t=0} = -[X, Y], \tag{2.37}$$

which can be used to compute ξ_Y everywhere on $\mathcal{O}_X$ using the G-invariance, and all G-invariant vector fields can be written in such a fashion.

The symplectic form on $\mathcal{O}_X$ can be defined by its action on vectors in $T_X \mathcal{O}_X$ as

$$\rho\left(\xi_Y, \xi_Z\right)\Big|_X \equiv \left\langle X, [Y, Z]\right\rangle = \mathrm{tr}\left\{X[Y, Z]\right\}, \tag{2.38}$$

which can be translated to $X' \in \mathcal{O}_X$ using the coadjoint G-action on $\mathcal{O}_X$. We show that ρ is closed and degenerate in the proof of Theorem A.1 in Appendix A. Given a Hamiltonian $H \colon \mathcal{O}_X \to \mathbb{R}$, we can use ρ to associate with it a Hamiltonian vector field $\mathfrak{d}H$ (see Sec 2.2) via

$$i_{\mathfrak{d}H}\rho = -dH \implies \rho(\mathfrak{d}H, \xi_Y) = -\xi_Y H \ \ \forall Y \in \mathfrak{g}. \tag{2.39}$$

A class of Hamiltonians can be defined on $\mathcal{O}_X$ using the Killing form on G, whose Hamiltonian vector fields take a particularly simple form. Given a fixed $Y \in \mathfrak{g}$, and some $Z \in \mathcal{O}_X$, set

$$H_Y \colon Z \mapsto \mathrm{tr}\left\{ZY\right\} \implies \mathfrak{d}H_Y = \xi_Y = -[\,.\,, Y], \tag{2.40}$$

with ξ_Y as defined in eq. (2.37). This can be derived explicitly using eq. (2.39), as shown in Appendix A.

Parametrization

Finally, we discuss an explicit parametrization of the regular coadjoint orbits of a group G using the exponential map. Consider then a basis $\{\lambda_i\}_{i=1}^{\dim(\mathfrak{g})}$ of the Lie algebra $\mathfrak{g}$, which is orthonormal under the Killing form and satisfy the commutation relation $[\lambda_a, \lambda_b] = i\mathfrak{f}_{bc}{}^a\lambda_c$, where $\mathfrak{f}_{bc}{}^a \in \mathbb{R}$ are the structure constants of the Lie algebra. The corresponding dual basis is denoted by $\{\lambda_i\}_{i=1}^{\dim(\mathfrak{g})}$. We have chosen the physics notation in defining the basis, where λ_i are Hermitian matrices and the structure constants are real.

A $Y \in \mathfrak{g}$ can then be expanded in this basis as[6] $Y = iY_a\lambda_a$. Since $Q = g^{-1}Xg \in \mathfrak{g}^*$, in our chosen basis, it can be expanded as $Q = iQ_a\lambda^a$ with $Q_a \equiv -i\mathrm{tr}\{Q\lambda_a\} \in \mathbb{R}$ and $\mathbf{Q} = \{Q_a\}_{a=1}^n$ are simply coordinates of the points of the Lie algebra in $\mathfrak{g} \cong \mathbb{R}_n$. Thus, this parametrization explicitly defines the coadjoint orbit as a subspace of $\mathfrak{g}^* \cong \mathbb{R}^n$.

Consider then the Hamiltonians $H_a = Q_a$, or equivalently, $H_a \colon Q \mapsto -i\mathrm{tr}\{Q\lambda_a\}$ for $Q \in \mathcal{O}_X$. Using eq. (2.40), the corresponding vector fields are simply $\xi_a = -[\,.\,, \lambda_a]$, so that using eqns 2.30 and 2.38, their Poisson bracket becomes

$$\{Q_a, Q_b\} = \rho(\xi_a, \xi_b) = \mathrm{tr}\{Q[\lambda_a, \lambda_b]\} = i\mathfrak{f}_{ab}^c\mathrm{tr}\{Q\lambda_c\} = \mathfrak{f}_{ab}^c Q_c. \tag{2.41}$$

Thus, we have defined a set of functions $Q_a \colon \mathcal{O}_X \to \mathbb{R}$, whose Poisson algebra is analogous to the Lie algebra of the group we started with.

Example: SU(2)

We demonstrate our formal constructions from this section explicitly using SU(2). Some details and lengthy computations have been relegated to Appendix A.4. Recall that SU(2) is the group defined as the set of 2×2 unitary matrices with unit determinant, under the usual matrix multiplication. It can be conveniently parametrized using the Pauli matrices as

$$\mathrm{SU}(2) = \left\{ g_0\mathbb{1} + i\,\mathbf{g}\cdot\boldsymbol{\sigma} \;\middle|\; (g_0, \mathbf{g}) \in \mathbb{R}^4,\ \sum_{\mu=0}^{3} g_\mu^2 = 1 \right\}. \tag{2.42}$$

[6] The i in the definition is required, since the basis elements are real, but the generators of curves on G must be skew-symmetric. Mathematicians usually define the basis of the Lie algebra as skew symmetric to avoid this i, while physicists, intending to identify the basis elements of $\mathfrak{g}$ as physical (and hence Hermitian) symmetry operator, put in the i explicitly instead.

The constraint on g_μ defines a 3-sphere in $\mathbb{R}^4$, so that as smooth manifolds, SU(2) is homeomorphic to S^3. The corresponding Lie algebra is

$$\mathfrak{su}(2) = \{X^a\sigma_a \,|\, \mathbf{X} \in \mathbb{R}^3\}, \qquad [\sigma_a, \sigma_b] = i\epsilon_{abc}\sigma_c, \tag{2.43}$$

with the basis $\lambda_i = \sigma_i$ being the Pauli matrices. In the following, we shall not distinguish between $\mathfrak{su}(2)$ and $\mathfrak{su}^*(2)$, so that we only use lower indices. The maximal torus $T \subset \mathrm{SU}(2)$ can be written as the set of diagonal matrices

$$T = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^* \end{pmatrix}, \;\; |\alpha|^2 = 1 \right\} \cong \mathrm{U}(1). \tag{2.44}$$

The regular adjoint/coadjoint orbits of SU(2) are then given by $\mathcal{O} = \mathrm{SU}(2)/T$. But as smooth manifolds, $\mathrm{SU}(2) \cong S^3$ and $\mathrm{U}(1) \cong S^1$, so that[7] $\mathcal{O} \cong S^3/S^1 \cong S^2$. Explicitly, all coadjoint orbits of SU(2) can be written as $|\mathbf{X}|^2 = r^2$, where $r = 0$ corresponds to the trivial orbit of $\mathbf{0} \in \mathfrak{su}(2)$, while all orbits for $r > 0$ are regular. Fixing an r and choosing an X on the 2-sphere of radius r ($X = r\sigma^3$, say), the corresponding coadjoint orbit is explicitly given by

$$\mathcal{O}_X = \{\mathbf{Q} \in \mathbb{R}^3 \,|\, |\mathbf{Q}|^2 = r^2\}. \tag{2.45}$$

This defines the embedding of $\mathcal{O}_X$ in $\mathbb{R}^3$.

To define the vector fields on $\mathcal{O}_X$, consider the Hamiltonians $H_a \colon Q \mapsto Q_a = -i\mathrm{tr}\,\{Q\sigma_a\}$. The corresponding vector field ξ_a generates the curve $\widetilde{X}(t) = e^{-it\sigma_a} X e^{it\sigma_a}$ on $\mathcal{O}_X$. Thus,

$$\xi_a(Q) = \frac{d}{dt}\left(e^{-it\sigma_a} Q e^{it\sigma_a}\right)\Big|_{t=0} = i[Q, \sigma_a] = \epsilon_{abc} Q_b \sigma_c, \tag{2.46}$$

which can be written in the traditional coordinate form as $\xi_a = \epsilon_{abc} Q_b\, \partial_c$, where $\partial_c = \frac{\partial}{\partial Q_c}$. Explicitly,

$$\xi_1 = Q_2\partial_3 - Q_3\partial_2, \qquad \xi_2 = Q_3\partial_1 - Q_1\partial_3, \qquad \xi_3 = Q_1\partial_2 - Q_2\partial_1. \tag{2.47}$$

These can be identified as the generators of rotations in $\mathbb{R}^3$ along the three orthogonal axes defined by Q_a. Finally, the symplectic form ρ can be explicitly written as

$$\rho = \frac{1}{2}\rho_{ab}\, dQ_a \wedge dQ_b. \tag{2.48}$$

[7] This can also be seen by a somewhat lengthy computation using explicit coordinates on $\mathfrak{g} \cong \mathbb{R}^3$, as shown in Appendix A.4.

Substituting the vectors and using eq. (2.41), we get

$$Q_1 = \rho\,(\xi_2, \xi_3) = Q_1\Big(Q_1\rho_{23} + Q_2\rho_{31} + Q_3\rho_{12}\Big), \tag{2.49}$$

and similar expressions for Q_2 and Q_3. Furthermore, since ρ is a 2-form on $\mathbb{R}^3$, its Hodge dual (using the Euclidean metric on $\mathbb{R}^3$) is simply a 3-vector $\mathbf{q}$, such that $\rho_{ab} = \epsilon_{abc}q_c$. Thus,

$$1 = Q_1\rho_{23} + Q_2\rho_{31} + Q_3\rho_{12} = \mathbf{Q}\cdot\mathbf{q} \implies \mathbf{q} = \frac{\mathbf{Q}}{|\mathbf{Q}|^2}. \tag{2.50}$$

Setting $|\mathbf{Q}| = r$, we get $\rho_{ab} = r^{-2}\,\epsilon_{abc}Q_c$. Thus, the Kirillov form on $\mathcal{O}_X$ can be written explicitly as

$$\rho = \frac{1}{r^2}\Big(Q_1 dQ_2\wedge dQ_3 + Q_2 dQ_3\wedge dQ_1 + Q_3 dQ_1\wedge dQ_2\Big) = R\Omega(S^2), \tag{2.51}$$

where $\Omega(S^2)$ is the volume form on a 2-sphere.

2.4 Time-dependent systems and extended phase space

The conventional Hamiltonian mechanics is defined for time-dependent Hamiltonians; however, in the symplectic formulation of classical mechanics, we assumed the Hamiltonian to be time-independent, so that time was only as a quantity parameterizing the trajectories. We now remedy this shortcoming by considering a general time-dependent Hamiltonian $H\colon \mathbb{R}\times\mathcal{M}\to\mathbb{R}$, as well as a time-dependent 2-form $\rho\colon \mathbb{R}\to\Omega^2(\mathcal{M})$ for some phase space $\mathcal{M}$, such each for each time t, $(\mathcal{M}, \rho(t))$ form a symplectic manifold. This situation is quite relevant for this thesis, since we seek a description of point particles interacting with external fields, which might be time-dependent.

Contact manifolds

The mathematical structure underlying this description is an extended phase space (or *contact manifold*) including time as one of the coordinates. Formally, a contact manifold (M_H, ρ_H) is an odd-dimensional manifold M_H equipped with a closed 2-form ρ_H of maximal rank. We shall refer to ρ_H as the *generalized symplectic form*. The rank of ρ_H is defined as the maximal number of linearly independent vector fields $\xi_i \in \mathrm{Vect}(M_H)$ for which $i_{\xi_i}\rho_H \neq 0$. Given coordinates $\widetilde{\boldsymbol{\zeta}}$ on M_H, we can write $\rho_H = \frac{1}{2}\,(\rho_H)_{ij}\,d\widetilde{\zeta}^i d\widetilde{\zeta}^j$, where $\boldsymbol{\rho}_H$ is a real skew-symmetric matrix. The rank of ρ_H is then simply the matrix rank of $\boldsymbol{\rho}_H$. However, since all eigenvalues of $\boldsymbol{\rho}_H$ are purely imaginary and occur in complex conjugate pairs, the matrix rank of $\boldsymbol{\rho}_H$

must be even. Thus, if $\dim(M_H) = 2m+1$, we must have $\operatorname{rank}(\rho_H) = 2m$, and the kernel of ρ_H, i.e, the set of vectors ξ satisfying $i_\xi \rho_H = 0$ is one dimensional.

We begin by simply rephrasing the time-independent Hamiltonian case using the extended phase space formalism. Consider a symplectic manifold $(\mathcal{M}, \rho)$ and a Hamiltonian $H\colon \mathcal{M} \to \mathbb{R}$ and set $M_H = \mathcal{M} \times \mathbb{R}$, where $\mathbb{R}$ corresponds to the time. Given coordinates $\boldsymbol{\zeta}$ on $\mathcal{M}$, define coordinates on M_H as $\widetilde{\boldsymbol{\zeta}} = (t, \boldsymbol{\zeta})$, so that $\widetilde{\zeta}_0 = t$. Define

$$\rho_H = \rho - dH \wedge dt. \tag{2.52}$$

Clearly, $d\rho_H = d\rho = 0$ and $\boldsymbol{\rho}_H$ is of maximal rank since $\boldsymbol{\rho}$ is nonsingular and hence of full rank, so that (M_H, ρ_H) forms a contact structure. To compute the null vector of ρ_H, we represent a vector field $\widetilde{\xi} \in \operatorname{Vect}(M_H)$ in these coordinates as

$$\widetilde{\xi} = \widetilde{\xi}^\mu \frac{\partial}{\partial \widetilde{\zeta}^\mu} = \widetilde{\xi}^0 \frac{\partial}{\partial t} + \widetilde{\xi}^i \frac{\partial}{\partial \zeta^i} = \widetilde{\xi}^0 \frac{\partial}{\partial t} + \xi. \tag{2.53}$$

Then,

$$\widetilde{\xi} \in \ker \rho_H \implies 0 = i_{\widetilde{\xi}} \rho_H = i_\xi \rho - (i_\xi dH)\, dt + \widetilde{\xi}^0 dH. \tag{2.54}$$

Using the equation of motion $i_{\mathfrak{d}H}\rho = -dH$, we get

$$0 = i_\xi \rho + (i_\xi i_{\mathfrak{d}H} \rho)\, dt - \widetilde{\xi}^0 i_{\mathfrak{d}H}\rho = \left(i_\xi - \widetilde{\xi}^0\, i_{\mathfrak{d}H}\right) \rho + \rho\,(\xi, \mathfrak{d}H)\, dt. \tag{2.55}$$

Since ρ does not contain a dt, both terms must vanish individually. The obvious solution is $\widetilde{\xi}^0 = 1$ and $\xi = \mathfrak{d}H$, which must be unique, since ρ_H has only one null vector. Thus, given H, the associated trajectory is simply the null vector of the associated contact structure M_H. The equation of motion can be written as

$$i_{\widetilde{\mathfrak{d}}H} \rho_H = 0, \tag{2.56}$$

where for extended phase space, we have defined the *suspension* of $\mathfrak{d}H \in \operatorname{Vect}(\mathcal{M})$ as $\widetilde{\mathfrak{d}}H = \frac{\partial}{\partial t} + \pi_{H*}\mathfrak{d}H$.

Generalization to time-dependent systems

Given the Liouville 1-form η such that $d\eta = \rho$, we can define the a 1-form $\eta_H = \eta - H dt$ on M_H such that $\rho_H = d\eta_H$. This provides the most straightforward route to the generalization to time-dependent cases: given $M_H = \mathbb{R} \times \mathcal{M}$ and a time-dependent Liouville 1-form η (not necessarily globally defined), we can define the corresponding generalized Liouville form as $\eta_H = \eta - H dt$. The generalized symplectic form

is then simply

$$\rho_H = d\eta_H = d\eta - dH \wedge dt = \rho - \left(dH - \frac{\partial \eta}{\partial t}\right) \wedge dt, \tag{2.57}$$

where $\rho(t)$, containing contains $d\zeta^i$, would be the symplectic form on the t-slice of M_H. The equation of motion is simply given by eq. (2.56), with the suspension again defined as earlier, except that $\mathfrak{d}H$ is now time-dependent, owing to the time-dependence of H.

Thus, the extended phase space formalism expresses time-dependent Hamiltonian dynamics compactly as $i_{\widetilde{\xi}_H}\rho_H = 0$. In order to unpack this result, given coordinates $(t, \boldsymbol{\zeta})$ on M_H, we 'define' the Hamiltonian vector field as $\widetilde{\xi}_H = \frac{\partial}{\partial t} + \dot{\zeta}^i \frac{\partial}{\partial \zeta^i}$. Substituting in the equation of motion, we get

$$\begin{aligned}
0 &= i_{\widetilde{\xi}_H}\left[\rho - \left(dH - \frac{\partial \eta}{\partial t}\right) \wedge dt\right] \\
&= i_{\widetilde{\xi}_H}\left[\frac{1}{2}\rho_{ij} d\zeta^i \wedge d\zeta^j - \left(\frac{\partial H}{\partial \zeta^i} - \frac{\partial \eta_i}{\partial t}\right) d\zeta^i \wedge dt\right] \\
&= -\rho_{ij}\dot{\zeta}^j\, d\zeta^i - \left(\frac{\partial H}{\partial \zeta^i} - \frac{\partial \eta_i}{\partial t}\right)\left(\dot{\zeta}^i dt - d\zeta^i\right) \\
&= \left(-\rho_{ij}\dot{\zeta}^j + \frac{\partial H}{\partial \zeta^i} - \frac{\partial \eta_i}{\partial t}\right) d\zeta^i - \dot{\zeta}^i \left(\frac{\partial H}{\partial \zeta^i} - \frac{\partial \eta_i}{\partial t}\right) dt
\end{aligned} \tag{2.58}$$

The coefficients of $d\zeta^i$ are the equations of motion

$$\rho_{ij}\dot{\zeta}^j = \frac{\partial H}{\partial \zeta^i} - \frac{\partial \eta_i}{\partial t} \implies \dot{\zeta}^i = \left(\boldsymbol{\rho}^{-1}\right)^{ij}\left(\frac{\partial H}{\partial \zeta^i} - \frac{\partial \eta_i}{\partial t}\right), \tag{2.59}$$

which is the generalization of eq. (2.18) to time-dependent H and η. The coefficient of dt does not give us anything new, since

$$\dot{\zeta}^i \left(\frac{\partial H}{\partial \zeta^i} - \frac{\partial \eta_i}{\partial t}\right) = \dot{\zeta}^i \rho_{ij} \dot{\zeta}^j = 0. \tag{2.60}$$

Finally, the contact manifolds also admit "canonical" coordinates as a generalization[27] of Darboux's theorem, which states that any contact structure (M_H, ρ_H) is locally isomorphic to $(\mathbb{R}^{2m+1}, \rho_0)$ with coordinates $(T, X^1, \ldots X^m, P_1, \ldots P_m)$, so that $\rho_0 = d\eta_0$, where $\eta_0 = P_i \wedge dX^i + dT$.

Action principle

Following the symplectic case, the extended phase space formalism can be recast as an action principle; the action being simply a line integral of η_H. Given coordinates $(t, \boldsymbol{\zeta})$, this becomes

$$S = \int_{\mathcal{C}} \eta_H = \int_{\mathcal{C}} \left[\eta_i d\zeta^i - H dt\right], \tag{2.61}$$

where $\mathcal{C}$ is now a curve on M_H, parametrized by τ such that $\dot{t} = 1$. Extremizing the action,

$$\begin{aligned}
0 &= \delta \int_{t_0}^{t_1} \left(\eta_i(\boldsymbol{\zeta}) \dot{\zeta}^i - H \right) d\tau \\
&= \int_{t_0}^{t_1} \left[\left(\frac{\partial \eta_i}{\partial t} \delta t + \frac{\partial \eta_i}{\partial \zeta^j} \delta \zeta^j \right) \dot{\zeta}^i + \eta_i(\boldsymbol{\zeta}) \delta \dot{\zeta}^i - \left(\frac{\partial H}{\partial \zeta^i} \delta \zeta^i + \frac{\partial H}{\partial t} \delta t \right) \right] d\tau \\
&= \int_{t_0}^{t_1} \left\{ \dot{\zeta}^i \left[\frac{\partial \eta_j}{\partial \zeta^i} \dot{\zeta}^j - \frac{d\eta_i}{d\tau} - \frac{\partial H}{\partial \zeta^i} \delta \zeta^i \right] + \delta t \left[\frac{\partial \eta_i}{\partial t} - \frac{\partial H}{\partial t} \right] \right\} d\tau \\
&= \int_{t_0}^{t_1} \left\{ \delta \zeta^i \left[\left(\frac{\partial \eta_j}{\partial \zeta^i} - \frac{d\eta_i}{d\zeta^j} \right) \dot{\zeta}^j - \frac{\partial H}{\partial \zeta^i} - \frac{\partial \eta_i}{\partial t} \right] + \delta t \left[\frac{\partial \eta_i}{\partial t} - \frac{\partial H}{\partial t} \right] \right\} d\tau,
\end{aligned} \tag{2.62}$$

which needs to be true for any variation $(\delta t(\tau), \delta \boldsymbol{\zeta}(\tau))$. We recover the equations of motion, as expected.

Example: Charged particle in an electromagnetic field

In this section, we illustrate some of the formal ideas developed in this section by applying them to derive the classical dynamics (Lorentz force) for a charged point particle in an external electromagnetic field. The underlying configuration space is $\mathbb{R}^d$, so that the phase space becomes $\mathcal{M} = T^*\mathbb{R}^d \cong \mathbb{R}^{2d}$, and the extended phase space is $M_H = \mathbb{R} \times \mathbb{R}^{2d}$.

In classical mechanics, the interaction with the electromagnetic field is encoded by minimally coupling the electromagnetic potential as $p_i \to p_i - qA_i$ and $\varepsilon \to \varepsilon - qA_0$, where $\varepsilon = |\mathbf{p}|^2 / 2m$ is the kinetic energy. Then, we can write the presymplectic form as

$$\eta_H = (p_i + qA_i) dx^i - \left(\frac{|\mathbf{p}|^2}{2m} - qA_0 \right) dt = p_i dx^i + \frac{|\mathbf{p}|^2}{2m} dt + qA, \tag{2.63}$$

where we have defined the 1-form $A = A_\mu dx^\mu$. Clearly, A, and hence η_H, may be time-dependent, so that we need the extended phase space formalism. The generalized symplectic form becomes

$$\eta = dp_i \wedge dx^i - \frac{p^i}{m} dp_i \wedge dt + qF, \tag{2.64}$$

where $F = dA = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu$ is the Maxwell curvature 2-form. Clearly, ρ_H is not of the Darboux form, so that $(t, \mathbf{x}, \mathbf{p})$ are not the canonical coordinates on M_H.

We can compute the equations of motion as

$$i_{\widetilde{\xi}} \rho_H = 0, \qquad \widetilde{\xi} = \dot{x}^\mu \frac{\partial}{\partial x^\mu} + \dot{p}_i \frac{\partial}{\partial p_i}, \tag{2.65}$$

where the trajectories are parametrized by τ, $\dot{f} = \frac{df}{d\tau}$ and $\dot{t} = \dot{x}^0 = 1$. Substituting eq. (2.64), this becomes

$$\begin{aligned} 0 &= \dot{p}_i dx^i - \dot{x}^i dp_i - \frac{p^i}{m}\left(\dot{p}_i dt - dp_i\right) + qF_{\mu\nu}\dot{x}^\mu dx^\nu \\ &= \left(\dot{p}_i - qF_{i0} - qF_{ij}\dot{x}^j\right) dx^i + \left(-\dot{x}^i + \frac{p^i}{m}\right) dp_i + \left(-\frac{p_i\dot{p}_i}{m} + qF_{i0}\dot{x}^i\right) dt \end{aligned} \tag{2.66}$$

From the coefficients of dx's and dp's, the equation of motion becomes

$$\dot{x}^i = \frac{p^i}{m}, \qquad \dot{p}_i = q\left(F_{i0} + F_{ij}\dot{x}^j\right). \tag{2.67}$$

For $d = 3$, we define the electric and magnetic field 3-vectors as $F_{i0} = E_i$ and $F_{ij} = \epsilon_{ijk}B^k$. The equations of motion become

$$\dot{\mathbf{x}} = \frac{\mathbf{p}}{m}, \qquad \dot{\mathbf{p}} = q\left(\mathbf{E} + \dot{\mathbf{x}} \times \mathbf{B}\right), \tag{2.68}$$

which simply describes the dynamics of a charged massive point particle in presence of a Lorentz force[69].

To illustrate the noncanonical nature of the coordinates further, we consider a simplified case, *viz*, the dynamics of a charged particle in a time-independent magnetic field $\mathbf{B} = B\mathbf{e}_z$, which can be described by the conventional symplectic formulation, as described in Sec 2.2. Thus, $\mathcal{M} = \mathbb{R}^{2d}$, equipped with the symplectic form

$$\begin{aligned} \rho &= dp_i \wedge dx^i + \frac{1}{2}qF_{ij}dx^i \wedge dx^j \\ &= dp_x \wedge dx + dp_y \wedge dy + dp_z \wedge dz + qB\, dx \wedge dy. \end{aligned} \tag{2.69}$$

We now compute the Poisson bracket $\{x, y\}$, by first computing the corresponding Hamiltonian vector fields using $i_{\xi_f}\rho = -df$ for $f = x, y$, as

$$\xi_x = \frac{\partial}{\partial p_x} - \frac{1}{qB}\frac{\partial}{\partial y}, \qquad \xi_y = \frac{\partial}{\partial p_y} + \frac{1}{qB}\frac{\partial}{\partial x}. \tag{2.70}$$

Thus,

$$\{x, y\} = \rho\left(\xi_x, \xi_y\right) = \frac{1}{qB}, \tag{2.71}$$

so that the coordinates x, y fail to satisfy the usual Poisson bracket $\{x, y\} = 0$ in presence of a magnetic field normal to the x-y plane. Physically, this is the setup for the Hall effect, so that we have simply obtained a classical analogue of the noncanonical commutation relation $[\hat{x}, \hat{p}] = -\frac{i\hbar}{qB}$ associated with the lowest Landau level problem in quantum Hall effect[74], which is a physical realization of noncommutative geometry.

2.5 Kinetic theory

Kinetic theory is a framework to derive an effective description of the classical mechanics of a system containing a large ($\sim 10^{23}$) number of point particles, each following the classical equations of motion and interacting with each other only weakly and at short ranges. The weak interaction implies that the motion of the individual particles is *uncorrelated*, except during a *collision*, i.e, when they are close enough to interact[8]. The formalism was originally developed by Maxwell and Boltzmann to describe gases at low pressures and densities, where the molecules, being neutral, do not exert any forces on each other at large distances, while at small distances, they interact via van der Walls forces[75].

In kinetic theory, instead of studying the motion of each individual particles, one studies the time-evolution of a coarse-grained density of particles in the phase space. This can further be used to derive macroscopic quantities such as currents, which are presumably accessible at macroscopic scales. For instance, in case of gases, one is interested in deriving the macroscopic pressures and volumes, while for other systems, one might be interested in various currents.

In this section, we formulate the conventional aspects of kinetic theory, *viz*, phase space volume, Liouville's theorem, Boltzmann equation and the derivation of macroscopic currents using the geometric formulation of classical mechanics discussed so far in this chapter. This formulation shall form the basis of our computations in Ch 5.

2.5.1 Phase space volume and Liouville's theorem

In order to define a 'density' on the phase space, one needs a notion of 'volume', that is, in some sense, invariant under Hamiltonian evolutions. For instance, in 1-dimensional classical mechanics, one defines a measure on the phase space $\mathbb{R}^2$ simply as

$$\Omega = \frac{dp \wedge dx}{h} = \frac{dp \wedge dx}{2\pi}, \tag{2.72}$$

where we have normalized[9] the conventional area form on $\mathbb{R}^2$ by $h = 2\pi\hbar,\ \hbar = 1$ to make it dimensionless.

[8] More precisely, we have two time-scales in the system, *viz*, the collision time, t_{coll} and the relaxation time, t_{rel}. Kinetic theory relies on the assumption that $t_{\text{coll}} \ll t_{\text{rel}}$.

[9] This normalization can be motivated by the *correspondence principle* between classical and quantum mechanics, by demanding that the quantum and classical counting of microstates give the same result. For instance, consider the harmonic oscillator $H = \frac{1}{2}\left(p^2 + x^2\right)$, and the quantum mechanical spectrum is given by $\varepsilon_n = \left(n + \frac{1}{2}\right)\hbar$. Let a phase space volume measure be defined as $\Omega = \mathcal{N} dp\, dx$, $\mathcal{N} \in \mathbb{R}$, and count the number of states with $\varepsilon \leq \varepsilon_0 \gg \hbar$:

$$N_{\text{classical}} = \mathcal{N} \int_{\mathcal{S}} dp\, dx = 2\pi\mathcal{N}\varepsilon_0, \qquad \mathcal{S} = \{(x,p) \in \mathbb{R}^2 \,|\, x^2 + p^2 < 2\varepsilon_0\},$$

and $N_{\text{quantum}} = \left\lfloor \frac{\varepsilon_0}{\hbar} - \frac{1}{2} \right\rfloor \approx \frac{\varepsilon_0}{\hbar}$. Demanding $N_{\text{classical}} = N_{\text{quantum}}$, we get $\mathcal{N} = (2\pi\hbar)^{-1} = h^{-1}$.

The symplectic manifolds are naturally equipped with a top form ρ^m, which can be used to define a normalized volume form as

$$\Omega = \frac{1}{(2\pi)^m m!}\rho^m \equiv \frac{1}{(2\pi)^m n!}\underbrace{\rho \wedge \cdots \wedge \rho}_{n \text{ times}}, \tag{2.73}$$

We are interested in the variation of this volume form under the flow corresponding to an arbitrary vector field $\xi \in \text{Vect}(\mathcal{M})$. Taking a cue from eq. (2.23), we compute

$$\mathcal{L}_\xi \Omega = (d\, i_\xi + i_\xi\, d)\,\frac{\rho^m}{(2\pi)^m n!} = \frac{1}{(2\pi)^m (m-1)!} d\left(i_\xi \rho \wedge \rho^{m-1}\right), \tag{2.74}$$

where $d\Omega = 0$, since Ω is a top form. But if ξ is a Hamiltonian vector field, i.e, $\xi = \mathfrak{d}H$ for some Hamiltonian H, then we can use the equations of motion to get

$$\mathcal{L}_{\mathfrak{d}H}\Omega = -\frac{1}{(2\pi)^m (n-1)!} d\left(dH \wedge \rho^{m-1}\right) = \frac{1}{(2\pi)^m (n-2)!}\, dH \wedge d\rho \wedge \rho^{n-2} = 0, \tag{2.75}$$

since ρ is closed. Thus, under a Hamiltonian evolution, the phase space volume is conserved. This is simply the statement of Liouville's theorem, which physically implies that under a Hamiltonian evolution, the volume of a region in phase space remains unchanged, while its shape may change[10]. Thus, Liouville's theorem is a direct consequence of the fact that Hamiltonian flows are symplectomorphisms.

Explicitly, given coordinates $\boldsymbol{\zeta}$ on $\mathcal{M}$, the volume form can be explicitly computed as

$$\Omega = \frac{1}{(2\pi)^m m!}\left(\frac{1}{2}\rho_{ij} d\zeta^i \wedge d\zeta^j\right)^m = \frac{1}{(2\pi)^m}\left[\frac{1}{2^m m!}\sum_{\sigma \in \mathcal{S}_{2n}} \text{sgn}\,(\sigma) \prod_{j=1}^{m} \boldsymbol{\rho}_{\sigma(2j-1),\sigma(2j)}\right] \bigwedge_{i=1}^{2m} d\zeta^i, \tag{2.76}$$

where we have used the fact that each nonzero term in the expansion will contain all $d\zeta^i$'s exactly once, the volume form being a top form, and the sign of the term depends on the number of exchanges to get the product of differential forms in the correct order, i.e, to the form $d\zeta^1 \wedge \cdots \wedge d\zeta^{2m}$. The sum over terms in the square brackets is the definition of the Pfaffian[11] of ρ, which we shall, with slight abuse of notation, denote as $\sqrt{\rho} = \sqrt{\det(\boldsymbol{\rho})}$. Thus, concisely, $(2\pi)^m \Omega = \sqrt{\rho}\, d^{2m}\zeta$ is the invariant (under symplectomorphisms) volume measure on $\mathcal{M}$. Using $d\Omega = 0$ and $i_{\mathfrak{d}H} d\zeta^j = \dot{\zeta}^j$, we get Liouville's theorem can be written using the

[10] There are further restrictions on how drastically the 'shape' can change, known as *Gromov's non-squeezing theorem*. For details, see Ref [71].

[11] The determinant of an even dimensional skew symmetric matrix can always be written as squares of a polynomial in its entries with integer coefficients. The Pfaffian is then defined as the square root of the determinant of the matrix. The determinant of odd dimensional skew-symmetric matrices always vanishes.

coordinates in its conventional form as

$$0 = \star\mathcal{L}_{\mathfrak{d}H}\Omega = \star d\, i_{\mathfrak{d}H}\left(\sqrt{\rho}\bigwedge_{i=1}^{2m} d\zeta^i\right) = \star d\left[\sqrt{\rho}\sum_{j=1}^{2m}(-1)^j\dot{\zeta}^j\bigwedge_{i\neq j} d\zeta^i\right] = \sum_{j=1}^{2m}\frac{\partial}{\partial\zeta^j}\left(\sqrt{\rho}\dot{\zeta}^j\right), \tag{2.77}$$

where $\star$ denotes the Hodge dual[12] and the $(-1)^j$ arises from the fact that both i_ξ and d are derivations[13].

For the extended phase space, we can define a volume form $\Omega_H = \Omega \wedge dt$. Explicitly,

$$\Omega_H = \frac{1}{(2\pi)^m m!}\rho^m \wedge dt = \frac{1}{(2\pi)^m m!}\rho_H^m \wedge dt, \tag{2.78}$$

where in the second equality, we can replace ρ with $\rho_H = \rho - (dH - \partial\eta_H/\partial t)dt$ as on expanding, we can have only one dt which is the one outside ρ^m, hence $\rho^m dt = \rho_H^m dt$.

Again, we can derive

$$\begin{aligned}\mathcal{L}_{\tilde{\mathfrak{d}}H}\Omega_H =& \left(d\, i_{\tilde{\mathfrak{d}}H} + i_{\tilde{\mathfrak{d}}H}\, d\right)\frac{1}{(2\pi)^m m!}\rho_H^m \wedge dt \\ =& \frac{1}{(2\pi)^m m!}\, d\left(n\,(i_{\tilde{\mathfrak{d}}H}\rho_H)\wedge\rho_H^{m-1}\wedge dt + \rho_H^m \wedge (i_{\tilde{\mathfrak{d}}H}dt)\right) \\ =& \frac{1}{(2\pi)^m (m-1)!}\, d\rho_H \wedge \rho_H^{m-1} = 0.\end{aligned} \tag{2.79}$$

This is the analogue of the Liouville's theorem for extended phase space.

Again, given coordinates $(t, \boldsymbol{\zeta})$ on M_H, the volume form on M_H can be written as $\Omega_H = \sqrt{\rho}\, d^{2m+1}\zeta$, where $d^{2m+1}\zeta$ also includes dt and $\sqrt{\rho}$ may now depend on t. The Liouville theorem can be written using these coordinates as

$$0 = \star\mathcal{L}_{\tilde{\mathfrak{d}}H}\Omega_H = \frac{\partial}{\partial t}\sqrt{\rho} + \sum_{j=1}^{2m}\frac{\partial}{\partial\zeta^j}\left(\sqrt{\rho}\dot{\zeta}^j\right), \tag{2.80}$$

since $\dot{t} = 1$. This is the form of Liouville's theorem that we shall use in the rest of this paper. Another convenient result is

$$\star\,\mathcal{L}_{\tilde{\mathfrak{d}}H}(\varphi\,\Omega_H) = \frac{\partial}{\partial t}(\sqrt{\rho}\,\varphi) + \sum_{j=1}^{2m}\frac{\partial}{\partial\zeta^j}\left(\sqrt{\rho}\,\varphi\,\dot{\zeta}^j\right), \tag{2.81}$$

where $\varphi\colon M_H \to \mathbb{R}$. This denotes the time-evolution under Hamiltonian flow of a quantity φ associated with the phase space.

[12]Recall that the coordinates $(t, \boldsymbol{\zeta})$ locally map M_H to $\mathbb{R}^{1,2m}$, which has a Minkowski metric. We use this metric to define the Hodge duals.

[13] Given a p-form θ and q-form φ, a derivation $\mathcal{D}$ acts as $\mathcal{D}(\theta\wedge\varphi) = (\mathcal{D}\theta)\wedge\varphi + (-1)^p\theta\wedge(\mathcal{D}\varphi)$.

2.5.2 Boltzmann equation

Consider a set of N particles, governed by Hamiltonian dynamics on a phase space $\mathcal{M}$ with Hamiltonian H. At a given time t, the state of the system can then be completely described by a set of N phase space coordinates $\boldsymbol{\zeta}_i(t)$. We define a phase space density corresponding to this configuration as

$$\widetilde{f}(t,\boldsymbol{\zeta}) = \frac{1}{N}\sum_{i=1}^{N}\delta\left(\boldsymbol{\zeta}-\boldsymbol{\zeta}_i(t)\right), \tag{2.82}$$

where f denotes the 1-particle probability distribution and $\delta(\boldsymbol{\zeta})$ is the Dirac delta distribution in $\dim(\mathcal{M})$ dimensions. The probability of finding a particle in a given phase space volume $V \subseteq \mathcal{M}$ at time t is given by

$$P(t,V) = \int_V \widetilde{f}(t,\boldsymbol{\zeta})\Omega = \int_V d\boldsymbol{\zeta}\,\sqrt{\rho(\boldsymbol{\zeta})}\widetilde{f}(t,\boldsymbol{\zeta}) = \int_V d\mathbf{X}\,d\mathbf{P}\,f(t,\mathbf{X},\mathbf{P}), \tag{2.83}$$

where Ω is the volume measure and $(\mathbf{X},\mathbf{P})$ are local Darboux coordinates on $V \subseteq \mathcal{M}$, which may not always exist. Integrating over the entire $\mathcal{M}$, we simply get $1 = \int_{\mathcal{M}}\Omega\,f(t,\boldsymbol{\zeta})$ for any t.

As we are usually interested in the $N\to\infty$ limit, we can "smooth out" $\widetilde{f}$ to get a smooth 1-particle density function $f:\mathbb{R}\times\mathcal{M}\to\mathbb{R}$. Physically, this smoothing involves a process of "coarse-graining", i.e, we forgo our exact knowledge of the positions of particles in favor of an average number of particles in a volume element. Of course, we choose to look only at volume elements which are big enough to contain a statistically significant number of particles, but they must also be small enough to count as an "infinitesimal" volume element[14]. More precisely, we only demand that $\widetilde{f}$ and f agree on volumes much larger than the volume occupied by single particles. Formally, given a $\epsilon > 0$,

$$\forall\, V\subseteq\mathcal{M}, \int_V \Omega > \epsilon, \quad \exists\,\delta \text{ such that } \left|\int_V \left(f(t,\boldsymbol{\zeta})-\widetilde{f}(t,\boldsymbol{\zeta})\right)\Omega\right| < \delta. \tag{2.84}$$

In kinetic theory, instead of tracking the trajectories of single particles, one is only interested in the time-evolution of the phase space probability distribution. Since classical point particles can neither be created nor be destroyed, the number of particles in a given volume can change either due to a flux of particles or due to collisions. The former involves particles following Hamiltonian dynamics, which is encoded in the Liouville's theorem, while the latter is encoded in a *collision* integral $C[f]$, which can be computed using

[14] This somewhat strange condition is actually quite natural for physical systems; for instance, for a gas under environmental conditions with a particle density of $\sim 10^{24}\,\mathrm{m}^{-3}$, one can choose the "infinitesimal" volume element as $1\,\mu m^3$, which contains $\sim 10^6$ particles.

the 2-body interactions of the Hamiltonian. Thus, we arrive at the Boltzmann equation[15]

$$\frac{\partial}{\partial t}\left(\sqrt{\rho} f\right) + \frac{\partial}{\partial \zeta^i}\left(\sqrt{\rho}\, f\, \dot{\zeta}^i\right) = C[f], \tag{2.85}$$

where the probability of finding the particle in a given volume is the integral of $\sqrt{\rho}\, f$ over the volume. If the system is dilute, we can ignore the collisions and set $C[f] = 0$, which we shall do from now on. The collisionless Boltzmann equation can be written compactly using eq. (2.81) as

$$\mathcal{L}_{\widetilde{\mathfrak{d}}H}\left(f\, \Omega_H\right) = 0. \tag{2.86}$$

But since $\mathcal{L}_{\widetilde{\mathfrak{d}}H}\left(\Omega_H\right) = 0$ by Liouville's theorem, we can set $\mathcal{L}_{\widetilde{\mathfrak{d}}H} f = 0$. Physically, this implies that f is simply advected with the Hamiltonian flow. In terms of coordinates on M_H, this gives us the conventional form of Boltzmann equation

$$0 = \star\mathcal{L}_{\widetilde{\mathfrak{d}}H} f = \frac{\partial f}{\partial t} + \dot{\zeta}^i \frac{\partial f}{\partial \zeta^i}. \tag{2.87}$$

From now on, we shall assume that the phase space density satisfies this equation.

2.5.3 Currents

The 1-particle probability distribution, which satisfy the continuity equations, can be used to compute the averages of quantities of physical interest, viz, "densities" and "currents". Given any function $\mathcal{Q}\colon \mathcal{M} \to \mathbb{R}$, we can compute the corresponding "density" J_0 and current J_i over $V \subseteq \mathcal{M}$ as

$$J^0 = \int_V \mathcal{Q}(t, \boldsymbol{\zeta})\, \Omega_H, \qquad J^i = \int_V \dot{\zeta}^i\, \mathcal{Q}(t, \boldsymbol{\zeta})\, \Omega_H. \tag{2.88}$$

A special case of interest is when the spacetime is a submanifold of M_H, so that one can integrate over the remaining phase space variables to obtain the currents as a function of spacetime. Thus, set $M_H = \mathbb{R}^{n,1} \times \mathcal{M}'$, with $\mathbb{R}^{n,1}$ parametrized by t and $\zeta^i,\ 1 \leq i \leq n$, while $\mathcal{M}'$ is parametrized by the remaining $\zeta^i,\ n{+}1 \leq i \leq 2m$. Then we define the current associated with $\mathcal{Q}$, which is independent of the spacetime coordinate, as

$$J^0 = \int_{\mathcal{M}'} \left(\mathcal{Q}\sqrt{\rho}\right) \bigwedge_{i=n+1}^{2m} d\zeta^i, \quad J^i = \int_{\mathcal{M}'} \left(\mathcal{Q}\, \dot{\zeta}^i\right) \bigwedge_{i=n+1}^{2m} d\zeta^i. \tag{2.89}$$

A few relevant examples are the number current($\mathcal{Q} = f$), the charge current($\mathcal{Q} = q\, f$, q being the charge of a particle) and the energy-current($\mathcal{Q} = \varepsilon f$, which typically depends on the momentum).

[15] For a more careful derivation using the BBKGY hierarchy, see Ref [75].

The spacetime gradient of the current defined in eq. (2.89) is

$$\begin{aligned}\partial_\mu J^\mu &= \int_{\mathcal{M}'} \mathcal{Q}(t,\boldsymbol{\zeta}) \left[\frac{\partial}{\partial t}\sqrt{\rho} + \sum_{i=1}^{n} \frac{\partial}{\partial \zeta^i}\left(\sqrt{\rho}\,\dot{\zeta}^i\right)\right] \bigwedge_{i=n+1}^{2m} d\zeta^i \\ &= \int_{\mathcal{M}'} \mathcal{Q}(t,\boldsymbol{\zeta}) \left[\star\mathcal{L}_{\widetilde{\mathfrak{d}}H}\Omega_H - \sum_{i=n+1}^{2m} \frac{\partial}{\partial \zeta^i}\left(\sqrt{\rho}\,\dot{\zeta}^i\right)\right] \bigwedge_{i=n+1}^{2m} d\zeta^i \end{aligned} \tag{2.90}$$

where we have used eq. (2.81). Integrating the second term on the RHS by parts and taking the Hodge dual of both sides, we get

$$d\star J - \int_{\mathcal{M}'} \left(\sum_{i=n+1}^{2m} \dot{\zeta}^i \frac{\partial \mathcal{Q}}{\partial \zeta^i}\right) \Omega_H = \int_{\mathcal{M}'} \mathcal{Q}\,\mathcal{L}_{\widetilde{\mathfrak{d}}H}\Omega_H. \tag{2.91}$$

In certain cases, this would turn out to be a covariant conservation law, as shown in Sec 5.3. For particle number current, $\mathcal{Q} = f$, and the conservation law simplifies to

$$d\star J = \int_{\mathcal{M}'} f\,\mathcal{L}_{\widetilde{\mathfrak{d}}H}\Omega_H. \tag{2.92}$$

Using Liouville's theorem $\mathcal{L}_{\widetilde{\mathfrak{d}}H}(f\,\Omega_H) = 0$, we deduce that J is conserved. In cases where Liouville's theorem breaks down, (See Ch 5) we can use eq. (2.91) to deduce the corresponding (non-)conservation laws.

3 Gauge anomalies

Gauge anomalies imply a breakdown of the classical conservation laws. One particular example, anomalies occur when a classical field theory of Weyl spinor fields coupled to nondynamical gauge/gravitational fields, when the associated classical conservation law breaks down when the theory is quantized.

In this chapter, we shall attempt only a brief exposition of gauge theories and anomalies, to provide some context to what we are trying to calculate in the rest of the thesis. A proper description of their origins would demand forays into quantum field theory that are, for the most part, irrelevant to this thesis, and so we cite the standard references for them. We shall spend considerably more time setting up anomalous hydrodynamics, for which some of the relevant quantities are computed using the semiclassical techniques in Ch 5.

3.1 Gauge theories

We begin with a brief description of classical field theories with a gauge symmetry. As classical field theory is not particularly relevant to this dissertation, we shall primarily be interested in a description of the gauge transformation and their geometric aspects.

Crudely put, 'gauge' refers to a redundancy in the description of the physical system. More precisely, given a theory on a smooth manifold X with a degree of freedom in some other manifold T, a field is a map $\phi\colon X \mapsto T$, and a 'gauge' can then be thought of as[1] a choice of coordinates on T that depends on the coordinates in X. A *gauge transformation* is then a continuous local coordinate transformation of T. These transformations usually form a Lie group, which is referred to as the *gauge group* of the theory. Finally, since the gauge freedom is unphysical, we demand that all physical theories be *gauge invariant*, which often provides strong constraints on the theory.

In the first subsection, we shall take this slightly unconventional approach to introducing gauge, talking only about gauge fields. Only in the next subsection shall we describe the usual origin of gauge symmetry in classical field theory, which usually comes about when one "promote a global symmetry to a local one".

[1] A particularly clear and insightful introduction to this perspective is an article by Terence Tao[76].

3.1.1 Fiber bundles

Fiber bundles are the most natural setup to study gauge theories. Formally, they are a trio of topological spaces $(\mathscr{F}, L, M)$ and a continuous surjective map $\pi\colon \mathscr{F} \to M$ such that $\forall\, x \in M$, the preimage $\pi^{-1}(x) \subset M$ is homeomorphic to $\mathscr{F}$. M is termed the *base space*, L the *total space*, π the *projection* and $\mathscr{F}$ the *fiber* of the fiber bundle. A fundamental property of fiber bundles is *local trivialization*, i.e, on open sets of M, the fiber bundle can be written as the product space $M \times \mathscr{F}$ with the projection simply projecting to the first component.

Two fiber bundles shall be relevant for our purposes: the *principle bundle* and the *associated bundle* on a smooth manifold M for a given group G. The former is simply a bundle $P \xrightarrow{\pi} M$ whose base space is M and whose fibers are copies of G (or, more precisely, *torsors* of G). Its sections are the space of all possible *gauges*. Given V, a vector space with a G-action, the associated bundle is formally defined as $P \times_G V$. Its sections are the wavefunctions of the problem, with the G-action on V describing the gauge transformations. The relevant features of fiber bundles used in the following are the related notions of *connections*, *covariant derivatives* and *curvatures*. In the following, we shall simply quote the needed results; for details we refer to Sec 16.3.2 of Ref [70].

The connection on a fiber bundle can be defined using the notion of a *parallel transport*. For instance, for the principle bundle, given a curve $\mathbf{x}(t)$ on the base space M and a point $(\mathbf{x}(0), g) \in \pi^{-1}(\mathbf{x}(0))$, the problem is to define a curve on L such that the tangent vector stays *horizontal*. Given a local trivialization around $x(0)$, the lift can then be written as $(\mathbf{x}(t), g(t))$, where given $\mathbf{x}(t)$, we seek to derive $g(t)$. One gets a differential equation of the form

$$\frac{\partial g}{\partial t} + \frac{\partial x^i}{\partial t}\, \mathscr{A}_i(\mathbf{x}) g = 0, \tag{3.1}$$

where $\mathscr{A}_i(\mathbf{x})$ is the connection on P, and the vector field along x is then called the *covariant derivative* on the principle bundle P.

The connection on the principle bundle induces a connection A on the associated bundle, which is of more interest to us. It can be used to define the covariant derivative $\mathcal{D} \equiv d - iA$, which acts on the wavefunctions. The curvature of the bundle is then defined as

$$F_{ij} \equiv i[\mathcal{D}_i, \mathcal{D}_j]. \tag{3.2}$$

The connection and curvature, being derivatives on G, are valued in the Lie algebra. The curvature 2-form encodes the local triviality of the fiber bundle. More precisely, a bundle is trivial iff the curvature vanishes.

3.1.2 Maxwell and Yang-Mills

Arguably, the earliest classical field theory with a gauge symmetry was the Maxwell's theory of electromagnetism, described in $d+1$ dimensions by the action

$$S[F] = \int_{\mathbb{R}^{d,1}} F \wedge \star F = \frac{1}{4} \int_{\mathbb{R}^{d,1}} d^{d+1}x F_{\mu\nu} F^{\mu\nu}, \tag{3.3}$$

where $F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu$ is the Maxwell's tensor, in whose dynamics the physical content of the theory is encoded. The corresponding equation of motion, *viz*, the Maxwell equations, can be written compactly using differential forms as

$$dF = 0, \qquad d \star F = 0. \tag{3.4}$$

where J is a current 1-form, which can be thought of as an external source. In $3+1$ dimensions, the components of F are the electric and magnetic fields, defined as $E_i = F_{0i}$ and $B_i = \frac{1}{2}\epsilon_{ijk} F^{jk}$. The conventional Maxwell's equations are written in terms of these fields, so that the physics of the system can be entirely described in terms of these quantities.

In order to couple this to other fields, one uses Poincaré's lemma to locally define a 1-form $A = A_\mu dx^\mu$ such that $F = dA$. Clearly, this choice is not unique, since $A \mapsto A + d\alpha$ leaves F invariant for any smooth function α. Thus, if we think of A_μ's as the degrees of freedom of the system, the description is redundant, since A and $A + d\alpha$ describe precisely the same physical state. The redundancy is the additional *gauge* freedom, under which the physical dynamics of the system must remain invariant. For instance, in presence of external charge current described by a 1-form $J = J_\mu dx^\mu$, where J_0 is the charge density and J_μ the charge current, the action becomes

$$S[A, J] = \int_{\mathbb{R}^{d,1}} (F \wedge \star F - A \wedge \star J) = \int_{\mathbb{R}^{d,1}} d^{d+1}x \left(\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J_\mu A^\mu \right), \tag{3.5}$$

the equations of motion become

$$dF = 0, \qquad d \star F = \star J. \tag{3.6}$$

Furthermore, setting $A \to A + d\alpha$ and using the fact that F is invariant under this transformations, we get

$$S[A + d\alpha, J] - S[A, J] = -\int_{\mathbb{R}^{d,1}} d\alpha \wedge \star J = \int_{\mathbb{R}^{d,1}} \alpha d \star J, \tag{3.7}$$

which must hold true for any $\alpha\colon \mathbb{R}^{d,1} \to \mathbb{R}$. Thus, the currents must satisfy

$$d \star J = 0 \iff \partial^\mu J_\mu = 0. \tag{3.8}$$

This is the conservation law associated with electromagnetism, which we simply interpret as a continuity equation, i.e, a local conservation of charge.

We seek to interpret the Maxwell connection A as the connection on an associated G-bundle, on which the connection should be $\mathfrak{g}$-valued. But since the Maxwell connection is real valued, we seek a group G for which $\mathfrak{g} \cong \mathbb{R}$. The commonly used group is $G = \mathrm{U}(1)$, as it is compact[2]. Thus, a natural generalization of electromagnetism would be to consider arbitrary Lie group G, which we do next. Particular examples are the classical Yang-Mills theories, where one takes $G = \mathrm{SU}(n)$, the Salam-Weinberg electroweak theory with $G = \mathrm{SU}(2) \times \mathrm{U}(1)$ and the standard model of particle physics, with $G = \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$.

Thus, we now consider the gauge connection A valued in $\mathfrak{g}$, for which the covariant derivative becomes $\mathcal{D} = d - iA$, so that using eq. (3.2), the curvature becomes

$$F = i[d - iA, d - iA] = dA - iA \wedge A. \tag{3.9}$$

Under a nonabelian gauge transformation,

$$A \to \widetilde{A} = gAg^{-1} - i\, dg\, g^{-1}, \qquad F \to \widetilde{F} = gFg^{-1}, \tag{3.10}$$

where the latter can be explicitly derived as

$$\begin{aligned}
\widetilde{F} &= d\widetilde{A} - i\widetilde{A} \wedge \widetilde{A} \\
&= \left(dg \wedge Ag^{-1} + g\, dAg^{-1} + gAg^{-1} \wedge dg\, g^{-1} - i\, dg\, g^{-1} \wedge dg\, g^{-1}\right) \\
&\quad - i\left(gAg^{-1} \wedge gAg^{-1} - i\, gAg^{-1} \wedge dg\, g^{-1} - i\, dg\, g^{-1} \wedge gAg^{-1} - dg\, g^{-1} \wedge dg\, g^{-1}\right) \\
&= gFg^{-1}
\end{aligned} \tag{3.11}$$

Given the Lie algebra $\mathfrak{g}$ of G, for any $g \in G$ which lies in the identity component of G, it can be written as e^{iX} for some $X \in \mathfrak{g}$. Choose a basis of the Lie algebra as λ_a, $a = 1, \ldots \dim(\mathfrak{g})$, whose commutators are given by $[\lambda_a, \lambda_b] = i\mathfrak{f}_{ab}^{\ c}\lambda_c$, where $\mathfrak{f}_{ab}^{\ c}$ are a set of real *structure constants* of the Lie algebra. The gauge connection and curvature are differential forms valued in the Lie algebra, so that we can expand them, given the bases

[2] One could also choose $G = \mathbb{R}$ under addition, a a setup sometimes known as *noncompact electrodynamics*.

of the underlying space and the Lie algebra, as $A = A^a\lambda_a = A^a_\mu dx^\mu \lambda_a$. Then, explicitly, $F = dA - iA \wedge A$ can be written in the given coordinates as

$$\begin{aligned}\frac{1}{2}F^a_{\mu\nu}\lambda_a dx^\mu \wedge dx^\nu &= \Big(\partial_\mu A^a_\nu \lambda_a - iA^b_\mu A^c_\nu [\lambda_b, \lambda_c]\Big) dx^\mu \wedge dx^\nu \\ &= \frac{1}{2}\Big(\partial_\mu A^a_\nu \lambda_a - \partial_\nu A^a_\mu \lambda_a + A^b_\mu A^c_\nu \mathfrak{f}^a_{bc}\Big) dx^\mu \wedge dx^\nu, \end{aligned} \tag{3.12}$$

so that $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + \mathfrak{f}^a_{bc} A^b_\mu A^c_\nu$.

3.1.3 Coupling to matter fields

We now come to one of the primary *raison d'être* of gauge theories: promoting global symmetries to local ones[3]. Consider then arguably the simplest possible theory with a continuous global symmetry: a complex scalar field in $n+1$ spacetime dimensions. Mathematically, $\phi\colon \mathbb{R}^{n,1} \to \mathbb{C}$, whose dynamics is governed by the action functional

$$S[\phi] = \int_{\mathbb{R}^{n,1}} d^{n+1}x \left[(\partial_\mu \phi^*)(\partial^\mu \phi) + V(|\phi|^2)\right], \tag{3.13}$$

where $V : \mathbb{R} \to \mathbb{R}$ is a smooth function. There is a natural action of U(1) on the field as $\phi(x) \mapsto e^{i\theta}\phi(x)$, $\theta \in S^1$, under which the action is manifestly invariant. To 'gauge' this symmetry is to demand that the action be invariant under $\phi(x) \mapsto e^{i\theta(x)}\phi(x)$, where $\theta\colon \mathbb{R}^{n,1} \to S^1$. Clearly, the action is not invariant under such a transformation, as $\partial_\mu\phi$ will yield a term where the derivative acts on θ. The way around is to introduce a *covariant derivative*, defined as $\mathcal{D}_\mu \equiv \partial_\mu - iA_\mu$, such that under the U(1) transform, $A_\mu \mapsto A_\mu + \partial_\mu\theta$. Then, under a gauge transformation,

$$\mathcal{D}_\mu\phi \mapsto e^{i\theta(x)}\left(\partial_\mu\phi + i\phi\,\partial_\mu\theta\right) - i(A_\mu + \partial_\mu\theta)\phi = \mathcal{D}_\mu\phi. \tag{3.14}$$

Here, A is the gauge field. We can use covariant derivatives

$$S[\phi, A] = \int_{\mathbb{R}^{n,1}} d^{n+1}x \left[(\mathcal{D}_\mu \phi^*)(\mathcal{D}^\mu \phi) + V(|\phi|^2)\right], \tag{3.15}$$

which is invariant under a gauge (or a local U(1)) transformation. In physical terms, this action, after quantization, describes *scalar QED*.

A more useful field theory is one coupling a U(1) gauge field to spinor fields, which, when quantized, can be used to describe fermions (for instance, electrons) interacting electromagnetically, in which case, the

[3] The question of a reasonable *apriori* motivation to do this (barring *post hoc* allusions to their usefulness) is something that has always eluded me.

gauge field is interpreted as a photon. Explicitly, the QED action can be written as

$$S[\psi, A] = \int_{\mathbb{R}^{n,1}} d^{n+1}x \left[\overline{\psi}(i\gamma^\mu \mathcal{D}_\mu - m)\psi\right], \qquad \overline{\psi} = \psi^\dagger \gamma^0, \tag{3.16}$$

where ψ is a spinor, i.e, it transforms under the spinor representation of the Lorentz group $SO(n,1)$, and γ^μ are the Dirac matrices, which satisfy $[\gamma^\mu, \gamma^\nu]_+ = \eta_{\mu\nu}$, and thus form a representation of the Clifford algebra.

We now move on to gauge symmetries more complicated than U(1). Consider then a Lie group G with a q-dimensional complex unitary representation R, i.e, to each $g \in G$ we associate $\mathfrak{R}_g \colon \mathbb{C}^q \to \mathbb{C}^q$. Then, we can define an action invariant under this transformation as

$$S[\psi, A] = \int_{\mathbb{R}^{n,1}} d^{n+1}x \left[\overline{\psi}_\alpha (i\gamma^\mu \partial_\mu - m)\psi_\alpha\right] = \int_{\mathbb{R}^{n,1}} d^{n+1}x \left[\bar{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi\right], \tag{3.17}$$

where Ψ is a q-component spinor, whose components transform as $\psi_\alpha \mapsto [\mathfrak{R}_g]_{\alpha\beta}\psi_\beta$, so that $\bar{\psi_\alpha} \mapsto [\mathfrak{R}_g^\dagger]_{\alpha\beta}\overline{\psi}_\beta$. Since $\mathfrak{R}_g$ is a unitary matrix, the action is invariant under this transformation. We again demand invariance under a local version, viz, under $\Psi \mapsto \mathfrak{R}_{g(x)}\Psi$, for which we need covariant derivatives $\mathcal{D}_\mu = \partial_\mu - iA_\mu$, where A_μ is a nonabelian gauge field, which transforms as

$$A_\mu \mapsto \mathfrak{R}_{g(x)} \left(A_\mu - i\,\mathfrak{R}_{g(x)}^{-1}\partial_\mu \mathfrak{R}_{g(x)}\right)\mathfrak{R}_{g(x)}^{-1}. \tag{3.18}$$

Thus, under a gauge transformation,

$$\mathcal{D}_\mu \Psi \mapsto \mathfrak{R}_{g(x)} \left[\partial_\mu \Psi + \left(\mathfrak{R}_{g(x)}^{-1}\partial_\mu \mathfrak{R}_{g(x)}\right)\Psi\right] - i\,\mathfrak{R}_{g(x)} \left(A_\mu - i\,\mathfrak{R}_{g(x)}^{-1}\partial_\mu \mathfrak{R}_{g(x)}\right)\Psi = \mathfrak{R}_{g(x)}\,\mathcal{D}_\mu \Psi, \tag{3.19}$$

so that the covariant derivative indeed transforms *covariantly*!

3.2 Anomalies

In the last section, we saw that one can define classical field theories invariant under a gauge transformations. In quantum field theory, one seeks to *quantize* the theory. It was quite a surprise to have discovered that the symmetries of the classical field theory do not necessarily imply that the corresponding quantum field theory is also symmetric; this is known as an anomaly. Thus, generically, anomalies imply a breakdown of the conservation law of the form derived in eq. (3.8), so that one can schematically write

$$d \star J = \mathcal{A}[A, F], \tag{3.20}$$

where J is a conserved current in the classical theory, and $\mathcal{A}[A, F]$ is termed the *anomaly polynomial*.

One way to quantize a classical field theory is to use Feynman's path integrals. Given an action $S[\phi]$, where ϕ denotes the relevant dynamical field of the problem, one defines the *path integral*, which is physically the amplitude to go from an initial configuration $\phi_1(x)$ to a final configuration $\phi_2(x)$ as

$$\mathcal{Z} = \int [d\phi] e^{iS[\phi]}, \tag{3.21}$$

where $[d\phi]$ is a measure that represents the "*sums over all paths*". Precisely what this notion means is often not entirely clear, so that the integration measure is not in general well defined. However, when it is defined, a classical symmetry of the action, which leaves the action invariant, may not leave the measure invariant. This origin of the anomaly from the variation of the path integral measure is generically termed a *Fujikawa calculation*, after the original calculation of the U(1) anomaly using this method by Fujikawa[77].

An alternative is the canonical quantization, where one "upgrades" the classical fields to quantum *field operators*, which are required to follow certain (anti-)commutation relations. The physical current is then given by the expectation value of the *current operator* $\hat{J}_\mu$. However, the classical conservation law for J as derived in eq. (3.8), leads to the vanishing of gradients of certain expectation value, a result known as a *Ward identity*[9]. The anomaly manifests itself in the canonical formalism as an explicit violation of a Ward identity, as one computes the relevant Feynman diagrams.

In the rest of this section, we simply list the commonly encountered examples of anomalies in $3 + 1$ dimensions:

Adler-Bell-Jackiw(ABJ) anomaly corresponds to the violation of the conservation law for the axial current in QED.Explicitly, one considers the fermionic action

$$S[\psi, A] = \int_{\mathbb{R}^{3,1}} d^{n+1}x \left[\overline{\psi}(i\gamma^\mu \mathcal{D}_\mu - m)\psi\right], \tag{3.22}$$

which has the vector(V) and axial(A) currents explicitly defined as $J_\mu = i\overline{\psi}\gamma_\mu\psi$ and $J^5_\mu = i\overline{\psi}\gamma_\mu\gamma^5\psi$, respectively. The corresponding conservation laws in the quantum field theory (or more precisely, the Ward identities) can be computed as the V-V-A vertex. At one-loop level, they computed the so-called *triangle diagram*, to get

$$\partial^\mu J^5_\mu = \frac{1}{16\pi^2}\epsilon^{\mu\nu\rho\lambda}F_{\mu\nu}F_{\rho\lambda}, \tag{3.23}$$

which can be written more concisely as

$$d \star J^5 = \frac{1}{(2\pi)^2 2!} F \wedge F = \frac{1}{2!} \left(\frac{F}{2\pi} \right)^2 . \tag{3.24}$$

Nonabelian singlet anomaly is a direct generalization of the ABJ anomaly, when one replaces the gauge group U(1) with some more general (and usually nonabelian) gauge group. Thus, setting $A = A^a_\mu \lambda_a dx^\mu$, the singlet anomaly becomes

$$d \star J^5 = \frac{1}{(2\pi)^2 2!} \mathrm{tr} \left\{ F \wedge F \right\} = \frac{1}{2!} \mathrm{tr} \left\{ \left(\frac{F}{2\pi} \right)^2 \right\} . \tag{3.25}$$

Nonabelian gauge anomaly is a violation of a *covariant* conservation law for a gauge current. It was originally computed by Bardeen as

$$\mathcal{D} \star J_a = \frac{1}{6\pi^2} d\mathrm{tr} \left\{ \lambda_a \left(A \wedge dA + \frac{1}{2} A^3 \right) \right\} = \frac{1}{24\pi^2} \tag{3.26}$$

This is known as the *consistent* anomaly, since it satisfies the Wess-Zumino consistency conditions. However, there also the *covariant* anomaly, which is usually written as

$$\mathcal{D} \star J_a = \frac{1}{8\pi^2} \mathrm{tr} \left\{ \lambda_a F \wedge F \right\} . \tag{3.27}$$

Clearly, this is covariant because the anomaly polynomial consists entirely of the gauge-covariant curvature 2-forms. These two expressions are equivalent upto the addition of a boundary term, which can be thought of as an addition of a (Bardeen) *counterterm*[78] to the action. In the present case, the two different forms of the gauge anomaly are different by

$$\frac{1}{12\pi^2} d\,\mathrm{tr} \left\{ \lambda_a \left(A \wedge F + F \wedge A - A^3 \right) \right\} , \tag{3.28}$$

which is indeed a boundary term. If original manifold $\mathbb{R}^{3,1}$ is thought of as the boundary of a manifold in one higher dimension, then the Bardeen counterterm can also be interpreted as an *anomaly inflow*[79] from the bulk, as described by the *Callan-Harvey mechanism*[80].

In higher dimensions, we have similar expressions for the anomaly polynomials, all of which turn out to be the Chern characters. Thus, they can all be computed using the generating function $e^{tF/2\pi}$, which one expands as a formal power series in F and picks out the coefficient of t^{2N} in $2N$ spacetime dimensions.

For manifolds which support a spin structure, integrating the anomaly polynomial gives the Chern numbers associated with the gauge curvature, which can also be thought of as an *instanton number*. Furthermore, the anomaly polynomials are also related to the index of the Dirac operator via the *Atiyah-Singer index theorem*[6, 73]. Thus, anomalies can be thought of as an *index density*, which reveals their topological nature.

Coupling the fermionic theory to a nondynamical curved spacetime background (i.e, gravity in the GR sense), we get the mixed gauge-gravitational anomaly. We shall not write out the polynomials explicitly, instead we shall just write down the corresponding generating function, which defines the anomaly in $2N$ spacetime dimensions in terms of the $\hat{A}$-genus and the Chern character as

$$d \star J = 2\pi \left[\hat{A}(R) \mathrm{ch} \left(\frac{F}{2\pi} \right) \right]_{2N} = 2\pi \left[\frac{R/2}{\sinh R/2} e^{-F/2\pi} \right]_{2N}, \tag{3.29}$$

where F and R are the gauge and Riemann curvature 2-forms, and the subscript $2N$ denotes extracting the $2N$-forms from the formal expansion in powers of F and R. This expression can be directly derived using techniques from supersymmetric quantum mechanics[81, 82].

3.3 Anomalous Relativistic Hydrodynamics

Relativistic hydrodynamics is a description of fluid mechanics that includes the relativistic effects, which arise when either the macroscopic velocities of the fluid or the 'velocity' of the microscopic particles constituting the fluid is comparable to the speed of light. In general, a fluid is described by its velocity field $u^\mu(x)$, and a set of thermodynamic variables, *viz*, the temperature field $T(x)$ and the chemical potential(s) $\mu_a(x)$. Here, $x \in \mathbb{R}^{n,1}$ is the spacetime coordinate[12], and the velocity is normalized as $u^\mu u_\mu = -1$. We shall term these *fluid fields*, and their time evolution is described by the equations of fluid mechanics.

A central assumption underlying all hydrodynamics is that the system always equilibrates locally over a finite time-scale[12, 83]. Globally, the system can fluctuate from the equilibrium, so that the thermodynamic variables are position dependent. Hydrodynamics is then useful in studying their long wavelength fluctuations[4], i.e, to study systems near equilibrium. A (local) form of the second law of thermodynamics then demands that there exists an *entropy current* S, whose divergence is nondecreasing everywhere(so that there are no "sinks" for the entropy current) under any fluid flow allowed by the theory. This, coupled with the first law of thermodynamics, serve to constrain various coefficients appearing in the hydrodynamic description.

[4]Long compared to some intrinsic length scale of the microscopic theory. If the underlying theory has well defined (quasi-)particle excitations, then this would simply be their mean free path.

The dynamical content of hydrodynamics is encoded in the local conservation laws of the system. Thus, in order to construct a hydrodynamic description of a quantum field theory, in principle we only need a list of currents (corresponding to continuous symmetries of the theory *á la* Noether) and their conservation laws, which might be broken in the case of a theory with anomalies. One then needs to construct the *constitutive relations*, which define the currents in terms of the dynamical fields, thereby reducing the conservation laws to the equations of fluid mechanics.

In this section, we shall consider the hydrodynamic description of a QFT with a single anomalous U(1) gauge symmetry in $(2N+1)+1$ dimensions. The Noether currents for the corresponding classical action are simply the symmetric energy momentum tensor ("energy current") $\mathcal{T}^{\mu\nu}$ corresponding to the diffeomorphism invariance and the particle current J^ν corresponding to the global U(1) invariance. On quantization, these current satisfy the anomalous conservation laws

$$\partial_\mu \mathcal{T}^{\mu\nu} = F^{\nu\lambda} J_\lambda, \quad \partial_\mu J^\mu = \mathcal{A}[F], \tag{3.30}$$

as described in Sec 3.2. We shall assume[5] that the presence of anomaly in the theory does not generate any entropy, in which case the anomalous currents can be derived completely using the second law constraint, as shown by Loganayagam[15].

A word about notation: To simplify the clutter of indices, we shall often use the shorthands $u \cdot v = u^\mu v_\mu$, $\partial \cdot u = \partial_\mu u^\mu$ and $(u \cdot \partial) v^\nu = u^\mu \partial_\mu v^\nu$. We shall also use a particularly neat notation due to Loganayagam of describing the hydrodynamic quantities using differential forms.

3.3.1 Constitutive relations and derivative expansion

A systematic way of generating the constitutive relations is provided by the *derivative expansion*[6]. The central idea is to assume that all fields are slowly varying w.r.t the spacetime, so that their spacetime derivatives are "small". In practice, the fluid fields are defined to be at the zeroth order of the derivative expansion, and the order of a given term is the total number of spacetime derivatives (∂_μ's) of the fluid fields. The constitutive relations at a given order are then sums over all terms allowed by the symmetries upto that order with arbitrary "transport" coefficients, and one uses the thermodynamic constraints to obtain these transport coefficients. We now construct the constitutive relations for our U(1)-anomalous fluid.

[5] For a rationale of this assumption, see Footnote 9 of Ref [15].

[6] This is inspired from the construction of effective field theory, where one writes down all terms with a given number of spacetime derivatives of fields, allowed by symmetries. The number of spacetime derivatives in a term then define the classical scaling dimension of the operator, so that terms with more derivatives are increasingly irrelevant in an RG sense.

Zeroth order

At the zeroth order in spacetime derivatives, there is only one vector[7], *viz*, u^μ, and there are only two symmetric tensors, *viz*, $\eta^{\mu\nu}$ and $u^\mu u^\nu$. Thus, at the zeroth order, we can write

$$\mathcal{T}^{\mu\nu} = \alpha u^\mu u^\nu + \beta \eta^{\mu\nu}, \quad J^\mu = \gamma u^\mu, \quad S^\mu = \eta u^\mu, \tag{3.31}$$

where α, β, γ and η are arbitrary coefficients. Since the constitutive relations are defined in a Lorentz invariant fashion, we consider the reference frame comoving with the fluid, in which $u^\mu = (1, 0, \cdots 0)$. Then, we can define the volume density of energy ε, pressure p, particle number n and entropy s as

$$\varepsilon = \mathcal{T}^{00}, \quad p = \mathcal{T}^{ii}, \quad n = J^0, \quad s = S^0. \tag{3.32}$$

These are "extrinsic" quantities, which can be expressed in a Lorentz invariant fashion as $\varepsilon = u_\mu u_\nu \mathcal{T}^{\mu\nu}$, $u_\mu J^\mu = -n$, etc. They depend on the position only via T and μ, i.e, $\varepsilon = \varepsilon(T(x), \mu(x))$ etc. Comparing with the constitutive relations, we can identify

$$\alpha = \varepsilon + p, \quad \beta = p, \quad \gamma = n, \quad \eta = s. \tag{3.33}$$

Thus, at zeroth order, the constitutive relations are

$$\mathcal{T}^{\mu\nu} = (\varepsilon + p) u^\mu u^\nu + p \eta^{\mu\nu}, \quad J^\mu = n u^\mu, \quad S^\mu = s u^\mu. \tag{3.34}$$

Higher orders:

The zeroth order terms describe an ideal fluid; however, they cannot encode the anomalous conservation laws of eq. (3.30). This can be deduced from a simple power counting: since A is at $O(\partial^0)$ in the derivative expansion, $F \sim O(\partial)$ and the anomaly polynomial $\mathcal{A}[F] \sim F^{N+1} \sim O(\partial^{N+1})$. But since $\partial J = \mathcal{A}[F]$, we shall need to add a term at $O(\partial^N)$ to J. Finally, since $\partial \mathcal{T} \sim FJ \sim O(\partial^{N+1})$, we shall also need to add a term at $O(\partial^N)$ to the constitutive relation for $\mathcal{T}$.

[7]We do not use A_μ as it is not gauge invariant. For any given point x, we can choose a gauge such that $A(x) = 0$. However, its gauge invariant derivatives ($F = dA$) will be used at higher orders.

Adding the anomalous contributions to the zeroth order constitutive relations, we set

$$\begin{aligned}\mathcal{T}^{\mu\nu} &= (\varepsilon + p)u^\mu u^\nu + p\eta^{\mu\nu} + (q^\mu u^\nu + u^\mu q^\nu) \\ J^\mu &= nu^\mu + \mathcal{J}^\mu, \\ S^\mu &= su^\mu + \mathcal{S}^\mu,\end{aligned} \tag{3.35}$$

where $q, \mathcal{J}$, $\mathcal{S}$ contain one or more spacetime derivatives of u_μ or A_μ. Since the energy, momentum and number of particles are 'physical' quantities, we shall demand that eq. (3.32) still hold, so that

$$u_\mu q^\mu = u_\mu j^\mu = u_\mu j_S^\mu = 0. \tag{3.36}$$

These expansions are by no means complete, as one could write down many more terms at different orders. Most of these terms will describe some sort of dissipative effects. Even restricting to nondissipative terms, there can potentially be other tensor corrections to $\mathcal{T}$ at $O(\partial^N)$, which cannot be written as $q^\mu u^\nu + u^\mu q^\nu$ for some vector q^μ. We shall ignore those corrections since they cannot be constrained by the second law of thermodynamics[15]. Thus, we essentially seek to add the minimal possible terms to the constitutive relations which can describe the conservation laws of eq. (3.30) and describe some macroscopic effects of the gauge anomaly in the microscopic.

Finally, there is the somewhat tricky issue of a *frame choice*, to address the fact that the velocity, temperature and chemical potentials are not uniquely defined away from equilibrium. In practice, this means that one can vary these fields and compensate for the variation by changing the anomalous contribution accordingly. For instance, the constitutive relations are invariant under

$$u_\mu \to u_\mu + \delta u_\mu, \qquad q_\mu \to q_\mu - (\varepsilon + p)\delta u_\mu, \qquad \mathcal{J}_\mu \to \mathcal{J}_\mu - n\,\delta u_\mu. \tag{3.37}$$

The origin of this ambiguity in relativistic hydrodynamics lies in the fact that unlike the nonrelativistic case, one cannot distinguish between the 'mass' and the 'energy'. One usually fixes this ambiguity by putting additional constraints on the anomalous components, e.g, $q = 0$ (*Landau frame*) or $\mathcal{J} = 0$ (*Eckart frame*). We shall not choose a particular frame beforehand; rather, we shall simply derive our results for the anomalous contribution and choose a frame in which they take simple forms.

The anomalous contributions are written explicitly in terms of derivatives of the fluid fields most compactly using differential forms corresponding to the currents, as we now describe.

Differential forms

Given a vector $\xi^\mu \in \text{Vect}(\mathbb{R}^{n,1})$, we can use the Minkowski metric to define a corresponding 1-form as $\xi = \xi_\nu dx^\nu = \xi^\mu \eta_{\mu\nu} dx^\nu$. Thus, we get 1-forms $u, q, \mathcal{J}, \mathcal{S}$, corresponding to the velocity field u_μ and the anomalous currents q^μ, $\mathcal{J}^\mu$ and $\mathcal{S}^\mu$, respectively. We also define their Hodge duals[70] $\star : \Omega^p(\mathbb{R}^{n,1}) \to \Omega^{n+1-p}(\mathbb{R}^{n,1})$, which we denote by an overbar as well as the usual $\star$, following Loganayagam[15]. Explicitly,

$$\bar{u} = u_\mu(\star dx^\mu) = \frac{1}{(n+1)!}\epsilon_{\mu\nu_1\nu_2\ldots\nu_n} u^\mu \left(\bigwedge_{i=1}^{n} dx^{\nu_i}\right). \tag{3.38}$$

Given 1-forms u and v, the inner product and gradient can be written as

$$\star\,(v_\mu u^\mu) = v \wedge \bar{u}, \quad \star(\partial_\mu u^\mu) = d\bar{u}, \tag{3.39}$$

where $\star 1 = \mathbb{V}$, the Euclidean volume form on $\mathbb{R}^{2N+1,1}$. This is the dictionary to go between the differential forms and vectors on $\mathbb{R}^{2N+1,1}$.

The fluid fields provide us with two 1-forms, *viz*, u and A. We define the corresponding "curvatures" as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \quad \Omega_{\mu\nu} = \partial_\mu u_\nu - \partial_\nu u_\mu. \tag{3.40}$$

In a reference frame specified by u, the electric field is defined as $E_\mu = F_{\mu\nu}u^\nu$. Since the acceleration is defined as $a_\mu = u^\nu \partial_\nu u_\mu$, we can use $u^\nu \partial_\mu u_\nu = \frac{1}{2}\partial_\mu (u_\nu u^\nu) = 0$ to express it in terms of u^ν and $\Omega_{\mu\nu}$ as

$$a_\mu = u^\nu (\partial_\nu u_\mu - \partial_\mu u_\nu) = -\Omega_{\mu\nu} u^\nu. \tag{3.41}$$

Thus, we can decompose the curvatures into their "electric" and "magnetic" components, with the former being along and latter transverse to the velocity field $u^\mu(x)$. In terms of differential forms, this decomposition[15] simply becomes

$$\begin{aligned} F = dA &= B + u \wedge E, \\ \Omega = du &= \omega - u \wedge a, \end{aligned} \tag{3.42}$$

where B and ω are 2-forms. Clearly,

$$u \wedge F = u \wedge B, \qquad u \wedge \Omega = u \wedge \omega. \tag{3.43}$$

The decompositions are particularly transparent in a frame where $u^\mu = (1, 0, \ldots 0)$, i.e, $u \equiv u_\mu dx^\mu = -dt$. Then, $E_0 = 0,\ E_i = F_{i0}$ and $a_0 = 0,\ a_i = -\Omega_{i0} = \partial_t u_i$, so that

$$F = B - F_{i0}\, dt \wedge dx^i \implies B = \frac{1}{2} F_{ij} dx^i \wedge dx^j,$$
$$\Omega = \omega - \Omega_{i0}\, dt \wedge dx^i \implies \omega = \frac{1}{2}\Omega_{ij} dx^i \wedge dx^j = \partial_i u_j dx^i \wedge dx^j.$$

In order to expose the mathematical symmetry between these decompositions, we have followed the traditional fluid mechanics convention [11] in defining the *vorticity* ω, so that $\vec{\omega} = \nabla \times \vec{u}$ in 3+1 dimensions. This is in contrast to the angular velocity, $\vec{\omega}_A = \frac{1}{2}\nabla \times \vec{u} = \frac{1}{2}\omega$, which is sometimes referred to as the "vorticity" in relativistic hydrodynamics (for instance, Refs [12, 15]).

The most general anomalous contributions to the constitutive relations in eq. (3.35) can be succinctly written as

$$\bar{q} = u \wedge \left(\sum_{k=1}^{N-1} \gamma_k^{(q)} \underbrace{F \wedge \cdots \wedge F}_{k} \wedge \underbrace{\Omega \wedge \cdots \wedge \Omega}_{N-1-k} \right), \tag{3.44}$$

and similar expressions for $\bar{\mathcal{J}}$ and $\bar{\mathcal{S}}$, where γ's are the transport coefficients. The task of hydrodynamics is then to constrain these transport coefficients using general principles such as those of thermodynamics.

3.3.2 Conservation laws and thermodynamic constraints

We next use the anomalous conservation laws (eq. (3.30)) and the constitutive relations constructed so far to derive the equations of fluid dynamics.

Ideal fluid

We begin with a description of ideal fluid, for which the conservation laws are

$$\partial_\mu \mathcal{T}_0^{\mu\nu} = 0, \quad \partial_\mu J_0^\mu = 0, \quad \partial_\mu S_0^\mu \geq 0, \tag{3.45}$$

and the constitutive relations terminate at the zeroth order. The equations of motion can be obtained by simply substituting the the constitutive relations in the conservation laws. We start off with the conservation of the U(1) current:

$$0 = \partial_\mu J_0^\mu = (u \cdot \partial) n + n\, \partial \cdot u. \tag{3.46}$$

This is simply the continuity (Bernoulli's) equation, since it indicates that the variation in $n(x)$ can happen only in presence of a source/sink for $u(x)$.

The conservation of the energy-momentum tensor becomes

$$\begin{aligned} 0 = \partial_\mu \mathcal{T}_0^{\mu\nu} &= u^\nu u^\mu \partial_\mu(\varepsilon + p) + (\varepsilon + p)\left(u^\nu \partial_\mu u^\mu + u^\mu \partial_\mu u^\nu\right) + \eta^{\mu\nu}\partial_\mu p \\ &= u^\nu (u\cdot\partial)(\varepsilon + p) + (\varepsilon + p)\left(u^\nu(\partial\cdot u) + a^\nu\right) + \partial^\nu p, \end{aligned} \tag{3.47}$$

which can be recast as

$$u^\nu(\partial\cdot u) + a^\nu = -\frac{1}{\varepsilon + p}\left[\partial^\nu p + u^\nu (u\cdot\partial)(\varepsilon + p)\right] \tag{3.48}$$

This is the relativistic version of the Euler's equation for the fluid flow, since the LHS is simply a relativistic version of $(\partial_t + \mathbf{v}\cdot\nabla)\,\mathbf{v}$. (See Ref [84] for details). Furthermore,

$$0 = u_\nu \partial_\mu \mathcal{T}_0^{\mu\nu} = -(u\cdot\partial)\varepsilon - (\varepsilon + p)\,(\partial\cdot u)\,, \tag{3.49}$$

since $u_\mu a^\mu = \frac{1}{2} u^\nu \partial_\nu (u_\mu u^\mu) = 0$.

Finally, the nonnegative divergence of the entropy current is

$$0 \le \partial_\mu S_0^\mu = (u\cdot\partial)s + s\,\partial\cdot u. \tag{3.50}$$

We can express the divergence of the entropy current in terms of the other two currents using the first law of thermodynamics, which reads

$$\varepsilon = -p + \mu n + Ts, \quad d\varepsilon = \mu dn + sdT. \tag{3.51}$$

Adding eqns 3.49, 3.46 and 3.50, we get

$$\begin{aligned} T\partial_\mu S_0^\mu &= T\partial_\mu S_0^\mu + \mu\partial_\mu J_0^\mu + u_\nu \partial_\mu \mathcal{T}_0^{\mu\nu} \\ &= (Ts + \mu n - \varepsilon - p)(\partial\cdot u) + u\cdot(T\partial s + \mu\partial n - \partial\varepsilon) = 0, \end{aligned} \tag{3.52}$$

where the last line follows from the first law of thermodynamics. We deduce that an ideal fluid does not produce entropy, as expected. This is the general strategy to use the second law of thermodynamics: we add scalars which vanish from conservation laws at each order to obtained a simplified expression for $T\partial\cdot S$, which we demand to be nonnegative.

Anomalous fluids

We can again compute the equations of fluid dynamics analogous to the zeroth order case. These are not very enlightening by themselves, as they do not correspond to any known equations of classical fluid dynamics. However, we can use them to derive the second law constraint for the anomalous fluid, *viz*, $\partial \cdot S = 0$. The (non-)conservation of energy-momentum tensor leads to,

$$\begin{aligned} 0 &= \partial_\mu \mathcal{T}^{\mu\nu} - F^{\nu\lambda} J_\lambda \\ &= \left[\partial_\mu \mathcal{T}_0^{\mu\nu} + u^\nu(\partial\cdot q) + (q\cdot\partial)u^\nu + (u\cdot\partial)q^\nu + q^\nu(\partial\cdot u)\right] - F^{\nu\lambda}\left[n u_\lambda + \mathcal{J}_\lambda\right]. \end{aligned} \tag{3.53}$$

from which it again follows that

$$0 = u_\nu \left(\partial_\mu \mathcal{T}^{\mu\nu} - F^{\nu\lambda} J_\lambda\right) = u_\nu \partial_\mu \mathcal{T}_0^{\mu\nu} - (\partial + a)\cdot q - E^\lambda \mathcal{J}_\lambda, \tag{3.54}$$

where we have used the fact that $u_\nu F^{\nu\lambda} u_\lambda = 0$ by antisymmetry of F, $u_\nu F^{\nu\lambda}\mathcal{J}_\lambda = E^\lambda \mathcal{J}_\lambda$ by definition, and

$$\begin{aligned} & u_\nu \left[u^\nu(\partial\cdot q) + (q\cdot\partial)u^\nu + (u\cdot\partial)q^\nu + q^\nu(\partial\cdot u)\right] \\ =& -\partial\cdot q + \frac{1}{2}(q\cdot\partial)(u_\nu u^\nu) + u_\nu u^\mu \partial_\mu q^\nu + (u\cdot q)(\partial\cdot u) \\ =& -\partial\cdot q + u^\mu \partial_\mu (u_\nu q^\nu) - q_\nu u^\mu \partial_\mu u_\nu \\ =& -(\partial + a)\cdot q, \end{aligned} \tag{3.55}$$

since $u\cdot q = 0$. The charge and entropy currents satisfy

$$0 = \partial_\mu J^\mu - \mathcal{A}[F] = \partial_\mu J_0^\mu + \partial_\mu \mathcal{J}^\mu - \mathcal{A}[F] \tag{3.56}$$

and

$$0 = \partial_\mu S^\mu = \partial_\mu S_0^\mu + \partial_\mu \mathcal{S}^\mu. \tag{3.57}$$

Again, adding eqns 3.54, 3.56 and 3.57, we get

$$\begin{aligned} T\partial_\mu S^\mu &= T\partial_\mu S^\mu + \mu\left(\partial_\mu J^\mu - \mathcal{A}\right) + u_\nu\left(\partial_\mu \mathcal{T}_0^{\mu\nu} - F^{\nu\lambda} J_\lambda\right) \\ &= \left[T\partial_\mu S_0^\mu + \mu\partial_\mu J_0^\mu + u_\nu \partial_\mu \mathcal{T}_0^{\mu\nu}\right] + \left[T\partial\cdot\mathcal{S} + \mu(\partial\cdot\mathcal{J} - \mathcal{A}) - (\partial + a)\cdot q - E\cdot\mathcal{J}\right]. \end{aligned} \tag{3.58}$$

Using eq. (3.52) and setting $T\partial_\mu S_0^\mu = 0$, we get the second law constraint involving only the anomalous contributions, as

$$T\,\partial\cdot\mathcal{S} + \mu(\partial\cdot\mathcal{J} - \mathcal{A}) - (\partial + a)\cdot q + E\cdot\mathcal{J} = 0. \tag{3.59}$$

Solutions and Replacement rules

We begin by taking Hodge dual of the second law constraint to express it in terms of differential forms as

$$T\,d\bar{\mathcal{S}} + \mu\,d\bar{\mathcal{J}} - d\bar{q} - a\wedge\bar{q} + E\wedge\bar{\mathcal{J}} = \mu\bar{\mathcal{A}}. \tag{3.60}$$

This equation was solved in general in [15], which we do not discuss in any details here. More intriguingly, Loganayagam and Surówka[16] showed that one can define a grand potential[8] current $\bar{\mathcal{G}}$ from which $\bar{q}$, $\bar{\mathcal{J}}$ and $\bar{\mathcal{S}}$ can be derived as

$$\bar{\mathcal{G}} = \bar{q} - \mu\bar{\mathcal{J}} - T\bar{\mathcal{S}}; \qquad \bar{\mathcal{J}} = -\frac{\partial\bar{\mathcal{G}}}{\partial\mu}, \quad \bar{\mathcal{S}} = -\frac{\partial\bar{\mathcal{G}}}{\partial T}. \tag{3.61}$$

Next, they propose an ansatz for $\bar{\mathcal{G}}$ as

$$\bar{\mathcal{G}} = \int_0^\infty \frac{d\varepsilon}{2\pi} g(\varepsilon)\bar{\mathcal{J}}_{\mathcal{G}}, \quad g(\varepsilon) = -T\,\ln\left(1 + e^{-\beta(\varepsilon-\mu)}\right), \tag{3.62}$$

where $g(\varepsilon)$ is the fermionic single particle grand potential(See Appendix C.1). This can be motivated by kinetic theory (Sec 2.5), where we would have written $\mathcal{G}^\mu$ as an integral of $g(\varepsilon)\dot{x}^\mu$ over all the phase space coordinates except x^μ. Thus, $\bar{\mathcal{J}}_{\mathcal{G}}$ can be thought of as the integral of $g(\varepsilon)\dot{x}^\mu$ over all phase coordinates except x^μ and $|\mathbf{p}| = \varepsilon$.

Using eqns (C.3) and (C.4) from Appendix C.1, the anomalous contributions to the constitutive relations can be explicitly written as

$$\bar{\mathcal{J}} = \int_0^\infty \frac{d\varepsilon}{2\pi} f(\varepsilon)\bar{\mathcal{J}}_{\mathcal{G}}, \quad \bar{\mathcal{S}} = \int_0^\infty \frac{d\varepsilon}{2\pi} h(\varepsilon)\bar{\mathcal{J}}_{\mathcal{G}}, \quad \bar{q} = \int_0^\infty \frac{d\varepsilon}{2\pi}\varepsilon f(\varepsilon)\bar{\mathcal{J}}_{\mathcal{G}}, \tag{3.63}$$

where $f(\varepsilon)$ and $h(\varepsilon)$ are the Fermi-Dirac distribution and the 1-particle entropy, respectively. In terms of

[8] In Ref [16], $\mathcal{G}$ is referred to as the "Gibbs free energy current". However, as the Gibbs free energy (per unit volume) is $G = \varepsilon + p - Ts = \mu n$, a Gibbs free energy current would more naturally be μJ.

these, the second law constraint becomes

$$\begin{aligned}\mu\bar{\mathcal{A}} &= \int_0^\infty \frac{d\varepsilon}{2\pi}\left[(Th(\varepsilon)+\mu f(\varepsilon)-\varepsilon f(\varepsilon))\,d\bar{\mathcal{J}}_{\mathcal{G}}+f(\varepsilon)\,(E-\varepsilon a)\wedge\bar{\mathcal{J}}_{\mathcal{G}}\right]\\ &= \int_0^\infty \frac{d\varepsilon}{2\pi}\left[g(\varepsilon)\,d\bar{\mathcal{J}}_{\mathcal{G}}+\frac{\partial g(\varepsilon)}{\partial\varepsilon}\bar{\mathcal{J}}^E_{\mathcal{G}}\right],\\ &= g(\varepsilon)\bar{\mathcal{J}}^E_{\mathcal{G}}\Big|^\infty_{\varepsilon=0}+\int_0^\infty \frac{d\varepsilon}{2\pi}g(\varepsilon)\left[d\bar{\mathcal{J}}_{\mathcal{G}}-\frac{\partial}{\partial\varepsilon}\bar{\mathcal{J}}^E_{\mathcal{G}}\right],\end{aligned} \tag{3.64}$$

where $\bar{\mathcal{J}}^E_{\mathcal{G}}=(E-\varepsilon a)\wedge\bar{\mathcal{J}}_{\mathcal{G}}$ and we have used eqns (C.3) and (C.4) again. The anomaly be accounted for by the $\varepsilon=0$ limit of the boundary term, i.e, $g(\varepsilon)\bar{\mathcal{J}}^E_{\mathcal{G}}(\varepsilon)|_{\varepsilon=0}=-\bar{\mathcal{A}}$. The physical interpretation is a spectral flow argument, that the anomalous contributions are simply injected at $\varepsilon=0$, the Weyl point, and then convected along the fluid flow. The term inside the integral vanishes using the continuity equation in the phase space.

In Ref [16], the equation for $\bar{\mathcal{J}}^E_{\mathcal{G}}(\varepsilon)$ and hence $\bar{\mathcal{J}}_{\mathcal{G}}(\varepsilon)$ is derived for all ε using its boundary value and the phase space continuity equation. Their final result is

$$\bar{\mathcal{J}}_{\mathcal{G}}=\frac{u}{2\pi\,N!}\wedge\left(\frac{B+\varepsilon\omega}{2\pi}\right)^N\implies\bar{\mathcal{G}}_{2N+2}=\frac{u}{2\pi\,N!}\wedge\int_0^\infty\frac{d\varepsilon}{2\pi}g(\varepsilon)\left(\frac{B+\varepsilon\omega}{2\pi}\right)^N. \tag{3.65}$$

It is convenient to define a generating function for $\bar{\mathcal{G}}$ as

$$\bar{\mathcal{G}}_\tau=\sum_{N=0}^\infty\bar{\mathcal{G}}_{2N+2}\tau^N=\frac{u}{2\pi}\wedge\int_0^\infty\frac{d\varepsilon}{2\pi}g(\varepsilon)e^{\frac{\tau}{2\pi}(B+\varepsilon\omega)} \tag{3.66}$$

We perform this integral explicitly in Sec 5.4. The result is

$$\bar{\mathcal{G}}_\tau=-u\wedge e^{\frac{\tau qB}{2\pi}}\frac{2\pi}{(\omega\tau)^2}\left[\frac{\frac{\omega\tau}{2\beta}}{\sin\left(\frac{\omega\tau}{2\beta}\right)}e^{\frac{\mu\omega\tau}{2\pi}}-\left(1+\frac{\mu}{2\pi}\omega\tau\right)\right], \tag{3.67}$$

In order to obtain the anomalous currents in $2N+2$ spacetime dimensions, we expand $\bar{\mathcal{G}}_\tau$ in a power series in τ and pick out the coefficient of τ^N, and then use eq. (3.61).

The generating function for $\bar{\mathcal{G}}$ remarkable because it is strongly reminiscent of the $\hat{\text{A}}$-genus and Chern character generating function for gauge and gravitational anomaly polynomials from eq. (3.29):

$$\hat{\text{A}}(R)\,\text{ch}(F)=\prod_i\frac{x_i/2}{\sinh(x_i/2)}\sum_j e^{y_j}, \tag{3.68}$$

where x_i and y_j are $i/2\pi$ times the form-valued formal eigenvalues of the Riemann curvature 2-form R and

the gauge field two form F, respectively. Thus, $\mathcal{G}$ in $2N+2$ dimensions can be obtained by the replacements $F \to \mu$ and $\mathrm{tr}\left\{R^{2n}\right\} \to 2(2\pi T)^{2n}\ \forall\, n \in \mathbb{Z}^{+}$ in the anomaly polynomial in $2N+4$ dimensions. These are the *replacement rules*, which is conjectured to hold even for theories with interaction, mirroring the fact that the anomalies are unaffected by addition of interactions.

4 Berry phase

In 1984, Michael Berry[85] pointed out that a nondegenerate quantum mechanical state subject to a cyclic adiabatic[86] process picks up an additional phase besides the dynamical phase from the Hamiltonian evolution. If the cyclic process is described as traversing a loop in a parameter space, where the Hamiltonian depends on the parameter, then the Berry's phase can be written as a line integral of a *Berry connection* 1-form over the path in the parameter space, or a surface integral of a *Berry curvature* 2-form over a region enclosed by the loop. It is this dependence on only the path, and not on the rate of traversal, that makes the Berry phase *geometric* in nature.

It is curious that the notion of Berry's phase was indirectly glimpsed many times in various context before Michael Berry's 1984 paper. Probably the earliest example(1956) is the Pancharatnam phase[87], acquired by polarized light beams passing through crystals. Around the same time, in trying to explain *anomalous Hall effect*, Karplus, Luttinger and Kohn[23, 24, 25] noted that the semiclassical equations of motion for electrons in a Bloch band include an *anomalous velocity*, proportional to a quantity which we now know[88, 89, 22, 90] as the Berry curvature.

Subsequently, Berry's original analysis has been analyzed and generalized in various directions. Almost contemporary to the original paper, Barry Simon[44] showed that Berry's phase is simply a holonomy on the line bundle of the eigenstates on the parameter space. Wilczek and Zee[91] removed the nondegeneracy condition, and thereby derived a *nonabelian* analogue of the Berry curvature. Finally, Aharonov and Anandan[92] lifted the adiabaticity condition on the cyclic process, which can then be interpreted as holonomies on the Hilbert space itself, thought of as a fiber bundle over the projective Hilbert space, i.e, the space of density matrices.

The Berry phase, owing to its geometric nature, is often a crucial ingredient of the topological invariants in condensed matter systems, a prime example being the TKNN invariant[43], which is geometrically the first Chern character of the Berry curvature over the Brillouin zone. It is also indispensable for a semiclassical description of Weyl fermions[26], where it encodes the gauge anomaly. In Ch 5, we shall see the corresponding nonabelian generalizations encode guage anomalies in all even dimensions.

In this chapter, we introduce the Berry's phase and its nonabelian generalization and compute the Berry

curvature and Chern numbers associated with various systems of interest in this thesis. We also describe the semiclassical dynamics of electrons in Bloch bands including the anomalous velocity term, and the interpretation of the equation of motion in the symplectic formalism discussed in Ch 2.

4.1 Basics

We start off with a derivation of the Berry phase[85] and its nonabelian generalization by Wilczek and Zee[91], using the language of differential forms. We also discuss their computation using projection operators.

4.1.1 Definition of Berry phase

Consider a quantum mechanical system with Hamiltonian $\mathcal{H}(\boldsymbol{\lambda})$, dependent on a parameter $\boldsymbol{\lambda} \in \mathscr{M}$, where $\mathscr{M}$ is a smooth manifold. The corresponding eigenvalue problem is $\mathcal{H}(\boldsymbol{\lambda})|n(\boldsymbol{\lambda})\rangle = \varepsilon_n(\boldsymbol{\lambda})|n(\boldsymbol{\lambda})\rangle$. Consider tuning the Hamiltonian *adiabatically*[1] along a closed curve $\boldsymbol{\lambda}\colon \mathbb{R} \to \mathscr{M}$. Since $\boldsymbol{\lambda}(0) = \boldsymbol{\lambda}(T)$ for some $T > 0$, the cyclic evolution must return an eigenstate to itself, upto an overall phase. If a given nondegenerate eigenstate $|n\rangle$ were independent of the parameter $\boldsymbol{\lambda}$, it would simply pick up the dynamical phase, given by $e^{-i\int_0^T dt\, \varepsilon_n(\boldsymbol{\lambda}(t))}$, where $\varepsilon_n(\boldsymbol{\lambda})$ is the corresponding eigenvalue. However, when $|n\rangle$ does depend on $\boldsymbol{\lambda}$, the system picks up an extra "geometric phase" $e^{i\gamma_n}$, termed the Berry phase. Crudely speaking, the Berry phase is "geometric" because it depends only on the path traversed in $\mathscr{M}$ and not on the time taken, unlike the time-dependent dynamical phase.

To compute the Berry phase explicitly for $|n\rangle$, a nondegenerate normalized eigenstate of $\mathcal{H}$, we begin by setting $|n_t\rangle = e^{-i\int_0^t dt'\, \varepsilon_n(\boldsymbol{\lambda}(t'))} e^{i\gamma_n(t)}|n_0\rangle$, where we have defined $|n_t\rangle = |n(\boldsymbol{\lambda}(t))\rangle$ to simplify the notation. The time-dependent Schrödinger equation, $i\partial_t|n_t\rangle = \mathcal{H}(\boldsymbol{\lambda}(t))|n_t\rangle$, then becomes

$$\varepsilon_n(\boldsymbol{\lambda}(t))|n_t\rangle - \dot{\gamma_n}(t)|n_t\rangle + i|\partial_t n_t\rangle = \varepsilon_n(\boldsymbol{\lambda}(t))|n_t\rangle. \tag{4.1}$$

Thus, taking inner product with $|n_t\rangle$, we get

$$\dot{\gamma_n}(t) = i\langle n_t|\partial_t n_t\rangle \implies \gamma_n(T) = i\int_0^T dt\, \langle n_t|\partial_t n_t\rangle. \tag{4.2}$$

[1] The change being adiabatic simply means that the associated time scale is much larger than that associated with the energy gap to other eigenstates of the Hamiltonian. This ensures that one does not transition to a different state during the variation of the parameter.

But since $|n_t\rangle$ depends on t only through $\boldsymbol{\lambda}$,

$$\gamma_n(T) = i \int_0^T dt \, \langle n_t | \partial_t n_t \rangle \cdot \frac{d\boldsymbol{\lambda}}{dt} = i \oint_{\mathcal{C}} d\boldsymbol{\lambda} \cdot \langle n(\boldsymbol{\lambda}) | \partial_{\boldsymbol{\lambda}} n(\boldsymbol{\lambda}) \rangle. \tag{4.3}$$

Thus, the Berry phase simply depends on an integral around the loop on the parameter space $\mathscr{M}$, independent of the time taken to traverse the loop. It is real by definition, since

$$\langle n(\boldsymbol{\lambda}) | n(\boldsymbol{\lambda}) \rangle = 1 \implies \langle \partial_{\boldsymbol{\lambda}} n(\boldsymbol{\lambda}) | n(\boldsymbol{\lambda}) \rangle = -\langle n(\boldsymbol{\lambda}) | \partial_{\boldsymbol{\lambda}} n(\boldsymbol{\lambda}) \rangle, \tag{4.4}$$

so that

$$\gamma_n^* = -i \oint_{\mathcal{C}} d\boldsymbol{\lambda} \cdot \langle n(\boldsymbol{\lambda}) | \partial_{\boldsymbol{\lambda}} n(\boldsymbol{\lambda}) \rangle^* = i \oint_{\mathcal{C}} d\boldsymbol{\lambda} \cdot \langle \partial_{\boldsymbol{\lambda}} n(\boldsymbol{\lambda}) | n(\boldsymbol{\lambda}) \rangle = \gamma_n. \tag{4.5}$$

If $\mathcal{C}$ is contractible, we can use Stokes theorem to write the Berry phase associated with $\mathcal{C}$ as an integral over an area $\mathcal{S}$ whose boundary is $\mathcal{C}$, as

$$\gamma_n(\mathcal{C}) = \oint_{\mathcal{C}} \mathfrak{a}_n = \int_{\mathcal{S}} \mathfrak{F}_n, \tag{4.6}$$

where we have defined[2] the Berry connection $\mathfrak{a}_n$, which is a 1-form on $\mathscr{M}$, and the corresponding Berry curvature[3] for the eigenstate $|n\rangle$ as

$$\mathfrak{a}_n = i \langle n | dn \rangle, \quad \mathfrak{F}_n = d\mathfrak{a}_n = i \langle dn | \wedge | dn \rangle. \tag{4.7}$$

The Berry connection computed above is defined only up to a choice of the phase of $|n(\boldsymbol{\lambda})\rangle$. To wit, if redefined our our eigenstates for $\mathcal{H}(\boldsymbol{\lambda})$ as $|n(\boldsymbol{\lambda})\rangle = e^{i\alpha(\boldsymbol{\lambda})} |\widetilde{n}(\boldsymbol{\lambda})\rangle$ and computed the Berry connection using the new eigenstates, we get

$$\widetilde{\mathfrak{a}}_n = i \langle \widetilde{n} | d\widetilde{n} \rangle = i \left(\langle n | e^{i\alpha} \right) d \left(e^{-i\alpha} | n \rangle \right) = \mathfrak{a}_n + d\alpha. \tag{4.8}$$

However, both $\gamma_n(\mathcal{C})$ and $\mathfrak{F}_n$ are well defined, the former because $\oint d\alpha = 0$ around any closed curve, and the latter because $d^2 = 0$.

[2] We use the sign convention conventionally used in condensed matter literature, which is opposite the one used by Michael Berry[85], who defines it as $\mathfrak{a} = -i\langle n | dn \rangle$.

[3] The terminology of *connections* and *curvatures* comes from the theory of fiber bundles, introduced in Sec 3.1. We discuss these geometric ideas in the context of Berry phase in Sec 4.2.1.

4.1.2 Nonabelian generalization

We now remove the nondegeneracy requirement, and consider an eigenvalue $\varepsilon_n(\boldsymbol{\lambda})$ of $\mathcal{H}$, whose eigenspace is r-fold degenerate throughout the loop $\mathcal{C}$. An adiabatic transport can now lead to rotation within this degenerate eigenspace. Let be $\{|n_\alpha(\boldsymbol{\lambda})\rangle\}_{\alpha=1}^r$ be an orthonormal basis for this eigenspace, which is related to any other basis $|\widetilde{n}_\alpha(\boldsymbol{\lambda})\rangle$ by a unitary transform. Again, we set $|n_{\alpha,t}\rangle = e^{-i\int_0^t dt'\,\varepsilon_n(\boldsymbol{\lambda}(t'))}e^{i\gamma_{n,\alpha\beta}(t)}|n_{\beta,0}\rangle$, where $e^{i\boldsymbol{\gamma}_n(t)}$ is now a $r\times r$ unitary matrix, so that $\boldsymbol{\gamma}_n(t)$ is a Hermitian matrix.

Like the abelian case, the Schrödinger equation leads to

$$\dot{\gamma}_{n,\alpha\beta}(t)|n_{\beta,t}\rangle = i|\partial_t n_{\alpha,t}\rangle \implies \dot{\gamma}_{n,\alpha\beta}(t) = i\langle n_{\alpha,t}|\partial_t n_{\beta,t}\rangle, \tag{4.9}$$

and the nonabelian Berry phase becomes

$$\gamma_{n,\alpha\beta}(T) = i\int_0^T dt\,\langle n_{\alpha,t}|\partial_t n_{\beta,t}\rangle = i\oint_{\mathcal{C}} d\boldsymbol{\lambda}\cdot\langle n_\alpha(\boldsymbol{\lambda})|\partial_{\boldsymbol{\lambda}} n_\beta(\boldsymbol{\lambda})\rangle. \tag{4.10}$$

We can again check that $\boldsymbol{\gamma}_n$ is Hermitian by definition, since

$$\left[\boldsymbol{\gamma}_n^\dagger(T)\right]_{\alpha\beta} = \gamma^*_{n,\beta\alpha}(T) = -i\oint_{\mathcal{C}} d\boldsymbol{\lambda}\cdot\langle n_\beta(\boldsymbol{\lambda})|\partial_{\boldsymbol{\lambda}} n_\alpha(\boldsymbol{\lambda})\rangle^* = i\oint_{\mathcal{C}} d\boldsymbol{\lambda}\cdot\langle n_\alpha(\boldsymbol{\lambda})|\partial_{\boldsymbol{\lambda}} n_\beta(\boldsymbol{\lambda})\rangle = \left[\boldsymbol{\gamma}_n(T)\right]_{\alpha\beta} \tag{4.11}$$

We can again define a Berry connection as $\mathfrak{a}_{n,\alpha\beta} = i\langle n_\alpha|dn_\beta\rangle$, which is a matrix valued 1-form. This matrix-valued Berry connection was first introduced by Wilczek and Zee[91].

Strictly speaking, the total "phase" picked up under the adiabatic transport should be $e^{i\gamma_n} = P\exp\left(\oint_{\mathcal{C}}\mathfrak{a}_n\right)$, where P denotes the path ordering. This was not an issue in the abelian case, but is crucial here since the matrices $\mathfrak{a}$ for different points $\boldsymbol{\lambda}\in\mathscr{M}$ do not commute. This path ordering makes a straightforward definition of Berry curvature using Stokes' theorem much trickier, so we shall use an alternative approach, *viz*, the identification of $\mathfrak{a}_n$ as a nonabelian guage field.

Recall that the Berry connection for the abelian case had a gauge dependence with a gauge group U(1), due to the nonuniqueness in the definition of the eigenstate upto a phase. To compute the nonabelian Berry phase, we had a nonunique choice of basis for the degenerate eigenspace corresponding to the eigenvalue ε_n. Using a unitary transform, defining a new basis as[4]

$$|\widetilde{n}\rangle = \mathcal{U}^*|n\rangle \implies |\widetilde{n}_\alpha\rangle = \mathcal{U}^*_{\alpha\beta}|n_\beta\rangle, \qquad \langle\widetilde{n}_\alpha| = \mathcal{U}_{\alpha\beta}\langle n_\beta|, \tag{4.12}$$

[4] This choice of unitary matrix is to derive a gauge variation that conforms with the definitions of Sec 3.1.

the new Berry connection becomes

$$\widetilde{\mathfrak{a}}_{n,ab} = i\,\mathcal{U}_{\alpha\gamma}\langle n_\gamma|dn_\delta\rangle\mathcal{U}^\dagger_{\delta\beta} + i\,\mathcal{U}_{\alpha\gamma}\delta_{\gamma\delta}\,d\mathcal{U}^\dagger_{\delta\beta} = \left[\mathcal{U}\mathfrak{a}_n\,\mathcal{U}^{-1} - i\,d\mathcal{U}\,\mathcal{U}^{-1}\right]_{\alpha\beta}, \tag{4.13}$$

where we have used $\mathcal{U}d\mathcal{U}^\dagger = -d\mathcal{U}\,\mathcal{U}^\dagger$ in the last step, since $\mathcal{U}\mathcal{U}^\dagger = \mathbb{1}$. We have simply recovered eq. (3.10) for a nonabelian gauge transformation, so that we identify $\mathfrak{a}_n$ as a nonabelian gauge field with the gauge group U(r). In Sec 3.1, we show that the corresponding nonabelian curvature is defined as $\mathfrak{F}_n = d\mathfrak{a}_n - i\mathfrak{a}_n \wedge \mathfrak{a}_n$, which follows from demanding that the curvature transform covariantly ($\mathfrak{F} \mapsto \mathcal{U}\,\mathfrak{F}\,\mathcal{U}^{-1}$) under a nonabelian gauge transformation.

4.1.3 Computation strategies

In practice, we are usually interested in computing the Berry curvature, which, unlike the Berry connection, is a gauge covariant quantity. We now discuss two strategies that would prove useful in explicitly computing the Berry curvature for the systems of interest.

The first approach involves replacing the differential of states with a differential on the Hamiltonian. Explicitly,

$$\mathcal{H}|n\rangle = \varepsilon_n|n\rangle \implies d\mathcal{H}\,|n\rangle + \mathcal{H}|dn\rangle = d\varepsilon_n|n\rangle + \varepsilon_n|dn\rangle. \tag{4.14}$$

Taking inner product with $|m\rangle$ which satisfies $\mathcal{H}|m\rangle = \varepsilon_m|m\rangle$ and using $\langle m|n\rangle = \delta_{mn}$, we get

$$(\varepsilon_n - \varepsilon_m)\langle m|dn\rangle = \langle m|d\mathcal{H}|n\rangle - \delta_{mn}d\varepsilon_n. \tag{4.15}$$

Since the Hamiltonian $\mathcal{H}$ is Hermitian by definition, its eigenvectors span the Hilbert space, so that we can expand $|dn\rangle$ in the eigenbasis of $\mathcal{H}$ as $|dn\rangle = \sum_m \langle m|dn\rangle$. If $|n\rangle$ is nondegenerate, then $\varepsilon_m \neq \varepsilon_n$ for any $m \neq n$. Using eq. (4.15) with $m \neq n$, we can compute the abelian Berry curvature as

$$\begin{aligned}
\mathfrak{F}_n &= i\langle dn| \wedge |dn\rangle = i\sum_{mm'}\langle m'|m\rangle\langle dn|m'\rangle \wedge \langle m|dn\rangle = i\sum_m \langle m|dn\rangle^* \wedge \langle m|dn\rangle \\
&= i\sum_{m\neq n}\frac{\langle m|d\mathcal{H}|n\rangle^*}{\varepsilon_n - \varepsilon_m} \wedge \frac{\langle m|d\mathcal{H}|n\rangle}{\varepsilon_n - \varepsilon_m} + i\langle n|dn\rangle^* \wedge \langle n|dn\rangle \\
&= i\sum_{m\neq n}\frac{\langle n|d\mathcal{H}|m\rangle \wedge \langle m|d\mathcal{H}|n\rangle}{(\varepsilon_n - \varepsilon_m)^2},
\end{aligned} \tag{4.16}$$

where $\langle n|dn\rangle^* \wedge \langle n|dn\rangle = \mathfrak{a}_n \wedge \mathfrak{a}_n = 0$, since $\mathfrak{a}_n$ is an abelian 1-form. Clearly, the Berry curvature diverges if ε_n becomes degenerate with any other ε_m at some point in the parameter space $\mathcal{M}$.

This expression for Berry curvature in eq. (4.16) is strongly reminiscent of the second order perturbation theory correction to the energy ε_n, if $d\mathcal{H}$ could be thought of as the 'perturbation'. Thus, the Berry curvature can be thought of as encoding the effect of all the virtual transitions to other eigenstates of the system.

The second approach defines the Berry curvature in terms of projectors. Consider the eigenspace of the Hamiltonian with eigenvalue $\varepsilon_n(\boldsymbol{\lambda})$ and degeneracy r, and define the projector $P_n(\boldsymbol{\lambda})$ to the eigenspace as

$$P_n = \sum_{\alpha=1}^{r} |n_\alpha\rangle\langle n_\alpha|, \quad P_n^2 = P_n. \tag{4.17}$$

Then,

$$dP_n \wedge dP_n = \sum_{\alpha\beta} \Big(|dn_\alpha\rangle\langle n_\alpha| + |n_\alpha\rangle\langle dn_\alpha| \Big) \wedge \Big(|dn_\beta\rangle\langle n_\beta| + |n_\beta\rangle\langle dn_\beta| \Big). \tag{4.18}$$

To reduce this down to the nonabelian Berry connection form, we shall need to contract $|dn_\alpha\rangle$ with some $\langle n_\gamma|$. We can multiply with P_n on left and $\mathbb{1} = \sum_m |m\rangle\langle m|$ on the right to get

$$\begin{aligned} P_n dP_n \wedge dP_n \mathbb{1} &= \sum_{\gamma,m} |n_\gamma\rangle \, \langle n_\gamma | dP_n \wedge dP_n | m\rangle \, \langle m| \\ &= \sum_{\alpha\beta\gamma m} |n_\gamma\rangle \Big[\langle n_\gamma | dn_\alpha\rangle \wedge \langle n_\alpha | dn_\beta\rangle \langle n_\beta | m\rangle + \langle n_\gamma | dn_\alpha\rangle \wedge \langle n_\alpha | n_\beta\rangle \langle dn_\beta | m\rangle \\ &\qquad + \langle n_\gamma | n_\alpha\rangle \langle dn_\alpha | \wedge | dn_\beta\rangle \langle n_\beta | m\rangle + \langle n_\gamma | n_\alpha\rangle \langle dn_\alpha | n_\beta\rangle \wedge \langle dn_\beta | m\rangle \Big] \langle m| \end{aligned} \tag{4.19}$$

The second and fourth terms cancel out using the orthogonality $\langle n_\alpha | n_\beta\rangle = \delta_{\alpha\beta}$, as

$$\sum_{\alpha\gamma,m} \langle n_\gamma | dn_\alpha\rangle \wedge \langle dn_\alpha | m\rangle + \sum_{\alpha\beta,m} \langle dn_\alpha | n_\beta\rangle \wedge \langle dn_\beta | m\rangle = \sum_{\alpha\beta,m} d\langle n_\alpha | n_\beta\rangle \wedge \langle dn_\beta | m\rangle = 0. \tag{4.20}$$

Using the fact that $\langle n_\beta | m\rangle = 1$ iff $n_\beta = m$ and zero otherwise to sum over m, we are left with

$$\begin{aligned} P_n dP_n \wedge dP_n &= \sum_{\alpha\beta\gamma} |n_\gamma\rangle \Big[\langle n_\gamma | dn_\alpha\rangle \wedge \langle n_\alpha | dn_\beta\rangle + \delta_{\alpha\gamma} \langle dn_\alpha | \wedge | dn_\beta\rangle \Big] \langle n_\beta| \\ &= \sum_{\alpha\beta\gamma} |n_\gamma\rangle \Big[(-i\mathfrak{a}_{n,\gamma\alpha}) \wedge (-i\mathfrak{a}_{n,\alpha\beta}) + \delta_{\alpha\gamma} d(-i\mathfrak{a}_{n,\alpha\beta}) \Big] \langle n_\beta| \end{aligned} \tag{4.21}$$

Thus,

$$iP_n dP_n \wedge dP_n = \sum_{\beta\gamma} |n_\gamma\rangle \left[d\mathfrak{a}_{n,\gamma\beta} - i \sum_\alpha \mathfrak{a}_{n,\gamma\alpha} \wedge i\mathfrak{a}_{n,\alpha\beta} \right] \langle n_\beta| = \sum_{\beta\gamma} |n_\gamma\rangle \, \mathfrak{F}_{n,\gamma\beta} \, \langle n_\beta|. \tag{4.22}$$

This expression can be used to compute the Berry curvature 2-form using the eigenspace projectors.

4.2 Berry curvature and Chern numbers

The Berry phase introduced in the last section is naturally associated with geometric and topological constructions. In the first subsection, we discuss its geometric formulation in the language of holonomies on complex vector bundles, initially proposed by Barry Simon[44] at the same time as Michael Berry's paper. We also discuss the associated topological invariants, *viz*, the Chern character. In the next three subsections, we shall used the ideas introduced in this and the previous sections to compute the Berry curvature and the associated Chern numbers for various quantum mechanical systems relevant to this dissertation.

4.2.1 Geometrical interpretation

The gauge theories are most naturally described in terms of connections and curvatures of fiber bundles, as discussed in Sec 3.1. We now discuss the fiber bundles associated with the eigenstates of a Hamiltonian $\mathcal{H}\colon \mathscr{H} \to \mathscr{H}$ dependent on parameters $\boldsymbol{\lambda} \in \mathscr{M}$, the associated connection and curvature being $\mathfrak{a}$ and $\mathfrak{F}$, respectively. In this language, Berry phase is simply the holonomy of the Berry connection[44]. More explicitly, a closed loop on the base space $\mathscr{M}$, when lifted to the total space (i.e, the fiber bundle), may not close anymore; this failure to close, i.e, failure to return to the original state after traversing a loop in the parameter space, is precisely the Berry phase.

Given the Hilbert space $\mathcal{H}$ and the parameter space $\mathscr{M}$, the space of all possible states in the theory is simply $\mathscr{M} \times \mathscr{H}$, which can be thought of as a (trivial) complex vector bundle $\mathfrak{C} = \mathscr{M} \times \mathscr{H} \xrightarrow{\pi} \mathscr{M}$, where $\pi(\boldsymbol{\lambda}, |\psi\rangle) \mapsto \boldsymbol{\lambda}$ is the projection to the base space. Given a basis $|\mathbf{e}_i\rangle$ of $\mathscr{H}$, we can define $\Psi\colon \mathscr{M} \to \mathbb{C}^n$ as $[\Psi(\boldsymbol{\lambda})]_i = \langle \mathbf{e}_i | \psi(\lambda)\rangle$. To parallel transport vectors $|\psi\rangle$ on this space, we define two vectors as "parallel" if their projection along each of the basis vectors $|\mathbf{e}_i\rangle$ is equal. Then, $|\psi(\boldsymbol{\lambda})\rangle$ is a parallel transport of $(\boldsymbol{\lambda}_0, |\psi_0\rangle \in \mathfrak{C}$ if $\langle \mathbf{e}_i | \psi(\boldsymbol{\lambda})\rangle = \langle \mathbf{e}_i | \psi(\boldsymbol{\lambda}')\rangle \,\forall\, i, \boldsymbol{\lambda}, \boldsymbol{\lambda}'$.

Equivalently, the parallel transport can be stated as $\mathcal{D}\Psi = 0$, where $\mathcal{D} = d$ in this case. The corresponding curvature is given by $\mathfrak{F} = \mathcal{D}^2$, which vanishes since $d^2 = 0$. Thus, the connection is flat, which is expected to be the case as $\mathfrak{C}$ is trivial. The covariant derivative $\mathcal{D}$ can also be thought of as acting on functions, i.e, local smooth section $\Psi\colon U \to \mathbb{C}^r$, where $U \in \mathscr{M}$ is an open set on which $\mathfrak{C}_n$ is trivialized. With a slight abuse of notation, we shall hereafter denote the condition for parallel transport simply as $d|\psi\rangle = 0$.

Consider the eigenspace of $\mathcal{H}(\boldsymbol{\lambda})$ corresponding to energy $\varepsilon_n(\boldsymbol{\lambda})$ which is r-fold degenerate for all $\boldsymbol{\lambda} \in \mathscr{M}$, and define the corresponding projector $P_n(\boldsymbol{\lambda})$, as in eq. (4.17). We construct a rank-r subbundle of $\mathfrak{C}_n \subset \mathfrak{C}$ by simply projecting applying the projectors $P_n(\boldsymbol{\lambda})$ on each fiber $\pi^{-1}(\boldsymbol{\lambda})$. The new fiber bundle, $\mathfrak{C}_n \xrightarrow{\pi_n} \mathscr{M}$, has fibers $\pi_n^{-1}(\boldsymbol{\lambda}) \cong \mathbb{C}^r$. The connection on $\mathfrak{C}$ defined above induces a connection on $\mathfrak{C}_n$, which, in general,

is not flat. This is precisely the (nonabelian, in general) Berry connection $\mathcal{D} = d - i\mathfrak{a}$. The corresponding curvature is given by

$$i\mathcal{D}^2 = i\,(d - i\mathfrak{a}) \wedge (d - i\mathfrak{a}) = d\mathfrak{a} - i\mathfrak{a} \wedge \mathfrak{a} = \mathfrak{F}. \tag{4.23}$$

Given the curvature 2-form $\mathfrak{F}$, one could integrate $\mathfrak{F}^n$ over $\mathscr{M}' \subset \mathscr{M}$, a nontrivial $2n$-dimensional cycle of $\mathscr{M}$, to get a number. These are the Chern numbers, which turn out to be integers and are topological invariants of the bundle. A nonzero Chern number is an obstruction to the definition of a flat connection for the given curvature. If that is the case, then a parallel transport of any vector along a closed loop would lead to a nontrivial holonomy, i.e, a Berry phase. Explicitly, for $\mathscr{M}' \in Z^{2n}(\mathscr{M})$, the Chern character for subbundle corresponding to energy ε_n is defined as

$$C_n = \int_{\mathscr{M}'} \mathrm{tr}\left\{\left(\frac{\mathfrak{F}}{2\pi}\right)^n\right\}. \tag{4.24}$$

In the next sections, we shall compute the Berry curvature for the given examples.

By definition, if a bundle supports a flat connection, then $\mathfrak{F} = 0$, so that the associated Chern character is zero. Thus, the Chern number of the original bundle $\mathfrak{C} = \mathscr{M} \times \mathscr{H}$ is zero, since $\mathfrak{C}$ is trivial. However, $\mathfrak{C}$ can be written as a direct sum of subbundles corresponding to all eigenspaces of a Hamiltonian, and the Chern characters simply add under this direct sum. Thus, we arrive at the important result that the sum of Chern numbers for all eigenspaces must vanish.

4.2.2 Two band models

The simplest situation where we can get a Berry curvature is a 2-dimensional Hilbert space $\mathscr{H} \cong \mathbb{C}^2$. The Hamiltonian $\mathcal{H}\colon \mathscr{H} \to \mathscr{H}$ is then a Hermitian operator, which can be written as a 2×2 matrix

$$\mathcal{H} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad \alpha, \delta \in \mathbb{R}, \;\; \beta^* = \gamma. \tag{4.25}$$

The Hamiltonian is conveniently parametrized using the Pauli matrices as

$$\mathcal{H} = h_0 \mathbb{1}_2 + \mathbf{h} \cdot \boldsymbol{\sigma}, \qquad \sigma\left[\mathcal{H}\right] = h_0 \pm |\mathbf{h}|\,, \tag{4.26}$$

where we have set $h_0 = \frac{1}{2}(\alpha + \delta)$, $h_1 = \mathrm{Re}[\beta]$, $h_2 = \mathrm{Im}[\beta]$ and $h_3 = \frac{1}{2}(\alpha - \delta)$, which depend on a parameter $\boldsymbol{\lambda} \in \mathscr{M}$. Since there are only two bands, the sum of their curvatures must vanish, so that we shall only compute the curvature associated with $|\psi\rangle$, the eigenstate with eigenvalue $\varepsilon_+ = h_0 + |\mathbf{h}|$. We shall find it

convenient to write $\mathbf{h}\colon \mathscr{M} \to \mathbb{R}^3$ in spherical polar coordinates as

$$\mathbf{h} = |\mathbf{h}|\,(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \iff \tan\theta = \frac{\sqrt{h_1^2+h_2^2}}{h_3}, \quad \tan\phi = \frac{h_1}{h_2}, \tag{4.27}$$

where $(\theta,\phi)\colon \mathscr{M} \to S^2$. The Hamiltonian can explicitly be written as

$$\mathcal{H} = \begin{pmatrix} h_0+h_3 & h_1 - ih_2 \\ h_1+ih_2 & h_0-h_3 \end{pmatrix} = h_0\mathbb{1}_2 + |\mathbf{h}| \begin{pmatrix} \cos\theta & \sin\theta\, e^{-i\phi} \\ \sin\theta\, e^{i\phi} & -\cos\theta \end{pmatrix}. \tag{4.28}$$

In this case, we can compute the Berry curvature directly from its definition. Setting $(\mathcal{H} - h_0 - |\mathbf{h}|)\,|\psi\rangle = 0$, the normalized eigenvector is given by

$$|\psi\rangle = \frac{1}{\sqrt{\sin^2\theta + (1-\cos\theta)^2}} \begin{pmatrix} \sin\theta\, e^{-i\phi} \\ 1-\cos\theta \end{pmatrix} = \begin{pmatrix} \cos\frac{\theta}{2}\, e^{-i\phi} \\ \sin\frac{\theta}{2} \end{pmatrix}. \tag{4.29}$$

Thus,

$$|d\psi\rangle = \frac{1}{2}\begin{pmatrix} -\sin\frac{\theta}{2}\, e^{-i\phi} \\ \cos\frac{\theta}{2} \end{pmatrix} d\theta - i \begin{pmatrix} \cos\frac{\theta}{2}\, e^{-i\phi} \\ 0 \end{pmatrix} d\phi, \tag{4.30}$$

and the Berry connection and curvature becomes

$$\begin{aligned} \mathfrak{a}_+ &= i\langle\psi|d\psi\rangle = \frac{i}{2}\left(-\cos\frac{\theta}{2}\sin\frac{\theta}{2} + \sin\frac{\theta}{2}\cos\frac{\theta}{2}\right) d\theta + \cos^2\frac{\theta}{2}\, d\phi = \frac{1}{2}\,(1+\cos\theta)\, d\phi, \\ \mathfrak{F}_+ &= d\mathfrak{a}_+ = -\frac{1}{2}\sin\theta\, d\theta\wedge d\phi = -\frac{1}{2}\Omega(S^2), \end{aligned} \tag{4.31}$$

where $\Omega(S^2)$ is the volume form on S^2. To rewrite this in terms of the components of $\mathcal{H}$, we use the definition of the polar coordinates from eq. (4.27) to get

$$\begin{aligned} d\theta &= \cos^2\theta\, d(\tan\theta) = \frac{h_3^2}{|\mathbf{h}|^2} d\left(\frac{\sqrt{h_1^2+h_2^2}}{h_3}\right), \\ d\phi &= \cos^2\phi\, d(\tan\phi) = \frac{h_1^2}{|\mathbf{h}|^2 - h_3^2} d\left(\frac{h_2}{h_1}\right), \end{aligned} \tag{4.32}$$

with $|\mathbf{h}|^2 = h_1^2+h_2^2+h_3^2$. The Berry connection and curvature become

$$\begin{aligned} \mathfrak{a}_+ &= \frac{h_1^2}{2\,|\mathbf{h}|\,(|\mathbf{h}|+h_3)} d\left(\frac{h_2}{h_1}\right) = \frac{h_1 dh_2 - h_2 dh_1}{2\,|\mathbf{h}|\,(|\mathbf{h}|+h_3)}, \\ \mathfrak{F}_+ &= -\frac{1}{2\,|\mathbf{h}|^3}\,(h_1 dh_2\wedge dh_3 + h_2 dh_3\wedge dh_1 + h_3 dh_1\wedge dh_2), \end{aligned} \tag{4.33}$$

where the latter is again ($-1/2$ times) the area form for a 2-sphere in Cartesian coordinates.

If $\mathscr{M}$ is 2-dimensional, the (first) Chern number associated with $\mathfrak{F}$ is given by

$$C_1 = \frac{1}{2\pi}\int_{\mathscr{M}} \mathfrak{F} = -\frac{1}{4\pi}\int_{\mathscr{M}} \varphi^*\Omega(S^2), \tag{4.34}$$

where $\varphi\colon \mathscr{M} \to S^2 : \boldsymbol{\lambda} \mapsto (\theta, \phi)$. Thus, geometrically, the Berry curvature is the pullback of half the volume form on S^2 under φ, and the Chern number is simply the *Brouwer degree* ("winding number") of this map, which must be an integer[70].

4.2.3 Weyl fermions

The Weyl Hamiltonian in $(2N+1)+1$ dimensions are described by the Hamiltonian

$$\mathcal{H}(\mathbf{p}) = \chi\, p_i \Gamma^i, \qquad \sigma\left[\mathcal{H}\right] = \{\pm|\mathbf{p}|\}, \tag{4.35}$$

where $\chi = \pm 1$ is the chirality and $\Gamma^i \in \mathrm{Mat}(2^N, \mathbb{C})$ are a set of anticommuting traceless matrices. Here, Γ_i, $i = 1, \dots 2N$ can be taken as the Dirac matrices in $2N$ spacetime dimensions, while $\Gamma_{2N+1} \equiv (-i)^N \prod_{i=1}^{2N} \Gamma_N$ is the analogue of "γ_5". The Hilbert space is $\mathscr{H} = \mathbb{C}^{2^N}$, so that for $N > 1$, each state is 2^{N-1}-fold degenerate. Furthermore, the two bands touch at $\mathbf{p} = 0$, where the Berry curvature diverges.

Physically, the Weyl Hamiltonian describes the dynamics of a fermion whose spin is locked to its momentum. This is particularly transparent for the $N = 1$, i.e, the $3+1$ dimensional case, where the Weyl Hamiltonian becomes $\mathcal{H} = \mathbf{p}\cdot\boldsymbol{\sigma} = 2\mathbf{p}\cdot\mathbf{S}$, where $\mathbf{S}$ is the spin operator. Thus, restricting to the positive energy subspace is equivalent to demanding that $\mathbf{S}$ be "parallel to" $\mathbf{p}$. Then, as the momentum unit vector $\hat{\mathbf{p}}$ traces a closed curve on S^2, the spin precesses along with it, thereby acquiring a phase, which is precisely the Berry's phase. We also deduce that the gauge group must be the space of phases, i.e, U(1).

For $N > 1$, the situation is more complicated, as in principle, the gauge group should be $\mathrm{U}(2^{N-1})$, the space of all possible choices of basis for the degenerate subspace, but in practice, a parallel transport does not explore the entirety of that space. Instead, since the wavefunctions of the Weyl Hamiltonian are spinors, under a parallel transport of $\mathbf{p}$ they can only transform under some representation of $\mathrm{Spin}(2N) \subset \mathrm{U}(2^{N-1})$. A more physical interpretation is to notice that $\mathbb{R}^{2N+1}$ is invariant under $\mathrm{SO}(2N+1)$, and fixing a "spin vector" along a specific $\hat{\mathbf{p}}$ direction on S^{2N} still leaves it invariant under $\mathrm{SO}(2N)$. Since the spinors transform under $\mathrm{Spin}(2N)$, the double cover of $SO(2N)$, that must be the gauge group of the nonabelian Berry curvature for $N > 1$. We do not have this distinction for $N = 1$ owing to the accidental isomorphism $\mathrm{Spin}(2) \cong \mathrm{SO}(2) \cong \mathrm{U}(1)$.

The parameter space is $\mathbb{R}^{2N+1}\backslash\{0\}$, which has a deformation retract to S^{2N}. Thus, $S^{2N}(R) \in \mathscr{M}$, a sphere of radius R centered at $\mathbf{p} = \mathbf{0} \in \mathbb{R}^{2N+1}$, is the only nontrivial 2-cycle in $\mathscr{M}$ on which we can integrate $\mathfrak{F}^N$ to compute a Chern number. However, under a rescaling $\mathbf{p} \to \lambda\mathbf{p},\ \lambda \in \mathbb{R}$, the eigenvalues change as $\chi\,|\mathbf{p}| \to \lambda\chi\,|\mathbf{p}|$, but the eigenstates stay constant, so that the Berry curvature, which only depends on the geometry of the vector bundle of states on $S^{2N}(R)$, is independent of R. Thus, we shall compute the integral of $\mathfrak{F}^N$ over the unit $2N$-sphere.

For $N = 1$, i.e, in $3+1$ dimensions, the Weyl Hamiltonian is simply $\chi\mathbf{p}\cdot\boldsymbol{\sigma}$, which is a two band model. Physically, the Hamiltonian describes the spin of the Weyl particle, which is locked parallel/antiparallel to the momentum $\mathbf{p}$, and thus rotates as we change the momentum. For this case, the Berry curvature 2-form for positive energy band can be computed using the result for 2-band models, with $h_0 = 0,\ \mathbf{h}(\mathbf{p}) = \chi\mathbf{p}$, as

$$\mathfrak{F}_\chi = -\frac{\chi}{4|\mathbf{p}|^2}\epsilon_{kij}\,\hat{p}^k\,dp^i dp^j = -\frac{\chi}{4}\epsilon_{kij}\,\hat{p}^k\,d\hat{p}^i d\hat{p}^j = -\frac{\chi}{2}\,\Omega(S^2), \tag{4.36}$$

where we have used $p^j = |\mathbf{p}|\,\hat{p}^j$ and the antisymmetry of the Levi-Civita tensor, and identified $\Omega(S^2)$, the area form for the 2-sphere, in the last step. Alternatively, since $\mathscr{M} = \mathbb{R}^3$ is 3-dimensional, the Hodge dual of $\mathfrak{F}$ is a 1-form, which corresponds to a 3-vector $\mathfrak{b}_\chi$, defined as

$$\mathfrak{F}_\chi = \frac{1}{2}\mathfrak{F}_{\chi,ij}dp^i \wedge dp^j = \frac{1}{2}\epsilon_{kij}\mathfrak{b}_{\chi,k}dp^i \wedge dp^j, \qquad \mathfrak{b}_\chi = -\chi\frac{\hat{\mathbf{p}}}{2|\mathbf{p}|^2}, \tag{4.37}$$

which corresponds to the field of a monopole[85] of strength $-1/2$, centered at the origin in $\mathbb{R}^3$. We can explicitly compute the Chern number associated with S^2 as

$$C_1 = \int_{S^2}\frac{\mathfrak{F}}{2\pi} = -\frac{1}{4\pi}\int_{S^2}\Omega(S^2) = -\chi. \tag{4.38}$$

For $N > 1$, the explicit computation of Berry curvature would be quite ugly; however, we can readily compute the Chern number using the definition of the Berry curvature in terms of projectors P_χ to the $\varepsilon = \chi\,|\mathbf{p}|$ subspace, defined as

$$P_\chi = \sum_{\alpha=1}^{2^{N-1}} |\chi,\mathbf{p},\alpha\rangle\langle\mathbf{p},\chi,\alpha| = \frac{1}{2}\left(\mathbb{1} + \chi\hat{p}^i\Gamma_i\right). \tag{4.39}$$

From eq. (4.22), the components of the Berry curvature are given by

$$iP_\chi\,dP_\chi \wedge dP_\chi = \sum_{\alpha\beta}|\mathbf{p},\chi,\alpha\rangle\mathfrak{F}_{\alpha\beta}\langle\mathbf{p},\chi,\beta|, \tag{4.40}$$

from which we can directly compute

$$\begin{aligned}\mathrm{tr}\left\{\left(iP_\chi\, dP_\chi \wedge dP_\chi\right)^N\right\} &= \sum_{\chi'=\pm 1}\sum_{\gamma=1}^{2^{N-1}} \langle \mathbf{p},\chi'\gamma|\left(iP_\chi\, dP_\chi \wedge dP_\chi\right)^N|\mathbf{p},\chi'\gamma\rangle, \\ &= \sum_{\alpha_i,\gamma=1}^{2^{N-1}} \mathfrak{F}_{\chi,\gamma\alpha_1}\wedge\mathfrak{F}_{\chi,\alpha_1\alpha_2}\wedge\cdots\wedge\mathfrak{F}_{\chi,\alpha_{N-1}\gamma} = \mathrm{tr}\left\{\mathfrak{F}_\chi^N\right\}. \end{aligned} \tag{4.41}$$

where we have used the orthogonality $\langle \mathbf{p},\chi,\alpha|\mathbf{p},\chi',\beta\rangle = \delta_{\chi\chi'}\delta_{\alpha\beta}$. We can use the properties of the Dirac matrices to commute them across $dP\chi\wedge dP_\chi$, since

$$dP_\chi\, P_\chi = P_{-\chi}\, dP_\chi \implies dP_\chi\wedge dP_\chi\, P_\chi = P_\chi\, dP_\chi\wedge dP_\chi. \tag{4.42}$$

Explicitly, this follow from

$$dP_\chi\, P_\chi = \frac{1}{4}d\hat{p}^i\,\Gamma_i\left(\mathbb{1}+\hat{p}^j\Gamma_j\right) = \frac{1}{4}\left(d\hat{p}^i\Gamma_i + \Gamma_i\Gamma_j\hat{p}^j d\hat{p}^i\right) = \frac{1}{4}\Big(d\hat{p}^i\Gamma_i + \left(2\delta_{ij}-\Gamma_j\Gamma_i\right)\hat{p}^j d\hat{p}^i\Big). \tag{4.43}$$

since the Dirac matrices satisfy $[\Gamma_i\Gamma_j]_+ = 2\delta_{ij}$, and $\delta_{ij}\hat{p}^j d\hat{p}^i = \hat{p}_i d\hat{p}^i = d\,|\hat{p}|^2 = 0$. Thus, using $P_\chi^2 = P_\chi$, we get

$$\mathrm{tr}\left\{\mathfrak{F}_\chi^N\right\} = i^N\mathrm{tr}\left\{\left(P_\chi\, dP_\chi\wedge dP_\chi\right)^N\right\} = i^N\mathrm{tr}\left\{P_\chi\left(dP_\chi\wedge dP_\chi\right)^N\right\}. \tag{4.44}$$

Substituting the explicit form of P_χ and using $\chi^2=1$, this becomes

$$\begin{aligned}\mathrm{tr}\left\{\mathfrak{F}_\chi^N\right\} &= i^N\mathrm{tr}\left\{\frac{1}{2}\left(\mathbb{1}+\chi\,\hat{p}^i\Gamma_i\right)\left[\frac{\chi}{4}\Gamma_j d\hat{p}^j\wedge\ \frac{\chi}{4}\Gamma_k d\hat{p}^k\right]^N\right\} \\ &= \frac{1}{2}\left(\frac{i}{4}\right)^N\left[\mathrm{tr}\left\{\prod_{\ell=1}^{2N}\Gamma_{i_\ell}\right\} + \chi\,\hat{p}^{i_0}\mathrm{tr}\left\{\prod_{\ell=0}^{2N}\Gamma_{i_\ell}\right\}\right]\bigwedge_{\ell=1}^{2N} d\hat{p}^{i_\ell}. \end{aligned} \tag{4.45}$$

The measure demands that Γ^{i_ℓ} be all different for $\ell = 1,\ldots 2N$, since $dp^i\wedge dp^j = 0$ if $i=j$. For a product of $1<k<2N+1$ different Γ matrices, using their antisymmetry as well as the cyclic property of trace, we get

$$T_k = \mathrm{tr}\left\{\Gamma_{i_1}\Gamma_2\ldots\Gamma_{i_k}\right\} = \mathrm{tr}\left\{\Gamma_2\ldots\Gamma_{i_k}\Gamma_{i_1}\right\} = (-1)^{k-1}\mathrm{tr}\left\{\Gamma_{i_1}\Gamma_2\ldots\Gamma_{i_k}\right\} = (-1)^{k-1}T_k, \tag{4.46}$$

so that $T_k = -T_k \implies T_k = 0$ if k is even. Thus, the first trace in eq. (4.45) must vanish. For the second trace, we must have one copy of each Γ matrix, and using

$$\prod_{i=1}^{2N+1}\Gamma_i = \left(\prod_{i=1}^{2N}\Gamma_i\right)\Gamma_{2N+1} = \left(\prod_{i=1}^{2N}\Gamma_i\right)(-i)^N\left(\prod_{j=1}^{2N}\Gamma_j\right) = (-i)^N(-1)^{N(2N-1)}\prod_{i=1}^{2N}\Gamma_i^2 = i^N\mathbb{1}_{2^N}, \tag{4.47}$$

we get

$$\mathrm{tr}\left\{\prod_{\ell=0}^{2N}\Gamma_{i_\ell}\right\} = \epsilon_{i_0 i_1 \ldots i_{2N}} \mathrm{tr}\left\{i^N \mathbb{1}_{2^N}\right\} = (2i)^N \epsilon_{i_0 i_1 \ldots i_{2N}}. \tag{4.48}$$

Thus,

$$\mathrm{tr}\left\{\mathfrak{F}_\chi^N\right\} = \frac{\chi}{2}\left(\frac{i}{4}\right)^N (2i)^N \epsilon_{i_0 i_1 \ldots i_{2N}} \bigwedge_{\ell=0}^{2N} \hat{p}^{i_\ell} = \chi(-1)^N \frac{(2N)!}{2^{N+1}}\, \Omega(S^{2N}), \tag{4.49}$$

where we have identified the volume form on S^{2N}, defined as

$$\Omega(S^{2N}) = \sum_{j=1}^{2N+1} (-1)^j d\hat{p}^j \bigwedge_{i\neq j} dp^i = \frac{1}{(2N)!}\epsilon_{i_0 i_1 \ldots i_{2N}} \hat{p}^{i_0}\, d\hat{p}^{i_1} \wedge \cdots \wedge \hat{p}^{i_{2N}}. \tag{4.50}$$

Thus, the Chern number over $\mathscr{M} = S^{2N}$ is given by

$$C_\chi = \frac{1}{N!}\int_{S^{2N}} \mathrm{tr}\left\{\left(\frac{\mathfrak{F}_\chi}{2\pi}\right)^N\right\} = \chi \frac{(-1)^N}{2}\, \frac{(2N)!}{(4\pi)^N N!}\, \mathrm{Vol}(S^{2N}), \tag{4.51}$$

But the volume of S^{2N} is given by

$$\mathrm{Vol}(S^{2N}) = \frac{2\,\pi^{N+\frac{1}{2}}}{\mathbf{\Gamma}\left(N+\frac{1}{2}\right)} = \frac{2^{2N}\,\pi^N\,\mathbf{\Gamma}(N)}{\mathbf{\Gamma}(2N)} = (4\pi)^N \frac{(N-1)!}{(2N-1)!} = 2(4\pi)^N \frac{N!}{(2N)!} \tag{4.52}$$

where $\mathbf{\Gamma}(.)$ is the Euler's gamma function, and we have used the duplication formula

$$\mathbf{\Gamma}(z)\mathbf{\Gamma}\left(z+\frac{1}{2}\right) = 2^{1-2z}\sqrt{\pi}\,\mathbf{\Gamma}(2z), \tag{4.53}$$

Substituting eq. (4.52) in eq. (4.51), the Chern number becomes

$$C_\chi = \frac{1}{N!}\int_{S^{2N}} \mathrm{tr}\left\{\left(\frac{\mathfrak{F}_\chi}{2\pi}\right)^N\right\} = \chi(-1)^N. \tag{4.54}$$

As expected, $C_+ + C_- = 0$, since the two subspaces add up to a trivial bundle over S^{2N}, as discussed at the end of Sec 4.2.1.

4.2.4 Circularly polarized light

As our last example, we compute the Berry connection and curvature for circularly polarized light in $3+1$ dimensions, which can be thought of as a 'classical' geometric phase. For a wave propagating along the $\hat{\mathbf{k}}$

direction, the complex electric field is given by

$$\mathcal{E}(\mathbf{r},t) = \hat{\mathbf{e}}^{\sigma} e^{i\omega t - i\mathbf{k}\cdot\mathbf{r}}, \quad \hat{\mathbf{e}}^{\sigma} = \frac{1}{\sqrt{2}}(\hat{\mathbf{n}}_1 + i\sigma\,\hat{\mathbf{n}}_2), \tag{4.55}$$

where $\sigma = \pm 1$ for right/left circular polarization, and $\{\hat{\mathbf{n}}_1, \hat{\mathbf{n}}_2, \hat{\mathbf{k}}\}$ form a right-handed orthonormal basis such that $\hat{\mathbf{n}}_1 \times \hat{\mathbf{n}}_2 = \hat{\mathbf{k}}$. We seek to compute the Berry connection

$$\begin{aligned}
\mathfrak{a}^{\sigma\sigma'} &= i\hat{\mathbf{e}}^{\sigma *} \cdot d\hat{\mathbf{e}}^{\sigma'} = \frac{i}{2}(\hat{\mathbf{n}}_1 - i\sigma\,\hat{\mathbf{n}}_2)\cdot(d\hat{\mathbf{n}}_1 + i\sigma'\,d\hat{\mathbf{n}}_2) \\
&= \frac{1}{2}\left(\sigma\hat{\mathbf{n}}_2 \cdot d\hat{\mathbf{n}}_1 - \sigma'\hat{\mathbf{n}}_1 \cdot d\hat{\mathbf{n}}_2\right) + \frac{i}{2}\left(\hat{\mathbf{n}}_1 \cdot d\hat{\mathbf{n}}_1 + \sigma\sigma'\hat{\mathbf{n}}_2 \cdot d\hat{\mathbf{n}}_2\right) \\
&= \frac{1}{2}\left(\sigma\hat{\mathbf{n}}_2 \cdot d\hat{\mathbf{n}}_1 - \sigma'\hat{\mathbf{n}}_1 \cdot d\hat{\mathbf{n}}_2\right),
\end{aligned}$$

where the last step follows from $\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_i = 1 \implies \hat{\mathbf{n}}_i \cdot d\hat{\mathbf{n}}_i = 0$. The Berry curvature becomes

$$\mathfrak{F}^{\sigma\sigma'} = -\frac{\sigma + \sigma'}{2} d\hat{\mathbf{n}}_1 \wedge d\hat{\mathbf{n}}_2 = -\delta^{\sigma\sigma'}\sigma\, d\hat{\mathbf{n}}_1 \wedge d\hat{\mathbf{n}}_2, \qquad [\sigma, \sigma' = \pm 1] \tag{4.56}$$

where we have a wedge product over the 1-forms (dk_i's) as well as a dot product over $\hat{\mathbf{n}}_i$'s. Clearly, $\mathfrak{F}^{\sigma\sigma'}$ is diagonal in polarization. Henceforth we shall simply write $\mathfrak{F}^{\sigma} = \sigma\, d\hat{\mathbf{n}}_1 \wedge d\hat{\mathbf{n}}_2$.

To compute $\mathfrak{F}^{\sigma}$ explicitly, using $\hat{\mathbf{n}}_i \cdot d\hat{\mathbf{n}}_i = 0$, we can write

$$\begin{aligned}
d\hat{\mathbf{n}}_1 &= (\hat{\mathbf{k}} \cdot d\hat{\mathbf{n}}_1)\hat{\mathbf{k}} + (\hat{\mathbf{n}}_2 \cdot d\hat{\mathbf{n}}_1)\hat{\mathbf{n}}_2 \\
d\hat{\mathbf{n}}_2 &= (\hat{\mathbf{k}} \cdot d\hat{\mathbf{n}}_2)\hat{\mathbf{k}} + (\hat{\mathbf{n}}_1 \cdot d\hat{\mathbf{n}}_2)\hat{\mathbf{n}}_1,
\end{aligned}$$

But

$$\hat{\mathbf{n}}_i \cdot \hat{\mathbf{k}} = 0 \implies \hat{\mathbf{k}} \cdot d\hat{\mathbf{n}}_i = -\hat{\mathbf{n}}_i \cdot d\hat{\mathbf{k}} = -\hat{\mathbf{n}}_i \cdot d\left(\frac{\mathbf{k}}{k}\right) = -\frac{1}{k}\hat{\mathbf{n}}_i \cdot \left(d\mathbf{k} - \frac{\mathbf{k}\cdot d\mathbf{k}}{k^2}\mathbf{k}\right) = -\frac{1}{k}\hat{\mathbf{n}}_i \cdot d\mathbf{k}, \tag{4.57}$$

so that

$$\begin{aligned}
\mathfrak{F}^{\sigma} &= -\sigma\,(\hat{\mathbf{k}} \cdot d\hat{\mathbf{n}}_1) \wedge (\hat{\mathbf{k}} \cdot d\hat{\mathbf{n}}_2) = -\frac{\sigma}{k^2}(\hat{\mathbf{n}}_1 \cdot d\mathbf{k}) \wedge (\hat{\mathbf{n}}_2 \cdot d\mathbf{k}) \\
&= -\frac{\sigma}{k^2} n_{1,i} n_{2,j} dk^i \wedge dk^j = -\frac{\sigma}{2k^2}\epsilon_{ijk}[\hat{\mathbf{n}}_1 \times \hat{\mathbf{n}}_2]_k dk^i \wedge dk^j, \\
&= -\frac{\sigma}{2k^2}\epsilon_{kij} k^k dk^i \wedge dk^j = -\sigma\,\Omega(S^2),
\end{aligned} \tag{4.58}$$

which corresponds to the strength of a monopole field with the first Chern number $C_1 = 2\sigma$. This corresponds

to the fact that the Weyl fermions had spin 1/2, while the photon has spin 1. In presence of both clockwise and counterclockwise polarization, the Chern numbers cancel, but we can still define a spin Chern number

$$C_1 = \sum_{\sigma=\pm 1} C_1^\sigma = 0, \qquad C_{1,\text{spin}} = \sum_{\sigma=\pm 1} \sigma C_1^\sigma = -4. \tag{4.59}$$

4.3 Anomalous dynamics

In condensed matter systems, the Berry curvature, discussed for generic systems in the last section, is realized in the electronic wavefunctions of crystalline solids. Thus, their band structures provide a useful platform to investigate the effects of the Berry's curvature on the physics of the system[22].

Berry curvature of Bloch bands

The quantum mechanical description of the electrons in a lattice in d-dimensions involves the Schrödinger equation for electrons, with a periodic background potential arising from the positively charged ions. Ignoring the electron-electron interactions, the 1-particle Schrödinger equation becomes

$$\mathcal{H}\psi(\mathbf{x}) = \left(-\frac{1}{2m}\nabla_{\mathbf{x}}^2 + V(\mathbf{x})\right)\psi(\mathbf{x}) = \varepsilon\psi(\mathbf{x}), \tag{4.60}$$

where the background potential is periodic, i.e, $V(\mathbf{x}+\mathbf{R}) = V(\mathbf{x})$ for any lattice vector $\mathbf{R}$. Thus, under a translation by $\mathbf{x} \to \mathbf{x}+\mathbf{R}$, the Hamiltonian goes to itself, but the wavefunctions can change by a phase, i.e, $\psi(\mathbf{x}+\mathbf{R}) = e^{i\gamma(R)}\psi(\mathbf{x})$. This follows from the Bloch(Floquet) theory for periodic differential equations. The solutions can be written as Bloch waves $\psi_{n,\mathbf{k}}(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}}u_{n,\mathbf{k}}(\mathbf{x})$, where n is the band index, $\mathbf{k}$ is the (quasi-)momentum and $u_{n,\mathbf{k}}(\mathbf{x}+\mathbf{R}) = u_{n,\mathbf{k}}(\mathbf{x})$. Under a lattice translations, the extra phase picked up is $e^{i\gamma(R)} = e^{i\mathbf{k}\cdot\mathbf{R}}$, which is invariant under $\mathbf{k} \to \mathbf{k}+\mathbf{Q}$, where $\mathbf{Q}$ is a reciprocal lattice vector. Thus, $\mathbf{k} \in$ BZ, the first Brillouin zone, which is simply a unit cell of the reciprocal lattice centered at $\mathbf{k} = \mathbf{0}$.

Given the form of the Bloch waves, one defines the *Bloch Hamiltonian*

$$\mathcal{H}_B(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{x}}\mathcal{H}e^{i\mathbf{k}\cdot\mathbf{x}} \implies \mathcal{H}_B(\mathbf{k})u_{n,\mathbf{k}}(\mathbf{k}) = \varepsilon_n(\mathbf{k})u_{n,\mathbf{k}}(\mathbf{k}). \tag{4.61}$$

Thus, we have defined a family of Hamiltonians, $\mathcal{H}_B(\mathbf{k})$ parametrized by $\mathbf{k} \in$ BZ, which has a discrete spectrum $\varepsilon_n(\mathbf{k})$ for each $\mathbf{k}$. Using the ideas from the last two sections and assuming nondegenerate bands,

we can define the Berry connection 1-form as

$$\mathfrak{a}_n(\mathbf{k}) = \langle u_{n,\mathbf{k}}| \left(i\frac{\partial}{\partial k_i} \right) |u_{n,\mathbf{k}}\rangle dk_i = \int_{\mathbb{R}^d} d\mathbf{x} u^*_{n,\mathbf{k}}(\mathbf{x}) \left(i\frac{\partial}{\partial k_i} \right) u_{n,\mathbf{k}}(\mathbf{x})\, dk_i, \tag{4.62}$$

and the corresponding curvature is $\mathfrak{F} = d\mathfrak{a}$.

Transport

A particular application of the Berry curvature in the physics of crystalline systems (in 3 space dimensions) is the electron transport in presence of an external electromagnetic field.

One could define semiclassical equations of motion, governing the dynamics of electrons as the time evolution of the expectation value of the position and momentum operators in various Bloch states. One particularly useful choice is that of wavepackets[22, 89, 88], which are finite-sized in both position and momentum, and can therefore be thought of as "classical" particles. This lets one use the classical kinetic theory approach (Sec 2.5) to study transport in crystals. Conventionally, the equations of motion for such electron wavepackets are described by the equations of motion[5]

$$\dot{\mathbf{x}} = \mathbf{v}_g, \qquad \dot{\mathbf{k}} = \mathbf{E} + \dot{\mathbf{x}} \times \mathbf{B}, \tag{4.63}$$

where $\mathbf{v}_g = -\nabla_{\mathbf{k}} \varepsilon_n(\mathbf{k})$ is the usual group velocity, and the second line is simply the classical Lorentz force. These equations have been useful in describing the electronic physics of metal and semiconductors.

However, including the effect of interband coupling, one obtains an extra term in $\dot{\mathbf{x}}$, and the equations of motion become

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{v}_g + \dot{\mathbf{k}} \times \mathfrak{b}, \\ \dot{\mathbf{k}} &= \mathbf{E} + \dot{\mathbf{x}} \times \mathbf{B}, \end{aligned} \tag{4.64}$$

where $\mathfrak{b}(\mathbf{k})$ is the Berry curvature of the band. The $\dot{\mathbf{k}} \times \mathfrak{b}$ is termed the *anomalous velocity*. The invoking of the Berry curvature in order to understand interband couplings can be motivated by eq. (4.16) and the comment after that, where Berry curvature can be thought of as encoding the residual effects of the other energy levels, which are inaccessible as one tunes $\mathbf{k}$, owing to the adiabatic approximation.

[5] As we shall restrict ourselves to dynamics in a single band, we shall omit explicit mention of the band index from now on.

Symplectic structure

In Ref [93], the semiclassical equations of motion of eq. (4.64) are shown to violate Liouville's theorem for the conservation of phase space volume using the conventional volume form d^3xd^3k on the $(\mathbf{x}, \mathbf{k})$ phase space. However, Liouville's theorem is restored if one instead redefines the volume form as $\sqrt{\rho}\, d^3xd^3k$, where $\sqrt{\rho} = 1 + \mathfrak{b} \cdot \mathbf{B}$. In [90], Duval and Horvathy show that this system can be described using the symplectic formulation of classical mechanics; however, the symplectic form is not given by $\rho = dk_i \wedge dx^i$. Thus, $\mathbf{x}$ and $\mathbf{k}$ are not the canonical variables in the phase space.

To derive the symplectic form, they begin by writing the original equations of motion as

$$\dot{x}^i - \epsilon^{ijk}\dot{k}^j\mathfrak{b}^k = v_g^i, \qquad \dot{k}^i - \epsilon^{ijk}\dot{x}^j B^k = E^i. \tag{4.65}$$

Defining $F^{ij}(\mathbf{x}) = \epsilon^{ijk}B^k$ and $\mathfrak{F}^{ij}(\mathbf{k}) = \epsilon^{ijk}\mathfrak{b}^k$ as 3×3 antisymmetric matrices corresponding to the (spatial part of) Maxwell and Berry curvature, and $\phi(\mathbf{x})$ as the Maxwell scalar potential, these become

$$\delta^{ij}\dot{x}^j - \mathfrak{F}^{ij}\dot{k}^j = \frac{\partial}{\partial k^i}\varepsilon, \qquad F^{ij}\dot{x}^j - \delta^{ij}\dot{k}^j = \frac{\partial}{\partial x^i}\phi. \tag{4.66}$$

These can finally be rewritten as

$$\begin{pmatrix} F & -\mathbb{1} \\ \mathbb{1} & -\mathfrak{F} \end{pmatrix}\begin{pmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{k}} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \mathbf{x}} \\ \frac{\partial}{\partial \mathbf{k}} \end{pmatrix}\mathcal{H}(\mathbf{x}, \mathbf{k}), \tag{4.67}$$

where the Hamiltonian is defined as $\mathcal{H}(\mathbf{x}, \mathbf{k}) = \varepsilon(\mathbf{k}) + \phi(\mathbf{x})$. Defining $\boldsymbol{\zeta} = (\mathbf{x}, \mathbf{k})$ and comparing with eq. (2.18), we identify the symplectic form as[6]

$$\rho = \frac{1}{2}\rho_{ij}d\zeta^i d\zeta^j, \quad \boldsymbol{\rho} = \begin{pmatrix} F & -\mathbb{1}_3 \\ \mathbb{1}_3 & -\mathfrak{F} \end{pmatrix}. \tag{4.68}$$

The usual coordinates $\mathbf{x}$ and $\mathbf{k}$ are not canonical, since for canonical coordinates, the Hamilton's equation should simply have $\boldsymbol{\rho} = \mathcal{J}^{-1}$ (see eq. (2.7)), which is not the case. The symplectic form can be written in a more enlightening form as

$$\rho = dk_i \wedge dx^i + \frac{1}{2}F_{ij}dx^i \wedge dx^j - \frac{1}{2}\mathfrak{F}^{ij}dk_i \wedge dk_j = dk_i \wedge dx^i + F - \mathfrak{F}. \tag{4.69}$$

The presence of this nontrivial symplectic form modifies the volume form and hence the Liouville's theorem.

[6] Our expression for $\boldsymbol{\rho}$ is different from that of Horvathy et al[90] since our Berry and Maxwell curvatures have opposite signs from theirs, the former by definition and the latter because we have set the electronic charge to $+1$.

For traditional classical mechanics, the volume form on phase space is simply

$$\Omega_0 = \frac{\omega_0^3}{3!} = \frac{1}{3!}(dk_i \wedge dx^i)^3 = d^3k\, d^3x \tag{4.70}$$

However, using the symplectic form of eq. (4.69), we can compute

$$\Omega = \frac{1}{3!}\left(dk_i \wedge dx^i + F - \mathfrak{F}\right)^3 = \frac{1}{3!}\Big((dk_i \wedge dx^i)^3 - 3\, dk_i \wedge dx^i \wedge F \wedge \mathfrak{F}\Big). \tag{4.71}$$

But

$$F = \frac{1}{2}F_{ij}dx^i \wedge dx^j = \frac{1}{2}\epsilon_{ijk}B^k\, dx^i \wedge dx^j, \qquad \mathfrak{F} = \frac{1}{2}\epsilon^{ijk}\mathfrak{b}_k\, dk_i \wedge dk_j. \tag{4.72}$$

Substituting and using $\epsilon_{ijk}dx^i \wedge dx^j \wedge dx^k = 3!\, d^3x$ etc, the volume form becomes

$$\Omega = \frac{1}{3!}\left(-3! - 3\,\epsilon_{ijk}B^j\epsilon^{ij\ell}\mathfrak{b}_\ell\right) d^3k \wedge d^3x = (1 + \mathfrak{b}\cdot\mathbf{B})\, d^3x \wedge d^3k, \tag{4.73}$$

which is precisely the volume form proposed in Ref [93]. Finally, the new Poisson brackets are

$$\{x^i, x^j\} = \frac{\epsilon^{ijk}\mathfrak{b}_k}{1 + \mathfrak{b}\cdot\mathbf{B}}, \quad \{x^i, k_j\} = \frac{\delta^i_j + B^i\mathfrak{b}_j}{1 + \mathfrak{b}\cdot\mathbf{B}}, \quad \{k_i, k_j\} = -\frac{\epsilon_{ijk}B^k}{1 + \mathfrak{b}\cdot\mathbf{B}}, \tag{4.74}$$

which explicitly shows that $(\mathbf{x}, \mathbf{k})$ are not the canonical coordinates on the phase space anymore.

5 Chiral Kinetic Theory

There has been much recent interest on the influence of Berry phases on the electronic property of solids, a number of which provide fruitful analogies for relativistic field theories. An interesting example occurs when there is a net flux of Berry curvature through a disconnected part of the Fermi surface, where an analogue of abelian axial anomaly occurs, manifesting itself in the nonconservation of particle number in the conservation band.

The axial anomaly is usually derived via sophisticated quantum field theory computation. However, Stephanov and Yin[26] showed that it can also be derived from a classical Hamiltonian phase space dynamics, with the quantum inputs being the $\hbar$ in the phase space volume and the Berry curvature effects (anomalous velocity) in the equations of motion. In this chapter, we reformulate their calculation using the symplectic formulation of classical mechanics as discussed in Ch 2, using which we generalize their calculation to compute nonabelian singlet and gauge anomalies in arbitrary even spacetime dimensions. Finally, we use semiclassical kinetic theory to explicitly compute the *grand potential current*, which determines all the anomalous contributions to the relativistic hydrodynamic currents, as discussed in Sec 3.3.

5.1 U(1) anomaly in $3+1$ dimensions

In Ref [26], Stephanov and Yin showed that in $3+1$ dimensions, the positive energy, positive helicity Weyl fermion with Hamiltonian $H = \sigma \cdot \mathbf{p}$ can be described by the semiclassical action

$$S[\mathbf{x}, \mathbf{p}] = \int dt(\mathbf{p} \cdot \dot{\mathbf{x}} + \mathbf{A} \cdot \dot{\mathbf{x}} - |\mathbf{p}| - \phi - \mathfrak{a} \cdot \dot{\mathbf{p}}), \tag{5.1}$$

where $\mathfrak{a}$ and ϕ are the Maxwell scalar and vector potentials, respectively. The Berry connection, $\mathfrak{a}$, corresponds to a monopole of unit strength at origin, and the corresponding curvature is given by eq. (4.37) as

$$\mathfrak{b} = \nabla_{\mathbf{p}} \times \mathfrak{a} = -\frac{\hat{\mathbf{p}}}{2|\mathbf{p}|^2} \implies \nabla_{\mathbf{p}} \cdot \mathfrak{b} = -2\pi\delta^3(\mathbf{p}). \tag{5.2}$$

5.1.1 Original calculation

The equations of motion can be obtained by a variation of the action w.r.t $\mathbf{x}$ and $\mathbf{p}$, to get

$$\begin{aligned} 0 &= \int dt \Big(\delta\mathbf{p}\cdot\dot{\mathbf{x}} + \mathbf{p}\cdot\delta\dot{\mathbf{x}} + \dot{\mathbf{x}}\cdot(\delta\mathbf{x}\cdot\nabla_x)\mathbf{A} + \mathbf{A}\cdot\delta\dot{\mathbf{x}} - \hat{\mathbf{p}}\cdot\delta\mathbf{p} - \delta\mathbf{x}\cdot\nabla_x\phi - \dot{\mathbf{p}}\cdot(\delta\mathbf{p}\cdot\nabla_p)\mathfrak{a} - \mathfrak{a}\cdot\delta\dot{\mathbf{p}} \Big) \\ &= \int dt \Big[\delta x^i \left(-\dot{p}_i - \frac{\partial\phi}{\partial x^i} - \frac{\partial A^i}{\partial t} - \dot{x}^j \frac{\partial A_i}{\partial x^j} + \dot{x}^j \frac{\partial A_j}{\partial x^i} \right) + \delta p_i \left(\dot{x}^i - \hat{p}^i - \dot{p}_j \frac{\partial \mathfrak{a}^j}{\partial p_i} + \dot{p}_j \frac{\partial \mathfrak{a}^i}{\partial p_j} \right) \Big]. \end{aligned} \tag{5.3}$$

Defining the electric and magnetic field as $\mathbf{E} = -\nabla_{\mathbf{x}}\phi - \partial_t\mathbf{A}$, $\mathbf{B} = \nabla_{\mathbf{x}} \times \mathbf{A}$ and the Berry curvature as $\mathfrak{b} = \nabla_{\mathbf{p}} \times \mathfrak{a}$, the equations of motion become

$$\dot{\mathbf{x}} = \hat{\mathbf{p}} + \dot{\mathbf{p}} \times \mathfrak{b}, \qquad \dot{\mathbf{p}} = \mathbf{E} + \dot{\mathbf{x}} \times \mathbf{B}. \tag{5.4}$$

These equations are equivalent to eq. (4.64), which describes the semiclassical dynamics of electrons in a Bloch band. The energy spectrum is simply given by the positive eigenvalue ($\varepsilon = |\mathbf{p}|$) of the Weyl Hamiltonian $H = \mathbf{p}\cdot\boldsymbol{\sigma}$. In Sec 4.3, we showed that eq. (5.4) describes Hamiltonian flows on the phase space $\mathbb{R}^3 \times \mathbb{R}^3$ with an unconventional symplectic structure, such that the invariant volume form is given by $\sqrt{\rho}\, d^3x\, d^3p$, with $\sqrt{\rho} = 1 + \mathfrak{b}\cdot\mathbf{B}$.

Given a phase space distribution function $f(t, \mathbf{x}, \mathbf{p})$, the number of particles in an infinitesimal phase space volume at time t is given by $\sqrt{\rho}\, f(t, \mathbf{x}, \mathbf{p}) d^3x\, d^3p$. The collisionless Boltzmann equation is

$$\partial_t \left(\sqrt{\rho} f\right) + \nabla_{\mathbf{x}} \cdot \left(\sqrt{\rho} f\, \dot{\mathbf{x}}\right) + \nabla_{\mathbf{p}} \cdot \left(\sqrt{\rho} f \dot{\mathbf{p}}\right) \stackrel{?}{=} 0. \tag{5.5}$$

Let f is advected with the flow, i.e,

$$\left[\partial_t + \dot{\mathbf{x}}\cdot\nabla_{\mathbf{x}} + \dot{\mathbf{p}}\cdot\nabla_{\mathbf{p}}\right] f = 0. \tag{5.6}$$

Then, Stephanov et al compute the remaining terms on the LHS of eq. (5.5) explicitly using the equations of motion, which can be solved for $\dot{\mathbf{x}}$ and $\dot{\mathbf{p}}$ by substituting the latter in the former and using the vector product identity

$$(\dot{\mathbf{x}} \times \mathbf{B}) \times \mathfrak{b} = (\dot{\mathbf{x}}\cdot\mathfrak{b})\,\mathbf{B} - (\mathfrak{b}\cdot\mathbf{B})\,\dot{\mathbf{x}}$$

followed by $(\dot{\mathbf{p}} \times \mathfrak{b})\cdot\mathfrak{b} = 0$ in the first term, to get

$$\dot{\mathbf{x}} = \hat{\mathbf{p}} + \mathbf{E} \times \mathfrak{b} + (\hat{\mathbf{p}}\cdot\mathfrak{b})\,\mathbf{B} - (\mathfrak{b}\cdot\mathbf{B})\,\dot{\mathbf{x}}. \tag{5.7}$$

The equation for $\mathbf{p}$ can be solved in a similar fashion. We get

$$\begin{aligned}\sqrt{\rho}\,\dot{\mathbf{x}} &= \hat{\mathbf{p}} + \mathbf{E}\times\mathfrak{b} + (\mathfrak{b}\cdot\hat{\mathbf{p}})\mathbf{B},\\ \sqrt{\rho}\,\dot{\mathbf{p}} &= \mathbf{E} + \hat{\mathbf{p}}\times\mathbf{B} + (\mathbf{E}\cdot\mathbf{B})\mathfrak{b}.\end{aligned} \tag{5.8}$$

Substituting in the LHS of eq. (5.5),

$$\begin{aligned}&\partial_t\left(\sqrt{\rho}\right) + \nabla_{\mathbf{x}}\cdot\left(\sqrt{\rho}\,\dot{\mathbf{x}}\right) + \nabla_{\mathbf{p}}\cdot\left(\sqrt{\rho}\dot{\mathbf{p}}\right)\\ &= \partial_t\left[1+\mathfrak{b}\cdot\mathbf{B}\right] + \nabla_{\mathbf{x}}\cdot\left[\hat{\mathbf{p}} + \mathbf{E}\times\mathfrak{b} + (\mathfrak{b}\cdot\hat{\mathbf{p}})\mathbf{B}\right] + \nabla_{\mathbf{p}}\cdot\left[\mathbf{E} + \hat{\mathbf{p}}\times\mathbf{B} + (\mathbf{E}\cdot\mathbf{B})\mathfrak{b}\right]\\ &= \mathfrak{b}\cdot(\partial_t\mathbf{B}) + \nabla_{\mathbf{x}}\cdot(\mathbf{E}\times\mathfrak{b}) + (\mathfrak{b}\cdot\hat{\mathbf{p}})(\nabla_{\mathbf{x}}\cdot\mathbf{B}) + \nabla_{\mathbf{p}}\cdot(\hat{\mathbf{p}}\times\mathbf{B}) + (\mathbf{E}\cdot\mathbf{B})(\nabla_{\mathbf{p}}\cdot\mathfrak{b})\\ &= \mathfrak{b}\cdot(\partial_t\mathbf{B}) + \mathfrak{b}\cdot(\nabla_{\mathbf{x}}\times\mathbf{E}) + \mathbf{B}\cdot(\nabla_{\mathbf{p}}\times\hat{\mathbf{p}}) + (\mathbf{E}\cdot\mathbf{B})(\nabla_{\mathbf{p}}\cdot\mathfrak{b})\\ &= \mathfrak{b}\cdot(\partial_t\mathbf{B} + \nabla_{\mathbf{x}}\times\mathbf{E}) + \mathbf{B}\cdot(\nabla_{\mathbf{p}}\times\hat{\mathbf{p}}) + (\mathbf{E}\cdot\mathbf{B})(\nabla_{\mathbf{p}}\cdot\mathfrak{b})\\ &= (\mathbf{E}\cdot\mathbf{B})(\nabla_{\mathbf{p}}\cdot\mathfrak{b}) = -(\mathbf{E}\cdot\mathbf{B})\,2\pi\delta^3(\mathbf{p}),\end{aligned} \tag{5.9}$$

where we have used the vector identity $\nabla\cdot(\mathbf{v}\times\mathbf{w}) = \mathbf{w}\cdot(\nabla\times\mathbf{w}) - \mathbf{v}\cdot(\nabla\times\mathbf{w})$, the Maxwell equations $\nabla\cdot\mathbf{B} = 0$ and $\partial_t\mathbf{B} = -\nabla\times\mathbf{E}$ and the fact that $\nabla_{\mathbf{p}}\times\hat{\mathbf{p}} = 0$, since $\hat{\mathbf{p}}$ is radial. Thus, Liouville's theorem, which demands that the LHS vanish, is violated at $\mathbf{p} = 0$. However, strictly speaking, this semiclassical description is invalid at the Dirac point $\mathbf{p} = 0$, since the band gap for the Weyl Hamiltonian vanishes and the adiabatic approximation needed to compute the Berry curvature breaks down. However, remarkably, the U(1) anomaly is encoded in the symplectic structure, and can be recovered by considering the Boltzmann equation far from the Dirac point.

Thus, the Boltzmann equation becomes

$$\partial_t\left(\sqrt{\rho}f\right) + \nabla_{\mathbf{x}}\cdot\left(\sqrt{\rho}f\,\dot{\mathbf{x}}\right) + \nabla_{\mathbf{p}}\cdot\left(\sqrt{\rho}f\dot{\mathbf{p}}\right) = (\mathbf{E}\cdot\mathbf{B})\,2\pi f\,\delta^3(\mathbf{p}). \tag{5.10}$$

The charge density and current associated with the Weyl fermions is defined as (Sec 2.5)

$$\rho(\mathbf{x},t) = \int\frac{d^3p}{(2\pi)^3}\sqrt{\rho}\,f(\mathbf{x},\mathbf{p},t), \quad \mathbf{J}(\mathbf{x},t) = \int\frac{d^3p}{(2\pi)^3}\sqrt{\rho}\,\dot{\mathbf{x}}\,f(\mathbf{x},\mathbf{p},t). \tag{5.11}$$

Integrating eq. (5.10) over the entire momentum space , we get

$$\partial_t\int\frac{d^3p}{(2\pi)^3}\sqrt{\rho}\,f + \nabla_{\mathbf{x}}\cdot\int\frac{d^3p}{(2\pi)^3}\sqrt{\rho}\,f\,\dot{\mathbf{x}} + \int\frac{d^3p}{(2\pi)^3}\nabla_{\mathbf{p}}\cdot\left(\sqrt{\rho}f\dot{\mathbf{p}}\right) = -(\mathbf{E}\cdot\mathbf{B})\int\frac{d^3p}{(2\pi)^2}f\,\delta^3(\mathbf{p}). \tag{5.12}$$

The third integral vanishes, as it is a surface term. Substituting the expressions for the density and currents leads to the anomalous conservation law

$$\partial_t \rho + \nabla \cdot \mathbf{J} = -\frac{1}{(2\pi)^2} f(\mathbf{x}, \mathbf{0}, t)(\mathbf{E} \cdot \mathbf{B}). \tag{5.13}$$

If the zero energy state is (at $\mathbf{p} = 0$) is filled up, then $f(\mathbf{x}, \mathbf{0}, t) = 1$. This would be true, for instance, at $T = 0$, if the Fermi level is positive. Then, we recover the expression for chiral anomaly in $3 + 1$ dimensions as[1]

$$\partial_\mu J^\mu = -\frac{1}{(2\pi)^2} \mathbf{E} \cdot \mathbf{B} = \frac{1}{2!\,(2\pi)^2} \star (F \wedge F)\,. \tag{5.14}$$

Thus, the anomaly is a direct consequence of the breakdown of Liouville theorem in the presence of a Berry monopole.

5.1.2 A symplectic formulation

We can conveniently reformulate the above calculation in terms of the extended phase space formalism, discussed in Sec. 2.4. The relevant extended phase space is simply $M_H = \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R}^3$, with coordinates $(t, \mathbf{x}, \mathbf{p})$. The semiclassical action of eq. (5.1) can be written as

$$S[\mathbf{x}, \mathbf{p}] = \int \eta_H; \quad \eta_H = p_i dx^i - |\mathbf{p}|dt + A - \mathfrak{a}, \tag{5.15}$$

so that the (generalized) symplectic form becomes

$$\rho_H = d\eta_H = dp_i \wedge dx^i - \hat{p}_i dp_i \wedge dt + F - \mathfrak{F}, \tag{5.16}$$

with

$$F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \qquad \mathfrak{F} = \frac{1}{2} \mathfrak{F}_{ij} dp^i \wedge dp^j. \tag{5.17}$$

The Berry curvature is singular at $\mathbf{p} = 0$, the band touching point. As a consequence,

$$d\rho_H = dF - d\mathfrak{F} = -d\mathfrak{F} = -2\pi\delta^3(\mathbf{p})\, d^3p, \tag{5.18}$$

[1] Note that on $\mathbb{R}^{d,1}$ with Minkowski metric, $\star dt \wedge d^d x = -1$.

since $dF = 0$ as the electromagnetic field satisfies the Maxwell's equations, while

$$d\mathfrak{F} = \frac{1}{2}\frac{\partial \mathfrak{F}_{jk}}{\partial p_i} dp^i dp^j dp^k = \frac{\partial}{\partial p_i}\left(\frac{1}{2}\epsilon^{ijk}\mathfrak{F}_{jk}\right) d^3p = (\nabla_{\mathbf{p}} \cdot \mathfrak{b})\, d^3p = 2\pi\delta^3(\mathbf{p})\, d^3p. \tag{5.19}$$

In Sec 2.5.3, we used the geometric formulation of Liouville's theorem to define particle currents. Using eq. (2.92), the conservation law of particle number current becomes

$$d \star J = \int_{\mathbb{R}^3} \mathcal{L}_{\mathfrak{d}H}(f\Omega_H) = \frac{1}{(2\pi)^3 2!}\int_{\mathbb{R}^3} d\rho_H \wedge \rho_H^2, \tag{5.20}$$

where we have used the last line of eq. (2.79) for the Lie derivative. Using eq. (5.18), the integrand becomes

$$d\rho_H \wedge \rho_H^2 = -d\mathfrak{F} \wedge (dp_i \wedge dx^i - \hat{p}_i dp^i \wedge dt + F - \mathfrak{F})^2 = -d\mathfrak{F} \wedge F^2, \tag{5.21}$$

where $F \wedge F$ is the only term possible in the expansion of ρ_H^2, since $d\mathfrak{F}$ contains all three dp's. Substituting in eq. (5.20), we get the U(1) anomaly as

$$d \star J = -\frac{1}{2(2\pi)^3}\left(\int_{\mathbb{R}^3} d\mathfrak{F}\right) F^2 = \frac{1}{2(2\pi)^2}\left(\int_{\mathbb{R}^3} \delta^3(\mathbf{p})\, d^3p\right) F^2 = \frac{1}{2!}\left(\frac{F}{2\pi}\right)^2. \tag{5.22}$$

Thus, the anomaly is encoded in the symplectic form via $d\rho_H \neq 0$. We shall refer to such ρ_H as an *anomalous symplectic form*.

An alternative approach to compute the momentum integral is to split it as

$$\frac{1}{2\pi}\int_{\mathbb{R}^3} d\mathfrak{F} = \frac{1}{2\pi}\left(\int_{B^3(\epsilon)} + \int_{\mathbb{R}^3 \setminus B^3(\epsilon)}\right) d\mathfrak{F} = \frac{1}{2\pi}\int_{S^2(\epsilon)} \mathfrak{F} = -1, \tag{5.23}$$

where $B^3(\epsilon)$ is a ball of radius ϵ centered at $\mathbf{p} = \mathbf{0}$ and $S^2(\epsilon) = \partial B^3(\epsilon)$. The integral over $\mathbb{R}^3 \setminus B^3(\epsilon)$ vanishes since $d\mathfrak{F} = 0$ away from $\mathbf{p} = \mathbf{0}$. In the last step, we have used Gauss' law, since the integral is simply the flux across a sphere which encloses a Berry monopole with unit charge, located at $\mathbf{p} = \mathbf{0}$.

5.1.3 Discussion

The central idea of this computation of chiral anomaly using a semiclassical formalism is the interpretation of the conservation of the U(1) current as a continuity equation, and hence the interpretation of the chiral anomaly as a nonconservation of the number of Weyl fermion. Furthermore, the Dirac sea is incompressible, as encoded in the validity of Liouville's theorem far from the Dirac point, so that by simply following the

particles at a given Fermi surface semiclassicaly, we can deduce the *rate of production* of the particles due to the chiral anomaly.

In Sec 2.5, kinetic theory was introduced as a formalism to extract macroscopic (and hopefully observable) dynamical variables given an ensemble of particles described by Hamiltonian dynamics. Since the U(1) anomaly turns out to be encoded in the symplectic form, its macroscopic effects can be seen in the currents[26]. For instance, the charge current can be explicitly written as

$$\mathbf{J}(\mathbf{x},t) = \int \frac{d^3p}{(2\pi)^3} f(t,\mathbf{x},\mathbf{p}) \left(\hat{\mathbf{p}} + \mathbf{E} \times \frac{\hat{\mathbf{p}}}{2|\mathbf{p}|^2} + \frac{1}{2|\mathbf{p}|^2}\mathbf{B} \right) \tag{5.24}$$

The first two terms represent the conventional classical currents. The first term is the charge current due to the motion of the particles, while the second term is the classical Hall effect, *viz*, a transverse current in presence of an external electric field which leads to a sideways shift on the trajectory of Weyl particles. Both of these are nonvanishing only if $f(t,\mathbf{x},\mathbf{p})$ is spherically asymmetric. Thus, for a spherically symmetric phase space distribution function, the current becomes

$$\mathbf{J} = \frac{1}{(2\pi)^3} \int_0^\infty |\mathbf{p}|^2 d|\mathbf{p}| \int_{S^2} d\Omega_2 \, f(|\mathbf{p}|) \left(\hat{\mathbf{p}} + \mathbf{E} \times \frac{\hat{\mathbf{p}}}{2|\mathbf{p}|^2} + \frac{1}{2|\mathbf{p}|^2}\mathbf{B} \right) = \frac{1}{4\pi^2} \left(\int_0^\infty d|\mathbf{p}| \, f(|\mathbf{p}|) \right) \mathbf{B}. \tag{5.25}$$

This current represents the chiral magnetic effect(CME), a current generated along the magnetic field, even in the absence of an electric field. Given a phase space distribution function, one can explicitly compute the CME coefficient using the chiral kinetic theory. For instance, given the zero temperature Fermi-Dirac distribution with chemical potential $\mu > 0$, we get

$$\mathbf{J}_{\text{CME}} = \frac{1}{4\pi^2} \left(\int_0^\infty d|\mathbf{p}| \Theta(\mu - |\mathbf{p}|) \right) \mathbf{B} = \frac{1}{4\pi^2} \mu \mathbf{B}. \tag{5.26}$$

This result has been proposed for quark-gluon plasmas[35] in a high energy physics as well as for Weyl semimetals in the condensed matter physics[32].

An obvious objection to this calculation is the use of Hamiltonian dynamics, which treats time and space differently and is hence not manifestly Lorentz invariant. Indeed, even a way to reformulate it in a Lorentz covariant fashion is unclear, since the Berry curvature 2-form contains differentials of only the three spatial components of the energy-momentum 4-vector. On the other hand, the quantum field theory for Weyl fermions is manifestly Lorentz invariant, and any low energy effective theory describing the same physics must inherit that symmetry. In Ref [39], the chiral kinetic theory is shown to be Lorentz invariant, albeit

with the coordinates transforming under a boost $\boldsymbol{\beta}$ as

$$\mathbf{x} \mapsto \mathbf{x} + \boldsymbol{\beta}\, t + \frac{1}{2|\mathbf{p}|}\boldsymbol{\beta} \times \hat{\mathbf{p}}, \quad \mathbf{p} \mapsto \mathbf{p} + \boldsymbol{\beta}\, \varepsilon + \frac{1}{2|\mathbf{p}|}(\boldsymbol{\beta} \times \hat{\mathbf{p}}) \times \mathbf{B}. \tag{5.27}$$

Furthermore, for consistency, the dispersion relation for the Weyl fermions needs to include the magnetic moment as $\varepsilon = |\mathbf{p}| - \hat{\mathbf{p}}\cdot\mathbf{B}/2|\mathbf{p}|$. This unusual representation of the Lorentz group arises from the fact that the Weyl particles carry a nonzero angular momentum (spin), so that their '*position*' becomes frame dependent. Under a Lorentz transform, the 'trajectory' of the particle moves sideways by an amount proportional to the spin, an effect known as *Wigner translations*. We investigate these aspects further in Ref [3].

5.2 Nonabelian guage fields

We shall next seek to generalize the calculations of the last section to include nonabelian gauge fields in a semiclassical formalism. In Sec 3.1, we discussed that the coupling to a (in general nonabelian) guage group G requires that the Weyl fermion then transforms under a nontrivial representation Λ of G, which is the "charge" of the Weyl fermion. Thus, in order to describe these internal degrees of freedom, we need to "dequantize" them[2]. In more practical terms, since the gauge fields are now matrix valued and the action is required to be a c-number, we need an analogue of the minimal couplings $\mathbf{p} \to \mathbf{p} - \mathbf{A}$ and $\mathbf{x} \to \mathbf{x} - \mathfrak{a}$. In this section, we describe such a dequantization procedure.

Dequantization

Given a symplectic manifold $(\mathcal{M}, \rho)$, let $\{\ ,\ \}$ be the induced Poisson algebra (defined by eq. (2.30)) on $C^\infty(\mathcal{M}, \mathbb{R})$, the space of smooth functions on $\mathcal{M}$. The problem of quantization is then to associate with this setup a Hilbert space $\mathscr{H}$ and a Hermitian operator $\hat{f}\colon \mathscr{H} \to \mathscr{H}$ for each $f \in C^\infty(\mathcal{M}, \mathbb{R})$, such that the operators satisfy the Dirac quantization condition

$$\{f, g\} = h \implies [\hat{f}, \hat{g}] = i\hat{h} \quad \forall\, f, g, h. \tag{5.28}$$

Thus, quantization maps the Poisson algebra of the functions on $\mathcal{M}$ to a Lie algebra of operators on $\mathscr{H}$.

For dequantization, we seek to go the other way round, i.e, given a Hilbert space $\mathscr{H}$ and a set of operators, we seek a symplectic manifold such that the corresponding Poisson algebra can be derived from the operator

[2] We did not have this problem with U(1) since the Weyl fermions transform under the fundamental (defining) representations of U(1) which simply changes the phase of the fermion field, and classical mechanics is oblivious to absolute phases of the wavefunctions.

algebra on $\mathscr{H}$ by using the Dirac quantization condition backwards. There is no solution to this problem in general; however, we only seek a special case. Given a finite dimensional unitary irreducible representation of a compact semisimple Lie group G on some complex Hilbert space $\mathscr{H}$, we seek the corresponding finite dimensional compact symplectic manifold.

Precisely such a correspondence was described in eq. (2.41) of Sec 2.3.2 for the regular coadjoint orbits of G with the Kirillov form. The quantum mechanical system can be recovered from the coadjoint orbit using *geometric quantization*[94, 95], which, for coadjoint orbits, turns out to be identical to the problem of constructing the unitary irreducible representations of Lie groups. This connection, and the use of geometric quantization to construct unitary irreps, is usually termed the *orbit method* and was developed by Kirillov[72], Kostant and Souriau.

Quantum mechanics and symmetries

Consider a Hilbert space $\mathscr{H}$, on which a representation Λ of G is defined. Let $\hat{X} = iX^a\hat{\lambda}_a$ be a generator of a symmetry on $\mathscr{H}$ with $X^a \in \mathbb{R}$, and define $X = iX^a\lambda_a \in \mathfrak{g}$. Given a state $|\psi\rangle \in \mathscr{H}$, we can use it to define a functional $F_\psi \colon \mathfrak{g} \to \mathbb{R}$ as

$$F_\psi \colon X \mapsto -i\langle\psi|\hat{X}|\psi\rangle = X^a\langle\psi|\hat{\lambda}_a|\psi\rangle \equiv X^a\alpha_{\psi,a}, \tag{5.29}$$

where $\alpha_{\psi,a} \in \mathbb{R}$ since it is a diagonal element of a Hermitian matrix $\hat{\lambda}_a$. Since F_ψ is linear, by definition, $F_\psi \in \mathfrak{g}^*$. Defining $\alpha_\psi = \alpha_{\psi,a}\lambda^a \in \mathfrak{g}^*$, where λ^a is a basis of $\mathfrak{g}^*$ dual to $\mathfrak{g}$, we can alternatively write $F_\psi(X) = (\alpha, X) = \mathrm{tr}\,\{\alpha X\}$, where in the last step, we have used the Killing form.

Under an adjoint action of the group, $X \mapsto gXg^{-1}$, F_ψ transforms under the coadjoint representation. Explicitly, we can write the orbit as

$$X \mapsto gXg^{-1} \implies F_\psi(X) \mapsto \mathrm{tr}\left\{\alpha_\psi XgXg^{-1}\right\} = \mathrm{tr}\left\{g^{-1}\alpha_\psi g\, X\right\} = \mathrm{tr}\left\{QX\right\}, \tag{5.30}$$

where $Q = g^{-1}\alpha_\psi g$ defines coordinates on the coadjoint orbit $\mathcal{O}_{\alpha_\psi}$ of α_ψ. Thus, symmetry transformations of X as $X \mapsto gXg^{-1}$ can alternatively be thought of as a translation on $\mathcal{O}_\psi$ under the group action $Q \mapsto g^{-1}Qg$.

This is a classical version of the original operator algebra, as given $X, Y \in \mathfrak{g}$, using eq. (2.41),

$$\{X, Y\} = X^aY^b\{Q_a, Q_b\} = X^aY^bf_{ab}^{\ c}Q_c \longleftrightarrow [X, Y] = X^aY^b[\hat{\lambda}_a, \hat{\lambda}_b] = X^aY^bf_{ab}^{\ c}\hat{\lambda}_c. \tag{5.31}$$

We should be able to quantize it to obtain the representation Λ back; however, that does not work for any

choice of $|\psi\rangle$. For a compact semisimple Lie groups, the unitary irreps are completely characterized by their highest weights[72], and one needs to take the highest weight state, $|\Lambda\rangle$, in order to be able to recover Λ on geometric quantization. Thus, hereafter, we set $F(X) = X^a\langle\Lambda|\Lambda_*(\lambda_a)|\Lambda\rangle$, and the corresponding coadjoint orbit is denoted by $\mathcal{O}_\Lambda$.

Recall that the regular coadjoint orbits of a compact semisimple Lie group G are equivalent to G/T, where $T \subseteq G$ is a maximal torus of G. A suitable set of coordinate on $\mathcal{O}_\Lambda$ is defined as $Q = g^{-1}\alpha_\Lambda g = Q^a\lambda_a$, where $g \in G$, $\alpha_\Lambda \in \mathfrak{g}$, $Q^a \in \mathbb{R}$, λ's are the generators of $\mathfrak{g}$ and $a = 1, 2, \ldots \dim(G)$.

Dequantizing nonabelian gauge fields

We shall use this construction to encode the nonabelian gauge fields which can be denoted by differential forms valued in $\mathfrak{g}$, i.e, $A \in \Omega^1(\mathbb{R}^{n,1}) \otimes \mathfrak{g}$. The corresponding curvature is defined as $F = dA - iA \wedge A$. The classical version is defined on $\mathbb{R}^{n,1} \times \mathcal{O}_\Lambda$ with coordinates $(t, \mathbf{x}, Q)$. Demanding that Q also transform under a gauge transformation, we seek to construct gauge invariant combinations involving A and F. Using the notation of Sec 3.1, under a nonabelian gauge transformation by $h(t, \mathbf{x}) \in G$,

$$A \mapsto h^{-1}Ah + ih^{-1}dh, \qquad F \mapsto h^{-1}Fh, \qquad Q \mapsto h^{-1}Qh, \tag{5.32}$$

so that A and F transform under the adjoint action by h^{-1} (instead of h). Then, $\widetilde{F} = \mathrm{tr}\,\{QF\}$ is invariant under this gauge transformation. However, $\mathrm{tr}\,\{QA\}$ does not work, since we need to cancel off the $h^{-1}dh$ term. However, including the Liouville form[3] $w_R = dg\,g^{-1}$ for the coadjoint orbit, we get a gauge-invariant 1-form

$$\begin{aligned}\mathrm{tr}\left\{Q\left(A + idg\,g^{-1}\right)\right\} \mapsto \mathrm{tr}\left\{h^{-1}Qh\left[(h^{-1}Ah + ih^{-1}dh) + id(h^{-1}g)g^{-1}h\right]\right\} \\ = \mathrm{tr}\left\{h^{-1}Qh\left[h^{-1}Ah + ih^{-1}dh - ih^{-1}dh + ih^{-1}dg\,g^{-1}h\right]\right\} \\ = \mathrm{tr}\left\{Q\left(A + idg\,g^{-1}\right)\right\}.\end{aligned} \tag{5.33}$$

Thus, our dequantization prescription is to upgrade

$$\mathcal{M} \mapsto \mathcal{M} \times \mathcal{O}_\Lambda, \qquad A \mapsto \mathrm{tr}\{QA\left(A + idg\,g^{-1}\right)\}, \qquad F \mapsto \mathrm{tr}\{QF\}, \tag{5.34}$$

which encodes gauge transformations in the translations on $\mathcal{O}_\Lambda$ by the coadjoint group action.

[3] This is reminiscent of the right-invariant Maurer-Cartan form on G. This is not a coincidence; the Maurer-Cartan form Θ_R is simply the pullback of w_R under the projection $G \to G/T = \mathcal{O}_X$.

Traces and integrals

In nonabelian gauge theories, one often encounters traces over the given representation of the gauge group (Sec 3.1). We now seek the corresponding classical analogue, which should be integrals over the coadjoint orbits. The symplectic form on $\mathcal{O}_\Lambda$ can be written as $\rho = dw_R = w_R \wedge w_R$, where we have used Cartan's structure equation. Using eq. (2.73), the volume measure on $\mathcal{O}_\Lambda$ is then given by

$$d\mu_\Lambda = \frac{1}{(2\pi)^{m_\Lambda} m_\Lambda!} \left[-i\mathrm{tr}\left\{Q w_R^2\right\}\right]^{m_\Lambda}, \qquad m_\Lambda = \frac{1}{2}\dim\left(\mathcal{O}_\Lambda\right). \tag{5.35}$$

Next, we seek the classical analogues of the traces of the form $\mathrm{tr}_\Lambda\left\{\hat{\lambda}_{a_1}\dots\hat{\lambda}_{a_k}\right\}$. The first guess would be

$$\int_{\mathcal{O}_Q} d\mu_\Lambda Q_{a_1}\dots Q_{a_N} = \mathrm{Str}\{\lambda_{a_1}\dots\lambda_{a_N}\}. \tag{5.36}$$

However, this integral is only approximate, with an error that tends to zero as we increase the dimension of the representation. For instance, the simplest such integral contains no Q_a's, which we expect to be the correspondence between the dimension of the representation and the volume of the corresponding coadjoint orbit

$$\dim\left(\Lambda\right) = \mathrm{tr}_\Lambda\{\mathbb{1}\} \longleftrightarrow \mathrm{vol}\left(\mathcal{O}_\Lambda\right) = \int_{\mathcal{O}_\Lambda} d\mu_\Lambda, \tag{5.37}$$

However, to recover the dimension of Λ exactly, one needs to expand the coadjoint orbit slightly and integrate $\mathcal{O}_{\Lambda+\mathfrak{W}}$ instead of $\mathcal{O}$, where we have shifted the highest weight vector Λ by the Weyl vector $\mathfrak{W}$ (See Appendix C of Ref. [2] for details). This "fudge factor" $\mathfrak{W}$ can be thought of as a *quantum correction*, which also improves the approximation for the remaining traces. Hereafter we shall assume that all integrals are over the *Weyl-shifted* coadjoint orbits.

Example: SU(2)

To illustrate the dequantization, we again consider SU(2), for which we constructed the coadjoint orbits explicitly in Sec 2.3.2. Recall that the unitary irreps of SU(2) are labeled by positive half-integers (*spin*) $j \in \mathbb{Z}^+/2$, and are $2j+1$ dimensional. Given j, a complete basis of $\mathscr{H} = \mathbb{C}^{2j+1}$ is then given by the eigenstates of $S_3 = \hat{\sigma_3}$, as

$$S_3|j,m\rangle = |j,m\rangle, \quad -j \le m \le j. \tag{5.38}$$

Here, the Cartan subalgebra is 1-dimensional, spanned by σ_3. The only root is 1, so that the Weyl vector, i.e. half the sum of roots, $\mathfrak{W} = \frac{1}{2}$. Also, m is the weight of a given state, and the highest weight state is

simply $|j,j\rangle$. Thus,

$$\langle j,j|S_i|j,j\rangle = j\delta_{i3} \implies \alpha_j = j\sigma_3, \tag{5.39}$$

so that the relevant coadjoint orbit[4] is a 2-sphere of radius j, with the symplectic form $j\Omega(S^2)$. Its volume can be computed by

$$\mathrm{vol}(\mathcal{O}_j) = \frac{j}{2\pi}\int_{S^2}\Omega(S^2) = \frac{j}{2\pi}\int_0^\pi d\theta\int_0^{2\pi} d\theta\,\sin\theta = 2j, \tag{5.40}$$

where we have parametrized S^2 in polar coordinates as

$$(Q_1,Q_2,Q_3) = j\,(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta) \tag{5.41}$$

The volume of $\mathcal{O}_j$ is only an approximation to $2j+1 = 2j(1+j^{-1})$, with an error of order $1/j$ which tends to zero as $j\to\infty$. To recover the dimension precisely, one needs to integrate over a sphere of radius $j+\frac{1}{2}$, which is precisely the Weyl shift $j\to j+\mathfrak{W}$.

As an illustration of integrals containing more Q's, we note that the integrals containing a single Q_a vanish, since the measure is symmetric under $Q_a\to -Q_a$, while those of Q_a^2 can again be evaluated in polar coordinates on S^2. For instance, the integral of $Q_3^2 = j^2\cos^2\theta$ can be computed as

$$\int_{\mathcal{O}_j} d\mu_j Q_3^2 = \frac{j^3}{2\pi}\int_0^\pi d\theta\int_0^{2\pi} d\theta\,\sin\theta\cos^2\theta = -j^3\int_0^\pi d(\cos\theta)\cos^2\theta = -\frac{j^3}{3}\cos^3\theta\Big|_0^\pi = \frac{2j^3}{3}. \tag{5.42}$$

The corresponding quantum traces can be computed using the basis of S_3 as

$$\mathrm{tr}\left\{S_3^2\right\} = \sum_{m=-j}^{j}\langle j,m|S_3^2|j,m\rangle = \sum_{m=-j}^{j} m^2, \tag{5.43}$$

which can be computed using Feynman's trick

$$\begin{aligned}
\mathrm{tr}\left\{S_3^2\right\} &= \frac{1}{4}\frac{\partial^2}{\partial\alpha^2}\sum_{m=-j}^{j} e^{2\alpha m}\bigg|_{\alpha\to 0} = \frac{1}{4}\frac{\partial^2}{\partial\alpha^2}\frac{\sinh(2j+1)\alpha}{\sinh\alpha}\bigg|_{\alpha\to 0}\\
&= \frac{1}{4}\lim_{\alpha\to 0}\left[(2j+1)^2\frac{\cosh(2j+1)\alpha}{\sinh\alpha} - 2(2j+1)\frac{\sinh(2j+1)\alpha\cosh\alpha}{\sinh^2\alpha} + \frac{\sinh(2j+1)\alpha}{\sinh^3\alpha}\left(1+\cosh^2\alpha\right)\right]\\
&= \frac{j}{3}\left(2j^2+3j+1\right) = \frac{2j^3}{3}\left(1+\frac{3}{2j}+\frac{1}{2j^2}\right),
\end{aligned} \tag{5.44}$$

which again agrees with the integral upto errors of order j^{-1}.

[4] Recall that for SU(2), we showed that the coadjoint orbit for $\mathbf{X}\cdot\boldsymbol{\sigma}$ is simply a 2-sphere of radius $|\mathbf{X}|$.

5.3 Nonabelian anomaly in arbitrary even dimensions

The symplectic formulation of the chiral kinetic theory, as discussed in Sec 5.1.2, can be generalized in two directions: coupling to nonabelian gauge fields[1], and considering dynamics in higher dimensions[2]. As Weyl fermions (and chiral anomaly) are defined only in even spacetime dimensions, we consider a Weyl fermion on $\mathbb{R}^{n,1},\ n = 2N+1$, coupled to a nonabelian gauge field with gauge group G, which we assume to be a compact semisimple Lie group. In Sec 4.2.3, we computed the Berry curvature for the Weyl Hamiltonian, and showed that it has the gauge group $\mathrm{Spin}(2N)$, which is nonabelian for $N > 1$. Thus, the positive energy wavefunction for the Weyl Hamiltonian transforms under a representation Λ of G, when parallel transported along $\mathbf{x}$, and under a representation $\mathfrak{Q}$ of $\mathrm{Spin}(2N)$ when parallel transported along $\mathbf{p}$. We now discuss the computation of anomalies in a semiclassical setup, using the dequantization prescription of Sec 5.2 to couple the nonabelian gauge fields to the classical particle.

5.3.1 Generalized symplectic form

We wish to upgrade the Liouville 1-form for the $3+1$ dimensional abelian case

$$\eta_H = p_i dx^i - |\mathbf{p}|dt + A - \mathfrak{a} \tag{5.45}$$

to a more general case where A and $\mathfrak{a}$ are both nonabelian. Following our prescription from eq. (5.34), we define the extended phase space as $M_H = \mathbb{R} \times \mathbb{R}^{2N+1} \times \mathbb{R}^{2N+1} \times \mathcal{O}_\Lambda \times \mathcal{O}_{\mathfrak{L}}$, on which we propose the presymplectic form

$$\eta_H = p_i dx^i - |\mathbf{p}|dt + \mathrm{tr}\left\{Q\left(A + i\, w_R\right)\right\} - \mathrm{tr}\left\{\mathfrak{Q}\left(\mathfrak{a} + i\, \mathfrak{w}_R\right)\right\}, \tag{5.46}$$

where Q and $\mathfrak{Q}$ denote the coordinates on $\mathcal{O}_\Lambda$ and $\mathcal{O}_{\mathfrak{L}}$, respectively, and w_R and $\mathfrak{w}_R$ are the corresponding Liouville 1-forms[5]. To compute ρ_H, we need the Cartan structure equation $dw_R = w_R \wedge w_R$, and

$$dQ = d(g\alpha g^{-1}) = dg\alpha g^{-1} - g\alpha g^{-1} dg g^{-1} = [dg\, g^{-1}, Q] = -[Q, w_R].$$

Thus, for a general matrix-valued r-form X, we can evaluate

$$d\,\mathrm{tr}\left\{QX\right\} = \mathrm{tr}\left\{dQ \wedge X + QdX\right\} = \mathrm{tr}\left\{-[Q, w_R] \wedge X + QdX\right\} = \mathrm{tr}\left\{Q\left(dX - [w_R, X]_\chi\right)\right\}, \tag{5.47}$$

[5] The Liouville 1-form is related to the Maurer-Cartan 1-form on the Lie group manifold. Explicitly, the pullback of w_R under the projection map $G \to G/T = \mathcal{O}_\Lambda$ is simply the (right) Maurer-Cartan form on G.

where $\chi = (-1)^{r+1}$. For $X = A - iw_R$, we get

$$\begin{aligned} d\,\mathrm{tr}\,\{Q\,(A + i\,w_R)\} &= \mathrm{tr}\left\{Q\left(dA + iw_R^2 - [w_R, A]_+ - i[w_R, w_R]_+\right)\right\} \\ &= \mathrm{tr}\,\{Q\,(dA - iA \wedge A + i(A \wedge A + iw_R \wedge A + iA \wedge w_R - w_R \wedge w_R)\} \\ &= \mathrm{tr}\left\{Q\left(F + i(A + iw_R)^2\right)\right\}, \end{aligned} \tag{5.48}$$

where $F = dA - iA^2$. With a similar computation for $d\,\mathrm{tr}\,\{\mathfrak{Q}\mathfrak{a}\}$, the generalized symplectic form becomes

$$\rho_H = dp_i \wedge dx^i - d|\mathbf{p}| \wedge dt + \widetilde{F} - \widetilde{\mathfrak{F}} + i\,\mathrm{tr}\left\{Q\,(A + iw_R)^2\right\} - i\,\mathrm{tr}\left\{\mathfrak{Q}\,(a + i\mathfrak{w}_R)^2\right\}, \tag{5.49}$$

where we have defined $\widetilde{F} = \mathrm{tr}\,\{QF\}$ and $\widetilde{\mathfrak{F}} = \mathrm{tr}\,\{\mathfrak{Q}\mathfrak{F}\}$. Using the fact that the gauge curvatures transform covariantly under gauge transformations, as well as eq. (5.33) for the last two terms, we deduce that all terms in the symplectic form are gauge invariant under both G and $\mathrm{Spin}(2N)$.

Note that $\widetilde{F}$ and $\widetilde{\mathfrak{F}}$ can be thought of as analogues of F and $\mathfrak{F}$ from the abelian case. They are closed on M_H(analogous to $dF = d\mathfrak{F} = 0$), which follows from $dF - i[A, F] = 0$, the nonabelian generalization of Maxwell's equations. Explicitly, using eq. (5.47),

$$\begin{aligned} d\,\mathrm{tr}\left\{Q\left(F + i(A + iw_R)^2\right)\right\} &= \mathrm{tr}\left\{Q\left(dF - i[A, dA] + [dA, w_R] + [\{A, w_R\}, w_R] - [dA, w_R] + [(w_R)^2, A]\right)\right\} \\ &= \mathrm{tr}\left\{Q\left(dF - i[A, dA] + [dA, w_R] + [\{A, w_R\}, w_R] - [dA, w_R] + [(w_R)^2, A]\right)\right\} \\ &= \mathrm{tr}\,\{Q\,(dF - i[A, F])\} = 0. \end{aligned} \tag{5.50}$$

Thus, we can schematically write $d\widetilde{\mathfrak{F}} \sim 2\pi\,\delta^n(\mathbf{p})d^n p$, which again represents the Berry monopole at $\mathbf{p} = \mathbf{0}$.

The extended phase space is endowed with the volume form

$$\Omega_H = \frac{1}{(2\pi)^m\,m!}\,\rho_H^m \wedge dt, \tag{5.51}$$

where

$$m = \frac{1}{2}\left(\dim\left(M_H\right) - 1\right) = n + m_\Lambda + m_{\mathfrak{L}}, \qquad n = 2N + 1, \quad m_a = \frac{1}{2}\dim\left(\mathcal{O}_a\right), \tag{5.52}$$

Recall that for a given $x \in \mathbb{R}^{2N+1}$, the gauge field $A(x)$ (but not its derivatives) can be set to zero by gauge transformations. Thus, by suitable gauge transformations, at a given point on $(t, \mathbf{x}, \mathbf{p}) \in \mathbb{R}^{4N+3}$, the last two terms of eq. (5.49) can be made into $-i\mathrm{tr}\,\{Qw_R \wedge w_R\}$ and $i\mathrm{tr}\,\{\mathfrak{Q}\mathfrak{w}_R \wedge \mathfrak{w}_R\}$, which are simply the symplectic forms on $\mathcal{O}_\Lambda$ and $\mathcal{O}_{\mathfrak{L}}$, respectively. Thus, we can factor out the volume forms on $\mathcal{O}_\Lambda$ and $\mathcal{O}_{\mathfrak{L}}$ to

write the volume form Ω_H as

$$\Omega_H = \frac{1}{(2N+1)!}\left(dp_i \wedge dx^i - d|\mathbf{p}| \wedge dt + \widetilde{F} - \widetilde{\mathfrak{F}}\right)^{2N+1} \wedge dt \wedge \mu_\Lambda \wedge \mu_{\mathfrak{L}}, \tag{5.53}$$

where μ_Λ and $\mu_{\mathfrak{L}}$ are the normalized invariant volume measures on $\mathcal{O}_\Lambda$ and $\mathcal{O}_{\mathfrak{L}}$, respectively.

5.3.2 Anomaly calculation

We can now compute the nonabelian singlet and gauge anomaly, analogous to the abelian case of Sec 5.1.2. We start with the singlet anomaly, i.e, the nonconservation of particle number current:

$$J^\mu = \int_{\mathscr{P}} f\, \dot{x}^\mu\, \Omega_H \tag{5.54}$$

where the integral is over $\mathscr{P} = \mathbb{R}^{2N+1} \times \mathcal{O}_\Lambda \times \mathcal{O}_{\mathfrak{L}}$. Using eq. (2.92) and assuming that the distribution function, f, is advected with the flow, i.e, $\mathcal{L}_{\widetilde{\mathfrak{d}}H} f = 0$, the conservation law for the particle number current becomes

$$d \star J = \int_{\mathscr{P}} \mathcal{L}_{\widetilde{\mathfrak{d}}H}(f\Omega_H) = \frac{1}{(2\pi)^m (m-1)!} \int_{\mathscr{P}} f\, d\rho_H^m, \tag{5.55}$$

where we have used eq. (2.79) for the Lie derivative. Using eq. (5.53), we are left with the integral

$$d \star J = \frac{1}{(2\pi)^{2N+1}(2N+1)!} \int_{\mathcal{O}_\Lambda} \mu_\Lambda \int_{\mathcal{O}_{\mathfrak{L}}} \mu_{\mathfrak{L}} \int_{\mathbb{R}^{2N+1}} d\left(dp_i \wedge dx^i - d|\mathbf{p}| \wedge dt + \widetilde{F} - \widetilde{\mathfrak{F}}\right)^{2N+1} \tag{5.56}$$

Since the RHS vanishes far from $\mathbf{p} = \mathbf{0}$, we use the same trick as eq. (5.23) for the abelian case to write this integral as a boundary integral over $S^{2N}(\epsilon)$. To compute the integral over $S^{2N}(\epsilon)$, define

$$I_N = \frac{1}{(2N+1)!} \int_{S^{2N}} \left(dp_i \wedge dx^i - \hat{p}_i dp_i \wedge dt + \widetilde{F} - \widetilde{\mathfrak{F}}\right)^{2N+1}. \tag{5.57}$$

Since the integrand is a $(4N+2)$-form, we need terms in its expansion containing exactly $2N$ dp's and $(2N+2)$ dx's. All such terms are of the form

$$\left(dp_i \wedge dx^i - d|\mathbf{p}| \wedge dt\right)^{2(N-k)} \wedge \widetilde{\mathfrak{F}}^k \wedge \widetilde{F}^{k+1}; \quad 0 \le k \le N.$$

But since $\widetilde{\mathfrak{F}}$ is globally defined and closed on $S^{2N}(\epsilon)$, for $k < N$ we can rewrite these terms as

$$d\left[\left(p_i dx^i - |\mathbf{p}|dt\right) \wedge \left(dp_i \wedge dx^i - d|\mathbf{p}| \wedge dt\right)^{2(N-k)-1} \wedge \widetilde{F}^{k+1} \wedge \widetilde{\mathfrak{F}}^k\right]; \quad 0 \le k < N.$$

This is a boundary term and $\partial S^{2N} = 0$, so that these terms integrate out to zero[6]. This trick does not work for $k = N$, since $[2(N-k)-1]_{k=N} = -1$. Thus, only the term with $k = N$ can potentially integrate to a nonzero value over S^{2N}, and we are left with

$$I_N = \frac{1}{(2N+1)!}\binom{2N+1}{N+1}\int_{S^{2N}} \widetilde{F}^{N+1} \wedge \left(-\widetilde{\mathfrak{F}}\right)^N = \frac{1}{(N+1)!}\widetilde{F}^{N+1}\int_{S^{2N}} \frac{1}{N!}\left(-\widetilde{\mathfrak{F}}\right)^N. \tag{5.58}$$

The conservation law becomes

$$d\star J = \left[\frac{1}{(N+1)!}\int_{\mathcal{O}_\Lambda} \mu_\Lambda \left(\frac{\widetilde{F}}{2\pi}\right)^{N+1}\right] \cdot \left[\frac{(-1)^N}{N!}\int_{S^{2N}(\epsilon)}\int_{\mathcal{O}_\mathfrak{L}} \mu_\mathfrak{L}\left(\frac{\widetilde{\mathfrak{F}}}{2\pi}\right)^N\right]. \tag{5.59}$$

The integral over $S^{2N}\times\mathcal{O}_\mathfrak{L}$ can be computed using the integral over the coadjoint orbit(eq. (5.36)) as

$$\int_{\mathcal{O}_\mathfrak{L}} \widetilde{\mathfrak{F}}^N d\mu_\mathfrak{L} = \mathfrak{F}^{a_1}\ldots\mathfrak{F}^{a_N}\int_{\mathcal{O}_\mathfrak{L}} d\mu_\Lambda \mathfrak{Q}_{a_1}\ldots\mathfrak{Q}_{a_N} = \mathfrak{F}^{a_1}\ldots\mathfrak{F}^{a_N}\mathrm{Str}\{\lambda_{a_1}\ldots\lambda_{a_N}\} = \mathrm{Str}_\mathfrak{L}\left(\mathfrak{F}^N\right), \tag{5.60}$$

where $\mathrm{Str}_\mathfrak{L}(\,.\,)$ denotes the symmetrized trace over the representation $\mathfrak{Q}$. Thus,

$$\frac{(-1)^N}{N!}\int_{S^{2N}(\epsilon)}\int_{\mathcal{O}_\mathfrak{L}} \mu_\mathfrak{L}\left(\frac{\widetilde{\mathfrak{F}}}{2\pi}\right)^N = \frac{(-1)^N}{N!}\int_{S^{2N}(\epsilon)} \mathrm{Str}_\mathfrak{L}\left(\frac{\widetilde{\mathfrak{F}}}{2\pi}\right)^N = 1, \tag{5.61}$$

where we identify the last integral as the Chern character of $\mathfrak{F}$, and use eq. (4.54) to set $C_+ = (-1)^N$ for the positive helicity Weyl fermion. Similarly,

$$\int_{\mathcal{O}_\Lambda} \widetilde{F}^{N+1} d\mu_\Lambda = \mathrm{Str}_\Lambda\left(F^{N+1}\right). \tag{5.62}$$

Substituting eqns (5.61) and (5.62) in eq. (5.59),

$$d\star J = \frac{1}{(N+1)!}\,\mathrm{Str}_\Lambda\left(\frac{\widetilde{F}}{2\pi}\right)^{N+1}, \tag{5.63}$$

which is the nonabelian singlet anomaly for the particle number current.

Similarly, to compute the gauge anomaly, we define the gauge current as

$$J^\mu_a = \int_{\mathscr{P}} f\, Q_a \dot{x}^\mu\, \Omega_H. \tag{5.64}$$

[6] Alternatively, we can use the homology $H^k\left(S^{2N},\mathbb{Z}\right) = \mathbb{Z}$, $k = 0$, $2N$ and 0 otherwise[73, 70] to infer that S^{2N} has no nontrivial cycles in any dimensions except 0 and $2N$, so that the only closed-but-not-exact forms that can integrate to a nonzero value are top forms on S^{2N}.

Using eq. (2.91), the conservation law for the gauge current becomes

$$d \star J_a - \int_{\mathscr{P}} \underbrace{\left(\sum_{i=n+1}^{2m} \dot{\zeta}^i \frac{\partial Q_a}{\partial \zeta^i} \right)}_{= \dot{Q}_a} f \, \Omega_H = \int_{\mathscr{P}} Q_a \mathcal{L}_{\tilde{\mathfrak{d}} H}(f \Omega_H). \tag{5.65}$$

Using $\dot{Q}_a = \mathfrak{f}_{ab}^c A_\mu^b \dot{x}^\mu Q^c$, the second term on the LHS becomes $\mathfrak{f}_{ab}^c A_\mu^b J_c^\mu$. Thus,

$$\mathcal{D} \star J \equiv d \star J_a + \star \left[\mathfrak{f}_{ab}^c A_\mu^b \dot{x}^\mu Q^c \right] = \frac{1}{(N+1)!} \operatorname{Str}_\Lambda \left(\lambda_a \frac{\widetilde{F}}{2\pi} \right)^{N+1}, \tag{5.66}$$

which is the expression for the nonabelian gauge anomaly in $2N+2$ spacetime dimensions.

5.3.3 Under the hood

In this section, we expand out the general but somewhat formal computation of nonabelian gauge anomaly from the last section to the more familiar form for the case of $3+1$ dimensions. This calculation is a direct generalization of the abelian anomaly calculation by Stephanov and Yin, as described in Sec 5.1.1. It further exposes the connection between the 'classical' coordinates Q on the coadjoint orbit and its quantum analogue. The fundamental reason why this calculation is tractable in $3+1$ dimensions is that in 3 dimensions, the exterior product of two 1-forms is dual to another 1-form, or, in more plebeian terms, that we have '*vector product identities*'.

In $3+1$ dimensions, our proposed semiclassical action for Weyl fermions becomes

$$S[\mathbf{x}, \mathbf{p}, g] = \int dt \left[\mathbf{p} \cdot \dot{\mathbf{x}} + \operatorname{tr} \left\{ Q \left(A_\mu \dot{x}^\mu + i \dot{g} g^{-1} \right) \right\} - |\mathbf{p}| - \mathfrak{a} \cdot \dot{\mathbf{p}} \right]. \tag{5.67}$$

The Berry phase is abelian, as was the case in Sec 5.1.1, and $A_\mu \dot{x}^\mu = A_0 + \mathbf{A} \cdot \dot{\mathbf{x}}$, as $\dot{x}^0 = \dot{t} = 1$. By varying the action w.r.t $\mathbf{x}(t)$, $\mathbf{p}(t)$ and $g(t)$, or alternatively, by expanding out $i_{\tilde{\mathfrak{d}} H} \rho_H = 0$, we can express the equations of motion in a forms similar to eq. (5.4) as

$$\begin{aligned} \dot{\mathbf{x}} =& \hat{\mathbf{p}} + \dot{\mathbf{p}} \times \mathfrak{b}, \\ \dot{\mathbf{p}} =& \operatorname{tr} \{ Q \mathbf{E} \} + \dot{\mathbf{x}} \times \operatorname{tr} \{ Q \mathbf{B} \}, \\ \dot{Q} =& -i [Q, A_0 + \mathbf{A} \cdot \dot{\mathbf{x}}], \end{aligned} \tag{5.68}$$

where we have defined the matrix-valued *electric* and *magnetic* components of the Maxwell tensor as

$$E_i = F_{i0}, \quad B_i = \frac{1}{2}\epsilon_{ijk}F_{jk}; \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\mu A_\nu - i[A_\mu, A_\nu]. \tag{5.69}$$

The nonabelian version of Maxwell's equation is $dF - i[A, F] = 0$, which can be written in coordinates as

$$\partial_{[\mu}F_{\nu\lambda]} - i[A_{[\mu}, F_{\nu\lambda]}] = 0, \tag{5.70}$$

where the square brackets in the subscript denotes cyclic permutation over indices. For all spacelike indices, we get

$$\epsilon^{ijk}\partial_i F_{jk} = i\epsilon^{ijk}[A_i, F_{jk}] \implies \partial_i B^i = i[A_i, B^i], \tag{5.71}$$

while setting $\lambda = 0$ and μ, ν to be spacelike, we get

$$\partial_i F_{j0} + \partial_j F_{0i} + \partial_t F_{ij} = i\left([A_i, F_{j0}] + [A_j, F_{0i}] + [A_0, F_{ij}]\right), \tag{5.72}$$

which, on multiplication by $\frac{1}{2}\epsilon^{ijk}$, gives

$$\begin{aligned} \frac{1}{2}\epsilon^{ijk}\left(\partial_i E_j - \partial_j E_i\right) &= \frac{1}{2}\epsilon^{ijk}\left(-\partial_t F_{ij} + i[A_i, E_j] - i[A_j, E_i] + [A_0, F_{ij}]\right) \\ \implies \epsilon^{kij}\partial_i E_j &= -\partial_t B^k + i[A_0, B^k] + i\epsilon^{kij}[A_i, E_j]. \end{aligned} \tag{5.73}$$

These are the nonabelian analogues of Maxwell equations derived from $dF = 0$. In vector notation, these can be written as

$$\begin{aligned} \nabla \cdot \mathbf{B} &= i(\mathbf{A}\cdot\mathbf{B} - \mathbf{B}\cdot\mathbf{A}) \\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} + i(A_0\mathbf{B} - \mathbf{B}A_0) + i(\mathbf{A}\times\mathbf{E} + \mathbf{E}\times\mathbf{A}). \end{aligned} \tag{5.74}$$

The gauge anomaly arises from a breakdown of Liouville's theorem, $\mathcal{L}_{\tilde{\delta}_H}\Omega_H = 0$. Thus, we need to compute

$$\star\,\mathcal{L}_{\tilde{\delta}_H}\Omega_H = \frac{\partial}{\partial t}\left(\sqrt{\rho_H}\right) + \frac{\partial}{\partial x^i}\left(\sqrt{\rho_H}\,\dot{x}^i\right) + \frac{\partial}{\partial p^i}\left(\sqrt{\rho_H}\,\dot{p}^i\right) + \frac{\partial}{\partial Q^a}\left(\sqrt{\rho_H}\,\dot{Q}^a\right), \tag{5.75}$$

where we now have an extra term corresponding to the coadjoint orbit coordinate. But $Q = Q^a\lambda_a$ satisfies

$$\dot{Q} = \dot{Q}^a\lambda_a = -iQ^b(A_0 + \mathbf{A}\cdot\dot{\mathbf{x}})^c[\lambda_b, \lambda_c] \Rightarrow \dot{Q}^a = \mathfrak{f}^a_{bc}\,Q^b(A_0^c + \mathbf{A}^c\cdot\dot{\mathbf{x}}) = \mathfrak{f}^a_{bc}\,Q^b A^c_\mu \dot{x}^\mu, \tag{5.76}$$

so that

$$\frac{\partial}{\partial Q^a}\left(\sqrt{\rho_H}\,\dot{Q}^a\right) = \frac{\partial}{\partial Q^a}\left[\sqrt{\rho_H}\,\mathfrak{f}^a_{bc}\,Q^b(A^c_\mu \dot{x}^\mu)\right] = \mathfrak{f}^a_{bc}A^c_\mu\left[\sqrt{\rho_H}\,\dot{x}^\mu\delta^b_a + Q^b\frac{\partial}{\partial Q^a}\left(\sqrt{\rho_H}\,\dot{x}^\mu\right)\right]$$
$$= \mathfrak{f}^a_{bc}Q^b\,A^0_\mu\frac{\partial}{\partial Q^a}\left(\sqrt{\rho_H}\right) + \mathfrak{f}^a_{bc}Q^b\,\mathbf{A}^c\cdot\frac{\partial}{\partial Q^a}\left(\sqrt{\rho_H}\,\dot{\mathbf{x}}\right), \tag{5.77}$$

where we have used the antisymmetry of the structure constants to set $\mathfrak{f}^a_{bc}\,\delta^b_a = 0$. Substituting in eq. (5.75),

$$\star\,\mathcal{L}_{\tilde{\mathfrak{d}}_H}\Omega_H = \left(\frac{\partial}{\partial t} + \mathfrak{f}^c_{ab}\,Q^a A^b_0\frac{\partial}{\partial Q^c}\right)\sqrt{\rho_H} + \left(\frac{\partial}{\partial x^i} + \mathfrak{f}^c_{ab}Q^a A^b_i\frac{\partial}{\partial Q^c}\right)\left(\sqrt{\rho_H}\,\dot{x}^i\right) + \frac{\partial}{\partial p^i}\left(\sqrt{\rho_H}\,\dot{p}^i\right). \tag{5.78}$$

Comparing with eq. (5.5), we note that including the coadjoint orbit has essentially replaced the partial derivatives with classical analogues of the (gauge) covariant derivatives! Similarly, the assumption that f be advected with the flow can be written explicitly as

$$\left[\left(\frac{\partial}{\partial t} + \mathfrak{f}^c_{ab}\,Q^a A^b_0\frac{\partial}{\partial Q^c}\right) + \dot{x}_i\left(\frac{\partial}{\partial x_i} + \mathfrak{f}^c_{ab}Q^a A^b_i\frac{\partial}{\partial Q^c}\right) + \dot{p}_i\frac{\partial}{\partial p_i}\right] f = 0. \tag{5.79}$$

The equations of motion can be solved for $\dot{\mathbf{x}}$ and $\dot{\mathbf{p}}$ as

$$\sqrt{\rho_H}\,\dot{\mathbf{x}} = \hat{\mathbf{p}} + \mathrm{tr}\,\{Q\mathbf{E}\}\times\mathfrak{b} + (\mathfrak{b}\cdot\hat{\mathbf{p}})\,\mathrm{tr}\,\{Q\mathbf{B}\},$$
$$\sqrt{\rho_H}\,\dot{\mathbf{p}} = \mathrm{tr}\,\{Q\mathbf{E}\} + \hat{\mathbf{p}}\times\mathrm{tr}\,\{Q\mathbf{B}\} + (\mathrm{tr}\,\{Q\mathbf{E}\}\cdot\mathrm{tr}\,\{Q\mathbf{B}\})\,\mathfrak{b}, \tag{5.80}$$

with $\sqrt{\rho_H} = 1 + \mathfrak{b}\cdot\mathrm{tr}\,\{Q\mathbf{B}\}$. Thus,

$$\left(\frac{\partial}{\partial t} + \mathfrak{f}^c_{ab}\,Q^a A^b_0\frac{\partial}{\partial Q^c}\right)\sqrt{\rho_H} = \mathfrak{b}\cdot\mathrm{tr}\left\{Q(\dot{\mathbf{B}} - i[A_0,\mathbf{B}])\right\}$$
$$\left(\frac{\partial}{\partial x^i} + \mathfrak{f}^c_{ab}Q^a A^b_i\frac{\partial}{\partial Q^c}\right)\left(\sqrt{\rho_H}\,\dot{x}^i\right) = \mathfrak{b}\cdot\mathrm{tr}\,\{Q(\nabla\times\mathbf{E} - i(\mathbf{A}\times\mathbf{E} + \mathbf{E}\times\mathbf{A}))\}$$
$$+ (\hat{\mathbf{p}}\cdot\mathfrak{b})\mathrm{tr}\,\{Q(\nabla\cdot\mathbf{B} - i[\mathbf{A},\mathbf{B}])\}$$
$$\frac{\partial}{\partial p^i}\left(\sqrt{\rho_H}\,\dot{p}^i\right) = \mathrm{tr}\,\{Q\mathbf{E}\}\cdot\mathrm{tr}\,\{Q\mathbf{B}\}\,\nabla\cdot\mathfrak{b}. \tag{5.81}$$

Substituting in eq. (5.78) and using the nonabelian Maxwell's equations, we are again left with only the $\nabla\cdot\mathfrak{b}$ term. Hence, combining these terms with the Boltzmann equation, we get

$$\left(\frac{\partial}{\partial t} + \mathfrak{f}^c_{ab}\,Q^a A^b_0\frac{\partial}{\partial Q^c}\right)\left(\sqrt{\rho_H}\,f\right) + \left(\frac{\partial}{\partial x^i} + \mathfrak{f}^c_{ab}Q^a A^b_i\frac{\partial}{\partial Q^c}\right)\left(\sqrt{\rho_H}\,f\,\dot{x}^i\right) + \frac{\partial}{\partial p^i}\left(\sqrt{\rho_H}\,f\,\dot{p}^i\right)$$
$$= -f(Q,\mathbf{x},\mathbf{p})\,\mathrm{tr}\,\{Q\mathbf{E}\}\cdot\mathrm{tr}\,\{Q\mathbf{B}\}\,(\nabla\cdot\mathfrak{b}) \tag{5.82}$$

Integrating over the momentum space and setting $f(Q,\mathbf{x},\mathbf{0})=0$ yields

$$-\frac{1}{(2\pi)^2}\mathrm{tr}\,\{Q\mathbf{E}\}\cdot\mathrm{tr}\,\{Q\mathbf{B}\} = \int\frac{d^3p}{(2\pi)^3}\left(\frac{\partial}{\partial t}+\mathfrak{f}^c_{ab}\,Q^aA^b_0\frac{\partial}{\partial Q^c}\right)(\sqrt{\rho_H}\,f)+\left(\frac{\partial}{\partial x^i}+\mathfrak{f}^c_{ab}Q^aA^b_i\frac{\partial}{\partial Q^c}\right)\left(\sqrt{\rho_H}\,f\,\dot{x}^i\right)$$
$$=\partial_\mu\left[\int\frac{d^3p}{(2\pi)^3}\sqrt{\rho_H}\,f\,\dot{x}^\mu\right]+\frac{\partial}{\partial Q^c}\left[\mathfrak{f}^c_{ab}\,Q^aA^b_\mu\int\frac{d^3p}{(2\pi)^3}\sqrt{\rho_H}\,f\,\dot{x}^\mu\right], \tag{5.83}$$

where in the last step, we have taken $A^b_\mu Q_c$ inside the Q-derivative, since A is independent of Q and $\mathfrak{f}^c_{ab}\delta^a_c=0$.

To compute the anomalies, we define the particle number and the gauge 4-currents as

$$J^\mu(\mathbf{x},t) = \int\frac{d^3p}{(2\pi)^3}\mu_\Lambda\,f(Q,\mathbf{x},\mathbf{p})\sqrt{\rho_H}\,\dot{x}^\mu,$$
$$J^\mu_a(\mathbf{x},t) = \int\frac{d^3p}{(2\pi)^3}\mu_\Lambda\,Q_a\,f(Q,\mathbf{x},\mathbf{p})\sqrt{\rho_H}\,\dot{x}^\mu, \tag{5.84}$$

where μ_Λ is the invariant measure over $\mathcal{O}_\Lambda$. For the singlet anomaly, we simply need to integrate eq. (5.83) over $\mathcal{O}_\Lambda$ and use the fact that $\partial\mathcal{O}_\Lambda=0$ to get

$$\partial_\mu J^\mu=-\frac{1}{(2\pi)^2}\mathbf{E}^a\cdot\mathbf{B}^b\int_{\mathcal{O}_\Lambda}\mu_\Lambda\,Q_aQ_b=-\frac{1}{(2\pi)^2}\mathbf{E}^a\cdot\mathbf{B}^b\,\mathrm{Str}\{\lambda_a\lambda_b\}=-\frac{1}{(2\pi)^2}\mathrm{Str}\{\mathbf{E}\cdot\mathbf{B}\}. \tag{5.85}$$

On the other hand, for the conservation law of the gauge current, we need to multiply eq. (5.83) by Q_a and then integrate over $\mathcal{O}_\Lambda$. Here, Q_a is merely a weight in the integral and does not parametrize the trajectory, so that $\dot{Q}_a=0$. Thus, the first term of the integral is simply $\partial_\mu J^\mu_a$, while the second term, on integration by parts, becomes

$$\int_{\mathcal{O}_\Lambda}\mu_\Lambda\,Q_a\frac{\partial}{\partial Q^c}\left[\mathfrak{f}^c_{eb}\,Q^eA^b_\mu\int\frac{d^3p}{(2\pi)^3}\sqrt{\rho_H}\,f\,\dot{x}^\mu\right]=-\int_{\mathcal{O}_\Lambda}\mu_\Lambda\,\delta_{ac}\left[\mathfrak{f}^c_{eb}\,Q^eA^b_\mu\int\frac{d^3p}{(2\pi)^3}\sqrt{\rho_H}\,f\,\dot{x}^\mu\right]=\mathfrak{f}^c_{ab}\,A^b_\mu J^\mu_c.$$

where we have used the antisymmetry of $\mathfrak{f}^c_{ab}$ in the last step. Thus,

$$\partial_\mu J^\mu_a+\mathfrak{f}^c_{ab}A^b_\mu J^\mu_c=-\frac{1}{(2\pi)^2}\mathbf{E}^b\cdot\mathbf{B}^c\left(\int_{\mathcal{O}_\Lambda}\mu_\Lambda Q_aQ_bQ_c\right)=-\frac{1}{(2\pi)^2}\mathbf{E}^b\cdot\mathbf{B}^c\,\mathrm{Str}\,\{\lambda_a\lambda_b\lambda_c\}\,, \tag{5.86}$$

which is the covariant form of the nonabelian gauge anomaly.

Thus, using vector product identities (and some tedious algebra), we recover the expressions for the nonabelian anomalies in $3+1$ dimensions. The central point of this calculation is the intricate fashion in which the terms depending on Q assemble to form covariant derivative of $\dot{x}^\mu$ instead of the usual partial derivative.

5.4 Anomalous hydrodynamics using chiral kinetic theory

In this section, we construct a hydrodynamic description for positive chirality Weyl fermions in $2N+2$ spacetime dimensions coupled to a U(1) gauge field using a semiclassical description and compare it to the results derived from thermodynamic considerations in Sec 3.3. This involves using kinetic theory to construct the macroscopic currents using the equations of motion and a phase space density, as described in Sec 2.5.

Extended phase space

In Sec 5.3, we discussed the semiclassical Hamiltonian description of Weyl fermions in an inertial reference frame. However, for hydrodynamics, it is more natural to consider the co-moving frame, defined by the given velocity field $u^\mu(x)$. As the frame may in general possess a nonzero acceleration as well as vorticity ($\Omega = du \neq 0$), we need a way to include the inertial forces in our formalism. In Appendix C.2, we derive the generalized symplectic form in a noninertial reference frame, and show that for massless particles, it is reasonable to include the inertial forces in the symplectic form as $\rho_H \to \rho_H + \varepsilon\Omega$ (accurate upto the linear order in Ω). This is reminiscent of the minimal coupling to the electromagnetic field, with the energy ε serving as the "charge".

The relevant extended phase space is given by $\mathcal{M}_H = \mathbb{R}^n \times \mathbb{R} \times \mathcal{O}_{\mathfrak{L}}$, $n = 2N+1$, where $\mathcal{O}_{\mathfrak{L}}$ is the coadjoint orbit corresponding to the dequantization of the nonabelian Berry phase, and we do not need the corresponding coadjoint orbit since the gauge group is U(1). The phase space is $2M+1$ dimensional, with $M = n + \frac{1}{2}\dim(\mathcal{O}_{\mathfrak{L}}) \equiv n + m_\Lambda$. The Weyl fermions, in a reference frame comoving with a fluid of velocity u such that $\Omega = d\bar{u}$, is then described by the generalized symplectic form

$$\rho_H \equiv d\eta_H = dp_i \wedge dx^i - d\varepsilon \wedge dt + F + \varepsilon\,\Omega - \widetilde{\mathfrak{F}} - i\,\mathrm{tr}\left\{\mathfrak{Q}\left(\mathfrak{w}_R - i\mathfrak{a}\right)^2\right\}, \tag{5.87}$$

where we have locally set $u^\mu = (1, 0, \ldots 0)$ by a suitable Lorentz transform, so that $-dt = u$. We shall need both positive and negative energy sectors here, so that $\varepsilon = c|\mathbf{p}|,\ c = \pm 1$.

Grand potential current

In Sec 3.3, we showed that the anomalous contribution to macroscopic currents can all be derived from a grand potential current $\vec{\mathcal{G}}$. To derive this current from the microscopic description, we assume that the Weyl fermions are in equilibrium(and hence described the the Fermi-Dirac statistics) with respect to the reference frame in which the generalized symplectic form is given by eq. (5.87). Then $\mathcal{G}^\mu$ is simply the

current associated with the 1-particle grand potential $g(\varepsilon) = -T\,\ln\left(1+e^{-\beta(\varepsilon-\mu)}\right)$ (See Appendix C.1 for details), so that using eq. (2.89), we can write $\mathcal{G}^\mu$ as the phase space integral

$$\mathcal{G}^\mu = \int_{\mathscr{P}} \frac{d^n p}{(2\pi)^n}\, d\mu_\Lambda\, \sqrt{\rho_H}\, g(\varepsilon)\, \dot{x}^\mu, \tag{5.88}$$

where $\mathscr{P} = \mathbb{R}^n \times \mathcal{O}_{\mathfrak{L}}$. To compute this, in principle, one next needs to solve the equation of motion ($i_{\tilde{\mathfrak{d}}_H}\rho_H = 0$) for $\sqrt{\rho_H}\,\dot{x}^i$, which can then be integrated over the momentum space and the co-adjoint orbit. In 3+1 dimensions, this is straightforward[26], and one gets

$$\sqrt{\rho_H}\,\dot{\mathbf{x}} = c\,\hat{\mathbf{p}} + \mathfrak{b}\times\mathbf{E} + (\hat{\mathbf{p}}\cdot\mathfrak{b})\,(\mathbf{B} + c|\mathbf{p}|\boldsymbol{\omega})\,, \qquad \mathfrak{b}^i = \frac{1}{2}\epsilon^{ijk}\mathfrak{F}_{jk}. \tag{5.89}$$

However, the task is much more complicated in spacetime dimensions greater than 4. Thus, we follow an alternative approach using the symplectic formulation of classical mechanics, which lets us compute such currents without computing $\sqrt{\rho_H}\,\dot{x}^i$ explicitly.

Taking the Hodge dual of eq. (5.88), we get

$$\bar{\mathcal{G}} = \frac{1}{(2\pi)^n}\int_{\mathscr{P}} g(\varepsilon)\sqrt{\rho_H}\left(\sum_{i=1}^{n}(-1)^i dx^1\wedge\ldots dx^{i-1}\wedge\dot{x}^i dx^{i+1}\wedge\ldots dx^n\right)\frac{d^n p}{(2\pi)^n}\wedge dt\wedge d\mu_\Lambda, \tag{5.90}$$

where we have used overbars for Hodge duals, a notation introduced in eq. (3.38). The differential form in the parenthesis is simply

$$i_{\tilde{\mathfrak{d}}_H}\,(d^n x) = i_{\tilde{\mathfrak{d}}_H}\left(\bigwedge_{i=1}^{n} dx^i\right) = \sum_{i=1}^{n}(-1)^i dx^1\wedge\ldots dx^{i-1}\wedge\dot{x}^i dx^{i+1}\wedge\ldots dx^n,$$

so that the integration measure of eq. (5.90) can readily be obtained as an antiderivation of the symplectic volume form, with only one term at $O(d^{n-1}x)$:

$$\begin{aligned} i_{\tilde{\mathfrak{d}}_H}\Omega_H &= i_{\tilde{\mathfrak{d}}_H}\left(\sqrt{\rho_H}\, d^n x\wedge d^n p\wedge dt\wedge d\mu_\Lambda\right)\\ &= \sqrt{\rho_H}\left[i_{\tilde{\mathfrak{d}}_H}\,(d^n x)\wedge d^n p\wedge dt\wedge d\mu_\Lambda + \text{terms involving } d^n x\right]. \end{aligned} \tag{5.91}$$

Thus, to compute $\bar{\mathcal{G}}$, we simply need to integrate $i_{\tilde{\mathfrak{d}}_H}\Omega_H$ terms over $\mathscr{P} = \mathbb{R}^{2N+1}\times\mathcal{O}_{\mathfrak{L}} \cong \mathbb{R}^+\times S^{2N}\times\mathcal{O}_{\mathfrak{L}}$, where $\mathbb{R}^+$ denotes the radial $|\mathbf{p}|$ axis. This automatically picks out the relevant terms, as any terms containing all

dx's will not be a top form on $\mathscr{P}$. Using the equation of motion $i_{\tilde{\partial}_H}\rho_H = 0$, we get

$$i_{\tilde{\partial}_H}\Omega_H = i_{\tilde{\partial}_H}\left(\frac{1}{m!}\rho_H^m \wedge dt\right) = \frac{1}{m!}\rho_H^m, \tag{5.92}$$

so that

$$\begin{aligned}\bar{\mathcal{G}} &= \frac{1}{(2\pi)^m m!}\int_{\mathscr{P}} g(\varepsilon)\rho_H^m \\ &= \frac{1}{(2\pi)^n n!}\int_0^\infty d|\mathbf{p}|\, g(\varepsilon)\int_{\mathcal{O}_{\mathfrak{L}}}\mu_{\mathfrak{L}}\int_{S^{2N}}\left(dp_i\wedge dx^i - d\varepsilon\wedge dt + F + \varepsilon\,\Omega - \widetilde{\mathfrak{F}}\right)^n \wedge(-dt),\end{aligned} \tag{5.93}$$

where we have peeled off one $(-d\varepsilon\wedge dt)$ from ρ_H^m to get the measure on $\mathbb{R}^+$. Except for the integral over ε, this is identical to the expression for the conservation law in eq. (5.56). As in Sec 5.3, eq. (5.61) , we can perform the integral over $\mathcal{O}_{\mathfrak{L}}\times S^{2N}$ to get

$$\bar{\mathcal{G}} = \frac{c^2}{N!}(-dt)\wedge\int_0^\infty \frac{d|\mathbf{p}|}{2\pi}g(\varepsilon)\left(\frac{F+\varepsilon\Omega}{2\pi}\right)^N. \tag{5.94}$$

Finally, substituting $u = -dt$, $c^2 = 1$ and using eq. (3.43),

$$\bar{\mathcal{G}}_c = \frac{u}{N!}\wedge\int_0^\infty \frac{d|\mathbf{p}|}{2\pi}g(c|\mathbf{p}|)\left(\frac{qB + c|\mathbf{p}|\omega}{2\pi}\right)^N, \tag{5.95}$$

where the subscript $c = s\pm 1$ denotes the positive/negative energy sector. This is precisely obtained eq. (3.65), which was derived in Ref [16] using thermodynamic arguments.

At finite temperatures, there will be excitations from both positive and negative energy sectors of the positive chirality Weyl cone. In order to derive physically meaningful expressions at a finite temperature, we must include both of them, and define

$$\begin{aligned}\bar{\mathcal{G}} &= \bar{\mathcal{G}}_+ + \bar{\mathcal{G}}_- \\ &= \frac{u}{N!}\wedge\left[\int_0^\infty \frac{d|\mathbf{p}|}{2\pi}g(|\mathbf{p}|)\left(\frac{B+|\mathbf{p}|\omega}{2\pi}\right)^N + \int_0^\infty \frac{d|\mathbf{p}|}{2\pi}g(-|\mathbf{p}|)\left(\frac{B-|\mathbf{p}|\omega}{2\pi}\right)^N\right].\end{aligned} \tag{5.96}$$

Substituting $|\mathbf{p}| = \varepsilon$ in the first integral and $|\mathbf{p}| = -\varepsilon$ in the second, we get

$$\bar{\mathcal{G}} = \frac{u}{N!}\wedge\int_{-\infty}^\infty \frac{d\varepsilon}{2\pi}g(\varepsilon)\left(\frac{B+\varepsilon\omega}{2\pi}\right)^N. \tag{5.97}$$

This integral is clearly divergent, as $g(\varepsilon)\sim(\varepsilon-\mu)$ for $\varepsilon\to-\infty$. This is expected, as we are integrating over

an infinitely deep Dirac sea. In order to regularize this integral, we need to subtract off the zero temperature vacuum contribution, where we define the "vacuum" as the many-body state where all 1-particle states with $\varepsilon < 0$ (i.e, below the Weyl node) are filled up. Since at $T = 0$, $g(\varepsilon) = (\varepsilon - \mu)\Theta(\mu - \varepsilon)$, where $\mu > 0$, define the regularized grand potential current as

$$\bar{\mathcal{G}}^{reg} = \frac{u}{N!} \wedge \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} \left[g(\varepsilon) - (\varepsilon - \mu)\Theta(-\varepsilon) \right] \left(\frac{B + \varepsilon\omega}{2\pi} \right)^N . \tag{5.98}$$

Using the explicit form of $g(\varepsilon)$, we next need to compute this integral explicitly. We use a standard trick:

Generating function

As in Sec 3.3, we define a generating function for $\bar{\mathcal{G}}$ as

$$\begin{aligned} \bar{\mathcal{G}}^{reg}_{\tau} &= u \wedge \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} \left[g(\varepsilon) - (\varepsilon - \mu)\Theta(-\varepsilon) \right] \sum_{N=0}^{\infty} \frac{\tau^N}{N!} \left(\frac{B + \varepsilon\omega}{2\pi} \right)^N \\ &= u \wedge e^{\frac{\tau q B}{2\pi}} \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} \left[g(\varepsilon) - (\varepsilon - \mu)\Theta(-\varepsilon) \right] e^{\frac{\tau\varepsilon\omega}{2\pi}} , \end{aligned} \tag{5.99}$$

where the sum should be thought of as a formal sum, i.e, we treat ω and B as c-numbers instead of differential forms. The integral so obtained is simply a generating function for a Sommerfeld expansion. To evaluate the integral, consider

$$I(\sigma) = \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} \left[g(\varepsilon) - (\varepsilon - \mu)\Theta(-\varepsilon) \right] e^{\sigma\varepsilon}. \tag{5.100}$$

The Heaviside integral part is easily evaluated using Feynman's trick:

$$\int_{-\infty}^{0} \frac{d\varepsilon}{2\pi} (\varepsilon - \mu) e^{\sigma\varepsilon} = \left(\frac{\partial}{\partial\sigma} - \mu \right) \int_{-\infty}^{0} \frac{d\varepsilon}{2\pi} e^{\sigma\varepsilon} = \left(\frac{\partial}{\partial\sigma} - \mu \right) \frac{1}{2\pi\sigma} = -\frac{1 + \mu\sigma}{2\pi\sigma^2}. \tag{5.101}$$

For the remaining integral, substitute $s = e^{\beta(\varepsilon - \mu)}$ and integrate by parts:

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} g(\varepsilon) e^{\sigma\varepsilon} &= -\frac{1}{2\pi\beta} \int_0^{\infty} \frac{ds}{\beta s} \ln\left(1 + \frac{1}{s}\right) \left(s\, e^{\beta\mu} \right)^{\frac{\sigma}{\beta}} \\ &= -\frac{e^{\mu\sigma}}{2\pi\beta^2} \int_0^{\infty} ds\, s^{\frac{\sigma}{\beta} - 1} \ln\left(1 + \frac{1}{s}\right) \\ &= -\frac{e^{\mu\sigma}}{2\pi\beta^2} \left[\frac{\beta}{\sigma} s^{\frac{\sigma}{\beta}} \ln\left(1 + \frac{1}{s}\right) \Big|_0^{\infty} - \frac{\beta}{\sigma} \int_0^{\infty} ds\, s^{\frac{\sigma}{\beta}} \left(-\frac{1}{s(s+1)} \right) \right] \\ &= -\frac{e^{\mu\sigma}}{2\pi\sigma\beta} \int_0^{\infty} ds \frac{s^{\frac{\sigma}{\beta} - 1}}{s + 1} = -\frac{e^{\mu\sigma}}{2\sigma\beta \sin\left(\pi\sigma/\beta\right)}, \end{aligned} \tag{5.102}$$

where in the last line, assuming $\sigma < \beta$, we have used the integral

$$\int_0^\infty ds\, \frac{s^{\alpha-1}}{1+s} = \frac{\pi}{\sin(\pi\alpha)}, \quad 0 < \alpha < 1. \tag{5.103}$$

From eq. (5.102) and eq. (5.101), we get

$$I(\sigma) = -\frac{1}{2\pi\sigma^2}\left[\frac{\frac{\pi\sigma}{\beta}}{\sin\left(\frac{\pi\sigma}{\beta}\right)}e^{\mu\sigma} - (1+\mu\sigma)\right]. \tag{5.104}$$

We also note that

$$\frac{1}{2\pi\sigma^2}\frac{\frac{\pi\sigma}{\beta}}{\sin\left(\frac{\pi\sigma}{\beta}\right)}e^{\mu\sigma} = \frac{1}{2\pi\sigma^2} + \frac{\mu}{2\pi\sigma} + \left(\frac{\mu^2}{4\pi} + \frac{\pi}{12\beta^2}\right)\sigma + O(\sigma^2), \tag{5.105}$$

so that the integral of the 'regulator' $(\varepsilon-\mu)\Theta(-\varepsilon)e^{\sigma\varepsilon}$ precisely subtracts off the singularities of the divergent integral of $g(\varepsilon)e^{\sigma\varepsilon}$.

Substituting the integral in eq. (5.99), the generator of the regularized grand potential current becomes

$$\bar{\mathcal{G}}_\tau^{reg} = -u \wedge e^{\frac{\tau q B}{2\pi}}\frac{2\pi}{(\omega\tau)^2}\left[\frac{\frac{\omega\tau}{2\beta}}{\sin\left(\frac{\omega\tau}{2\beta}\right)}e^{\frac{\mu\omega\tau}{2\pi}} - \left(1+\frac{\mu}{2\pi}\omega\tau\right)\right], \tag{5.106}$$

which is identical to the expression obtained[7] in Ref. [16], as discussed at the end of Sec 3.3. Thus, we get the magical similarity between the generating function for the grand potential function and that for the mixed anomaly for Weyl fermions, which essentially follows from the fact that the Sommerfeld integral is equal to the generating function for the $\hat{A}$-genus.

[7] After replacing $\omega \to 2\omega_A$, as they defines their vorticity as the angular velocity, ω_A.

Part II

Transfer matrices

6 Transfer matrices

Transfer matrices, also known as the monodromy matrix in dynamical systems literature[54, 55], arise naturally in discrete calculus as an alternative representation of finite order linear homogeneous difference equations (recurrence relation). Dividing the system into finite-sized blocks, they are linear operators that implement a first order shift on a block. While containing the same information as the original differential equation, for periodic systems, they are a useful tool to extract the global behavior of the solutions, since their spectra determine the asymptotics of the solutions.

For noninteracting tight-binding models of condensed matter systems, the wavefunctions are solutions of the Schrödinger equation, which can be recast as a linear homogeneous difference equation. The asymptotics of wavefunctions then determine the physical nature of the solutions, *viz*, the bulk and the boundary eigenstates. Thus, transfer matrices have been used to study lattice models for topological phases of matter, which exhibit a nontrivial bulk state topology as well as nontrivial edge states. However, the traditional construction of transfer matrix for such systems fails when a certain "hopping" matrix is singular, which happens to be the case for quite a few commonly encountered models (as we shall see in Ch 7).

In this chapter, we propose an alternative construction of transfer matrices which works even when the hopping matrix is singular. We discuss various features of this construction, as well as the general formalism to extract physically meaningful data, *viz*, the bulk and edge spectra, from the transfer matrix. These somewhat formal analyses are substantiated by explicit computations for an assortment of models of common interest in Ch 7.

6.1 Transfer matrix preliminaries

We begin by illustrating the idea of a transfer matrix using a simple example. Consider then the homogeneous linear difference equation

$$\mathbf{z}_n = \sum_{\ell=1}^{R} A_{n,\ell}\, \mathbf{z}_{n-\ell}; \quad n > 0, \tag{6.1}$$

where $\mathbf{z}\colon \mathbb{Z} \to \mathbb{C}^N$, $A_{n,m} \in \mathrm{Mat}(N, \mathbb{C})$ and the initial condition $\mathbf{z}_n = \mathbf{z}_n^0$ for $-R < n \leq 0$. Then, this difference equation can easily be recast as

$$\underbrace{\begin{pmatrix} \mathbf{z}_n \\ \mathbf{z}_{n-1} \\ \vdots \\ x_{n-R} \end{pmatrix}}_{\Phi_n} = \underbrace{\begin{pmatrix} A_{n,1} & A_{n,2} & \dots & A_{n,R-1} & A_{n,R} \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}}_{T_n} \underbrace{\begin{pmatrix} \mathbf{z}_{n-1} \\ \mathbf{z}_{n-2} \\ \vdots \\ x_{n-R-1} \end{pmatrix}}_{\Phi_{n-1}}. \tag{6.2}$$

This reduces the problem to the form $\Phi_n = T_n \Phi_{n-1}$, i.e, a first order difference equation for $\Phi\colon \mathbb{Z} \to \mathbb{C}^{RN}$, where T_n are the *transfer matrices*. We can "solve" the system by setting

$$\Phi_n = T_n T_{n-1} \dots T_1 \Phi_0; \qquad \Phi_0 = \begin{pmatrix} \mathbf{z}_0^0 & \mathbf{z}_{-1}^0 & \dots & \mathbf{z}_{-R+1}^0 \end{pmatrix}^{\mathrm{T}}. \tag{6.3}$$

Furthermore, if the system were periodic, i.e, if $A_{n,\ell}$ is independent of n for any ℓ, then the solution can simply be written as $\Phi_n = T^n \Phi_0$.

Thus, in general, given a linear homogeneous discrete equation with solution $\Phi\colon \mathbb{Z} \to \mathbb{C}^{\mathscr{N}}$, the transfer matrices are a set of linear operators $T_n\colon \mathbb{C}^{\mathscr{N}} \to \mathbb{C}^{\mathscr{N}}$, $n \in \mathbb{Z}$ such that we can write $\Phi_n = T_n \Phi_{n-1}$. In the following analysis, we shall always assume our difference equation to be translation invariant, so that $T_n = T$. Given the initial condition $\Phi_0 = \varphi$, the solution can then be computed explicitly as $T^n \varphi$ for $n > 0$, (and also for $n < 0$ if T is invertible). More importantly, however, the transfer matrix is an excellent tool to extract the asymptotic behavior of the solutions, which is our primary interest. We shall next show that the information about the asymptotics is encoded in $\sigma[T]$, the spectrum of T.

To begin with, if $\Phi_0 = \varphi$ is an eigenstate of T with eigenvalue ρ, then

$$\Phi_n = T^n \varphi = \rho^n \varphi, \tag{6.4}$$

so that the solution grows exponentially with base ρ. Thus, for $n \to \infty$, the solution grows exponentially for $|\rho| > 1$, decays exponentially for $|\rho| < 1$ and stays finite for $|\rho| = 1$. For arbitrary φ, let $|\,.\,|$ be a vector norm on $\mathbb{C}^{\mathscr{N}}$ and $|\varphi|$ be finite. Then,

$$\lim_{n\to\infty} |\Phi_n| = \lim_{n\to\infty} |T^n \varphi| \leq |\varphi| \lim_{n\to\infty} \|T^n\|\,; \qquad \|T^n\| = \sup_{|\Phi|=1} |T^n \Phi|\,, \tag{6.5}$$

where $\|\,.\,\|$ is the matrix norm on $\mathrm{Mat}(\mathscr{N}, \mathbb{C})$. If T is a normal matrix[1], it can be written as $T = U \cdot \Theta \cdot U^\dagger$,

[1]Recall that a complex square matrix X is termed *normal* if it commutes with its adjoint, i.e, if $X \cdot X^\dagger = X^\dagger \cdot X$. The

where $U \in \mathrm{U}(N)$, so that

$$\lim_{n\to\infty} |\Phi_n| \leq |\varphi| \lim_{n\to\infty} \left\| U \cdot \Theta^n \cdot U^\dagger \right\| \leq |\varphi| \lim_{n\to\infty} \|\Theta^n\| \leq |\varphi| \lim_{n\to\infty} \|\Theta\|^n , \tag{6.6}$$

since $\|U\| = 1$. Furthermore, since $\Theta = \mathrm{diag}\{\rho_1, \rho_2, \dots \rho_{\mathscr{N}}\}$, with the $\rho_i \in \sigma[T]$ indexed in an increasing order of magnitude, i.e, $|\rho_1| \leq |\rho_2| \leq \cdots \leq |\rho_{\mathscr{N}}|$, its matrix norm can be explicitly computed to be $|\rho_{\mathscr{N}}|$. This provides an upper bound on the growth rates of the solutions for arbitrary initial values φ. Thus, for a normal T,

$$\sigma[T] \subset B^2 \implies \text{All solutions decay as } n \to \infty, \tag{6.7}$$

where $B^2 \equiv \{z \in \mathbb{C} \,|\, |z| < 1\}$ is the open unit disk in the complex plane,

For a normal T, we can further refine our analysis by defining projectors $\mathcal{P}_s \colon \mathbb{C}^{\mathscr{N}} \to \mathbb{C}^{\mathscr{N}}$ to the eigenspace of T corresponding to the eigenvalue ρ_s. Explicitly, if ρ_s has multipliticy k, its eigenvalue problem $T\varphi_{s,i} = \rho_s \varphi_{s,i},\ i = 1, \dots k$ define the projectors explicitly as $\mathcal{P}_s = \sum_{i=1}^{k} \varphi_{s,i}\varphi_{s,i}^\dagger$. The diagonalization of T is equivalent to writing it as

$$T = \sum_s \rho_s \mathcal{P}_s, \qquad \mathcal{P}_s\mathcal{P}_{s'} = \mathcal{P}_{s'}\mathcal{P}_s = \delta_{ss'}\mathcal{P}_s, \tag{6.8}$$

with $\rho_s \in \sigma[T]$. Using these projectors, we define

$$\mathcal{P}_< \equiv \sum_{|\rho_s|<1} \mathcal{P}_s, \quad \mathcal{P}_> \equiv \sum_{|\rho_s|>1} \mathcal{P}_s, \quad \mathcal{P}_= \equiv \sum_{|\rho_s|=1} \mathcal{P}_s, \tag{6.9}$$

which satisfy $\mathcal{P}_< + \mathcal{P}_> + \mathcal{P}_= = \mathbb{1}$. Defining

$$T_< = T\mathcal{P}_< = \sum_{s,s',|\rho_s|<1} \rho_{s'}\mathcal{P}_{s'}\mathcal{P}_s = \sum_{s,s',|\rho'_s|<1} \rho_{s'}\delta_{ss'}\mathcal{P}_s = \sum_{|\rho_s|<1} \rho_s\mathcal{P}_s, \tag{6.10}$$

we by definition have $\sigma[T_<] \subset B^2$. Thus, from eq. (6.7), a sufficient condition for $\varphi \in \mathbb{C}^{\mathscr{N}}$ to decay for $n \to \infty$ is simply $\mathcal{P}_<\varphi = \varphi$.

We finally show that the last statement, with suitable definitions of projectors, holds even when T is not normal. Consider the most general case when T may not be diagonalizable at all, i.e, including cases when $\mathrm{rank}(T) < \dim(T)$, so that the set of eigenvectors of T do not span $\mathbb{C}^{\mathscr{N}}$; however, the set of generalized eigenvectors of T does span $\mathbb{C}^{\mathscr{N}}$. Given $\rho \in \sigma[T]$, such that its geometric multiplicity is less than its algebraic multiplicity k, we define a set of left(φ) and right(ϕ) generalized eigenvectors, which satisfy the

matrix X is diagonalizable by a unitary matrix iff it is normal[96].

generalized eigenvalue equations

$$(T-\rho\mathbb{1})^k\varphi=0,\qquad \phi^\dagger(T-\rho\mathbb{1})^k=0. \tag{6.11}$$

The generalized eigenvectors define a basis of $\mathbb{C}^{\mathscr{N}}$, in which T can be expressed in an "almost diagonal" form, which is the so-called *Jordan canonical form*[97]:

$$T=\sum_s[\rho_s\mathcal{P}_s+\mathcal{D}_s],\quad \mathcal{P}_s\mathcal{P}'_s=\delta_{ss'}\mathcal{P}_s,\quad \mathcal{P}_{s'}\mathcal{D}_s=\mathcal{D}_s\mathcal{P}_{s'}=\mathcal{D}_s\delta_{ss'}, \tag{6.12}$$

where $\rho_s\in\sigma[T]$ and $\mathcal{D}_s$ are nilpotent operators of order equal to the algebraic multiplicity of ρ_s. We can again define the projectors as in eq. (6.9), and construct

$$T_<\equiv T\mathcal{P}_<=\sum_{s,s',|\rho_s|<1}[\rho_{s'}\mathcal{P}_{s'}+\mathcal{D}_{s'}]\,\mathcal{P}_s=\sum_{|\rho_s|<1}[\rho_s\mathcal{P}_s+\mathcal{D}_s] \tag{6.13}$$

which satisfies $\sigma[T_<]\subset B^2$ by the spectral radius formula[98]. Thus, from eq. (6.7), a sufficient condition for $\varphi\in\mathbb{C}^{\mathscr{N}}$ to decay for $n\to\infty$ is simply $\mathcal{P}_<\varphi=\varphi$.

Finally, note that if T is nonsingular, a given Φ_0 can also be propagated backwards, i.e, for $n<0$. Then we can also study the asymptotic behavior for $n\to-\infty$, which is identical to the $n\to\infty$ behavior with T replaced by T^{-1}, i.e,

$$\lim_{n\to-\infty}|\Phi_n|=\lim_{n\to-\infty}|T^n\varphi|\le|\varphi|\lim_{n\to-\infty}\left\|T^{-|n|}\right\|\le|\varphi|\lim_{n\to-\infty}\left\|T^{-1}\right\|^n. \tag{6.14}$$

Thus, for a normal T, the growth rate for $n\to-\infty$ is bounded by $|\rho_1|$, the smallest eigenvalue (in magnitude), while that for $n\to\infty$ is bounded by $|\rho_{\mathscr{N}}|$, the largest eigenvalue (in magnitude).

6.2 Tight-binding lattice models

Given a tight-binding Hamiltonian on a lattice, the Schrödinger equation (i.e, the eigenvalue problem of the Hamiltonian) can be written as a complex linear homogeneous difference equation, which could be represented using transfer matrices. Conventionally, one needs to invert a *hopping matrix* to reduce it to the form of eq. (6.1), after which the transfer matrix can be constructed as in eq. (6.2). However, the construction breaks down if the hopping matrix turns out to be singular, which, as we show in Ch 7, happens for many systems of common interest. In this section, we present an alternative construction of transfer matrices, which works

equally well for both singular and nonsingular *hopping matrices*.

6.2.1 The fermionic Hilbert space

We begin with a formal (and somewhat unconventional[2]) exposition of tight-binding models and the associated mathematical objects, and construct the many-body Hilbert space. As we are using only the single-particle sector of this general construction and hence do not necessarily need fermionic statistics, this subsection can be skipped without a loss of continuity.

Tight-binding models are defined on a lattice graph $\mathbf{\Lambda}$ whose sites are indexed by $\mathbf{n} = (n_1, n_2, \dots n_N) \in \mathbb{Z}^N$. Physically, the lattice is embedded in the real space $\mathbb{R}^d$, with a set of "lattice vectors" $\mathbf{a}_i \in \mathbb{R}^d,\ i = 1, \dots N$, so that the position of site $\mathbf{n}$ is given by $\mathbf{r_n} = n_i \mathbf{a}_i \in \mathbb{R}^n$. To each lattice site, we associated a (single) fermionic Fock space. Mathematically, to each $\mathbf{n} \in \mathbf{\Lambda}$ we associate a two dimensional complex Hilbert space[3] $\mathscr{H}_n \cong \mathbb{C}^2$, spanned by orthonormal basis vectors $|0_\mathbf{n}\rangle$ and $|1_\mathbf{n}\rangle$. Physically, these vectors correspond to the quantum states where the site contains either zero or one electron, following Pauli's exclusion principle for fermions.

Two linear operators, $\mathfrak{c}_\mathbf{n}$ ("annihilation operator") and $\mathfrak{c}^\dagger_\mathbf{n}$("creation operator") on $\mathscr{H}_\mathbf{n}$, can be defined on $\mathscr{H}_\mathbf{n}$, which act on the basis elements as

$$\begin{aligned} \mathfrak{c}_\mathbf{n}|1_\mathbf{n}\rangle = |0_\mathbf{n}\rangle, \quad \mathfrak{c}_\mathbf{n}|0_\mathbf{n}\rangle = 0, \\ \mathfrak{c}^\dagger_\mathbf{n}|0_\mathbf{n}\rangle = |1_\mathbf{n}\rangle, \quad \mathfrak{c}^\dagger_\mathbf{n}|1_\mathbf{n}\rangle = 0. \end{aligned} \tag{6.15}$$

The composition of these operators, $\hat{n}_\mathbf{n} = \mathfrak{c}^\dagger_\mathbf{n}\mathfrak{c}_\mathbf{n} \colon \mathscr{H}_\mathbf{n} \to \mathscr{H}_\mathbf{n}$, is a Hermitian operator with eigenstates $|0_\mathbf{n}\rangle$ and $|1_\mathbf{n}\rangle$, the corresponding eigenvalues being 0 and 1. Thus, $\hat{n}_\mathbf{n}$ "counts" the number of electrons at site $\mathbf{n}$. Explicitly, we can also check that

$$[\mathfrak{c}_\mathbf{n}, \mathfrak{c}^\dagger_\mathbf{n}]_+ = \mathfrak{c}_\mathbf{n}\mathfrak{c}^\dagger_\mathbf{n} + \mathfrak{c}^\dagger_\mathbf{n}\mathfrak{c}_\mathbf{n} = \mathbb{1}_\mathbf{n}. \tag{6.16}$$

The space of linear transformations $\mathcal{H}_\mathbf{n} \to \mathcal{H}_\mathbf{n}$ is isomorphic to $\mathrm{Mat}(2, \mathbb{C})$, the space of 2×2 complex matrices, which is spanned by the set $\{\mathfrak{c}, \mathfrak{c}^\dagger, \mathfrak{c}^\dagger\mathfrak{c}, \mathfrak{c}\mathfrak{c}^\dagger\}$.

[2] Conventionally, to derive a tight binding model for a given system, one starts with the continuum single electron Schrödinger equation represented in suitable (*Wannier*) basis where the basis states are localized on lattice sites, and then performs second quantization to represent states in the occupation number (Fock) basis. Thus, one first constructs $\mathscr{H}^{(1)}$ and then uses it to construct many-body Hilbert spaces. We instead start from Fock bases on each site, which is somewhat unphysical, but conceptually cleaner, in my opinion.

[3] Strictly speaking, we only care about normalized wavefunctions, so this space should be $\mathbb{CP}^1$, the space of rays in $\mathbb{C}^2$. However, we shall consider all vectors in $\mathbb{C}^2$ for now, and 'normalize' our wavefunctions only at the very end.

The *many body Hilbert space* $\mathscr{H}$ is defined as the antisymmetric tensor product over the Hilbert spaces $\mathscr{H}_{\mathbf{n}}$ corresponding to single sites. The relevant operators on $\mathscr{H}$ are the fermionic creation and annihilation operators, $c^\dagger_{\mathbf{n}}$ and $c_{\mathbf{n}}$, which create and annihilate an electron at site $\mathbf{n}$, respectively, and satisfy the fermionic anticommutation relation

$$[c_{\mathbf{m}}, c_{\mathbf{n}}]_+ = [c^\dagger_{\mathbf{m}}, c^\dagger_{\mathbf{n}}]_+ = 0, \quad [c_{\mathbf{m}}, c^\dagger_{\mathbf{n}}]_+ = \delta_{\mathbf{mn}}, \tag{6.17}$$

which encode the fermionic statistics, i.e, the antisymmetry of many body wavefunctions under the exchange of two particles. The definition of these operators using tensor products of $\mathfrak{c}$'s and $\mathfrak{c}^\dagger$'s is quite nontrivial, so we start with an illustration of the relevant ideas in the simplest possible case: a lattice with just two sites.

Set $\mathbf{\Lambda} = \{1, 2\}$, with the corresponding Hilbert spaces $\mathscr{H}_i$, $i = 1, 2$. Recall that tensor product is noncommutative ($A \otimes B \neq B \otimes A$ in general), so that we define a direct sum of all possible orderings, *viz*, $\widetilde{\mathscr{H}} = (\mathscr{H}_1 \otimes \mathscr{H}_2) \oplus (\mathscr{H}_2 \otimes \mathscr{H}_1)$. The fermionic Hilbert space is then a subspace of $\widetilde{\mathscr{H}}$, which consists of only the antisymmetric linear combinations. Explicitly, given $|\psi_i\rangle \in \mathscr{H}_i$, we define the elements of $\mathscr{H}$ as the antisymmetrized tensors

$$|\psi_1\rangle \wedge |\psi_2\rangle = \frac{1}{2}\left(|\psi_1\rangle \otimes |\psi_2\rangle - |\psi_2\rangle \otimes |\psi_1\rangle\right). \tag{6.18}$$

A suitable basis for $\mathscr{H}$ is $\{|0_1\rangle \wedge |0_2\rangle, |1_1\rangle \wedge |0_2\rangle, |0_1\rangle \wedge |1_2\rangle, |1_1\rangle \wedge |1_2\rangle\}$. Next, we need to define the creation/annihilation operators as tensor products of $\mathfrak{c}_i$'s and $\mathbb{1}_i$'s. Let us try

$$c^\dagger_1 = \mathfrak{c}^\dagger_1 \otimes \mathbb{1}_2 + \mathbb{1}_2 \otimes \mathfrak{c}^\dagger_1 \equiv \mathcal{S}\left[\mathfrak{c}^\dagger_1 \otimes \mathbb{1}_2\right], \tag{6.19}$$

where $\mathcal{S}[\,.\,]$ denotes symmetrization. The action on the basis vectors of $\mathscr{H}$ is

$$c^\dagger_1|0_1\rangle \wedge |\mathbf{e}_2\rangle = |1_1\rangle \wedge |\mathbf{e}_2\rangle, \qquad c^\dagger_1|1_1\rangle \wedge |\mathbf{e}_2\rangle = 0; \qquad \mathbf{e} \in \{0, 1\} \tag{6.20}$$

However, we cannot define $c^\dagger_2$ in an identical fashion, since we need $[c^\dagger_1, c^\dagger_2] = 0$. More explicitly, let us define

$$c^\dagger_1|\mathbf{e}_1\rangle \wedge |0_2\rangle = \lambda_{\mathbf{e}}|\mathbf{e}_1\rangle \wedge |0_2\rangle, \qquad c^\dagger_1|\mathbf{e}_1\rangle \wedge |1_2\rangle = 0; \qquad |\lambda_{\mathbf{e}}| = 1. \tag{6.21}$$

Since the only state on which $c^\dagger_1 c^\dagger_2$ gives a nonzero answer is $|\Omega\rangle = |0_1\rangle \wedge |0_2\rangle$, consider

$$0 = [c^\dagger_1, c^\dagger_2]_+|\Omega\rangle = \lambda_0 c^\dagger_1|0_1\rangle \wedge |1_2\rangle + c^\dagger_2|1_1\rangle \wedge |0_2\rangle = (\lambda_0 + \lambda_1)\,|1_1\rangle \wedge |1_2\rangle, \tag{6.22}$$

so that $\lambda_1 = -\lambda_0$. Thus, the action of $c^\dagger_2$ on a state must depend on the occupancy of site 1. The way out

is to define the operators as

$$c_1 = \mathcal{S}\left[\mathfrak{c}_1 \otimes \mathbb{1}_2\right], \quad c_1^\dagger = \mathcal{S}\left[\mathfrak{c}_1^\dagger \otimes \mathbb{1}_2\right], \quad c_2 = \mathcal{S}\left[(-1)^{\hat{n}_1} \otimes \mathfrak{c}_2\right], \quad c_2^\dagger = \mathcal{S}\left[(-1)^{\hat{n}_1} \otimes \mathfrak{c}_2^\dagger\right], \tag{6.23}$$

so that $\lambda_0 = 1$ and $\lambda_1 = -1$. These satisfy the canonical anticommutation relations by construction.

We can now generalize to arbitrary lattices based on the 2-site case. We begin by choosing a total ordering[4] of $\mathbf{\Lambda}$ as $\mathbf{n}_1 < \mathbf{n}_2 < \ldots$, and defining $\widetilde{\mathscr{H}}$ analogous to the 2-site case as

$$\widetilde{\mathscr{H}} = \bigoplus_{\sigma \in S_{\mathbf{\Lambda}}} \left[\mathscr{H}_{\sigma(\mathbf{n}_1)} \otimes \mathscr{H}_{\sigma(\mathbf{n}_2)} \otimes \ldots\right], \tag{6.24}$$

where $S_{\mathbf{\Lambda}}$ is the permutation group on $\mathbf{\Lambda}$. The many-body Hilbert space $\mathscr{H}$ is simply defined as the set of antisymmetric tensors in $\widetilde{\mathscr{H}}$,

$$\mathscr{H} = \bigwedge_{\mathbf{n} \in \mathbf{\Lambda}} \mathscr{H}_{\mathbf{n}} = \mathrm{span}\left\{\bigwedge_{\mathbf{n} \in \mathbf{\Lambda}} |\mathbf{e}_{i_{\mathbf{n}},\mathbf{n}}\rangle\right\}, \qquad \mathbf{e}_{i,\mathbf{n}} \in \{0,1\}. \tag{6.25}$$

where the $\wedge$-product is defined as in eq. (6.21). The 1-particle creation/annihilation operators can then be constructed as

$$c_{\mathbf{n}}^\dagger = \mathcal{S}\left[\left(\bigotimes_{\mathbf{m}<\mathbf{n}} (-1)^{\hat{n}_{\mathbf{m}}}\right) \otimes \mathfrak{c}_{\mathbf{n}}^\dagger \otimes \left(\bigotimes_{\mathbf{m}>\mathbf{n}} \mathbb{1}_{\mathbf{m}}\right)\right], \tag{6.26}$$

and a similar definition of $c_{\mathbf{n}}$. Note that these operators are highly nonlocal, owing to the (potentially infinitely long) string[5] of $(-1)^{\hat{n}_{\mathbf{m}}}$ needed to ensure the fermionic statistics. Thus, in order to satisfy the canonical anticommutation relations, the fermions are required to be nonlocal objects, as this construction shows.

The total number operator $\hat{n}$ is $\hat{n} = \sum_{\mathbf{n} \in \mathbf{\Lambda}} c_{\mathbf{n}}^\dagger c_{\mathbf{n}}$, and $\sigma[\hat{n}]$ consists of positive integers. The vacuum state $|\Omega\rangle = \bigwedge_{\mathbf{n} \in \mathbf{\Lambda}} |0_{\mathbf{n}}\rangle$ can be alternatively defined using these operators either as $|\Omega\rangle \in \ker \hat{n}$, which is unique upto the choice of a phase, or as the state which satisfies $c_{\mathbf{n}}|\Omega\rangle = 0 \; \forall \, \mathbf{n} \in \mathbf{\Lambda}$. All other states in $\mathscr{H}$ can then be obtained by application of $c_{\mathbf{n}}^\dagger$'s on $|\Omega\rangle$. This lets us partition $\mathscr{H}$ into different particle number sectors as

$$\mathscr{H} = \bigoplus_{n=0}^{\infty} \mathscr{H}^{(n)}, \qquad \mathscr{H}^{(n)} \equiv \{|\Psi\rangle \in \mathscr{H};\ \hat{n}|\Psi\rangle = n|\Psi\rangle\}. \tag{6.27}$$

The zero particle sector of $\mathscr{H}$ is 1-dimensional, and spanned by $|\Omega\rangle$. In the rest of this chapter, we shall

[4] The fact that such a total order exists for any $\mathbf{\Lambda}$ follows from Zorn's lemma.

[5] This issue appears when one tries to represent a fermionic model as a spin model, as we are tacitly doing by associating $\mathbb{C}^2$ to each site, which is also the Hilbert space for the spin 1/2 representation of SU(2). In that context, the string of operators is known as a *Jordan string*.

only be interested in the 1-particle section $\mathscr{H}^{(1)}$, which is spanned by the states $c^\dagger_{\mathbf{n}}|\Omega\rangle$, which correspond to an electron localized at site $\mathbf{n}$. Thus, a generic wavefunction $|\psi\rangle \in \mathscr{H}^{(1)}$ can be expressed as

$$|\psi\rangle = \sum_{\mathbf{n}\in\mathbf{\Lambda}} \psi_{\mathbf{n}} c^\dagger_{\mathbf{n}}|\Omega\rangle; \quad \psi_{\mathbf{n}} \in \mathbb{C}. \tag{6.28}$$

The inner product on $\mathscr{H}$ can be defined using the pairing to the dual space $\mathscr{H}^*$, whose vacuum state is denoted by $\langle\Omega|$ and whose states can be constructed by applying $c_{\mathbf{n}}$'s it. Explicitly, the dual vector to $|\psi\rangle = c^\dagger_{\mathbf{n}_1}\dots c^\dagger_{\mathbf{n}_k}|\Omega\rangle \in \mathscr{H}$ is $\langle\Omega|c_{\mathbf{n}_k}\dots c_{\mathbf{n}_1} \in \mathscr{H}^*$. The inner product on $\mathscr{H}$ can then be defined using the pairing $\mathscr{H}^* \times \mathscr{H} \to \mathbb{R}$. Operationally, the inner product of $|\psi\rangle$ and $|\phi\rangle$ can be computed by writing out $\langle\psi|\phi\rangle$, using the anticommutation relations to reorder the operators, and finally use $\langle\Omega|\Omega\rangle = 1$ and $c_{\mathbf{n}}|\Omega\rangle = 0$. For instance, given $|\psi\rangle, |\phi\rangle \in \mathscr{H}^{(1)}$,

$$\langle\varphi|\psi\rangle = \sum_{\mathbf{n}\mathbf{n}'} \varphi^*_{\mathbf{n}'}\psi_{\mathbf{n}}\langle\Omega|c_{\mathbf{n}'}c^\dagger_{\mathbf{n}}|\Omega\rangle = \sum_{\mathbf{n}\mathbf{n}'} \varphi^*_{\mathbf{n}'}\psi_{\mathbf{n}}\langle\Omega|\delta_{\mathbf{n}\mathbf{n}'} - c^\dagger_{\mathbf{n}}c_{\mathbf{n}'}|\Omega\rangle = \sum_{\mathbf{n}} \varphi^*_{\mathbf{n}}\psi_{\mathbf{n}}. \tag{6.29}$$

To restrict to the set of physically distinct states, we consider only the normalized states, i,e, $|\psi\rangle \in \mathscr{H}^{(1)}$ such that $\langle\psi|\psi\rangle = 1$, and identify states that differ only by a global phase, i.e, $|\psi\rangle \sim e^{i\varphi}|\psi\rangle$. Thus, the physical states live in the projective space $\mathbb{P}\mathscr{H}^{(1)}$, which is the space of (complex) rays in $\mathscr{H}^{(1)}$.

This picture is often generalized by associating q quantum states with each site $\mathbf{n} \in \mathbf{\Lambda}$, so that $\mathscr{H}_{\mathbf{n}} \cong \mathbb{C}^{2q}$. Physically, these internal states may correspond to spin/orbital/sublattices degrees of freedom. The total Hilbert space on the lattice $\mathbf{\Lambda}$ is then given by $\mathscr{H} = \bigotimes_{\mathbf{n}\in\mathbf{\Lambda}} \mathbb{C}^q$, and the corresponding 1-particle sector is given by $\mathscr{H}^{(1)} = \text{span}\left(c^\dagger_{\mathbf{n},\alpha}|\Omega\rangle\right)$, where $\mathbf{n} \in \mathbf{\Lambda}$ and $\alpha \in \{1, 2, \dots q\}$, so that a generic wavefunction becomes

$$|\psi\rangle = \sum_{\mathbf{n},\alpha} \psi_{\mathbf{n},\alpha} c^\dagger_{\mathbf{n},\alpha}|\Omega\rangle. \tag{6.30}$$

The inner product is similar to the previous case, with a summation over α.

6.2.2 Schrödinger equation and recursion relations

A quantum mechanical description of electrons in a crystal involves a Hamiltonian describing the dynamics of electrons in a lattice formed by the positive ions/nuclei. In order to study the system, one intends to solve the eigenvalue problem for the Hamiltonian operator (i.e, the Schrödinger equation) in a convenient basis. Two common choices are the plane wave basis, consisting of a set of orthonormal states delocalized over the entire system, and the Wannier basis, consisting of a set of states localized at individual sites.

For tight-binding models, we represent the Hamiltonian in the Wannier basis, where the dynamics is generated by the tunneling between different sites. The tunneling between sites $\mathbf{n}$ and $\mathbf{n}'$ is then encoded in the Hermitian operators of the form $c^\dagger_{\mathbf{n}+\mathbf{m},\alpha} t_{\mathbf{n},\mathbf{m},\alpha\beta} c_{\mathbf{n},\beta} + \text{h.c}$, where $\alpha, \beta = 1, \dots d$ and $\mathbf{m} \in \mathbb{Z}^N$. The tight-binding Hamiltonian $\mathcal{H} : \mathscr{H}^{(1)} \to \mathscr{H}^{(1)}$ is generically a sum of such terms. The lattice translation invariance (or equivalently, periodicity) dictates that the hopping parameter t be a function only of the separation $\mathbf{m}$, independent of $\mathbf{n}$. Furthermore, we shall demand that the hoppings be finite range, i.e, that $\exists\, R \in \mathbb{R}^+$ such that $t_{\mathbf{m},\alpha\beta} = 0 \;\forall\, \mathbf{m}$ such that $|\mathbf{m}| > R$. Thus, generically, we shall be interested in 1-particle Hamiltonians of the form

$$\mathcal{H} = \sum_{\mathbf{m},\mathbf{n}\in\mathbf{\Lambda}} \sum_{\alpha,\beta=1}^{q} \left[c^\dagger_{\mathbf{n}+\mathbf{m},\alpha}\, t_{\mathbf{m},\alpha\beta}\, c_{\mathbf{n},\beta} + \text{h.c.} \right]. \tag{6.31}$$

where h.c. denotes the Hermitian conjugate of the first term. We can expand it out and relabel the indices ($\mathbf{n} \to \mathbf{n} + \mathbf{m}$ and $\alpha \leftrightarrow \beta$) to get

$$\mathcal{H} = \sum_{\mathbf{m},\mathbf{n}} \sum_{\alpha,\beta} \left[c^\dagger_{\mathbf{n}+\mathbf{m},\alpha}\, t_{\mathbf{m},\alpha\beta}\, c_{\mathbf{n},\beta} + c^\dagger_{\mathbf{n}-\mathbf{m},\alpha}\, t^*_{\mathbf{m},\beta\alpha}\, c_{\mathbf{n},\beta} \right]. \tag{6.32}$$

We next seek to solve the Schrödinger equation to obtain the spectrum $\sigma[\mathcal{H}]$. Acting on the wavefunction of eq. (6.30), we get

$$\begin{aligned}
\mathcal{H}|\psi\rangle &= \sum_{\mathbf{m},\mathbf{n}} \sum_{\alpha,\beta} \left[c^\dagger_{\mathbf{n}+\mathbf{m},\alpha}\, t_{\mathbf{m},\alpha\beta}\, c_{\mathbf{n},\beta} + c^\dagger_{\mathbf{n}-\mathbf{m},\alpha}\, t^*_{\mathbf{m},\beta\alpha}\, c_{\mathbf{n},\beta} \right] \sum_{\mathbf{n}',\alpha'} \psi_{\mathbf{n}',\alpha'} c^\dagger_{\mathbf{n}',\alpha'} |\Omega\rangle \\
&= \sum_{\mathbf{m},\mathbf{n},\mathbf{n}'} \sum_{\alpha,\beta,\alpha'} \psi_{\mathbf{n}',\alpha'} \left[c^\dagger_{\mathbf{n}+\mathbf{m},\alpha}\, t_{\mathbf{m},\alpha\beta} + c^\dagger_{\mathbf{n}-\mathbf{m},\alpha}\, t^*_{\mathbf{m},\beta\alpha} \right] \left(\delta_{\mathbf{n},\mathbf{n}'} \delta_{\beta\alpha'} \mathbb{1} - c^\dagger_{\mathbf{n}',\alpha'} c_{\mathbf{n},\beta} \right) |\Omega\rangle \\
&= \sum_{\mathbf{m},\mathbf{n}} \sum_{\alpha,\beta} \psi_{\mathbf{n},\beta} \left[c^\dagger_{\mathbf{n}+\mathbf{m},\alpha}\, t_{\mathbf{m},\alpha\beta} + c^\dagger_{\mathbf{n}-\mathbf{m},\alpha}\, t^*_{\mathbf{m},\beta\alpha} \right] |\Omega\rangle,
\end{aligned} \tag{6.33}$$

using the fermionic anticommutation relations and the definition of vacuum $c_{\mathbf{n},\alpha}|\Omega\rangle = 0$. Relabeling the indices, the the Schrödinger equation, $\mathcal{H}|\psi\rangle = \varepsilon|\psi\rangle$, becomes

$$\sum_{\mathbf{n},\alpha} \left(\sum_{\mathbf{m},\beta} \left[t_{\mathbf{m},\alpha\beta} \psi_{\mathbf{n}-\mathbf{m},\beta} + t^*_{\mathbf{m},\beta\alpha} \psi_{\mathbf{n}+\mathbf{m},\beta} \right] - \varepsilon \psi_{\mathbf{n},\alpha} \right) c^\dagger_{\mathbf{n},\alpha} |\Omega\rangle = 0. \tag{6.34}$$

As the states $c^\dagger_{\mathbf{n},\alpha}|\Omega\rangle$ are linearly independent, all their coefficients must vanish individually, so that we get the recursion relation

$$\sum_{\mathbf{m},\beta} \left[t_{\mathbf{m},\alpha\beta} \psi_{\mathbf{n}-\mathbf{m},\beta} + t^*_{\mathbf{m},\beta\alpha} \psi_{\mathbf{n}+\mathbf{m},\beta} \right] - \varepsilon \psi_{\mathbf{n},\alpha} = 0. \tag{6.35}$$

Defining $\boldsymbol{\psi}_{\mathbf{n}} = \left\{\psi_{\mathbf{n},\alpha}\right\}_{\alpha=1}^{q}$ and $\mathbf{t}_{\mathbf{m}} = \left\{t_{\mathbf{m},\alpha\beta}\right\}_{\alpha,\beta=1}^{q}$, we get

$$\sum_{\mathbf{m},\beta}\left[\mathbf{t}_{\mathbf{m}}\boldsymbol{\psi}_{\mathbf{n}-\mathbf{m}} + \mathbf{t}_{\mathbf{m}}^{\dagger}\boldsymbol{\psi}_{\mathbf{n}+\mathbf{m}}\right] - \varepsilon\boldsymbol{\psi}_{\mathbf{n}} = 0. \tag{6.36}$$

This is a representation of Schrödinger's equation for generic tight-binding models as a recursion relation.

Consider now the case of a $d = N = 1$, i.e, $\mathbf{\Lambda} \cong \mathbb{Z}$ embedded in $\mathbb{R}$, so that the $n \in \mathbb{Z}$ corresponds to the point $r = na \in \mathbb{R}$, where a is the lattice constant. The difference equation becomes

$$\sum_{\ell=0}^{R}\left(\mathbf{t}_{\ell}\boldsymbol{\psi}_{n+\ell} + \mathbf{t}_{\ell}^{\dagger}\boldsymbol{\psi}_{n-\ell}\right) = \varepsilon\,\boldsymbol{\psi}_{n}, \tag{6.37}$$

supplanted by suitable boundary conditions. We shall seek to solve this difference equation by using an ansatz of the form $\boldsymbol{\psi}_n = \lambda^n\boldsymbol{\psi}_0$. However, the boundary conditions put additional constraints on the parameter λ. Two commonly used boundary conditions are:

1. **Periodic boundary conditions:** The wavefunction is periodic with period equal to the system size, $L \in a\mathbb{Z}$, i.e, $\psi_{n+L/a} = \psi_n$. Thus, $\lambda^{L/a} = 1 \implies \lambda = e^{ika};\ k \in \frac{2\pi}{L}\mathbb{Z}$.

2. **Infinite system:** On physical grounds, we are only interested in solutions ψ_n that stay finite for $n \to \pm\infty$, i.e, $\lim_{n\to\pm\infty}|\boldsymbol{\psi}_n| < \infty$. Thus, $|\lambda| = 1 \implies \lambda = e^{ika};\ k \in \mathbb{R}$.

For both of these cases, substituting $\boldsymbol{\psi}_n = e^{ika\,n}\boldsymbol{\psi}_0 = e^{ikr}\boldsymbol{\psi}_0$ and collecting coefficients of e^{ikr}, we get

$$0 = \left[\sum_{\ell=0}^{R}\left(\mathbf{t}_{\ell}e^{ika\,\ell} + \mathbf{t}_{\ell}^{\dagger}e^{-ika\,\ell}\right) - \varepsilon\mathbb{1}_q\right]\boldsymbol{\psi}_0 \equiv \left[\mathcal{H}_B(k) - \varepsilon\mathbb{1}_q\right]\boldsymbol{\psi}_0, \tag{6.38}$$

which yields a nontrivial solution for $\boldsymbol{\psi}_0$ iff, following Cramers' rule, ε satisfies

$$\det\left[\mathcal{H}_B(k) - \varepsilon\mathbb{1}_q\right] = 0, \tag{6.39}$$

i.e, ε is an eigenvalue of the operator $\mathcal{H}_B$. This is defined as the *Bloch Hamiltonian* of the system. Given an eigenvector $\boldsymbol{\psi}_0(k)$ of the Bloch Hamiltonian with eigenvalue $\varepsilon(k)$, the corresponding wavefunction is given by $\boldsymbol{\psi}_{n,k} = e^{ikr}\boldsymbol{\psi}_0(k)$.

In condensed matter terminology, we have simply recovered the common knowledge that for infinite systems or periodic boundary conditions, we can diagonalize the tight-binding Hamiltonian in the plane wave basis $\{e^{ikr}\boldsymbol{\psi}_0(k)\}$, where the parameter k is simply the *(quasi-)momentum*. Now, for a d-dimensional system (on $\mathbb{R}^d$), if the system is infinite or periodic along $d' \leq d$ directions, then we can expand in the

plane wave basis along those directions, thereby reducing the problem to one on $\mathbb{R}^{d-d'}$ parametrized by the quasimomentum $\mathbf{k}_\perp = (k_1, k_2, \ldots k_{d'}) \in \mathbb{T}^{d'}$. In quantum mechanical language, this works because the system is translation invariant, so that the (quasi-)momentum $\mathbf{k}_\perp$ is a *good quantum number*[45, 46].

6.2.3 The conventional transfer matrix

For systems with topologically nontrivial bulk bands, we are particularly interested in the boundary states, where the translation symmetry is naturally broken in the direction normal to the edge, as the system is finite in that direction. We consider the simplest situation where such states could exist, viz, a slab geometry. Explicitly, given a lattice $\mathbf{\Lambda}$ embedded in $\mathbb{R}^d$, we shall consider a tight binding system with open (Dirichlet) boundary conditions(OBC) along one direction, which we can choose on will, and periodic boundary conditions(PBC) along $d' = d - 1$ remaining directions parallel to the edge. By choosing the direction with OBC, we can explore the presence of surface states along different surfaces.

Thus, we are starting with a family of 1D tight-binding models parametrized by $\mathbf{k}_\perp \in \mathbb{T}^{d-1}$, which can be written in the position space as

$$\mathcal{H} = \sum_{n=0}^{N} \sum_{\alpha,\beta=1}^{q} \sum_{\ell=0}^{R} \left[c^\dagger_{n+\ell,\alpha} t_{\ell,\alpha\beta} c_{n,\beta} + \text{h.c.} \right] = \sum_{n=0}^{N} \sum_{\ell=0}^{R} \left[c^\dagger_{n+\ell} \mathbf{t}_\ell c_n + \text{h.c.} \right], \tag{6.40}$$

where R is the range of the hopping (which we have assumed to be finite), and we have suppressed the explicit dependence on $\mathbf{k}_\perp$ to avoid notational clutter; however, all parameters should be assumed to depend on $\mathbf{k}_\perp$, unless stated otherwise. The corresponding recursion relation (eq. (6.36)) becomes

$$\sum_{\ell=0}^{R} \left(\mathbf{t}_\ell \boldsymbol{\psi}_{n+\ell} + \mathbf{t}^\dagger_\ell \boldsymbol{\psi}_{n-\ell} \right) = \varepsilon \, \boldsymbol{\psi}_n. \tag{6.41}$$

We can group the wavefunctions $\boldsymbol{\psi}_\ell$ into blocks to reduce this to a first order difference equation[6]. Physically speaking, we are constructing blocks consisting of these *sites*, so that the system is periodic in these blocks and the hopping between such blocks is restricted to nearest neighbor[49], as shown in Fig. 6.1). These blocks, hereafter referred to as *supercells*, can be thought of as sites of a *superlattice* $\widetilde{\mathbf{\Lambda}}$. In terms of these supercells, each containing $\mathscr{N} = qR$ degrees of freedom, the recursion relation can be written as

$$J\Psi_{n+1} + M\Psi_n + J^\dagger \Psi_{n-1} = \varepsilon \Psi_n. \tag{6.42}$$

[6]This is always possible as the hopping has a finite range, hence we can always choose a supercell consisting of R sites. Clearly, this grouping is not unique; however, it will not affect the size and spectral properties of our transfer matrix. We give a more detailed argument towards the end of this section.

Here, J is the *hopping matrix* connecting nearest neighbor supercells and M is the *on-site matrix*, which encodes the hopping between degrees of freedom inside the supercell as well as the on-site energies. The wavefunction for a supercell, Ψ_n, can be explicitly written as

$$\Psi_n = \begin{pmatrix} \psi_n \\ \psi_{n+1} \\ \vdots \\ \psi_{n+R-1} \end{pmatrix} \in \mathbb{C}^{\mathscr{N}}. \tag{6.43}$$

For a nonsingular J, the conventional transfer matrix construction works by noticing that[50]

$$\Psi_{n+1} = J^{-1}(\varepsilon \mathbb{1} - M)\Psi_n - J^{-1}J^{\dagger}\Psi_{n-1} \tag{6.44}$$

can be rewritten as

$$\begin{pmatrix} \Psi_{n+1} \\ \Psi_n \end{pmatrix} = \begin{pmatrix} J^{-1}(\varepsilon \mathbb{1} - M) & -J^{-1}J^{\dagger} \\ \mathbb{1} & 0 \end{pmatrix} \begin{pmatrix} \Psi_n \\ \Psi_{n-1} \end{pmatrix} \equiv T \begin{pmatrix} \Psi_n \\ \Psi_{n-1} \end{pmatrix}, \tag{6.45}$$

which is a matrix analogue of eq. (6.2).

6.3 The generalized transfer matrix

The conventional transfer matrix construction breaks down when the hopping matrix (J) is singular. Such cases have been studied in various context in an ad-hoc fashion; however, a unified treatment has so far been missing. In this section, we describe the construction of a *generalized transfer matrix*, which is well defined even for case with singular hopping matrices, and thus provides a general unified framework to discuss transfer matrices.

6.3.1 Diagonstics

We begin by further analyzing the reason for the breakdown of the conventional transfer matrix construction, viz, the singularity of J, which means that $r = \text{rank}\,(J) < \dim\,(J)$. But $\text{rank}\,(J)$ is the number of linearly independent rows in J, and hence the 'dimension' of the recursion relation in eq. (6.42). A more physical way to visualize $\text{rank}\,(J)$ is to think of the $\mathscr{N}$ degrees of freedom inside each supercell as $\mathscr{N}$ sites[7]. Then, $\text{rank}\,(J)$ denotes the number of *bonds* between adjacent supercells, and singularity of J implies that there are

[7]This is the opposite of the traditional way of doing things, where sites inside a supercell are effectively treated as *orbitals*.

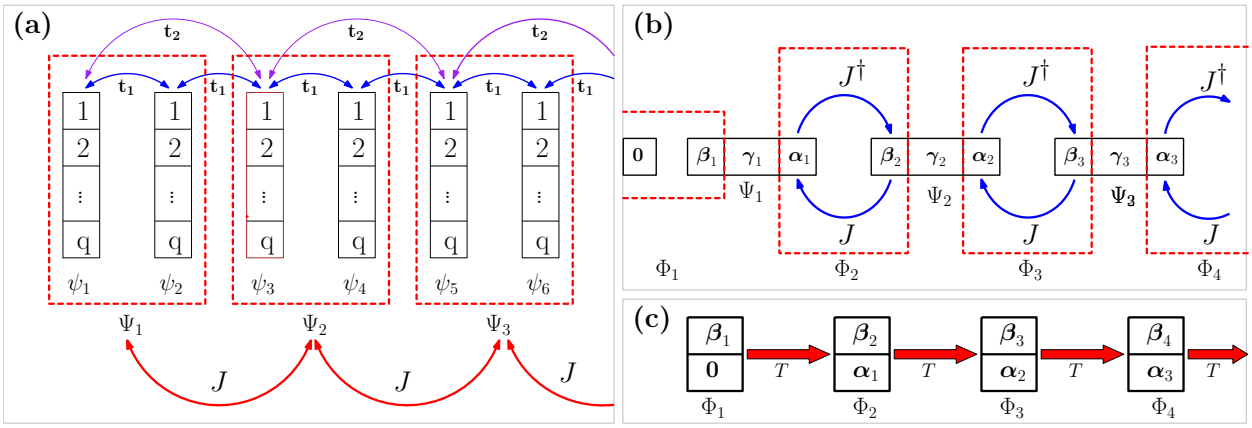


Figure 6.1: (a) A schematic depiction of the recursion relation, with q internal degrees of freedom, range of interaction $R = 2$ and Dirichlet boundary conditions at the left edge. We can form blocks (supercells) of such *sites* with 2 sites each, so the there is only nearest neighbor hopping between them. (b) A simplified depiction of the reduced recursion relation, with $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ corresponding to the coefficients of V, W and X subspaces(introduced in Sec 6.3), respectively. (c) We club together $\boldsymbol{\beta}_n$ with $\boldsymbol{\alpha}_{n-1}$ to obtain Φ_n, which is translated by one step using the transfer matrix.

sites in the supercell from which one cannot hop directly to a site in another supercell. We seek to compute a transfer matrix for singular J, where we, in some sense, *mod out* the redundant degrees of freedom, thereby inverting J on a reduced subspace to get a reduced transfer matrix. We will see that the corresponding transfer matrix[8] will be $2r \times 2r$.

Next, we note that $\mathcal{G} = (\varepsilon \mathbb{1} - M)^{-1}$ is the resolvent (or the Green's function) of a single supercell. Clearly, the matrix $\varepsilon \mathbb{1} - M$ is singular when ε is an eigenvalue of M. Physically, consider a system with uncoupled supercells, each with $\mathscr{N}$-degrees-of-freedom, obtaining by setting $J = 0$ in eq. (6.46). The corresponding spectrum consists of $\mathscr{N}$ degenerate levels, which broaden into bands as we turn J on. The eigenvalues of M can be interpreted as the centers of these bands. Since we are primarily concerned with the band gaps and the edge states therein, we can take $\varepsilon \mathbb{1} - M$ to be nonsingular as far as we do not venture deep inside the bulk bands[9].

Besides the finite-range hoppings, we shall make one more assumption, viz

$$J^2 = 0, \quad \text{so that } r < \frac{\mathscr{N}}{2}.$$

This might seem quite stringent, but recall that we are at liberty to choose our supercells, so that this condition can always be satisfied by choosing a large enough supercell. Physically, for $\mathscr{N} > 2$, this simply

[8] Indeed, if J had full rank, then we could have inverted it to get a $2\mathscr{N} \times 2\mathscr{N}$ transfer matrix, as computed in Sec 6.2.1.

[9] This breaks down if the bandwidth turns out to be zero, i.e, when the band pinches to a point(for instance, in case of graphene, or Chern insulator with $m = 1$). However, we can get around that using the well known trick of adding a small imaginary part to the Green's function in order to move the poles off the real line.

means that in a given supercell, the nodes in a supercell that are connected to the right neighboring supercell and the left neighboring supercell are not directly connected to each other.

6.3.2 Construction

We start off with the recursion relation

$$J\Psi_{n+1} + J^\dagger \Psi_{n-1} = (\varepsilon \mathbb{1} - M)\Psi_n, \tag{6.46}$$

where rank $(J) = r$, and perform a reduced singular value decomposition[96] (SVD) of J,

$$J = V \cdot \Xi \cdot W^\dagger, \tag{6.47}$$

where the matrices satisfy

$$V^\dagger \cdot V = W^\dagger \cdot W = \mathbb{1}, \quad V^\dagger \cdot W = 0. \tag{6.48}$$

The first two expressions follow from the definition of SVD, while $J^2 = 0$ implies the third. Hence, $J^2 = 0$ is required to ensure that the V and W subspaces are orthogonal and the corresponding coefficients can be extracted by taking suitable inner products. The SVD can equivalently be written as

$$J = \sum_{i=1}^{r} \xi_i \mathbf{v}_i \otimes \mathbf{w}_i \tag{6.49}$$

with

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \langle \mathbf{w}_i, \mathbf{w}_j \rangle = \delta_{ij}, \quad \langle \mathbf{v}_i, \mathbf{w}_j \rangle = 0, \tag{6.50}$$

where

$$V = (\mathbf{v}_1, \dots \mathbf{v}_r)_{\mathscr{N}\times r}, \quad W = (\mathbf{w}_1, \dots \mathbf{w}_r)_{\mathscr{N}\times r}, \quad \Xi = \mathrm{diag}\{\xi_1, \dots, \xi_r\}_{r\times r}. \tag{6.51}$$

We shall hereby refer to these vector pairs $(\mathbf{v}_i, \mathbf{w}_i)$ as *channels*. As we can still change the phases of $\mathbf{v}$ and $\mathbf{w}$ without violating the orthonormality, we choose their phases such that all the singular values (ξ_i) are positive. Thus, $\Xi^\dagger = \Xi$.

Morally speaking, the only directions in $\mathbb{C}^{\mathscr{N}}$ relevant for the problem are $\mathbf{v}_i$'s and $\mathbf{w}_i$'s, i.e, span$\{V\}$ and span$\{W\}$. Take a basis of $\mathbb{C}^{\mathscr{N}}$ as $\{\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_j\}$, where $i = 1, \dots r,\ j = 1, \dots \mathscr{N} - 2r$, and expand Ψ_n as

$$\Psi_n = \sum_{i=1}^{r} (\alpha_{n,i}\mathbf{v}_i + \beta_{n,i}\mathbf{w}_i) + \sum_{j=1}^{\mathscr{N}-2} \gamma_{n,j}\mathbf{x}_j, \tag{6.52}$$

with $\alpha_{n,i}, \beta_{n,i}, \gamma_{n,i} \in \mathbb{C}$, or, equivalently,

$$\Psi_n = V\boldsymbol{\alpha}_n + W\boldsymbol{\beta}_n + X\boldsymbol{\gamma}_n, \tag{6.53}$$

with $\boldsymbol{\alpha}_n, \boldsymbol{\beta}_n \in \mathbb{C}^r,\ \boldsymbol{\gamma}_n \in \mathbb{C}^{\mathcal{N}-2r}$. We have defined X analogous to V and W, so that

$$V^\dagger \cdot X = W^\dagger \cdot X = 0, \quad X^\dagger \cdot X = \mathbb{1}. \tag{6.54}$$

Also, $\boldsymbol{\alpha}_n = (\alpha_{n,1}, \alpha_{n,2}, \ldots \alpha_{n,r})$, with $\boldsymbol{\beta}_n$ and $\boldsymbol{\gamma}_n$ defined in a similar fashion.

We start by rewriting the recursion relation in eq. (6.46), in terms of the Green's function $\mathcal{G} = (\varepsilon\mathbb{1} - M)^{-1}$, as

$$\Psi_n = \mathcal{G} \cdot J\ \Psi_{n+1} + \mathcal{G} \cdot J^\dagger\ \Psi_{n-1}. \tag{6.55}$$

But

$$J\Psi_n = V \cdot \Xi\ \boldsymbol{\beta}_n, \quad J^\dagger\Psi_n = W \cdot \Xi\ \boldsymbol{\alpha}_n, \tag{6.56}$$

which follows from the SVD, eq. (6.48) and eq. (6.54). We can now premultiply eq. (6.55) by $V^\dagger$, $W^\dagger$ and $X^\dagger$ to extract the coefficients $\boldsymbol{\alpha}_n$, $\boldsymbol{\beta}_n$ and $\boldsymbol{\gamma}_n$, respectively. In order to simplify notation, we denote the restriction of $\mathcal{G}$ to V and W subspaces by $\mathcal{G}_{vv} = V^\dagger \cdot \mathcal{G} \cdot V$, $\mathcal{G}_{vw} = W^\dagger \cdot \mathcal{G} \cdot V$, etc[10]. Thus, the recursion relation reduces to

$$\begin{aligned}
\boldsymbol{\alpha}_n &= \mathcal{G}_{vv} \cdot \Xi\ \boldsymbol{\beta}_{n+1} + \mathcal{G}_{wv} \cdot \Xi\ \boldsymbol{\alpha}_{n-1}, \\
\boldsymbol{\beta}_n &= \mathcal{G}_{vw} \cdot \Xi\ \boldsymbol{\beta}_{n+1} + \mathcal{G}_{ww} \cdot \Xi\ \boldsymbol{\alpha}_{n-1}, \\
\boldsymbol{\gamma}_n &= \mathcal{G}_{vx} \cdot \Xi\ \boldsymbol{\beta}_{n+1} + \mathcal{G}_{wx} \cdot \Xi\ \boldsymbol{\alpha}_{n-1},
\end{aligned} \tag{6.57}$$

where the $\mathcal{G}_{ab},\ a, b \in \{v, w\}$ is a $r \times r$ matrix.

As $\mathcal{G}_{ab}$ are simply restrictions of the Green's functions, they are propagators connecting the a and b degrees of freedom for each supercell, while Ξ can be interpreted as encoding the relative strength of each channel, or the corresponding *tunneling probabilities*. The recursion equation in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (eq. (6.57)) has a simple diagrammatic interpretation as superpositions of possible nearest neighbor hopping processes, as illustrated in Fig. 6.2.

[10]Note that the order of V and W in subscript is opposite to that in the expression.

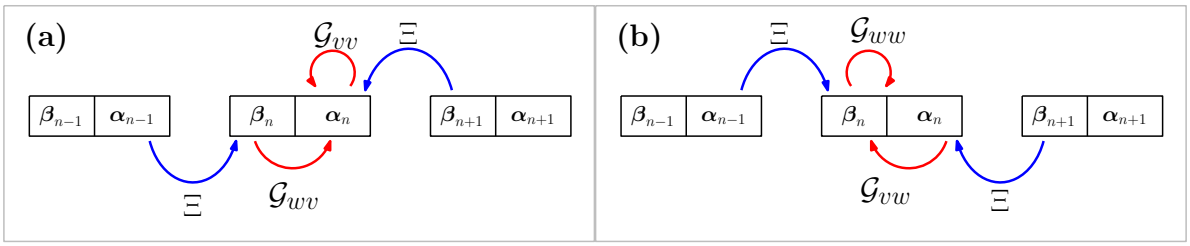


Figure 6.2: Diagrammatic representation of the recursion relations (eq. (6.57)) for (a) $\boldsymbol{\alpha}_n$ and (b) $\boldsymbol{\beta}_n$. Notice that the Green's functions $\mathcal{G}_{ab}$ express the propagation within a block, while Ξ expresses the tunneling probabilities between blocks.

To construct the transfer matrix, the first two equations of eq. (6.57) can be rewritten as

$$\begin{pmatrix} \mathcal{G}_{vv}\,\Xi & -\mathbb{1} \\ \mathcal{G}_{vw}\,\Xi & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{n+1} \\ \boldsymbol{\alpha}_n \end{pmatrix} = -\begin{pmatrix} 0 & \mathcal{G}_{wv}\,\Xi \\ -\mathbb{1} & \mathcal{G}_{ww}\,\Xi \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_n \\ \boldsymbol{\alpha}_{n-1} \end{pmatrix}, \tag{6.58}$$

which reduces to

$$\Phi_{n+1} = T\Phi_n, \quad \Phi_n \equiv \begin{pmatrix} \boldsymbol{\beta}_n \\ \boldsymbol{\alpha}_{n-1} \end{pmatrix}, \tag{6.59}$$

where

$$T = -\begin{pmatrix} \mathcal{G}_{vv}\,\Xi & -\mathbb{1} \\ \mathcal{G}_{vw}\,\Xi & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 & \mathcal{G}_{wv}\,\Xi \\ -\mathbb{1} & \mathcal{G}_{ww}\,\Xi \end{pmatrix}. \tag{6.60}$$

To proceed further, we need to invert the 2×2 block matrix. Using the block matrix identities from Appendix B.1, we proceed as

$$\begin{aligned} \begin{pmatrix} \mathcal{G}_{vv}\,\Xi & -\mathbb{1} \\ \mathcal{G}_{vw}\,\Xi & 0 \end{pmatrix}^{-1} &= \left[\begin{pmatrix} \mathcal{G}_{vv}\,\Xi & 0 \\ \mathcal{G}_{vw}\,\Xi & \mathbb{1} \end{pmatrix} \begin{pmatrix} \mathbb{1} & -\Xi^{-1}\,\mathcal{G}_{vv}^{-1} \\ 0 & \mathcal{G}_{vw}\,\mathcal{G}_{vv}^{-1} \end{pmatrix} \right]^{-1} \\ &= \begin{pmatrix} \mathbb{1} & -\Xi^{-1}\,\mathcal{G}_{vv}^{-1} \\ 0 & \mathcal{G}_{vw}\,\mathcal{G}_{vv}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{G}_{vv}\,\Xi & 0 \\ \mathcal{G}_{vw}\,\Xi & \mathbb{1} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \mathbb{1} & \Xi^{-1}\,\mathcal{G}_{vw}^{-1} \\ 0 & \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1} \end{pmatrix} \begin{pmatrix} \Xi^{-1}\,\mathcal{G}_{vv}^{-1} & 0 \\ -\mathcal{G}_{vw}\,\mathcal{G}_{vv}^{-1} & \mathbb{1} \end{pmatrix} \\ &= \begin{pmatrix} 0 & \Xi^{-1}\,\mathcal{G}_{vw}^{-1} \\ -\mathbb{1} & \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1} \end{pmatrix} \end{aligned} \tag{6.61}$$

Thus,

$$T = -\begin{pmatrix} 0 & \Xi^{-1}\,\mathcal{G}_{vw}^{-1} \\ -\mathbb{1} & \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1} \end{pmatrix} \begin{pmatrix} 0 & \mathcal{G}_{wv}\,\Xi \\ -\mathbb{1} & \mathcal{G}_{ww}\,\Xi \end{pmatrix} = \begin{pmatrix} \Xi^{-1}\,\mathcal{G}_{vw}^{-1} & -\Xi^{-1}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\,\Xi \\ \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1} & \left(\mathcal{G}_{wv} - \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\right)\Xi \end{pmatrix}. \tag{6.62}$$

Hence, we have managed to construct a closed form expression for a $2r\times2r$ transfer matrix $T(\varepsilon, \mathbf{k}_\perp)$ explicitly for the given recursion relation. This is one of our central results.

Defining $\mathscr{G}_{ab} = \mathcal{G}_{ab}\,\Xi$, we can also express this result as

$$T = \begin{pmatrix} \mathscr{G}_{vw}^{-1} & \mathscr{G}_{vw}^{-1}\,\mathscr{G}_{ww} \\ \mathscr{G}_{vv}\,\mathscr{G}_{vw}^{-1} & \mathscr{G}_{wv} - \mathscr{G}_{vv}\,\mathscr{G}_{vw}^{-1}\,\mathscr{G}_{ww} \end{pmatrix}. \tag{6.63}$$

This expression is somewhat cleaner, but it obscures the different physical significance associated with $\mathcal{G}$ and Ξ as depicted in Fig 6.2, as well as properties of $\mathcal{G}$, which we now state. As the Green's function is Hermitian, i.e, $\mathcal{G}^\dagger = \mathcal{G}$, we have

$$\mathcal{G}_{vv}^\dagger = \mathcal{G}_{vv}, \quad \mathcal{G}_{ww}^\dagger = \mathcal{G}_{ww}, \quad \mathcal{G}_{vw}^\dagger = \mathcal{G}_{wv}. \tag{6.64}$$

Using these properties and eq. (B.4), we can compute

$$\begin{aligned} \det(T) =&\ \det\left[\Xi^{-1}\,\mathcal{G}_{vw}^{-1}\right] \det\left[\left(\mathcal{G}_{wv} - \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\right)\,\Xi - \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{vw}\,\Xi\,\left(-\Xi^{-1}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\,\Xi\right)\right] \\ =&\ \det\left(\Xi^{-1}\right) \det\left(\mathcal{G}_{vw}^{-1}\right) \det\left(\mathcal{G}_{wv} - \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww} + \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\right) \det\left(\Xi\right) \\ =&\ \det\left(\mathcal{G}_{vw}^{-1}\right) \det\left(\mathcal{G}_{wv}\right) = \frac{\left(\det \mathcal{G}_{vw}\right)^*}{\det \mathcal{G}_{vw}} = e^{-2i\theta}, \end{aligned} \tag{6.65}$$

where $\theta = \arg\left(\det \mathcal{G}_{vw}\right)$. We can gauge this phase away by redefining

$$\Phi_n \to e^{in\theta/r}\Phi_n, \quad T \to e^{i\theta/r}T. \tag{6.66}$$

In the following, whenever we refer to the transfer matrix, we shall assume that we have gauged away the phase of the determinant of T so that $\det T = 1$. Thus, our final expression for the transfer matrix would be

$$T = e^{\frac{i}{r}\arg(\det \mathcal{G}_{vw})} \begin{pmatrix} \Xi^{-1}\,\mathcal{G}_{vw}^{-1} & -\Xi^{-1}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\,\Xi \\ \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1} & \left(\mathcal{G}_{wv} - \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\right)\,\Xi \end{pmatrix}, \tag{6.67}$$

Thus, we have constructed a transfer matrix in a basis independent fashion, as we have never referred to the explicit form of the J and M matrices. It reduces the computation of transfer matrix for a system to the identification of the J and M matrices, as everything else can be mechanized. We shall illustrate that with a plethora of examples in Chapter 7.

6.3.3 Properties

In the theory of discrete flows, there is a special place for symplectic monodromy (transfer) matrices, as they correspond to discrete versions of Hamiltonian flows. Recall that a matrix T is $\mathcal{J}$-unitary or complex-

symplectic ($T \in Sp(2r,\mathbb{C})$) if it satisfies $T^\dagger \mathcal{J} T = \mathcal{J}$, where $\mathcal{J}$ is a nonsingular antisymmetric matrix. Furthermore, it is symplectic ($T \in Sp(2r,\mathbb{R})$) if it is also real.

We seek the corresponding condition for our construction. For the conventional choice of $\mathcal{J}$, using the properties of $\mathcal{G}_{ab},\ a,b \in \{v,w\}$ and setting

$$\mathcal{S} = \mathcal{G}_{wv} - \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww} \implies \mathcal{S}^\dagger = \mathcal{G}_{wv}^\dagger - \mathcal{G}_{ww}^\dagger\,\left(\mathcal{G}_{vw}^\dagger\right)^{-1}\,\mathcal{G}_{vv}^\dagger = \mathcal{G}_{vw} - \mathcal{G}_{ww}\,\mathcal{G}_{wv}^{-1}\,\mathcal{G}_{vv},$$

we begin by computing

$$\begin{aligned}
T^\dagger \mathcal{J} T &= \begin{pmatrix} \left(\Xi^\dagger\,\mathcal{G}_{vw}^\dagger\right)^{-1} & \left(\mathcal{G}_{vw}^\dagger\right)^{-1}\,\mathcal{G}_{vv}^\dagger \\ -\Xi^\dagger\,\mathcal{G}_{ww}^\dagger\,\left(\mathcal{G}_{vw}^\dagger\right)^{-1} & \left(\Xi^\dagger\right)^{-1} \end{pmatrix} \begin{pmatrix} \left(\Xi^\dagger\,\mathcal{G}_{vw}^\dagger\right)^{-1} & \left(\mathcal{G}_{vw}^\dagger\right)^{-1}\,\mathcal{G}_{vv}^\dagger \\ -\Xi^\dagger\,\mathcal{G}_{ww}^\dagger\,\left(\mathcal{G}_{vw}^\dagger\right)^{-1} & \left(\Xi^\dagger\right)^{-1} \end{pmatrix} \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \begin{pmatrix} \Xi^{-1}\,\mathcal{G}_{vw}^{-1} & -\Xi^{-1}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\,\Xi \\ \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1} & \mathcal{S}\,\Xi \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{G}_{wv}^{-1}\,\Xi^{-1} & \mathcal{G}_{wv}^{-1}\,\mathcal{G}_{vv} \\ -\Xi\,\mathcal{G}_{ww}\,\mathcal{G}_{wv}^{-1}\,\Xi^{-1} & \Xi\,\mathcal{S}^\dagger \end{pmatrix} \begin{pmatrix} \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1} & \mathcal{S}\,\Xi \\ -\Xi^{-1}\,\mathcal{G}_{vw}^{-1} & \Xi^{-1}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\,\Xi \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{G}_{wv}^{-1}\,\left(\Xi^{-1}\,\mathcal{G}_{vv} - \mathcal{G}_{vv}\,\Xi^{-1}\right)\,\mathcal{G}_{vw}^{-1} & \mathcal{G}_{wv}^{-1}\,\left(\Xi^{-1}\,\mathcal{S} + \mathcal{G}_{vv}\Xi^{-1}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}\right)\Xi \\ -\Xi\left(\mathcal{G}_{ww}\,\mathcal{G}_{wv}^{-1}\,\Xi^{-1}\,\mathcal{G}_{vv} + \mathcal{S}^\dagger\,\Xi^{-1}\right)\mathcal{G}_{vw}^{-1} & \Xi\left(-\mathcal{G}_{ww}\,\mathcal{G}_{wv}^{-1}\,\Xi^{-1}\,\mathcal{S} + \mathcal{S}^\dagger\Xi^{-1}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww}s\right)\Xi \end{pmatrix} \\
&= \mathcal{J} + \begin{pmatrix} \mathcal{G}_{wv}^{-1}[\Xi^{-1},\mathcal{G}_{vv}]\mathcal{G}_{vw}^{-1} & \mathcal{G}_{wv}^{-1}[\Xi^{-1},\mathcal{G}_{wv}]\Xi - \mathcal{G}_{wv}^{-1}[\Xi^{-1},\mathcal{G}_{vv}]\mathcal{G}_{vw}^{-1}\mathcal{G}_{ww}\Xi \\ \Xi[\Xi^{-1},\mathcal{G}_{vw}]\mathcal{G}_{vw}^{-1} - \Xi\,\mathcal{G}_{ww}\mathcal{G}_{wv}^{-1}[\Xi^{-1},\mathcal{G}_{vv}]\mathcal{G}_{vw}^{-1} & \Xi\left\{[\Xi^{-1},\mathcal{G}_{ww}] - [\Xi^{-1},\mathcal{G}_{vw}]\mathcal{G}_{vw}^{-1}\mathcal{G}_{ww} - \mathcal{G}_{ww}\mathcal{G}_{wv}^{-1}[\Xi^{-1},\mathcal{G}_{wv}] + \mathcal{G}_{ww}\mathcal{G}_{wv}^{-1}[\Xi^{-1},\mathcal{G}_{vv}]\mathcal{G}_{vw}^{-1}\mathcal{G}_{ww}\right\}\Xi \end{pmatrix}.
\end{aligned}$$

To satisfy $T^\dagger\,\mathcal{J}\,T = \mathcal{J}$, we demand that all the commutators in this expression vanish. Thus,

$$[\mathcal{G}_{ab},\Xi] = 0\ \forall\, a,b \in \{v,w\} \implies T \in Sp(2r,\mathbb{C}). \tag{6.68}$$

Physically, this condition implies that the various channels that connect the nearest neighbor supercells are independent, so that the order of tunneling (Ξ) and propagation ($\mathcal{G}$) is irrelevant. The spectral properties of $\mathcal{J}$-unitary operators has been studied in great detail in the mathematics literature[99].

Furthermore, in the discussion on bulk bands, we show that if the transfer matrix is symplectic, it can effectively be decomposed into a set of chains, one corresponding to each channel. The conditions on $\mathcal{G}_{ab}$ obtained above can be thought of as physical manifestations of that fact. As Ξ is, by definition, a diagonal matrix, in order for it to commute with another matrix A, A, in general, must also be diagonal[11]. Hence, for T to be symplectic, $\mathcal{G}_{vv}$ and $\mathcal{G}_{vw}$ must also be diagonal.

Finally, we show that the size and spectral properties of our transfer matrix are independent of a choice of supercell. More precisely, we shall consider supercells which are $m \in \mathbb{Z}^+$ times a *minimal supercell*, where

[11] This breaks down if two or more diagonal entries of Ξ are equal, as Ξ then becomes proportional to identity in that subspace, so that the restriction of A to that subspace can be arbitrary.

we define the *minimal supercell* as one containing the minimum number of sites so that the hopping between the supercells is nearest neighbor and the corresponding hopping matrix is nilpotent. Consider m copies of the recursion relation of eq. (6.42) for $n = nm, nm-1, \ldots nm-m+1$, and define

$$\widetilde{\Psi}_n = (\Psi_{mn}, \Psi_{mn-1}, \ldots \Psi_{mn-m+1})^{\mathrm{T}}. \tag{6.69}$$

Then, $\widetilde{\Psi}_n$ follows the recursion relation

$$\widetilde{J}\widetilde{\Psi}_{n+1} + \widetilde{J}^{\dagger}\widetilde{\Psi}_{n-1} = \left(\varepsilon \mathbb{1} - \widetilde{M}\right)\widetilde{\Psi}_n, \tag{6.70}$$

where $\widetilde{J}$ and $\widetilde{M}$ can be written in terms of J and M as

$$\widetilde{J} = \begin{pmatrix} 0 & 0 & \ldots & J \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}, \quad \widetilde{M} = \begin{pmatrix} M & J^{\dagger} & \ldots & 0 \\ J & M & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & M \end{pmatrix}. \tag{6.71}$$

Clearly, $\mathrm{rank}\left(\widetilde{J}\right) = \mathrm{rank}\,(J) = r$, and the reduced SVD of $\widetilde{J}$ is given by

$$\widetilde{J} = \widetilde{V} \cdot \Xi \cdot \widetilde{W}^{\dagger}; \qquad \widetilde{V} = \begin{pmatrix} V \\ \vdots \\ 0 \end{pmatrix}_{\mathscr{N} m \times r}, \quad \widetilde{W} = \begin{pmatrix} 0 \\ \vdots \\ W \end{pmatrix}_{\mathscr{N} m \times r}, \tag{6.72}$$

where the singular values of $\widetilde{J}$ are same as those of J. Following the calculation in Sec 6.3, we compute the recursion relations for $\widetilde{\boldsymbol{\alpha}}_n$ and $\widetilde{\boldsymbol{\beta}}_n$, the coefficients of $\widetilde{\Psi}_n$ along $\widetilde{V}$ and $\widetilde{W}$, and construct a $2r \times 2r$ transfer matrix $\widetilde{T}$, so that

$$\widetilde{\Phi}_{n+1} = \widetilde{T}\widetilde{\Phi}_n, \quad \widetilde{\Phi}_n \equiv \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_n \\ \widetilde{\boldsymbol{\alpha}}_{n-1} \end{pmatrix}. \tag{6.73}$$

But using the definition of $\widetilde{\Psi}_n$, we get

$$\begin{aligned} \widetilde{\boldsymbol{\alpha}}_n &= \widetilde{V}^{\dagger}\widetilde{\Psi}_n = V^{\dagger}\Psi_{nm} = \boldsymbol{\alpha}_{nm}, \\ \widetilde{\boldsymbol{\beta}}_n &= \widetilde{W}^{\dagger}\widetilde{\Psi}_n = W^{\dagger}\Psi_{nm-m+1} = \boldsymbol{\beta}_{(n-1)m+1}, \end{aligned} \tag{6.74}$$

so that

$$\widetilde{\Phi}_n = \begin{pmatrix} \boldsymbol{\beta}_{(n-1)m+1} \\ \boldsymbol{\alpha}_{(n-1)m} \end{pmatrix} = \Phi_{(n-1)m+1}. \tag{6.75}$$

Using the old transfer matrix, T, we also have

$$\widetilde{\Phi}_{n+1} = \Phi_{nm+1} = T^m \Phi_{(n-1)m+1} = T^m \widetilde{\Phi}_n, \tag{6.76}$$

so that the action of $\widetilde{T}$ is identical to the action of T^m on all wavefunction. We conclude that $\widetilde{T} = T^m$.

Thus, we have shown that if we take a supercell $m \in \mathbb{Z}^+$ times the minimal supercell, the effective transfer matrix is simply $T \to T^m$, where $\dim(T) = 2\,\mathrm{rank}(J)$ stays invariant under this operation. As we are only concerned with the behavior of T^n for large n(See Sec 6.4), the band structure, as expected, stays invariant under such a transformation. Hence, we can always make the supercell bigger than the minimal supercell, while leaving the bands and edge states invariant.

6.4 Using the transfer matrix

Given a tight-binding model in $d \geq 1$ dimensions, we have represented the Schrödinger equation as a recursion relation, and hence a transfer matrix equation with $T(\varepsilon, \mathbf{k}_\perp)$, with $\varepsilon \in \mathbb{R}$ and $\mathbf{k}_\perp \in \mathbb{T}^{d-1}$ for $d > 1$. We now apply the ideas of Sec 6.1 to a family of transfer matrices, parametrized by $(\varepsilon, \mathbf{k}_\perp)$, to formally study their bulk and edge spectra. We shall illustrate them with explicit computations in Ch 7.

6.4.1 Bulk bands

The "bulk" states of a tight-binding model are the states that stay finite, and hence normalizable to a delta function, in the infinite system size limit. In a transfer matrix formalism, with $\Phi_n = T^n \Phi_0$, a solution Φ_n corresponds to a "bulk band" iff $|\Phi_n|$ stays finite as $n \to \pm\infty$. Specifically, if $\varphi = \Phi_0$ is an eigenvector of the transfer matrix with eigenvalue $\rho \in \mathbb{C}$, then $|\Phi_n| = |\rho|^n |\varphi|$, so that Φ_n is a bulk state iff $|\rho| = 1$, i.e, the ρ lies on the unit circle $S^1 \equiv \{z \in \mathbb{C} \mid |z| = 1\}$ in the complex plane. Hence, a given $(\varepsilon, \mathbf{k}_\perp) \in \mathbb{R} \times \mathbb{T}^{d-1}$ lies in the *bulk band* if and only if all eigenvalues of $T(\varepsilon, \mathbf{k}_\perp)$ lie on the unit circle. At the other extreme, a given $(\varepsilon, \mathbf{k}_\perp)$ lies in the *bulk gap* if and only if all eigenvalues of $T(\varepsilon, \mathbf{k}_\perp)$ lie off the unit circle.

More formally, we define the *bulk band*, $\mathscr{B} \subset \mathbb{R} \times \mathbb{T}^{d-1}$, as

$$\mathscr{B} = \left\{(\varepsilon, \mathbf{k}_\perp) \,|\, \sigma\left[T(\varepsilon, \mathbf{k}_\perp)\right] \subset S^1\right\}, \tag{6.77}$$

and the *bulk gap* as

$$\mathscr{G} = \left\{(\varepsilon, \mathbf{k}_\perp) \,|\, \sigma\left[T(\varepsilon, \mathbf{k}_\perp)\right] \subset \mathbb{C}\backslash S^1\right\}. \tag{6.78}$$

For $r > 1$, the possibility exists that there can be points $(\varepsilon, \mathbf{k}_\perp)$ for which some eigenvalues are on and some off the unit circle. We shall term such points *partial gaps*, $\mathscr{P}$, defined as

$$\mathscr{P} = \left(\mathbb{R} \times \mathbb{T}^{d-1}\right) \setminus (\mathscr{G} \cup \mathscr{B}) \,. \tag{6.79}$$

By construction, $\mathscr{B} \cup \mathscr{G} \cup \mathscr{P} = \mathbb{R} \times \mathbb{T}^{d-1}$, so that each $(\varepsilon, \mathbf{k}_\perp)$ falls in one of these sets.

To compute the bulk bands, one needs to compute the eigenvalues of the transfer matrix as a function of $(\varepsilon, \mathbf{k}_\perp)$, which can always be done numerically. However, if the transfer matrix is symplectic ($T \in \mathrm{Mat}(2r, \mathbb{R})$ and $T^{\mathrm{T}} \mathcal{J} T = \mathcal{J}$), its eigenvalues always occur in reciprocal pair. This is because given $T\varphi = \rho\varphi$, we also have

$$\mathcal{J}\varphi = T^{\mathrm{T}} \mathcal{J} T \varphi = T^{\mathrm{T}} \mathcal{J}(\rho\varphi) \implies T^{\mathrm{T}} (\mathcal{J}\varphi) = \rho^{-1} (\mathcal{J}\varphi) \,,$$

and since the spectra of T and T^{T} are identical, we conclude that if ρ is an eigenvalue of T, so is ρ^{-1}.

Let us start off with $r = 1$, where T being symplectic implies that the product of eigenvalues, $\det T = 1$. The eigenvalues can be solved for as

$$\Delta \equiv \mathrm{Tr}\, T = \rho + \rho^{-1} \implies \rho = \frac{1}{2}\left[\Delta \pm \sqrt{\Delta^2 - 4}\right]. \tag{6.80}$$

Hence, either both the eigenvalues lie on the real line ($\Delta^2 > 4$) or they form a complex conjugate pair($\Delta^2 < 4$) satisfying $\det T = \rho^* \rho = 1$, i.e, they lie on the unit circle. Consequently, a given $(\varepsilon, \mathbf{k}_\perp)$ either belongs to $\mathscr{G}$ or $\mathscr{B}$, so that $\mathscr{P} = \varnothing$.

For $r > 1$, consider the characteristic polynomial of T

$$P(\rho) = \det(\rho \mathbb{1} - A) = \sum_{n=0}^{2r} a_n \rho^n. \tag{6.81}$$

The eigenvalues of A are the zeros of $P(\rho)$. But $P(\rho) = 0 \implies P(\rho^{-1}) = 0$, so that

$$0 = P(\rho^{-1}) = \sum_{n=0}^{2r} a_n \rho^{-n} = \rho^{-2r} \left(\sum_{n=1}^{2r} a_{2r-n} \rho^n \right). \tag{6.82}$$

As $\rho \neq 0$, we conclude that $a_n = a_{2r-n}$, i.e, the characteristic polynomial is *palindromic*[59, 100], which can be used to compute the eigenvalues. To wit, rewrite the eigenvalue condition as

$$0 = P(\rho) = \rho^r \left(a_r + \sum_{n=1}^{r} a_{r-n} \left(\rho^n + \rho^{-n}\right) \right). \tag{6.83}$$

Defining $\Delta = \rho + \rho^{-1}$, we can express $\rho^n + \rho^{-n}$ as Chebyshev polynomials of the first kind in Δ. Explicitly, for $\rho = e^{i\theta}$,

$$\rho^n + \rho^{-n} = 2\cos(n\theta) = 2\,U_n(\cos\theta) = 2\,U_n\left(\frac{\Delta}{2}\right). \tag{6.84}$$

The first few examples are

$$\rho^2 + \rho^{-2} = \Delta^2 - 2, \quad \rho^3 + \rho^{-3} = \Delta^3 - 3\Delta, \quad \rho^4 + \rho^{-4} = \Delta^4 - 4\Delta^2 + 2, \text{ etc.}$$

Thus, the characteristic polynomial for T can be written as a polynomial of order r in Δ, as

$$a_r + 2\sum_{n=1}^{r} a_{r-n} U_n\left(\frac{\Delta}{2}\right) = 0 \tag{6.85}$$

The roots $\Delta_1, \ldots \Delta_r \in \mathbb{C}$ are *generalized Floquet discriminants* of T, which can be expressed in terms of traces of powers of T. We can now solve for $\rho \in \sigma\,[T]$ as

$$\rho + \rho^{-1} = \Delta_n \implies \rho = \frac{1}{2}\left[\Delta_n \pm \sqrt{\Delta_n^2 - 4}\right], \tag{6.86}$$

where $n = 1, 2 \ldots r$. Hence, if the transfer matrix is symplectic, we can essentially decompose it into a set of independent $r = 1$ systems!

In the following, we work out the case of $r = 2$ explicitly. The eigenvalue condition becomes

$$a_0(\rho^2 + \rho^{-2}) + a_1(\rho + \rho^{-1}) + a_2 = 0,$$

with

$$a_0 = 1, \quad a_1 = -\mathrm{Tr}A, \quad a_2 = \frac{1}{2}\left((\mathrm{Tr}A)^2 - \mathrm{Tr}A^2\right).$$

In terms of Δ, we get

$$a_0(\Delta^2 - 2) + a_1\Delta + a_2 = 0,$$

which implies that

$$\Delta_\pm = \frac{1}{2a_0}\left[-a_1 \pm \sqrt{a_1^2 - 4a_0(a_2 - 2a_0)}\right].$$

Substituting a_n's, we get the Floquet discriminants as

$$\Delta_\pm = \frac{1}{2}\left[\mathrm{Tr}A \pm \sqrt{2\mathrm{Tr}A^2 - (\mathrm{Tr}A)^2 + 8}\right]. \tag{6.87}$$

For the cases where the transfer matrix is symplectic, we can compute the band edges directly, without having to diagonalize T for all possible $(\varepsilon, \mathbf{k}_\perp)$, which one would need to do in general. This follows from the fact that since the eigenvalues appear in reciprocal pairs, they either appear on the unit circle or on the real line (See eqns (6.80) and (6.86)). Thus, as we move around in the $(\varepsilon, \mathbf{k}_\perp)$ space, the only way to go from a bulk band to a gap is to have a pair of eigenvalues collide at ± 1 and go off the unit circle onto the real line. The primary advantage of computing the Floquet discriminants is that it provides us with a convenient condition for this eigenvalue collision: it happens when $|\Delta_i| = 2$ for some i.

In conclusion, if the transfer matrix is symplectic, then for a rank r problem, the behavior of a state simply depends on the r Floquet discriminants $\Delta_i(\varepsilon, \mathbf{k}_\perp)$'s. we have an oscillating (normalizable) state for $|\Delta_i| \leq 2$ and a growing/decaying state for $|\Delta_i| > 2$. We can alternatively define the bulk band and the band-gap as

$$\begin{aligned} \mathscr{G} &= \{(\varepsilon, \mathbf{k}_\perp) \mid |\Delta_i(\varepsilon, \mathbf{k}_\perp)| > 2 \; \forall \; i = 1, \dots r\}, \\ \mathscr{B} &= \{(\varepsilon, \mathbf{k}_\perp) \mid |\Delta_i(\varepsilon, \mathbf{k}_\perp)| \leq 2 \; \forall \; i = 1, \dots r\}, \end{aligned} \tag{6.88}$$

with the band edges given by the conditions $|\Delta_i| = 2, \; i = 1, 2, \dots r$. In practice, we can simply solve this conditions for $\varepsilon(\mathbf{k}_\perp)$, numerically if needed, to compute the band edges, without having to diagonalize $T(\varepsilon, \mathbf{k}_\perp)$ for all values of $(\varepsilon, \mathbf{k}_\perp)$.

6.4.2 Edge states

For an infinite system, the only physically sensible states (i.e, normalizable to a delta function) are the bulk states. However, for a system with an edge[12], for instance, a system defined on a half-plane, additional states may occur, most of whose weights are concentrated near the edge and which decay exponentially into the bulk. Thus, in order to study the edge/surface modes, we need to restrict our Hamiltonian operator to a half space, with a suitable (for instance, Dirichlet) boundary condition at the surface.

We "define" an edge state to be an eigenstate of the half-space Hamiltonian with the following two features:

1. The state must be normalizable, or equivalently, it must decay into the bulk away from the edge/surface.
2. The state must satisfy the given boundary conditions.

By definition, these states must reside outside the bulk band, i.e, only for $(\varepsilon, \mathbf{k}_\perp) \in \mathscr{G} \cup \mathscr{P}$. Typically, one is interested in the existence of these states, and, should they exist, in the *edge spectrum*, i.e, the energy of

[12] We use the term '*edge*' more generally to mean a boundary of the system.

the edge state, $\varepsilon_{edge}(\mathbf{k}_\perp)$, as a function of the transverse momentum $\mathbf{k}_\perp$. In the subsequent analysis, given a $\Phi \in \mathbb{C}^{2r}$, we are interested in deriving the conditions on it corresponding to these constraints

For the former (decay) condition, note that given a Hamiltonian for a quasi-1D chain, we have two inequivalent choices of half-spaces, viz, $n < 0$ and $n > 0$. Corresponding to these, we can define two kinds of edges states: a $\Phi \in \mathbb{C}^{2r}$ forms a *left* edge state if $|T^n\Phi| \to 0$ as $n \to \infty$, while it forms a *right* edge state if $|T^n\Phi| \to 0$ as $n \to -\infty$. These are related under $T \leftrightarrow T^{-1}$. In Sec 6.1, we defined $\mathcal{P}_<$, the projector to the decaying subspace, and showed that a necessary condition for a $T^n\Phi$ to decay as $n \to \infty$ is simply $\mathcal{P}_<\Phi = \Phi$, and the equivalent condition for $n \to -\infty$ is $\mathcal{P}_>\Phi = \Phi$. Thus, our *decay conditions* are simply

$$\text{Left edge: } \mathcal{P}_<\Phi = \Phi, \qquad \text{Right edge: } \mathcal{P}_>\Phi = \Phi. \tag{6.89}$$

We next consider the boundary condition. Note that the edge spectrum can get modified quite drastically by local terms at the boundary, an effect commonly known as *edge reconstruction*. In order to consider the most general case, we should take a Hamiltonian $\tilde{\mathcal{H}} = \mathcal{H} + \delta\mathcal{H}$, where $\delta\mathcal{H}$ is an operator localized at the edge, which can account for the edge reconstruction, for instance, due to an impurity[101] or lattice deformation[102]. Such a boundary condition imposes additional conditions[49, 103, 104] on the eigenvectors of the transfer matrix. Since in the present case, we only intend to expound the geometry and topology associated with the band structure which is independent of such local deformations, we shall restrict ourselves to open(Dirichlet) boundary conditions at sites $n = 0$ and N, with no boundary terms[13].

For OBC, at the left edge, we demands that[60] $\Psi_0 = (0, 0)$, leading to $\alpha_0 = 0$ in Φ_1. Similarly, at the right edge, we demand that $\Psi_{N+1} = (0, 0)^{\mathrm{T}}$, leading to $\beta_{N+1} = 0$ in Φ_N. Thus, in terms of Φ, the boundary condition for the left and right edges becomes

$$\Phi_1 = \begin{pmatrix} \boldsymbol{\beta}_1 \\ 0 \end{pmatrix}, \qquad \Phi_N = \begin{pmatrix} 0 \\ \boldsymbol{\alpha}_N \end{pmatrix}, \tag{6.90}$$

respectively. Note that $\boldsymbol{\beta}_1, \boldsymbol{\alpha}_1 \in \mathbb{C}^r$ are still undetermined on their respective edges. We shall use the decay conditions to fix these in the next section.

We can also define projectors to write the boundary condition in a way similar to the decay conditions. Define the $2r \times r$ matrices

$$\mathcal{Q}_{\boldsymbol{\alpha}} = \begin{pmatrix} 0_{r\times r} \\ \mathbb{1}_{r\times r} \end{pmatrix}, \quad \mathcal{Q}_{\boldsymbol{\beta}} = \begin{pmatrix} \mathbb{1}_{r\times r} \\ 0_{r\times r} \end{pmatrix}, \tag{6.91}$$

as the injectors into the $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ subspaces, respectively. In terms of these operators, the OBC on the left

[13] Most of the following is a restatement of the results by Lee and Joannopoulos[49] in our formalism

edge is equivalent to the statement that $\Phi \in \text{range}(\mathcal{Q}_{\boldsymbol{\beta}})$, while the right edge is equivalent to $\Phi \in \text{range}(\mathcal{Q}_{\boldsymbol{\alpha}})$. Finally, define the projectors

$$\mathcal{P}_R = \mathcal{Q}_{\boldsymbol{\alpha}} \mathcal{Q}_{\boldsymbol{\alpha}}^{\dagger}, \quad \mathcal{P}_L = \mathcal{Q}_{\boldsymbol{\beta}} \mathcal{Q}_{\boldsymbol{\beta}}^{\dagger}. \tag{6.92}$$

Our *boundary conditions* become

$$\text{Left edge: } \mathcal{P}_L \Phi = \Phi, \qquad \text{Right edge: } \mathcal{P}_R \Phi = \Phi. \tag{6.93}$$

Thus, we have obtained two sets of conditions, viz, the decay conditions and the boundary conditions, that we need to solve simultaneously in order to obtain the *physical* edge states, i.e, the edge states that would be observed in an exact diagonalization of the lattice models on finite size lattices. However, before we attempt to do so, we can ask a somewhat perverse question, which turns out to have important consequences: What if we chose the *wrong* decay condition for a given boundary? We tabulate the situation as follows:

	$\mathcal{P}_< \Phi = \Phi$	$\mathcal{P}_> \Phi = \Phi$
$\mathcal{P}_L \Phi = \Phi$	Left edge	*Unphysical*
$\mathcal{P}_R \Phi = \Phi$	*Unphysical*	Right edge

Table 6.1: Boundary(rows) vs decay(column) conditions.

The *wrong* choice of decay condition implies that the corresponding state grows (instead of decaying) exponentially in the bulk, and is hence not normalizable and unphysical. However, we shall see that in order to account for all the windings corresponding to the edge state, we shall need to take the unphysical states into account. Furthermore, these should not be thought of as a complete fantasy, since they can be revealed by changing the boundary condition, as we shall demonstrate explicitly in Sec 7.1.4

Taking the *correct* decay conditions, for the left edge state, we need to simultaneously solve

$$\mathcal{P}_< \Phi = \Phi = \mathcal{P}_L \Phi, \tag{6.94}$$

or, alternatively,

$$\mathcal{P}_< \Phi_1 = \Phi_1; \quad \Phi_1 = \begin{pmatrix} \boldsymbol{\beta}_1 \\ 0 \end{pmatrix}. \tag{6.95}$$

Note that as $\text{rank}\,(\mathcal{P}_<) \leq r$, this is a homogeneous linear system of up to r equations for the r variables, viz, the coefficients of $\boldsymbol{\beta}_1$. But for a nontrivial state, we demand that $\boldsymbol{\beta}_1 \neq 0$, from which we can obtain a Cramer's condition, which can be numerically solved to obtain the physical edge spectrum.

7 Applications

In this chapter, we discuss a few applications of our generalized transfer matrix formalism to the noninteracting tight-binding models commonly arising in condensed matter contexts. We shall restrict ourselves to the models corresponding to *topological phases of matter*, which are characterized by a topological invariant on the bulk Brillouin zone and often exhibit nontrivial surface states. We shall see that besides essentially mechanizing the process of computing the bulk and edge spectra, our construction also provides further insight into the topological nature of the edge states in terms of windings around the noncontractible loops of a complex energy Riemann surfaces.

For condensed matter systems, the lattice models are often specified simply by their Bloch Hamiltonians (unlike the position space Hamiltonians described in Sec 6.2.2). In order to construct the transfer matrix, we need to *inverse Fourier transform* the Hamiltonian in the direction along which the transfer matrix acts by translation. For instance, consider the simple case of a lattice model with only nearest neighbor hopping on a (hyper-c)cubic lattice in $\mathbb{R}^d$, where we seek to construct the transfer matrix which translates along one of the lattice vectors ($\hat{x}$, say). Using the periodicity of the Bloch Hamiltonian along x, we can express it as

$$\mathcal{H}_B(k_x, \mathbf{k}_\perp) = \sum_{\alpha,\beta=1}^{q} \left[\widetilde{J}(\mathbf{k}_\perp) e^{ik_x a_x} + \widetilde{M}(\mathbf{k}_\perp) + \widetilde{J}^\dagger(\mathbf{k}_\perp) e^{-ik_x a_x} \right], \tag{7.1}$$

where a_x is the lattice constant along x. Transforming to position space along x leads to

$$\mathcal{H}(\mathbf{k}_\perp) = \sum_n \sum_{\alpha,\beta=1}^{q} \left[c^\dagger_{n+1,\alpha} \widetilde{J}_{\alpha\beta}(\mathbf{k}_\perp) c_{n,\beta} + c^\dagger_{n-1,\alpha} \widetilde{J}^\dagger_{\alpha\beta}(\mathbf{k}_\perp) c_{n,\beta} + c^\dagger_{n,\alpha} \widetilde{M}_{\alpha\beta}(\mathbf{k}_\perp) c_{n,\beta} \right], \tag{7.2}$$

which renders the Hamiltonian of the form in eq. (6.40) with $R = 1$, and we identify $J = \widetilde{J}$ and $M = \widetilde{M}$. We discuss the more general case (for arbitrary lattices and/or directions of translation) in Sec 7.1.5.

Once we have computed the generalized transfer matrix, we shall use the results from Sec 6.4 to compute the bulk/edge spectra, and compare them to the results obtained from exact diagonalization on finite lattices. Furthermore, we shall expose the topological nature of the edge states in terms of windings on certain manifolds that we can construct based on the algebraic structure of the generalized transfer matrices.

7.1 The case of $r = 1$

We begin with the simplest case for our construction, *viz*, that of $r = 1$. Even though the transfer matrices are merely 2×2 in this case, we shall see that it encompasses quite a few lattice models commonly encountered. Furthermore, many simplifications which work for $r = 1$ make this case particularly analytically tractable, so that we shall be able to expose the workings of the somewhat formal computation from the last chapter with detailed calculations.

7.1.1 The generalized transfer matrix

The expression for the transfer matrix for $r = 1$ admits simplifications, owing to the fact that the individual blocks of eq. (6.62) are simply 1×1 matrices, i.e, complex numbers. The generalized transfer matrix becomes

$$T = \frac{1}{|\mathcal{G}_{vw}|} \begin{pmatrix} 1 & -\mathcal{G}_{ww} \\ \mathcal{G}_{vv} & -\det\left(\mathcal{G}|_{\mathrm{span}(\mathbf{v},\mathbf{w})}\right) \end{pmatrix}, \tag{7.3}$$

where we have set $\Xi \in \mathbb{R}$ to 1 by a suitable rescaling[1], and we have defined the restricted determinant

$$\det\left(\mathcal{G}|_{\mathrm{span}(\mathbf{v},\mathbf{w})}\right) = \begin{vmatrix} \mathcal{G}_{vv} & \mathcal{G}_{vw} \\ \mathcal{G}_{wv} & \mathcal{G}_{ww} \end{vmatrix} = \mathcal{G}_{vv}\mathcal{G}_{ww} - \mathcal{G}_{vw}\mathcal{G}_{wv}.$$

The prefactor becomes $|\mathcal{G}_{vw}|$ after we gauge away the phase of T by setting $T \to e^{i \arg(\mathcal{G}_{vw})} T$ (See eq. (6.66)). The conditions on the Green's function in eq. 6.64 reduce to $\mathcal{G}_{vv},\ \mathcal{G}_{ww} \in \mathbb{R}$ and $\mathcal{G}^*_{vw} = \mathcal{G}_{wv}$. Finally, since the transfer matrix is real and has unit determinant, $T \in Sp(2, \mathbb{R}) \cong SL(2, \mathbb{R})$. Hence, by construction[2], all transfer matrices for $r = 1$ are symplectic.

We can write out the Floquet discriminant as

$$\Delta = \mathrm{tr}\,\{T\} = \frac{1}{|\mathcal{G}_{vw}|}\left[1 - \det\left(\mathcal{G}|_{\mathrm{span}(\mathbf{v},\mathbf{w})}\right)\right]. \tag{7.4}$$

The band edges are given by $\Delta(\varepsilon, \mathbf{k}_\perp) = \pm 2$, which can be used to solve for $\varepsilon(\mathbf{k}_\perp)$, at least locally. Note that ε enters the calculation only as $\mathcal{G} = (\varepsilon \mathbb{1} - M)^{-1}$, which is a rational function of ε, so that solving for the band edges is equivalent to finding the zeros of a polynomial in ε.

Computing the edge is also particularly simple for $r = 1$. Recall from 6.4.2 that a given vector is a physical edge state if it satisfies both a decay condition and a boundary condition. Considering the left edge, for

[1] This also rescales ε, but that is analogous to writing the energy in *units* of the hopping energy, a practice that is quite standard.

[2] This may not be true for higher ranks, as $Sp(n, \mathbb{R}) \subset SL(n, \mathbb{R})$ is a proper subset for $n > 2$.

a $\Phi_1 \in \mathbb{C}^2$, the decay condition was given by $\mathcal{P}_< \Phi_1 = \Phi_1$, where $\mathcal{P}_< = \sum_{|\rho_s|<1} \mathcal{P}_s$. But since T has only two eigenvalues and $\det T = 1$, at most one of them (ρ_1, say) can lie *inside* the unit disk. Thus, $\mathcal{P}_< = \mathcal{P}_1$ becomes a projector to an eigenspace of T, and the decay condition implies that Φ_1 is an eigenvector of T. A similar analysis for the right edge, with the corresponding eigenvalue now lying *outside* the unit disk. The boundary conditions simply demand that one of the components of Φ vanish. Thus, the conditions for $\Phi_{1,N}$ to be a physical left/right edge state can be written concisely as

$$\begin{aligned} T\Phi_1 &= \rho\,\Phi_1,\ |\rho| < 1, \qquad \Phi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ T\Phi_N &= \rho\,\Phi_N,\ |\rho| > 1, \qquad \Phi_N = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \end{aligned} \tag{7.5}$$

where we have exercised our right to scale $\Phi_{1,N}$ by an arbitrary complex number. Explicitly, using eq. (7.3), the condition for a physical left edge state is

$$T\Phi_1 = \frac{1}{|\mathcal{G}_{vw}|} \begin{pmatrix} 1 \\ \mathcal{G}_{vv} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \implies \mathcal{G}_{vv} = 0, \quad |\mathcal{G}_{vw}| > 1. \tag{7.6}$$

Similarly, the condition for a physical right edge state is

$$T\Phi_N = \frac{1}{|\mathcal{G}_{vw}|} \begin{pmatrix} -\mathcal{G}_{ww} \\ |\mathcal{G}_{vw}|^2 - \mathcal{G}_{vv}\mathcal{G}_{ww} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \implies \mathcal{G}_{ww} = 0, \quad |\mathcal{G}_{vw}| > 1. \tag{7.7}$$

The equations $\mathcal{G}_{vv} = 0$ and $\mathcal{G}_{vw} = 0$ describe curves in the $(\varepsilon, \mathbf{k}_\perp)$ space, to which we can associate a winding number, as discussed in Sec 7.2.

The condition for the edge states can also be expressed as an Evans function[56] known in the dynamical system literature. We simply demand that a given $\varphi \in \mathbb{C}^2$ be an eigenvector of T (hereby referred to as the *eigenvalue condition*). Physically, this is equivalent to the statement that φ satisfies *either* the left *or* the right decay condition, i.e, it lies entirely in the growing or the decaying subspace. We can later check whether these states are physical by computing the corresponding eigenvalues. Given an antisymmetric matrix $\mathcal{J} \in \mathrm{Mat}(2, \mathbb{C})$, φ satisfies $\varphi^T\, \mathcal{J}\, \varphi = 0$, so that the eigenvalue condition ($T\varphi \propto \varphi$) can then be equivalently expressed as

$$f(\varepsilon, \mathbf{k}_\perp) \equiv \varphi^T\, \mathcal{J}\, T(\varepsilon, \mathbf{k}_\perp)\, \varphi = 0, \qquad \mathcal{J} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{7.8}$$

where we have chosen a specific $\mathcal{J}$. This provides a closed form condition for the existence of edge states.

7.1.2 Hofstadter model

We start off by repeating Hatsugai's[47] calculation in our formalism. The Hofstadter Hamiltonian models a 2-dimensional electron gas in the presence of a magnetic field normal to the plane. For a lattice Hamiltonian, the magnetic field simply correspond to the fact that there is a magnetic flux threading each plaquette, so that the electrons pick up a phase while going around the plaquettes. These phases are incorporated in the tight binding model by adding extra phases to the hopping strengths, the choice of which is nonunique and equivalent to the gauge choice for the vector potential in the continuum (Sec 3.1). The Hofstadter Hamiltonian on a square lattice can be written as

$$\mathcal{H} = -\sum_{m,n} \left[c^\dagger_{m+1,n} c_{m,n} + c^\dagger_{m,n+1} e^{2\pi m\phi} c_{m,n} + \text{h.c.} \right], \tag{7.9}$$

where ϕ denotes the magnetic flux associated with each plaquette. We have chosen the gauge such that all phases arising due to the flux depend on m. Thus, the system is translation invariant along y, so that a partial Fourier transform along y lead to

$$\mathcal{H}(k_y) = -\sum_{m} \left[c^\dagger_m c_{m+1} + c^\dagger_{m+1} c_m + 2\cos(k_y - 2\pi m\phi) c^\dagger_m c_m \right] \tag{7.10}$$

To construct the generalized transfer matrix for this system, we first need to construct a supercell, so that the system is periodic under supercell translations. If ϕ is rational, i.e, $\phi = p/q$, where $p, q \in \mathbb{Z}^+$ are coprime, then the system is periodic along m with period q, so that we can club together q physical sites to make a supercell. Thus, M would be a a $q \times q$ matrix. Since the eigenvalues of M form the band centers, we conclude that we have q bulk bands. This strong dependence of the number of bulk bands on the denominator of ϕ for rational ϕ, leads to the intricate fractal structure of the Hofstadter butterfly[105]. Furthermore, for even q, the system is gapless, while for odd q, the system is gapped and exhibits nontrivial edge states.

The identification of J and M matrices are then straightforward (See 6.3). There is only one link between adjacent supercells, *viz*, site q of supercell n connects with site 1 of supercell $n+1$, so that the hopping J has all entries equal to zero except $J_{1q} = 1$. The on-site matrix M has $2\cos(k_y - 2\pi n\phi)$ on its main diagonal entries and 1's on its first diagonals. For instance, for the simplest nontrivial case of $\phi = 1/3$, these matrices are

$$J = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad M = 2 \begin{pmatrix} \cos\left(k_y - \frac{2\pi}{3}\right) & 1 & 0 \\ 1 & \cos\left(k_y + \frac{2\pi}{3}\right) & 1 \\ 0 & 1 & \cos(k_y) \end{pmatrix}. \tag{7.11}$$

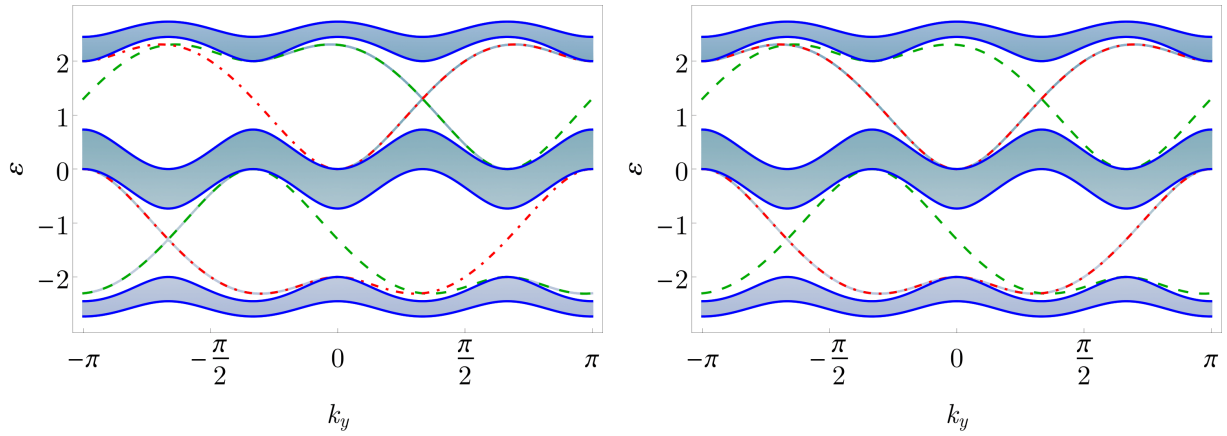


Figure 7.1: The spectrum of Hofstadter model, with the band edges (dark blue) computed using the transfer matrix formalism and the left and right edge state dispersion (dashed and dashed-dot) from the Evans equation (7.8), overlaid on the spectrum computed using exact diagonalization for a (top) commensurate and (bottom) incommensurate system. Note that in the latter case, the edge states seen in exact diagonalization exactly follow the winding right edge state obtained from the transfer matrix.

The actual computation of the transfer matrix is quite tedious[3] and not very enlightening; however, the results can sometime be written in a compact form. For instance, for $\phi = 1/3$, the Floquet discriminant is a polynomial in ε, given by

$$\Delta(\varepsilon, k_y) = \varepsilon^3 - 6\varepsilon - 2\cos(3k_y), \tag{7.12}$$

As expected, $q = 3$, and $\Delta(\varepsilon) = \pm 2$ is a cubic equation, which has three real solutions for each k_y corresponding to the upper and lower edges of the three bulk bands, as shown in Fig 7.1. For the Hofstadter model, $\Delta(\varepsilon)$ is always polynomial in ε, for reasons discussed in the next subsection.

The edge state calculation is identical to Hatsugai's, so we do not describe it in any detail. However, we emphasize his remark that if the total number of sites is *commensurate* with the flux ϕ, i.e, a multiple of q, then for a given k_y, we either get physical edge states on both left and right edges, or no physical edge states at all. In order to have a physical edge state for all k_y, which will have an associated winding, we need to consider a system with the number of sites *incommensurate* with the flux[4] (see Fig 7.1). Thus, it is useful to look at both *physical* or *unphysical* edge states, i.e, those satisfying all combinations of the boundary and decay conditions, to expose their topological nature, which manifests itself only partly in numerics for a given edge and system size.

In our picture, for the case of incommensurate system size, the number of supercells is not an integer. For the Hofstadter model, the $\mathscr{N}$ degrees of freedoms per supercell are *physical* sites, so that a fractional

[3] We use Mathematica™ to perform the analytic manipulations for most examples in this chapter. For an explicit computation, see the example of Chern insulator.

[4] Our usage of the words "commensurate" and "incommensurate" seems to be opposite that of Hatsugai!

number of supercells is physically sensible. In general, the degrees of freedom constituting a supercell are not physical sites; however, we shall see that the number of supercells being fractional still formally makes sense, and we can contrive the corresponding (potentially unphysical) boundary conditions for exact diagonalization computations, which will exhibit the winding of the edge states. We shall hereafter use the word *incommensurate* (w.r.t the superlattice) to refer to such computations, where the number of supercells is not an integer.

7.1.3 Natural basis and "unfolding"

Hatsugai's original calculation[47, 60] of the transfer matrix worked because of the fact that in the Hofstadter model, only the nearest neighbor hoppings are nonzero. Before we proceed to further examples, we stop to consider the implications of only having *nearest neighbor hopping*[5] inside a single supercell, which implies, in our notation, that M is tridiagonal, and $J = \mathbf{e}_1 \otimes \mathbf{e}_{\mathscr{N}}$, as in the Hofstadter model. Explicitly, if

$$J = \begin{pmatrix} 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, \qquad M = \begin{pmatrix} \mu_1 & \tau_1 & \dots & 0 \\ \tau_1 & \mu_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mu_{\mathscr{N}} \end{pmatrix}, \tag{7.13}$$

with $\mu_n, \tau_n \in \mathbb{R}\ \forall\, n = 1, 2, \dots \mathscr{N}$, where we have defined $\tau_{\mathscr{N}} = 1$, then we can write the recursion relation as

$$\tau_n \phi_{n+1} + \mu_n \phi_n + \tau_n \phi_{n-1} = \varepsilon \phi_n, \tag{7.14}$$

where τ_n and μ_n are periodic with period $\mathscr{N}$. Following Hatsugai and others[48, 49], we can compute the transfer matrix for translation by a supercell as the product

$$T = \prod_{n=1}^{\mathscr{N}} \mathcal{T}_n, \qquad \mathcal{T}_n = \begin{pmatrix} -\frac{1}{\tau_n}(\varepsilon - \nu_n) & -1 \\ 1 & 0 \end{pmatrix}, \tag{7.15}$$

where $\mathcal{T}_n$ is the transfer matrix from the site n to the site $n+1$, satisfying the periodicity $\mathcal{T}_{n+\mathscr{N}} = \mathcal{T}_n$. This construction always results in the transfer matrix T being polynomial in ε, as it is defined as a product of matrices $\mathcal{T}_n$, each of which is linear in ε. Subsequently, the Floquet discriminant, $\Delta = \mathrm{tr}\,\{T\}$ is a polynomial in ε, as was the case for the Hofstadter model.

A natural question to ask is whether a given system can be reduced to this *nearest-neighbor-hopping* form. More explicitly, given a particular J and M, we ask whether they can be reduced to the form of eq. (7.13)

[5] Recall that we are thinking of different degrees of freedom inside a supercell as "sites", even if physically, they might refer to spin/orbit/something else.

by some unitary transformations, or equivalently, by a suitable choice of basis of $\mathbb{C}^{\mathscr{N}}$. We start off by noting that J provides a natural orthonormal basis for $\mathbb{C}^{\mathscr{N}}$, viz, $\{\mathbf{v}, \mathbf{x}_j, \mathbf{w}\}$, in which J takes the form required in eq. (7.13), since $J = \mathbf{v} \cdot \mathbf{w}^\dagger$, so that $\langle \mathbf{v} J \mathbf{w} \rangle = 1$ and all other matrix elements vanish. Furthermore, M becomes

$$M = \begin{pmatrix} \mathbf{v}^\dagger M \mathbf{v} & \mathbf{v}^\dagger M X & \mathbf{v}^\dagger M \mathbf{w} \\ X^\dagger M \mathbf{v} & X^\dagger M X & X^\dagger M \mathbf{w} \\ \mathbf{w}^\dagger M \mathbf{v} & \mathbf{w}^\dagger M X & \mathbf{w}^\dagger M \mathbf{w} \end{pmatrix}, \tag{7.16}$$

where X is unitary since $\{\mathbf{x}_i\}$ are orthonormal. We have some freedom in choosing X, since we can choose any basis of $\text{span}\{\mathbf{x}_i\} \subset \mathbb{C}^{\mathscr{N}}$. We seek to use this freedom to attempt to reduce M to the form of eq. (7.13).

For tridiagonalization of M, we must have $\mathbf{w}^\dagger M \mathbf{v} = 0$, which simply means that the degrees of freedom in a supercell connected to the next and the previous supercells are not directly connected to each other (except for when $\mathscr{N} = 2$). This can always be arranged by taking a big enough supercell. Also, since M is Hermitian, $X^\dagger M X$ can always be reduced to a tridiagonal form by a suitable nonunique choice of $X \in \text{U}(\mathscr{N} - 2)$ using the Lanczos/Householder algorithm[106, 107]. Finally, we are left with the conditions

$$X^\dagger M \mathbf{v} \propto (1, 0 \ldots 0)^T, \quad X^\dagger M \mathbf{w} \propto (1, 0 \ldots 0)^T, \tag{7.17}$$

which should be checked explicitly for the case at hand. If such an X exists, then we shall refer to such a transformation as *unfolding* the model to a 1D chain. A quick survey of the matrix M in this basis reveals various restrictions on the transfer matrices. For instance, if $\mathbf{v}^\dagger M \mathbf{v}$ and $\mathbf{w}^\dagger M \mathbf{w}$ are real, it immediately follows that the entries of the transfer matrix are real. Furthermore, we can often glean information about the edge states by looking at the hopping of the resulting 1D chain, as discussed explicitly in Sec 7.1.4

7.1.4 Chern insulator

We next consider the Chern insulator, which is defined on a 2-dimensional square lattice by the Bloch Hamiltonian

$$\mathcal{H}_B = a \sin k_x \sigma^x + a \sin k_y \sigma^y + b(2 - m - \cos k_x - \cos k_y)\sigma^z, \tag{7.18}$$

where $a, b, m \in \mathbb{R}$. This model is a discrete (lattice) version of the 2-dimensional Dirac Hamiltonian, given by $\mathcal{H}_{\text{Dirac}} = \sigma_x \hat{p}_x + \sigma_y \hat{p}_y + m\sigma_z$, where $\hat{p}_i = -i\partial_i$ are the momentum operators and m is the *mass* of the fermion. The bulk spectrum for the Chern insulator is given by

$$\varepsilon = \pm \sqrt{a^2(\sin^2 k_x + \sin^2 k_y) + b^2(2 - m - \cos k_x - \cos k_y)^2}, \tag{7.19}$$

so that the system is gapless ("*semimetallic*") for some $\mathbf{k}$ only if $m = 0, 2, 4$, and is gapped ("*insulating*") otherwise.

The 'topological' nature of this model refers to an obstruction to a definition of the eigenstates of $\mathcal{H}_B(\mathbf{k})$ which vary smoothly with $\mathbf{k}$. More explicitly, consider the eigenvalue problem $\mathcal{H}(\mathbf{k})u(\mathbf{k}) = \varepsilon_+ u(\mathbf{k})$ with ε_+ corresponds to choosing the positive sign in eq. (7.19). Since the vectors $u(\mathbf{k}) \in \mathbb{C}^2$ are only defined upto a phase, we must choose their phases in order to define a continuous $u(\mathbf{k})$. Mathematically, this choice of phase corresponds to choosing a smooth global section of a U(1) bundle defined on the Brillouin zone $\mathbb{T}^2$, and no such section exists if the bundle is twisted (For more details, see Sec 4.2). One way to measure this twist is the first Chern number $C_1 \in \mathbb{Z}$, whose nonvanishing corresponds to a nontrivial bundle. This is our *bulk topological invariant*.

The first Chern number of a given band $\varepsilon_\pm$ is proportional to the integral of the Berry-curvature 2-form $\mathfrak{F}_\pm$ over the parameter space $\mathbb{T}^2 \ni \mathbf{k}$, i.e, the first Brillouin zone. Since the Bloch Hamiltonian is a two-level system, we can use the results from Sec 4.2.2 to compute the Berry curvature. Furthermore, using eq. (4.34), the Chern number is simply the winding number associated with the map $\mathbb{T}^2 \to S^2$ explicitly given by

$$\mathbf{k} \mapsto \frac{1}{\varepsilon_\pm}\Big(\sin k_x, \sin k_y, 2 - m - \cos k_x - \cos k_y\Big), \tag{7.20}$$

where the RHS is in terms of the Cartesian coordinates on S^2 embedded in $\mathbb{R}^3$. Thus, the Chern number becomes[40]

$$C_1 = \frac{1}{2\pi}\int_{\mathbb{T}^2} \mathfrak{F} \;=\; \begin{cases} 1, & 0 < m < 2 \\ -1, & 2 < m < 4 \\ 0, & \text{otherwise.} \end{cases} \tag{7.21}$$

We say that the system is in a *topological phase* for $m \in (0, 2) \cup (2, 4)$ and a *trivial phase* for $m \in (-\infty, 0) \cup (4, \infty)$. Clearly, one always encounters a gapless point when tuning m between topologically distinct phases, i.e, phases with different Chern numbers.

The Chern insulator model turns out to be the *drosophila melanogaster* (fruit fly) of our analysis of topological states using the generalized transfer matrix[6]. This is essentially because the transfer matrix is quadratic in ε, so that the corresponding energy Riemann surface would turn out to be a torus, while almost all other models correspond to surfaces of higher genera. which we discuss in Sec 7.2. We begin by explicitly constructing the transfer matrix for translations along x.

[6]The inspiration for our general construction was, in fact, an attempt to repeat Hatsugai's analysis of the Hofstadter model for the case of Chern insulator. We present that original computation in Appendix C.3.

Transfer matrix

We begin by inverse Fourier transforming the Chern insulator Bloch Hamiltonian along x:

$$\begin{aligned}\mathcal{H}(k_y) &= \sum_{n=0}^{N}\left[\frac{a}{2i}\left(c_{n+1}^{\dagger}\sigma^x c_n - c_n^{\dagger}\sigma^x c_{n+1}\right) + \frac{b}{2}\left(c_{n+1}^{\dagger}\sigma^z c_n + c_n^{\dagger}\sigma^z c_{n+1}\right) + c_n^{\dagger}\left(a\sin k_y\sigma^y + b\Lambda(k_y)\sigma^z\right)c_n\right], \\ &= \sum_{n}\left[c_{n+1}^{\dagger}\left(\frac{-ia\sigma^x + b\sigma^z}{2}\right)c_n + c_n^{\dagger}\left(a\sin k_y\sigma^y + b\Lambda(k_y)\sigma^z\right)c_n + c_n^{\dagger}\left(\frac{ia\sigma^x + b\sigma^z}{2}\right)c_{n+1}\right], \quad (7.22)\end{aligned}$$

where $\Lambda(k_y) = 2 - m - \cos k_y$. Comparing with eq. (7.2), we can readily identify

$$J = \frac{1}{2}\left(-ia\sigma^x + b\sigma^z\right), \quad M = a\sin k_y\sigma^y + b\,\Lambda(k_y)\sigma^z. \quad (7.23)$$

Clearly, $\det(J) = (b^2 - a^2)/4$, so that J becomes singular when $a = b$, which is precisely the case that we are interested in. We shall hereafter set $a = b = 1$, so that

$$J = \frac{1}{2i}\left(\sigma^x - i\sigma^z\right), \quad M = \sin k_y\sigma^y + \Lambda(k_y)\sigma^z. \quad (7.24)$$

We compute the SVD of J as

$$J = \mathbf{v}\cdot\mathbf{w}^{\dagger}, \qquad \mathbf{v} = \frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad \mathbf{w} = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 1 \end{pmatrix}. \quad (7.25)$$

Next,

$$\mathcal{G} = \left(\varepsilon\mathbb{1}_2 - \sin k_y\sigma^y - \Lambda(k_y)\sigma^z\right)^{-1} = A\left(\varepsilon\mathbb{1}_2 + \sin k_y\sigma^y + \Lambda(k_y)\sigma^z\right), \quad (7.26)$$

where $A = \left(\varepsilon^2 + \Lambda^2 - \sin^2 k_y\right)^{-1}$. The individual components are

$$\mathcal{G}_{vv} = A(\varepsilon + \sin k_y), \quad \mathcal{G}_{ww} = A(\varepsilon - \sin k_y), \quad \mathcal{G}_{vw} = \mathcal{G}_{wv} = -A\,\Lambda(k_y). \quad (7.27)$$

Thus, using eq. (7.3), the transfer matrix is given by[7]

$$T = \frac{1}{|\Lambda(k_y)|}\begin{pmatrix} -\varepsilon^2 + \Lambda^2(k_y) + \sin^2 k_y & \varepsilon - \sin k_y \\ -(\varepsilon + \sin k_y) & 1 \end{pmatrix}. \quad (7.28)$$

[7]Note that this is not identical to the transfer matrix obtained in eq. (C.26), but is related by a similarity transform, as they both have the same determinant and trace.

The condition for the bulk band edges is $\Delta = \pm 2$, where the Floquet discriminant is given by

$$\Delta(\varepsilon, k_y) = \mathrm{tr}\,\{T(\varepsilon, k_y)\} = \frac{1 - \varepsilon^2 + \Lambda^2(k_y) + \sin^2 k_y}{|\Lambda(k_y)|}. \tag{7.29}$$

Thus, the bulk band edges are explicitly given by

$$\varepsilon^2 = \sin^2 k_y + (2 \mp 1 - m - \cos k_y)^2. \tag{7.30}$$

The bands are symmetric under $\varepsilon \to -\varepsilon$, and stretch between $\varepsilon_{min} < |\varepsilon| < \varepsilon_{max}$, with

$$\varepsilon_{min} = \sqrt{\sin^2 k_y + (1 - m - \cos k_y)^2}, \qquad \varepsilon_{max} = \sqrt{\sin^2 k_y + (3 - m - \cos k_y)^2}. \tag{7.31}$$

for $m < 2$, while ε_{min} and ε_{max} switch roles for $m > 2$.

We can compute the edge spectrum explicitly using eqns (7.6) and (7.7), with the explicit definitions of $\mathcal{G}_{ab}$ for the Chern insulator in eq. (7.27), as

$$\varepsilon_L(k_y) = -\sin k_y, \qquad \varepsilon_R(k_y) = \sin k_y,$$

alongwith the condition for it to be physical, *viz*,

$$1 < |A\Lambda(k_y)| = \frac{|\Lambda(k_y)|}{\varepsilon^2 - \sin^2 k_y + \Lambda^2(k_y)} = \frac{1}{|\Lambda(k_y)|} \implies |\Lambda(k_y)| < 1, \tag{7.32}$$

since $\varepsilon_{L,R}^2 = \sin^2 k_y$. We plot the bulk and edge spectra computed above, alongwith the spectrum computed using exact diagonalization, in Fig. 7.2(a).

Unfolding the 1D chain and SSH model

For the Chern insulator, the unfolding to a 1D chain is particularly neat, as it leads to an alternating bond model, a quintessence of which is the Su-Schrieffer-Hieger(SSH) model[108] for polyacetylene. Since M is a 2×2 matrix and hence, by definition, *tridiagonal*, the unfolding requires a unitary operator which takes J to the desired form of eq. (7.13). Take a unitary operator $\mathcal{U}$, defined as

$$\mathcal{U} = \frac{1}{\sqrt{2}}(i\mathbb{1} + \sigma^x), \tag{7.33}$$

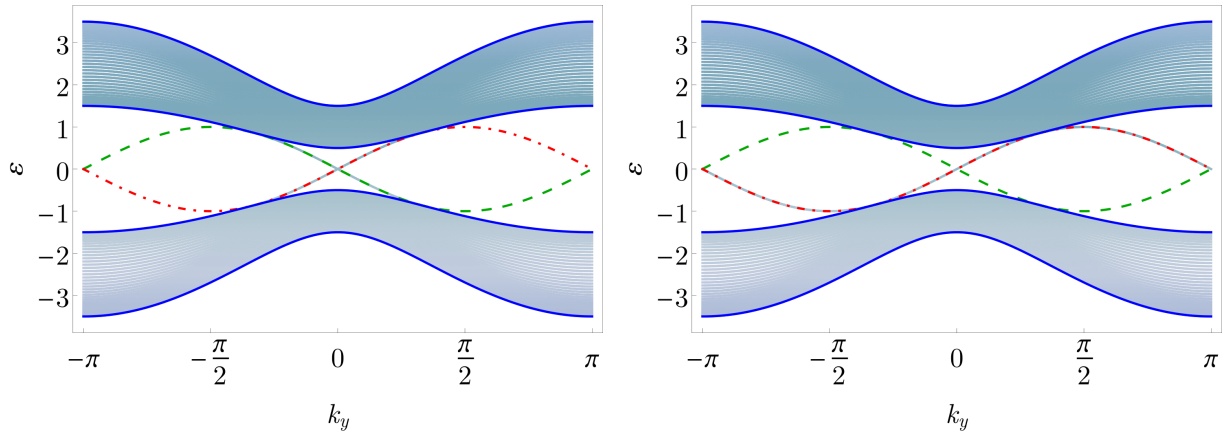


Figure 7.2: The spectrum of Chern insulator for $m = 0.8$, with the band edges (dark blue) computed using the transfer matrix formalism and the left and right edge state dispersion (dashed and dashed-dot) from the Evans equation (7.8), overlaid on the spectrum computed using exact diagonalization for a (top) commensurate and (bottom) incommensurate system. Note that in the latter case, the edge states seen in exact diagonalization exactly follow the winding right edge state obtained from the transfer matrix.

so that $\mathcal{U}\,\mathbf{v} = \mathbf{e}_1$ and $\mathcal{U}\,\mathbf{w} = \mathbf{e}_2$. Transforming the Bloch Hamiltonian as $\mathcal{H}_B \to \mathcal{H}'_B = \mathcal{U}\,\mathcal{H}_B\,\mathcal{U}^\dagger$, we get

$$\mathcal{H}'_B = \sin k_x \sigma^x + \sin k_y \sigma^z - (2 - m - \cos k_x - \cos k_y)\sigma^y \tag{7.34}$$

Inverse Fourier transforming along x, this becomes

$$\mathcal{H}'(k_y) = \sum_n \left[c^\dagger_{n+1} \left(\frac{-i\sigma^x + \sigma^y}{2} \right) c_n - c^\dagger_n \left(\frac{i\sigma^x + \sigma^y}{2} \right) c_{n+1} + c^\dagger_n \left(\sin k_y \sigma^z - \Lambda(k_y)\sigma^y \right) c_n \right], \tag{7.35}$$

where $c_n \equiv (c_n, \bar{c}_n)^T$. Redefining $\bar{c}_n = b_{2n}, c_n = b_{2n+1}$ and expanding the products, we get

$$\mathcal{H}'(k_y) = \sum_n \Big[\left(-i\,\tau_n b^\dagger_{n+1} b_n + \text{h.c.} \right) + \mu_n b^\dagger_n b_n \Big], \tag{7.36}$$

where

$$\mu_n = (-1)^n \sin(k_y), \quad \tau_n = \begin{cases} \Lambda(k_y) & ; n = \text{even}, \\ 1 & ; n = \text{odd}. \end{cases}$$

Hence, by a basis transformation on the Chern insulator, we have obtained the Hamiltonian for a 1D chain with alternating bond strengths 1 and $\Lambda(k_y)$. This is analogous to the case of the Su-Schrieffer-Hieger(SSH) model[41], with the addition of an alternating on-site energy term.

The SSH picture provides a straightforward interpretation for the emergence of the edge states: whenever one opens a boundary, one gets an edge state if the boundary cuts open a strong bond. Scanning as a

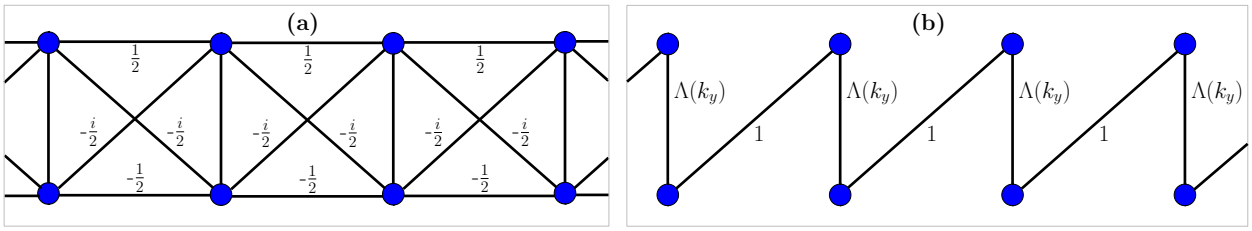


Figure 7.3: Unfolding the Chern insulator: In (a), we see the Chern insulator in the usual basis, treating the two degrees of freedom as sites. A change of basis in (b) transforms the model to a 1D chain with alternating hopping.

function of k_y, we can see that the edge states vanish when the bonds change their relative strength, i.e, when $|\Lambda(k_y)| = 1$, which is what one obtains from more elementary means[46] or sees in exact diagonalization. In the SSH model, the edge state appears at zero energy[41]. However, for the Chern insulator, we also have an on-site energy term $\mu_n = (-1)^n \sin k_y$. Hence, the spectrum of the edge state is given by $\varepsilon(k_y) = -\sin k_y$ for the left edge ($n = 1$) and $\varepsilon(k_y) = \sin k_y$ for the right edge ($n = 2 \times$ number of supercells), which is also what we got from a direct computation.

If the SSH chain has an even number of sites, so that the number of sites is commensurate with the size of the supercell, the edge states always occur in pairs, i.e, either both at the left and right end or not at all. This corresponds to the physical situation, as the aforementioned *sites* correspond to local spin/orbital degrees of freedom and hence always occur in pairs, which explains why the left and the right edges always *switch off* at the same k_y in the computation above. However, if one considers the incommensurate case where the SSH chain has an odd number of sites, there is an edge state for every k_y. If we allow such an (unphysical) boundary condition[8], we can expose the entire edge state in an exact diagonalization calculation, as shown in Fig 7.2(b). Thus, using Dirichlet boundary conditions for different "system sizes", we can expose edge spectrum throughout the Brillouin zone.

7.1.5 Further examples

We finally consider lattice models with $r = 1$ on nonsquare lattices and/or nontrivial directions of translation, where identifying the J and M matrix is not as straightforward. Consider then a tight binding model with the Bloch Hamiltonian given by $\mathcal{H}_B(\mathbf{k})$, $\mathbf{k} \in \mathbb{T}^d$, defined on a lattice with lattice vectors $\mathbf{a}_\alpha$, $\alpha = 1, \dots N$ embedded in $\mathbb{R}^N$, so that the physical lattice sites are at coordinates $\mathbf{r_n} = \sum_\alpha n_\alpha \mathbf{a}_\alpha$, $\mathbf{n} \in \mathbb{Z}^N$. Since the Bloch Hamiltonian is periodic under translations by reciprocal vectors[9], it can be written in terms of periodic

[8] In practice, we numerically construct the $2N \times 2N$ Hamiltonian for a system of width N along x, parametrized by k_y, and literally delete the last row and column to get a $(2N-1) \times (2N-1)$ Hamiltonian, which we diagonalize to get the exact diagonalization plot of Fig 7.2(b)

[9] Recall that given a set of lattice vectors $\mathbf{a}_\alpha$, the the reciprocal lattice vectors $\mathbf{G}_\alpha$ are defined such that $\mathbf{G}_\alpha \cdot \mathbf{a}_\beta = \delta_{\alpha\beta}$.

functions of the scalars $\mathbf{k}\cdot\mathbf{a}_\alpha$.

To construct the transfer matrix which translates along the direction in $\mathbb{R}^d$ specified by the unit vector $\hat{\mathbf{f}}$, we begin by constructing a set of orthonormal coordinates $\{\mathbf{e}_i\}_{i=1}^d$ such that $\mathbf{e}_1 = \hat{\mathbf{f}}$. In this basis, the lattice vectors become $\mathbf{a}_\alpha = a_{\alpha,i}\mathbf{e}_i$, so that $\mathbf{k}\cdot\mathbf{a}_\alpha = a_{\alpha,i}k_i$, where we have defined $k_i = \mathbf{k}\cdot\mathbf{e}_i$. The periodicity of $\mathcal{H}_B$ then lets us write it as sum of matrices dependent only on $\mathbf{k}_\perp = \{k_j,\ j = 2,\dots d\}$, with prefactors $e^{i\,a_{\alpha,1}k_1}$. Thus, in order to construct the transfer matrix, we demand that all $a_{\alpha,1}$ are commensurate with some γ, i.e,

$$\exists\,\gamma\in\mathbb{R}\text{ s.t }\gamma = M_\alpha a_{\alpha,1}\quad \forall\, a_{\alpha,1}\neq 0,\quad M_\alpha\in\mathbb{Z}. \tag{7.37}$$

If such a γ can be found, then set $R = \max\gamma/a_{\alpha,i}$, so that the Bloch Hamiltonian becomes

$$\mathcal{H}_B = M_0(\mathbf{k}_\perp) + \sum_{\ell=1}^{R} J_\ell(\mathbf{k}_\perp)e^{i\,a_{\alpha,1}k_1}. \tag{7.38}$$

This can now be inverse Fourier transformed to obtain a lattice model with range R hopping. Choosing a suitable supercell, we can identify the J and M matrix using the definitions from Sec 6.2.

Note that the existence of γ defined above is crucial for the definition of a transfer matrix. For instance, given a square lattice with $\mathbf{a}_1 = (1,0)^{\mathrm{T}}$ and $\mathbf{a}_2 = (0,1)^{\mathrm{T}}$, one can choose $\hat{\mathbf{f}}\propto(1,r)^T$, and define orthonormal axes in $\mathbb{R}^2$ as ($\mathcal{N} = \sqrt{1+r^2}$)

$$\mathbf{e}_1 = \mathcal{N}\begin{pmatrix}1\\ r\end{pmatrix},\quad \mathbf{e}_2 = \mathcal{N}\begin{pmatrix}-r\\ 1\end{pmatrix},\implies \frac{a_{2,1}}{a_{1,1}} = -r. \tag{7.39}$$

Thus, if $r = p/q$, we must take $\gamma = pa_{1,1} = qa_{2,2}$, so that $R = \max\{p,q\}$. This makes intuitive sense, since if one moves in the direction specified by (q,p) on a square lattice, the system is periodic with periodicity R. However, if r is irrational, we cannot satisfy eq. (7.37) for any γ, since the system is aperiodic along $\hat{\mathbf{f}}$. Thus, demanding the existence of a γ satisfying eq. (7.37), we are essentially stating that the system be periodic along the direction in which we intend to translate using our transfer matrix. This is essentially tautological, since we defined our translation matrix only for systems which are translation invariant.

To demonstrate this formal construction, take a tight-binding model with the Bloch Hamiltonian[109, 110]

$$\mathcal{H}_B(\mathbf{k}) = 2\begin{pmatrix}0 & \cos(\mathbf{k}\cdot\mathbf{a}_1) & \cos(\mathbf{k}\cdot\mathbf{a}_3)\\ \cos(\mathbf{k}\cdot\mathbf{a}_1) & 0 & \cos(\mathbf{k}\cdot\mathbf{a}_2)\\ \cos(\mathbf{k}\cdot\mathbf{a}_3) & \cos(\mathbf{k}\cdot\mathbf{a}_2) & 0\end{pmatrix}, \tag{7.40}$$

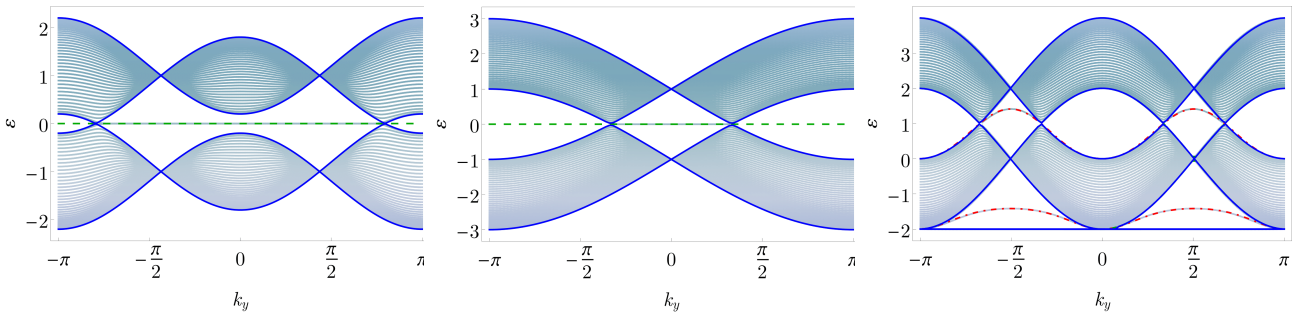


Figure 7.4: The spectrum of (top) Dirac Semimetal, (middle) Graphene and (bottom) Kagome semimetal. See Sec 7.1.5 and table 7.1 for details.

defined on the kagome lattice, with lattice vectors

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{a}_2 = \frac{1}{2}\begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}, \quad \mathbf{a}_3 = -\frac{1}{2}\begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix}. \tag{7.41}$$

Thus, the lattice is $\mathbb{Z}^3$, embedded in $\mathbb{R}^2$. Let us define the orthonormal axes on $\mathbb{R}^2$ as

$$\mathbf{e}_1 = \frac{1}{2}\begin{pmatrix} \sqrt{3} \\ -1 \end{pmatrix}, \quad \mathbf{e}_2 = \frac{1}{2}\begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix}, \tag{7.42}$$

and demand that the transfer matrix translate along $\mathbf{e}_1$. The lattice vector in these coordinates become

$$\mathbf{a}_1 = \frac{\sqrt{3}}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2, \quad \mathbf{a}_2 = -\frac{\sqrt{3}}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2, \quad \mathbf{a}_3 = -\mathbf{e}_2. \tag{7.43}$$

Thus, $a_{1,1} = -a_{2,1} = \sqrt{3}/2$ and $a_{3,1} = 0$, so that we can choose $\gamma = \sqrt{3}/2$ and $R = 1$. This is a particularly simple case, as it can be reduced to the form of eq. (7.2). Substituting $\mathbf{k} \cdot \mathbf{a}_\alpha$ in terms of $k_i = \mathbf{k} \cdot \mathbf{e}_i$ and setting $k_1 = k_x$, $k_2 = k_y$, the Bloch Hamiltonian becomes

$$\mathcal{H}_B = \begin{pmatrix} 0 & e^{\frac{i}{2}\left(\sqrt{3}\,k_x + k_y\right)} + e^{-\frac{i}{2}\left(\sqrt{3}\,k_x + k_y\right)} & 2\cos k_y \\ e^{\frac{i}{2}\left(\sqrt{3}\,k_x + k_y\right)} + e^{-\frac{i}{2}\left(\sqrt{3}\,k_x + k_y\right)} & 0 & e^{\frac{i}{2}\left(-\sqrt{3}\,k_x + k_y\right)} + e^{-\frac{i}{2}\left(-\sqrt{3}\,k_x + k_y\right)} \\ 2\cos(k_y) & e^{\frac{i}{2}\left(-\sqrt{3}\,k_x + k_y\right)} + e^{-\frac{i}{2}\left(-\sqrt{3}\,k_x + k_y\right)} & 0 \end{pmatrix}$$

$$= e^{i\frac{\sqrt{3}}{2}k_x} \underbrace{\begin{pmatrix} 0 & e^{ik_y/2} & 0 \\ e^{ik_y/2} & 0 & e^{-ik_y/2} \\ 0 & e^{-ik_y/2} & 0 \end{pmatrix}}_{J(k_y)} + \underbrace{\begin{pmatrix} 0 & 0 & 2\cos k_y \\ 0 & 0 & 0 \\ 2\cos k_y & 0 & 0 \end{pmatrix}}_{M(k_y)} + e^{-i\frac{\sqrt{3}}{2}k_x} \underbrace{\begin{pmatrix} 0 & e^{-ik_y/2} & 0 \\ e^{-ik_y/2} & 0 & e^{ik_y/2} \\ 0 & e^{ik_y/2} & 0 \end{pmatrix}}_{J^\dagger(k_y)}.$$

Comparing to eq. (7.1) with $k_x \to \frac{\sqrt{3}}{2}k_x$, we can readily identify the J and M matrices.

Model	J	M
Chern Insulator	$\frac{1}{2i}(\sigma^x - i\sigma^z)$	$\sin k_y \sigma^y + \Lambda(k_y)\sigma^z$
Dirac Semimetal	$\frac{1}{2i}(\sigma^x - i\sigma^z)$	$\Lambda(k_y)\sigma^z$
Graphene	$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1-e^{ik_y} \\ 1-e^{-ik_y} & 0 \end{pmatrix}$
Kagome Semimetal	$\begin{pmatrix} 0 & e^{ik_y/2} & 0 \\ e^{ik_y/2} & 0 & e^{-ik_y/2} \\ 0 & e^{-ik_y/2} & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 2\cos k_y \\ 0 & 0 & 0 \\ 2\cos k_y & 0 & 0 \end{pmatrix}$

Table 7.1: A list of J and M matrix for some of the well-known topological and semimetal states. The corresponding spectra are plotted in Fig. 7.4

In conclusion, using the generalized transfer matrix construction, the calculation of bulk bands as well as edge states becomes simply a matter of identifying the J and M matrices. We list these matrices for some of the well known topological and semimetal phases in Table 7.1. The corresponding band structures and edge states, superimposed over the exact diagonalization result, are collected in Fig. 7.4.

7.2 Riemann surfaces and windings

One of the significant aspects of topological phases of matter are the edge/surface states, whose existence is determined by the bulk characteristics and which cannot be gapped out by adding boundary terms. These states often reflect the topological invariants of the bulk. For instance, despite the strong dependence of the edge spectrum on the precise boundary condition, the number of (signed) crossings of a given energy level in the band gap is a topological invariant, equal to the bulk Chern number. The proof of this so called *bulk-boundary correspondence* is highly nontrivial[62], and has been worked out in detail for the clean limit only in certain specific cases[68, 45]. However, an alternative perspective, due to Hatsugai, identifies the topological invariants of the edge states as winding numbers of the edge states around the holes in the (complex) energy Riemann surface. He also provides a proof of this correspondence[60].

In this section, we describe the geometrical structures associated with the transfer matrices. The central purpose of this analysis is to obtain a better understanding of the topological nature of the edge states.

7.2.1 The two complexifications

In the Bloch analysis of discrete periodic systems, we usually restrict ourselves to real energies and momenta, which correspond to plane wave eigenstates. However, in this section, we shall see that there is much to be gained by allowing them to be complex ("complexifying" them). In the following, we shall only describe the situation for $r = 1$. Furthermore, we shall restrict ourselves to a system in 2 spatial dimensions, with hard boundary conditions along x and periodic boundary conditions along y, so that the transverse momentum is $\mathbf{k}_\perp = k_y \in S^1$.

Consider, then, a 2×2 transfer matrix for a 2-dimensional system, $T(\varepsilon, k_y)$. The eigenvalues of the transfer matrix are

$$\rho_\pm = \frac{1}{2}\left[\Delta \pm \sqrt{\Delta^2 - 4}\right], \quad \Delta = \mathrm{tr}\,\{T\}, \tag{7.44}$$

which satisfy $\rho_+\rho_- = \det T = 1$. Following the Bloch ansatz, we can put $\rho_+ = e^{ik_x} \implies \rho_- = e^{-ik_x}$, so that k_x is a function of (ε, k_y). In the standard Bloch theory, $(\varepsilon, k_y) \in \mathscr{G}$, the band gap, if $|\rho_\pm(\varepsilon, k_y)| \neq 1$, i.e, when $\rho_+ = e^{ik_x}$ has no real solution in $k_x \in \mathbb{R}$. Physically, this simply means that there are no states propagating along x in the gap.

However, $\rho_\pm(\varepsilon, k_y) = e^{\pm ik_x}$ can always be solved in $\mathbb{C}$, as $\rho_+\rho_- = 1 \implies \rho_\pm \neq 0$. That is our first complexification. In terms of the Floquet discriminant,

$$\Delta(\varepsilon, k_y) = 2\cos k_x. \tag{7.45}$$

By solving this equation for $k_x \in \mathbb{C}$, we get the so called *complex band structure* of the system[57, 47], which can also be numerically computed and plotted in a 3-dimensional space $(\mathrm{Re}(k_x), \mathrm{Im}(k_x), \varepsilon)$ for a given k_y[48, 103]. The imaginary part of k_z is interpreted as the inverse penetration depth of the edge modes, with $\mathrm{Im}(k_x)$ negative (positive) corresponding to the left (right) edge.

Now on to the second, and much more interesting, complexification. We note that the expression for the eigenvalues involves $\sqrt{\Delta^2 - 4}$, which is not a genuine function until we choose a branch of the square root. For real ε, the argument of the square root is also real and the two branches are picked for $\rho_\pm$, respectively. However, if we allow ε to be complex, the square root becomes a genuine function from a two sheeted Riemann surface to the complex plane, with the two sheets corresponding to the two choices for a branch, connected at the branch cuts in the complex plane[57, 47]. For real eigenvalues, the two sheets correspond to the magnitude of the eigenvalue being greater than (less than) unity. It is this structure that we seek to expose in the following.

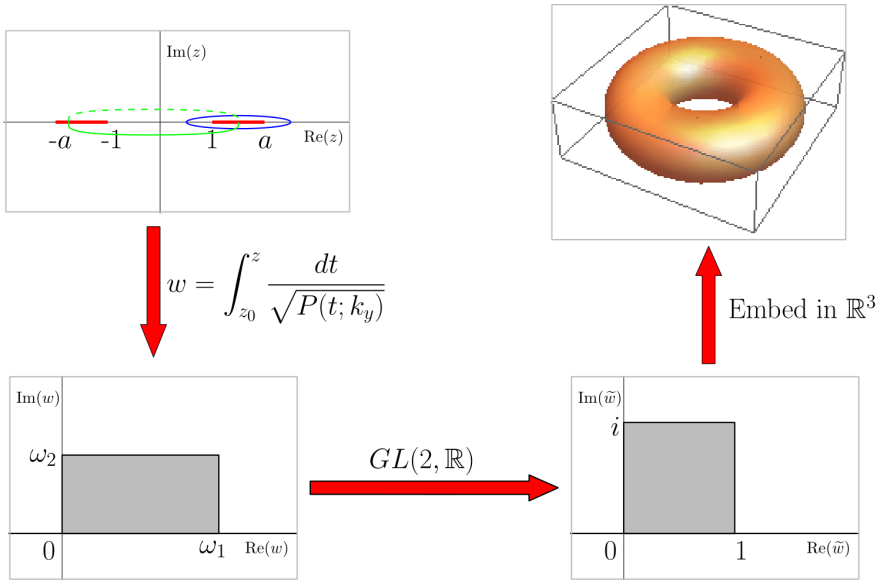


Figure 7.5: The schematic for plotting the Riemann sheet corresponding to Chern insulator.

The Floquet discriminant is, by construction, a rational function of ε and $\zeta = e^{ik_y}$. However, we shall restrict ourselves to the cases where it is a polynomial in ε, so that the denominator is independent of ε (see Sec 7.1.3 for relevant conditions for this to happen). Let us, then, define the discriminant of eq. (7.44) as

$$P(\varepsilon, k_y) = \Delta^2(\varepsilon, k_y) - 4. \tag{7.46}$$

We shall hereafter simply write $P(\varepsilon)$, tacitly assuming the dependence on k_y. For a given system with $\mathscr{N}$ degrees of freedom per supercell, the highest power of ε is that in $\det(\varepsilon \mathbb{1} - M)$, i.e, $\varepsilon^{\mathscr{N}}$, so that $P(\varepsilon)$ is a polynomial of order $2\mathscr{N}$ in ε. Thus, for a given k_y, $P(\varepsilon)$ has $2\mathscr{N}$ real roots, corresponding to the band edges for $\mathscr{N}$ bands. Allowing ε to be complex, we get a ε-Riemann surface consisting of two Riemann spheres connected along $\mathscr{N}$ branch cuts on the real axis, which is a surface with genus[70, 73] $\mathscr{N} - 1$.

In the following, we exhibit this structure explicitly for the case of the Chern insulator. Starting with eq. (7.29), we can write

$$P(z) = \varepsilon_{min}^4 (z - a)(z - 1)(z + 1)(z + a) \tag{7.47}$$

with

$$z(k_y) = \frac{\varepsilon}{\varepsilon_{min}(k_y)}, \quad a(k_y) = \frac{\varepsilon_{max}(k_y)}{\varepsilon_{min}(k_y)} > 1,$$

where $\varepsilon_{min}(k_y)$ and $\varepsilon_{max}(k_y)$ are band edges, as defined in eq. (7.31). The prefactor, ε_{min}^4, is nonzero for all k_y, except when the parameter $m = 0, 2, 4$, i.e, at the gapless points. Hence, as far as edge states are

concerned, we shall drop it in the subsequent discussion as it does not affect the roots of $P(z)$ and hence the branch-cut structure. On the other hand, for $m = 0, 2, 4$, the system becomes gapless and the topology of the Riemann sheet changes. In fact, for the gapless case, the polynomial can be written as

$$P(z) = z^2(z - \varepsilon_{max})(z + \varepsilon_{max}), \quad z = \varepsilon, \tag{7.48}$$

so that the Riemann surface now consists of two sheets connected at the single branch cut running between $-\varepsilon_{max}$ and ε_{max}, which has the topology of a sphere[70]. This is a general feature: the topology of the ε-Riemann surface changes at the gapless points.

For the gapped case, given $a(k_y)$, we can map the Riemann surface to a torus[10] explicitly, using the elliptic integral[70]:

$$w = \int_{z_0}^{z} \frac{dt}{\sqrt{P(t; k_y)}} \tag{7.49}$$

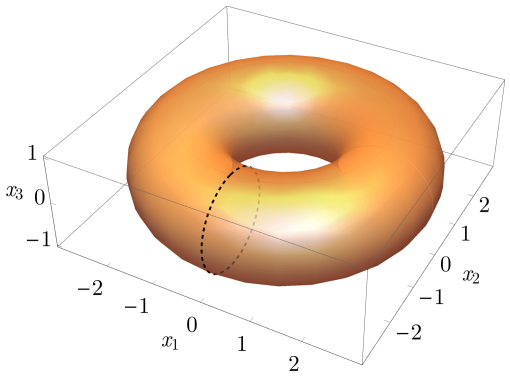
where the integral is independent of the path, as long as it does not wind around the branch cuts, corresponding to the two holonomies of the torus. On the other hand, such a winding gives the two periods of the torus, as

$$\omega_1(k_y) = \oint_\alpha \frac{dt}{\sqrt{P(t)}}, \quad \omega_2(k_y) = \oint_\beta \frac{dt}{\sqrt{P(t)}}. \tag{7.50}$$

Hence, the elliptic integral maps the coordinate z on the Riemann sheet to w on the rectangle formed by $0, \omega_1, \omega_1 + \omega_2$ and ω_2 in the complex plane, with the opposite edges identified. We can perform a $GL(2, \mathbb{R})$ transform $w \mapsto \widetilde{w}$ to map this rectangle to the square S bounded by the points $0, 1, 1 + i$ and i. Finally, given $\widetilde{w} = \theta + i\phi$, we can embed the torus in 3 dimensional Euclidean space as as

$$\mathbf{x} = \Big((R + \sin\phi)\cos\theta,\ (R + \sin\phi)\sin\theta,\ \cos\phi \Big),$$

where $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ and $R > 1$ is a fixed constant. Using the sequence of maps described above, any curve on the complex-ε plane can now be visualized as a curve on a torus. A schematic of this process is depicted in Fig. 7.5. We also show such a plot in the adjacent figure, with the dashed black loop enclosing one of the branch cuts.



[10] or a rectangle in the complex plane with opposite edges identified, to be precise.

7.2.2 Windings on the Riemann surface

The physical edge states for a topological Hamiltonian are required to satisfy two conditions, *viz*, the boundary conditions and the decay condition, which, for $r = 1$, turn out to be $\mathcal{G}_{vv,ww}(\varepsilon, k_y) = 0$ and $|\mathcal{G}_{vw}(\varepsilon, k_y)| > 1$, respectively. The former generically defines a curve in the complex-ε space, while the latter inequality selects out a part of this curve which corresponds to the *physical* edge state. However, it is to those curves in totality that we can associate a winding number, which will encode the topological nature of the edge states.

In the last subsection, we constructed a set of Riemann surfaces parametrized by k_y for a given Floquet discriminant $\Delta(\varepsilon, k_y)$, which was a polynomial of order $\mathscr{N}$. If the system stays gapped[11], then they all have the same genus, *viz*, $\mathscr{N} - 1$ independent of k_y, so that they can be deformed into each other by smooth transformations. Thus, we map them all to a generic surface $\mathcal{S}$ of genus $\mathscr{N} - 1$. A smooth level set $f(\varepsilon, k_y) = 0$ could then define a curve on $\mathcal{S}$, which would be closed since $k_y \in S^1$. The set of closed loops on $\mathcal{S}$ is classified by $\pi_1(\mathcal{S})$, the fundamental group of $\mathcal{S}$, which is nontrivial for a genus $\mathscr{N} - 1$ surface if $\mathscr{N} > 1$.

The set of curves defined by $\mathcal{G}_{vv,ww}(\varepsilon, k_y) = 0$, however, cannot fall into all homotopy classes in $\pi_1(\mathcal{S})$. The essential reason is that all energies ε corresponding to physical systems must stay positive, so that we are, in fact, stuck to (two copies of) the real line in the ε-Riemann sheet, connected at the branch cuts[12] Thus, the actual maps that we are concerned with are simply those which wind around one particular band gap, and hence, in essence, are $S^1 \to S^1$. This map is associated with just one winding number as $\pi_1(S^1) \cong \mathbb{Z}$, which is not the same as the fundamental group of the Riemann surface, as, for instance, $\pi_1(T^2) \cong \mathbb{Z} \times \mathbb{Z}$.

We now consider the case of Chern insulator, where $\mathcal{S} = \mathbb{T}^2$. The edge spectrum is explicitly given by $\varepsilon_{R,L}(k_y) = \pm \sin k_y$, as computed in Sec 7.1.4. If the associated curve, $\varepsilon_L(k_y)$, winds around a hole of the Riemann surface, it has to be on both the sheets. But the two sheets correspond to the eigenvalues of T being less than or greater than 1, i.e, for the modes to be decaying as $n \to \infty$ and $n \to -\infty$, respectively. Hence, in order to have a curve with a nontrivial winding, we need both the physical and unphysical states, as defined in Table 6.1. We point out that in Hatsugai's analysis, the winding was obtained using only the *physical* edge states by using a boundary condition such that $\Phi_1 = \Phi_N$ in Table 6.1, so that any given state is physical at at least one of the edges. This corresponds to the incommensurate case in our description.

In the discussion on the ε-Riemann surface, we remarked that its topology changes when the system becomes gapless. In particular, for the Chern insulator at $m = 0 = k_y$, the Riemann sheet is a 2-sphere,

[11] When the topology depends on k_y, for instance, if the gap closes for some k_y, then we shall need the language of cobordism to describe the family of Riemann surfaces.

[12] We could have done this computation by simply gluing together two branches of $\sqrt{P(z)}$ wherever $P(z) = 0$, where the branches yield the same result, but the language of Riemann surface is more familiar and hence less ad hoc.

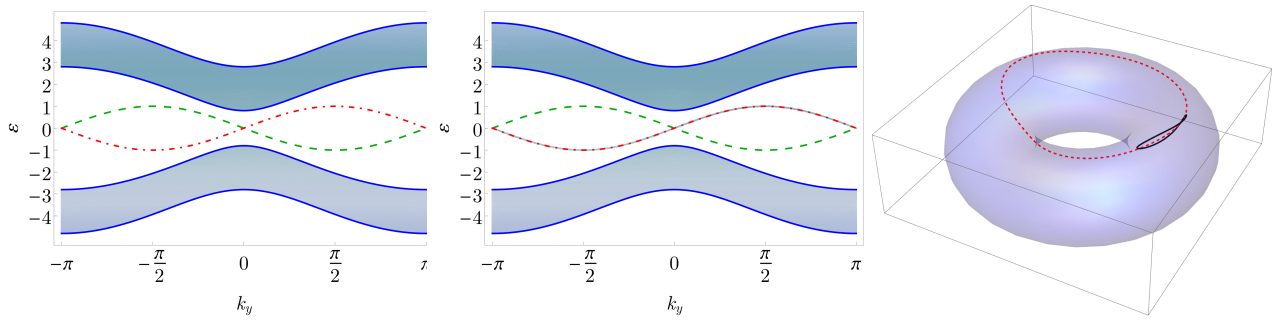


Figure 7.6: The spectrum of Chern insulator for $m = -0.8$, with the band edges (dark blue) computed using the transfer matrix formalism and the left and right edge state dispersion (dashed and dashed-dot), overlaid on the spectrum computed using exact diagonalization for a (left) commensurate and (center) incommensurate system. In the rightmost panel, we plot the transfer matrix corresponding to the left edge state for Chern insulator(solid black line), which is contractible, and compare it with the case of $m = +0.8$(red dashed curve), where it is incontractible. The underlying (solid) torus is the manifold $\mathrm{Sp}(2, \mathbb{R})$.

on which all loops are contractible. Hence, as one tunes m across one of these gapless points, the winding number (and hence the Chern number) can change, as the loops that were non-contractible on the torus can be contracted to a point on the sphere. This does not necessarily mean that there are no states anymore that satisfy the boundary and decay conditions; rather, it simply implies that the curves corresponding to such states are now contractible. Furthermore, we can also expose such a state in exact diagonalization by taking an incommensurate system, as shown in Fig. 7.6. Physically, this indicates that even when the bulk is trivial, there can still be states localized on the edge that decay into the bulk, but they are not topologically protected, and hence can be removed by adding a suitable boundary term.

7.2.3 Winding in $Sp(2, \mathbb{R})$

For $r = 1$, the corresponding transfer matrices $T \in Sp(2, \mathbb{R})$, a Lie group which, as a 3-dimensional smooth manifold, is homeomorphic to a solid 2-torus, i.e, $D^2 \times S^1$, where D^2 represents the 2-dimensional open disc. In Appendix B.2, we describe a particular parametrization of this Lie group. Given $\varepsilon(k_y)$ which is a continuous function of k_y, consider $T\left(\varepsilon(k_y), k_y\right)$. As $k_y \in S^1$, this describes a curve $\mathcal{C}$ on $Sp(2, \mathbb{R})$, so that edge spectra correspond to curves on $\mathrm{Sp}(2, \mathbb{R})$. Since[13] $\pi_1\left(\mathrm{Sp}(2, \mathbb{R})\right) \cong \mathbb{Z}$, the winding number of this curve provides another diagnostic of the topological nature of the edge states. For instance, in Fig 7.6, we plot the left edge state of the Chern insulator in both the topological and the trivial regime, where the winding number correctly determines the Chern number associated with the bulk in each case.

Note that this computation does not need any of the complexifications described in the previous sections. Another advantage of plotting these curves in $Sp(2, \mathbb{R})$ over the curves on the ε-Riemann surface is that

[13] See Appendix B.2 for a proof.

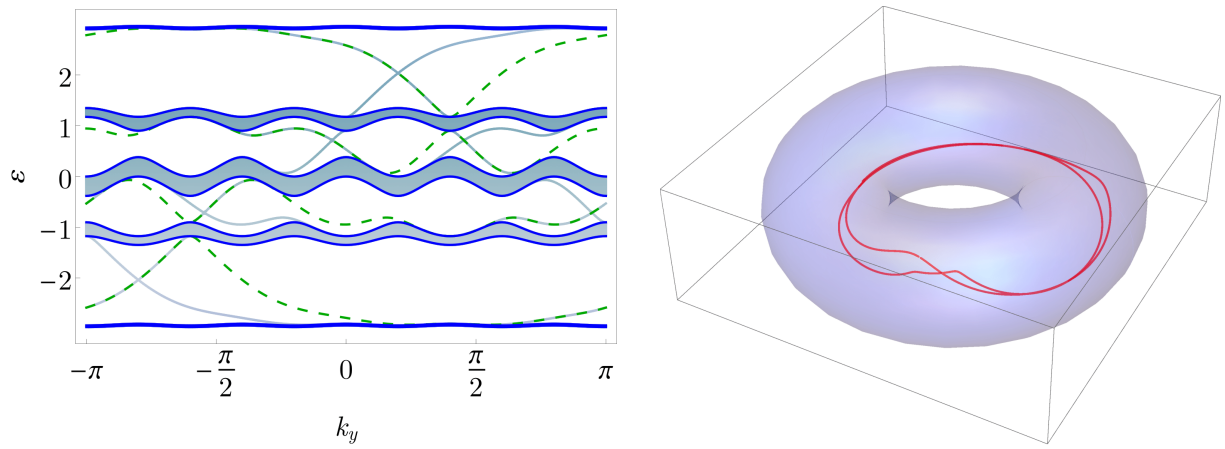


Figure 7.7: The spectrum of the Hofstadter model for $\phi = 1/5$ (top), and (bottom) the curve of transfer matrices on $Sp(2,\mathbb{R})$ corresponding to the left edge state in the second gap from the bottom. Note that the curve winds around twice in $Sp(2,\mathbb{R})$, as expected from the spectrum.

the curves described here is always a solid torus for all rank 1 systems, as opposed to the Riemann surface, which is a surface whose genus is a function of the number of bands. For instance, we plot the edge state for the Hofstadter model with $\phi = 1/5$ in Fig. 7.7, the Riemann surface corresponding to which has genus 4. The edge state shown has a winding number of 2, a fact that can be easily gleaned from the figure.

The edge spectra could be computed using the Evans condition of eq. (7.8), i.e, $\varphi^{\mathrm{T}}\mathcal{J}T(\varepsilon, k_y)\varphi = 0$. We now show that the corresponding winding number is independent of φ, and hence is, to some extent, independent of the specific choice of the boundary conditions. Since $\pi_1\left(Sp(2,\mathbb{R})\right) \cong \mathbb{Z}$, any curve $\mathcal{C}$ on $Sp(2,\mathbb{R})$ is associated with a winding number (also known as Maslov index[100, 71]), so that we have a map

$$\mu : Z_1(Sp(2,\mathbb{R})) \to \mathbb{Z}, \tag{7.51}$$

which associates a winding number with each loop, $\mathcal{C} \in Z_1(Sp(2,\mathbb{R}))$, where $Z_1(M)$ denotes the set of all closed loops on a smooth manifold M. Now, the Evans condition for a given φ is a continuous function of k_y, to which we can associate a curve $\mathcal{C}_\varphi$, with the corresponding winding number $\mu(\mathcal{C}_\varphi)$. Hence, for each $\varphi \in \mathbb{C}^2\backslash\{0\}$, we get a map $\varphi \mapsto \mu(\mathcal{C}_\varphi) \in \mathbb{Z}$. But as $\mu(\mathcal{C}_\varphi)$ is an integer, it cannot change continuously under a continuous change of φ. Thus, $\mu(\mathcal{C}_\varphi)$ must be independent of φ's for a given gap.

So far, we have not shown using our formalism that the winding number of a curve corresponding to an edge spectrum in $Sp(2,\mathbb{R})$ is same as the winding number of the corresponding curve on the ε-Riemann surface, even though we notice it to be so in all the examples that we checked[14]. The interpretation of the Chern number as a Maslov index can provide new ways of computing it numerically, as well as analytically[111, 50].

[14] A proof of a similar statement is discussed in Ref. [50] using K-theory.

7.3 An example for $r = 2$: TCI

The analytic computation of the transfer matrix naturally becomes more intricate for $r > 1$. However, if the transfer matrix turns out to be symplectic, we can take advantage of the additional structure (See Sec 6.4.1) for exact computations. Here, we compute the transfer-matrix for a $r = 2$ model in closed form and derive exact analytical expressions for its surface spectrum for such a case. The model we study is a C_4–invariant topological crystalline insulator (TCI), first introduced by Fu[112]. The model has topological surface states, which are protected by crystalline symmetries alongside time reversal symmetry.

The Fu model is defined on a 3-dimensional tetragonal lattice, with alternating layers of square lattices of A and B type along the z axis. The system has a 4-fold rotation symmetry in the plane normal to the z axis. The lattice model consists of nearest and next-nearest neighbor hoppings between two orbitals on each site (typically identified as p_x and p_y), with the strength of hopping being equal in magnitude but opposite in sign on the A and B sublattices. Thus, the model consists of 4 bands, with 2 orbitals and 2 sublattice degrees of freedom. The Bloch Hamiltonian is given by

$$\mathcal{H}(\mathbf{k}) = \begin{pmatrix} \mathcal{H}_A(\mathbf{k}) & \mathcal{H}_L(\mathbf{k}) \\ \mathcal{H}_L^\dagger(\mathbf{k}) & \mathcal{H}_B(\mathbf{k}) \end{pmatrix} \tag{7.52}$$

with the layer Hamiltonian $\mathcal{H}_\mathrm{a},\ \mathrm{a} \in \{A, B\}$ and the inter-layer hopping $\mathcal{H}_I$. The 2×2 blocks are given by

$$\begin{aligned} \mathcal{H}_\mathrm{a}(\mathbf{k}) &= 2t_1^\mathrm{a} \begin{pmatrix} \cos k_x & 0 \\ 0 & \cos k_y \end{pmatrix} + 2t_2^\mathrm{a} \begin{pmatrix} \cos k_x \cos k_y & \sin k_x \sin k_y \\ \sin k_x \sin k_y & \cos k_x \cos k_y \end{pmatrix} \\ &= \left[t_1^\mathrm{a}(\cos k_x + \cos k_y) + 2t_2^\mathrm{a} \cos k_x \cos k_y\right] \mathbb{1}_2 + 2t_2^\mathrm{a} \sin k_x \sin k_y \sigma_x + t_1^\mathrm{a}(\cos k_x - \cos k_y)\sigma_z, \\ \mathcal{H}_L(\mathbf{k}) &= \left[t_1' + 2t_2'(\cos k_x + \cos k_y) + t_z' e^{ik_z}\right] \mathbb{1}_2, \end{aligned} \tag{7.53}$$

where we take $t_i^A = -t_i^B \equiv t_i$ for $i = 1, 2$, so that $\mathcal{H}_A = -\mathcal{H}_B = \mathcal{H}_0$. The system is invariant under C_4 rotations, with the C_4 action defined by

$$\mathcal{C}_4\ \mathcal{H}(k_x, k_y, k_z)\ \mathcal{C}_4^{-1} = \mathcal{H}(-k_y, k_x, k_z), \tag{7.54}$$

where $\mathcal{C}_4 = i\mathbb{1}_2 \otimes \sigma_y$. Clearly, a cut normal to the z axis preserves the C_4 symmetry. We cut the system along[15] z, so that we shall need to inverse Fourier transform along z. Defining

$$\mathcal{H}_L(\mathbf{k}) = \mathcal{H}_1(\mathbf{k}_\perp) + t_z' e^{ik_z} \mathbb{1}_2 \implies \mathcal{H}_1(\mathbf{k}_\perp) = \left[t_1' + 2t_2'(\cos k_x + \cos k_y)\right] \mathbb{1}_2, \tag{7.55}$$

[15]This is opposed to the transfer matrix acting along x as in the previous sections, but it conforms to the notation in Ref. [112]

where $\mathbf{k}_\perp = (k_x, k_y)$, we identify

$$J = t'_z \begin{pmatrix} 0 & \mathbb{1}_2 \\ 0 & 0 \end{pmatrix}, \quad M = \begin{pmatrix} \mathcal{H}_0 & \mathcal{H}_1 \\ \mathcal{H}_1 & -\mathcal{H}_0 \end{pmatrix}. \tag{7.56}$$

In order to reduce the notational clutter, we set

$$\mathcal{H}_0 = a\,\mathbb{1}_2 + \mathbf{b}\cdot\boldsymbol{\sigma}, \quad \mathcal{H}_1 = m\mathbb{1}_2, \tag{7.57}$$

where we have defined

$$\begin{aligned} a &= t_1(\cos k_x + \cos k_y) + 2t_2 \cos k_x \cos k_y, \\ \mathbf{b} &= \begin{pmatrix} 2t_2 \sin k_x \sin k_y, & 0, & t_1(\cos k_x - \cos k_y) \end{pmatrix}, \\ m &= t'_1 + 2t'_2(\cos k_x + \cos k_y), \end{aligned} \tag{7.58}$$

and $b = |\mathbf{b}|$. We also normalize the parameters of the model so that $t'_z = 1$.

To compute the transfer matrix, we begin with the SVD of J as $J = V \cdot \Xi \cdot W^\dagger$, with

$$V = \begin{pmatrix} \mathbb{1}_2 \\ 0 \end{pmatrix}, \quad \Xi = \mathbb{1}_2, \quad W = \begin{pmatrix} 0 \\ \mathbb{1}_2 \end{pmatrix}. \tag{7.59}$$

The condition for the transfer matrix being complex-symplectic (eq. (6.68)) was that $[\mathcal{G}_{ab}, \Xi] = 0$, which is trivially true here. Furthermore, as M and J are both real, the transfer matrix will be real. Thus, we conclude that $T \in \mathrm{Sp}(4, \mathbb{R})$. Next, we need to compute

$$\mathcal{G} = \begin{pmatrix} (\varepsilon - a)\mathbb{1}_2 - \mathbf{b}\cdot\boldsymbol{\sigma} & -m\mathbb{1}_2 \\ -m\mathbb{1}_2 & (\varepsilon + a)\mathbb{1}_2 + \mathbf{b}\cdot\boldsymbol{\sigma} \end{pmatrix}^{-1} \equiv \begin{pmatrix} A & B \\ B & D \end{pmatrix}^{-1}, \tag{7.60}$$

where

$$A = (\varepsilon - a)\mathbb{1}_2 - \mathbf{b}\cdot\boldsymbol{\sigma}, \quad B = -m\mathbb{1}_2, \quad D = (\varepsilon + a)\mathbb{1}_2 + \mathbf{b}\cdot\boldsymbol{\sigma}. \tag{7.61}$$

As each block here is invertible for almost all ε, we use the eq. (B.8) from Appendix B.1 to get

$$\mathcal{G} = \begin{pmatrix} A^{-1} + A^{-1}B\mathcal{S}_{11}^{-1}BA^{-1} & -A^{-1}B\mathcal{S}_{11}^{-1} \\ -\mathcal{S}_{11}^{-1}BA^{-1} & \mathcal{S}_{11}^{-1} \end{pmatrix}, \tag{7.62}$$

where $\mathcal{S}_{11} = \mathcal{G}^{-1}/A = D - BA^{-1}B$. The computation of $\mathcal{G}_{ab},\ a, b \in \{v, w\}$ for the definition of V and W of

eq. (7.59) is simply extracting the correct submatrices, so that

$$\begin{pmatrix} \mathcal{G}_{vv} & \mathcal{G}_{wv} \\ \mathcal{G}_{vw} & \mathcal{G}_{ww} \end{pmatrix} = \begin{pmatrix} A^{-1} + A^{-1}B\mathcal{S}_{11}^{-1}BA^{-1} & -A^{-1}B\mathcal{S}_{11}^{-1} \\ -\mathcal{S}_{11}^{-1}BA^{-1} & \mathcal{S}_{11}^{-1} \end{pmatrix}. \tag{7.63}$$

Setting $\Xi = \mathbb{1}$ in our explicit expression for the generalized transfer matrix (eq. (6.62)), we need to compute

$$T = \begin{pmatrix} \mathcal{G}_{vw}^{-1} & -\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww} \\ \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1} & \mathcal{G}_{wv} - \mathcal{G}_{vv}\,\mathcal{G}_{vw}^{-1}\,\mathcal{G}_{ww} \end{pmatrix}. \tag{7.64}$$

Substituting $\mathcal{G}_{ab}$ from eq. (7.63), we get

$$\begin{aligned} T &= \begin{pmatrix} -AB^{-1}\mathcal{S}_{11} & \left(AB^{-1}\mathcal{S}_{11}\right)\mathcal{S}_{11}^{-1} \\ -\left(A^{-1} + A^{-1}B\mathcal{S}_{11}^{-1}BA^{-1}\right)\left(AB^{-1}\mathcal{S}_{11}\right) & -A^{-1}B\mathcal{S}_{11}^{-1} + \left(A^{-1} + A^{-1}B\mathcal{S}_{11}^{-1}BA^{-1}\right)\left(AB^{-1}\mathcal{S}_{11}\right)\mathcal{S}_{11}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} -AB^{-1}\mathcal{S}_{11} & AB^{-1} \\ -\left(B^{-1}\mathcal{S}_{11} + A^{-1}B\right) & B^{-1} \end{pmatrix} = \begin{pmatrix} -AB^{-1}\left(D - BA^{-1}B\right) & AB^{-1} \\ -B^{-1}\left(D - BA^{-1}B\right) - A^{-1}B & B^{-1} \end{pmatrix} \\ &= \begin{pmatrix} B - AB^{-1}D & AB^{-1} \\ -B^{-1}D & B^{-1} \end{pmatrix}. \end{aligned} \tag{7.65}$$

Finally substituting the blocks from eq. (7.61), we get

$$T = \frac{1}{m}\begin{pmatrix} \eta^2\mathbb{1}_2 - 2a\mathbf{b}\cdot\boldsymbol{\sigma} & (a-\varepsilon)\mathbb{1}_2 + \mathbf{b}\cdot\boldsymbol{\sigma} \\ (a+\varepsilon)\mathbb{1}_2 + \mathbf{b}\cdot\boldsymbol{\sigma} & -\mathbb{1}_2 \end{pmatrix}, \tag{7.66}$$

where $\eta^2 = \varepsilon^2 - a^2 - b^2 - m^2$. Thus,

$$\begin{aligned} \operatorname{tr}\{T\} &= \frac{1}{m}\left[\operatorname{tr}\left\{\eta^2\mathbb{1}_2 - 2a\mathbf{b}\cdot\boldsymbol{\sigma}\right\} + \operatorname{tr}\{-\mathbb{1}_2\}\right] \\ &= \frac{2}{m}\left(\eta^2 - 1\right), \\ \operatorname{tr}\left\{T^2\right\} &= \frac{1}{m^2}\left[\operatorname{tr}\left\{(\eta^2\mathbb{1}_2 - 2a\mathbf{b}\cdot\boldsymbol{\sigma})^2\right\} + 2\operatorname{tr}\{((a-\varepsilon)\mathbb{1}_2 + \mathbf{b}\cdot\boldsymbol{\sigma})((a+\varepsilon)\mathbb{1}_2 + \mathbf{b}\cdot\boldsymbol{\sigma})\} + \operatorname{tr}\left\{(-\mathbb{1}_2)^2\right\}\right] \\ &= \frac{2}{m^2}\left[\left(\eta^2 - 1\right)^2 + 4a^2b^2\right] - 4. \end{aligned} \tag{7.67}$$

Since T is symplectic, we deduce that the bulk bands are completely determined by the Floquet discriminants. Using eq. (6.87) for $r = 2$, the two Floquet discriminants Δ_μ, $\mu = \pm 1$ can be explicitly computed as

$$\begin{aligned} \Delta_\mu &= \frac{1}{2}\left[\operatorname{tr}T + \mu\sqrt{2\operatorname{tr}T^2 - (\operatorname{tr}T)^2 + 8}\right] = \frac{1}{2}\left[\frac{2}{m}\left(\eta^2 - 1\right) \pm \mu\sqrt{\frac{16a^2b^2}{m^2}}\right] \\ &= \frac{1}{m}\left[\varepsilon^2 - a^2 - b^2 - m^2 - 1 + 2\mu ab\right] = \frac{1}{m}\left[\varepsilon^2 - m^2 - 1 - (a + \mu b)^2\right], \end{aligned} \tag{7.68}$$

in terms of which the bulk energy spectrum becomes

$$\sigma\left[T\right]=\frac{1}{2}\left(\Delta_{\mu}\pm\sqrt{\Delta_{\mu}^{2}-4}\right),\quad\mu=\pm1. \tag{7.69}$$

The band edges are given by $|\Delta_{\mu}(\varepsilon,\mathbf{k}_{\perp})|=2$. Setting $\Delta_{\mu}=2\lambda$, $\lambda=\pm1$, we get

$$\varepsilon^{2}-m^{2}-1-(a+\mu b)^{2}=2\lambda\implies\varepsilon_{\text{bulk}}=\pm\sqrt{(m+\lambda)^{2}+(a+\mu b)^{2}}. \tag{7.70}$$

Clearly, we have 8 eigenvalues, and the spectrum is symmetric about $\varepsilon=0$.

For the edge states, given $\Phi=(\boldsymbol{\beta},0)^{T}$, which satisfies the boundary conditions for the left edge, we demand that $T\Phi$ is in the same subspace as Φ, spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$. But

$$T\left(\begin{array}{c}\boldsymbol{\beta}\\0\end{array}\right)=\left(\begin{array}{c}\left(\eta^{2}\mathbb{1}_{2}-2a\mathbf{b}\cdot\boldsymbol{\sigma}\right)\boldsymbol{\beta}\\\left((a+\varepsilon)\mathbb{1}_{2}+\mathbf{b}\cdot\boldsymbol{\sigma}\right)\boldsymbol{\beta}\end{array}\right). \tag{7.71}$$

Thus, for Φ to be a left edge state, we demand that

$$\left((a+\varepsilon)\mathbb{1}_{2}+\mathbf{b}\cdot\boldsymbol{\sigma}\right)\boldsymbol{\beta}=0, \tag{7.72}$$

From Cramers' rule, we get a nontrivial solution for $\boldsymbol{\beta}$ iff the matrix is singular, i.e, iff

$$\det\left[(a+\varepsilon_{L})\mathbb{1}_{2}+\mathbf{b}\cdot\boldsymbol{\sigma}\right]=(a+\varepsilon_{L})^{2}-b^{2}=0. \tag{7.73}$$

A similar analysis for the right edge state leads to the condition

$$\det\left[(a-\varepsilon_{R})\mathbb{1}_{2}+\mathbf{b}\cdot\boldsymbol{\sigma}\right]=(a-\varepsilon_{R})^{2}-b^{2}=0. \tag{7.74}$$

The left and right edge spectra are given by

$$\varepsilon_{L}=-a\pm b,\quad\varepsilon_{R}=a\pm b. \tag{7.75}$$

Thus, exploiting the symplectic structure of the transfer matrix, we have analytically obtained explicit expressions for the boundaries of the bulk bands and the edge spectra as a function of parameters a, b and m, as defined in eq. (7.58), which themselves are functions of $\mathbf{k}_{\perp}$. We plot the analytically computed spectra alongside the spectra computed using exact diagonalization in Fig 7.8.

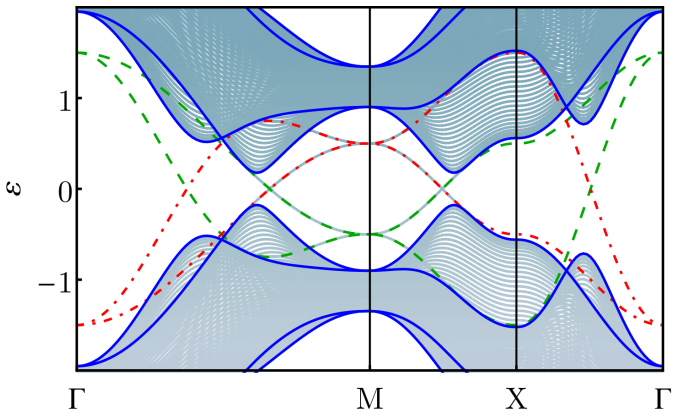


Figure 7.8: The spectrum of the topological crystalline insulator model due Fu[112], with the parameters $t_1 = 0.5$, $t_2 = 0.25$, $t_1' = 1.25$, $t_2' = 0.25$ and $t_z' = 1$ in eq. (7.53). The band edges (dark blue) and the left and right edge state dispersion (dashed and dashed-dot) computed using the transfer matrix formalism, overlaid on the spectrum computed using exact diagonalization equivalent to Fig, 2(b) of Ref. [112].

We notice that this model exhibits partial gaps (defined in Sec 6.4.1), which we expected to generically be the case for $r > 1$. These corresponds to the $(\varepsilon, \mathbf{k}_\perp)$ values where one pair of eigenvalues of the transfer matrix lie on the unit circle, while the other pair lies off it. The edge states always touch one of the band edges, but sometimes they can mean the edge to a partial gap, so that for a given $(\varepsilon, \mathbf{k}_\perp)$, there is an edge state as well as a bulk band state. This is clearly seen in Fig 7.8, for instance, between M and X points.

A striking feature of the band structure of TCI's is the existence of a quadratic band touching at the surface near the high symmetry point. In the case at hand, this occurs at the $\widetilde{M}$ point, i.e, the projection of the M point of the 3D Brillouin zone on a constant k_z plane. This corresponds to the transverse quasimomentum $\mathbf{k}_M = (\pi, \pi)$. We can see the band touching analytically by expanding the (left, say) edge spectrum in the vicinity of this point as $\mathbf{k}_\perp = \mathbf{k}_M + \delta\mathbf{k}$ upto the second order in δk, to get

$$\varepsilon_L \approx -2(t_2 - t_1) - \frac{t_1 - 2t_2}{2}(\delta k)^2 \pm \frac{t_1}{2}\sqrt{(\delta k)^4 + 4\left[1 - \left(\frac{2t_2}{t_1}\right)^2\right]\delta k_x^2 \delta k_y^2}\,. \tag{7.76}$$

For $t_1 = 2t_2 = t$, we get a radially symmetric quadratic band touching, with the spectrum given by

$$\varepsilon_L \approx -t\left[1 \mp \frac{1}{2}(\delta k)^2\right]. \tag{7.77}$$

Furthermore, we can analytically track the lifting of degeneracy of the surface states at the high symmetry point M on addition of a C_4-breaking term. For instance, we can add a term $\delta\mathcal{H} = \mu\sigma_z \otimes \sigma_z$ to the Hamiltonian, leading to $\mathbf{b} \mapsto \mathbf{b} + (0, 0, \mu)$. Physically, this corresponds to breaking the degeneracy of the p_x

and p_y orbitals. Then, at the $\widetilde{M}$ point, the left edge spectrum becomes

$$a = 2(t_2 - t_1), \quad b = (0, 0, \mu) \implies \varepsilon_L = 2(t_1 - t_2) \pm \mu.$$

The gap is clearly proportional to μ, the strength of the C_4 breaking term.

Thus, our analytically computed closed form expressions for bulk and surface bands agree with the exact diagonalization results. Furthermore, these expressions let us analytically study the fine-tuned nature of this surface quadratic band touching *vis à vis* the C_4 crystal symmetry, as well as derive the coefficients of a $\mathbf{k} \cdot \mathbf{p}$ expansion around that point, which was guessed on symmetry grounds in Ref. [112].

7.4 Large r: Disordered Systems

We have so far discussed analytical computations using our generalized transfer matrix for $r = 1, 2$, where it was analytically tractable. However, the power of this method lies in its generality, and the fact that it provides us with an algorithm to compute arbitrary band structures which can be easily mechanized. In this section, we do precisely that in order to study disordered systems. In particular, we investigate metal-insulator transitions in disordered tight-binding models.

Transfer Matrix Setup

Consider a generic d-dimensional tight binding lattice model with q degrees of freedom per unit cell. The simplest model of a disorder is the diagonal(also known as on-site or Anderson type) disorder. This simply involves adding a random on-site potential, i.e, adding to our tight binding Hamiltonian a term of the form $\mathcal{H}_{\text{disorder}} = \sum_{\mathbf{n},\alpha} V_{\mathbf{n},\alpha} c^\dagger_{\mathbf{n},\alpha} c_{\mathbf{n},\alpha}$, where $\{V_{\mathbf{n},\alpha}\}$ are independent and identically distributed (iid) real random variables, chosen from some fixed probability distribution. The disorder potential explicitly breaks the translational symmetry, so that the transverse momentum, $\mathbf{k}_\perp$ is not a good quantum number anymore. Instead, we consider the system on a *strip geometry*, i.e, infinite along x and finite (with open or periodic boundary conditions) along all the transverse directions.

Thus, we write our system in the position basis and construct the supercells from the sites corresponding to a constant x. For instance, for a 2 dimensional strip of width L_y, with sites indexed by $m = 1, \dots L_y$ and internal degrees of freedom at each site by $\alpha = 1, \dots, q$, the Hamiltonian takes the form

$$\mathcal{H} = \mathcal{H}_0 + \sum_{n=1}^{\infty} \sum_{m=1}^{L_y} \sum_{\alpha=1}^{q} V_{nm\alpha} c^\dagger_{nm\alpha} c_{nm\alpha} \tag{7.78}$$

where $\mathcal{H}_0$ is the translation-invariant Hamiltonian for the clean system, and the iid real random variables $\{V_{nm\alpha}\}$ are taken from a uniform distribution of width W centered at 0. Our supercells now consist of the $\mathscr{N} = qL_y$ degrees of freedom. Using the method described in Sec 6.4, we can identify the hopping matrix J and the on-site matrix M_n, where only M depends on n as the disorder is diagonal. We can construct the transfer matrix as a function of n, i.e, $\Phi_{n+1} = T_n\Phi_n$, where T_n now depends on the disorder realization. For a system with N sites along the x axis, we define the total transfer matrix as the product $\mathbf{T}_N \equiv \prod_{n=1}^{N} T_n$.

To investigate the existence of topological edge states, we note that for a clean system in a strip geometry in 2-dimensions, there are edge states localized at $m = 1, L_y$ along the y axis and strongly delocalized('*metallic*') along x. Since these modes are topological in nature, they are not expected to localize in presence of weak disorder; however, for a strong enough disorder, we may have a percolation transition. Since the eigenvalues of T determine the growth rate of the corresponding wavefunctions along x, for a fully delocalized state the corresponding eigenvalue ρ of the total transfer matrix must lie on the unit circle[16]. Conventionally, one computes the Lyapunov exponent defined as $\lambda = \ln|\rho|$, which must vanish for a delocalized state. We next discuss the recipe to compute the Lyapunov exponents numerically.

Lyapunov Exponents and Localization Lengths

The conventional approaches[113, 114, 115] to studying bulk phases of disordered non-interacting models and their Anderson transitions rely on obtaining the smallest Lyapunov exponent (in magnitude), or equivalently, the longest localization length in the x direction for a fixed energy ε. When the Fermi energy is set to ε, a further finite size scaling analysis of the longest localization length in the transverse directions discriminates between conducting and insulating phases of the bulk. In order to observe the quasi–$(d-1)$ dimensional metallic edge modes in a d-dimensional disordered topological phase, it is desirable to compute the multiset of all[17] Lyapunov exponents, hereafter termed the *Lyapunov spectrum*.

For a clean system, the eigenvalues ρ_i of the transfer matrix determine the growth/decay rate of the corresponding eigenstates, so that we can identify the Lyapunov exponents, or alternatively, the inverse localization length, as $\lambda_i = 1/l_i = \ln|\rho_i|$. Alternatively, we can define $\mathbf{\Lambda} = (T^\dagger T)^{1/2}$ with eigenvalues $\Lambda_i = |\rho_i|$, so that $\lambda_i = \ln\Lambda_i$. For the disordered case, the transfer matrices depend on n, so that we define

$$\mathbf{\Lambda} = \lim_{N\to\infty}\left[\mathbf{T}_N^\dagger\mathbf{T}_N\right]^{\frac{1}{2N}}; \qquad \mathbf{T}_N \equiv \prod_{n=1}^{N} T_n. \tag{7.79}$$

[16] If the transfer matrices are regarded as the evolution map of a dynamical system in time N, then the 'metallic' edge states correspond to stable limit cycles as $N \to \infty$.

[17] This is in contrast to studying the onset of localization in a metal, where one seeks only the largest Lyapunov exponent.

The fact that such a finite valued matrix exists is guaranteed by Oseledec's theorem[116]. The Lyapunov exponents are again given by $\lambda_i = \ln \Lambda_i$, where $\Lambda_i \in \mathbb{R}$ are the eigenvalue of $\mathbf{\Lambda}$.

In principle, given the transfer matrix, one could directly compute the matrix product in eq. (7.79), and hence the Lyapunov exponents, as a function of N. However, in practice, such a numerical matrix multiplication and diagonalization is usually plagued by numerical rounding and overflow errors, associated with the finite precision of the floating point representation of real numbers. In order to circumvent these issues, we follow the method described in Ref [116].

The key idea is to perform a QR decomposition[115] after every step involving a matrix multiplication. Recall that the QR decomposition for a complex nonsingular square matrix M involves writing it as $M = QR$, where Q is unitary and R is upper triangular with real, positive diagonal entries, sorted in descending order. The actual matrix multiplication of T_n's is then reduced to multiplying a set of upper triangular matrices R_n's. Explicitly, we begin by performing a QR decomposition of the first transfer matrix in the sequence as $T_1 = Q_1 R_1$, and setting $T_2' = T_2 Q_1$. Iterating, we get

$$\begin{aligned}
\mathbf{T}_N &= \left(\prod_{n=3}^{N} T_n\right) T_2 T_1 = \left(\prod_{n=3}^{N} T_n\right) T_2 (Q_1 R_1) \\
&= \left(\prod_{n=3}^{N} T_n\right) T_2' R_1 = \left(\prod_{n=4}^{N} T_n\right) T_3 (Q_2 R_2) R_1 \\
&= \ldots = T_N' \prod_{m=1}^{N} R_m = Q_N \prod_{m=1}^{N} R_m,
\end{aligned} \tag{7.80}$$

where we have defined $T_{n+1}' \equiv T_n Q_n$ and carried out its QR decomposition as $T_{n+1}' = Q_{n+1} R_{n+1}$ at each iteration. As $Q^\dagger Q = \mathbb{1}$ and $R_m^\dagger R_m = S_m$ is diagonal with the diagonal entries $S_{m,ii} = (R_{m,ii})^2$, we get

$$\mathbf{\Lambda} = \left[\prod_{m=1}^{N} S_m\right]^{\frac{1}{2N}} = \operatorname{diag}\left\{\left(\prod_{m=1}^{N} R_{m,ii}\right)^{\frac{1}{N}}\right\}_i \tag{7.81}$$

Only the diagonal elements of R_m are needed at each iteration, so that the computation of each eigenvalue of $\mathbf{\Lambda}$ now involves $O(N)$ multiplications, as opposed to $O(N^2)$ for a direct multiplication, thereby dramatically reducing the accumulation of numerical error. As $N \to \infty$, the Lyapunov exponents converge to

$$\lambda_i = \lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \ln[(R_m)_{ii}]. \tag{7.82}$$

Convergence to the true Lyapunov exponents can also be ascertained by studying the statistical fluctuations of the average on the right hand side of eq. (7.82).

7.4.1 Disordered Chern Insulator

We now specialize to the case of a Chern insulator with diagonal disorder. For a clean system, the on-site matrix M is given by

$$M = \frac{1}{2}\sum_{m=1}^{L_{\max}} \left(\mathbf{e}_{m+1}\cdot\mathbf{e}_m^\dagger\right)\otimes(i\sigma^y - \sigma^z) + \text{h.c.} + (2-m)\mathbb{1}_{L_y}\otimes\sigma^z. \tag{7.83}$$

Thus, in presence of a diagonal disorder V, the on-site Green's function becomes

$$\mathcal{G}_n = \left(\varepsilon\mathbb{1}_L\otimes\mathbb{1}_2 - M - V_n\right)^{-1}, \tag{7.84}$$

where $L_{\max} = L_y$ in the case of PBC and $L_{\max} = L_y - 1$ for the open boundary condition. It is worth remarking that for fixed ε and M, $\mathcal{G}_n^{-1}$ is non-invertible only for a set of measure V_n realizations, i.e, almost everywhere, and so we shall side step questions of its singularity.

The inter-layer hopping matrix J remains unchanged for this ensemble of disorder and takes the form of a $2L_y\times 2L_y$ matrix

$$J = \frac{1}{2i}\mathbb{1}_{L_y}\otimes(\sigma^x - i\sigma^z), \tag{7.85}$$

which, however, remains singular, with rank $r = L_y$. This conforms with the expectation of L_y independent channels in the non-disordered limit, which are explicitly coupled by disorder. The SVD for J remains virtually unchanged:

$$J = \mathbb{1}_{L_y}\otimes(\mathbf{v}\cdot\mathbf{w}^\dagger) = \sum_{y=1}^{L_y}\mathbf{V}_y\cdot\mathbf{W}_y^\dagger, \tag{7.86}$$

with $\mathbf{v}$ and $\mathbf{w}$ defined as in eq. (7.25), and we have defined the channels $\mathbf{V}_y := \mathbf{e}_y\otimes\mathbf{v}$ and $\mathbf{W}_y := \mathbf{e}_y\otimes\mathbf{w}$; $\{\mathbf{e}_y\}_{y=1}^{L_y}$ being the standard basis of $\mathbb{C}^{L_y}$. For each n, the transfer matrix T_n can now be numerically computed using eq. (6.62), which can be used to further compute the Lyapunov exponents using eq. (7.82). Since $\Xi = \mathbb{1}_{L_y}$, the transfer matrix is symplectic (eq. (6.68)), so that the eigenvalues occur in reciprocal pairs and the Lyapunov spectrum is symmetric about zero. We seek our edge mode, localized along y but completely delocalized along x, corresponding to a zero Lyapunov exponent (within numerical error).

In Fig 7.9, we show the Lyapunov spectrum as a function of N, the number of layers along x, for the Chern insulator in the topological phase, with $m = 1$ and strip width $L_y = 40$. We limit ourselves to the energy $\varepsilon = 0$, corresponding to the center of the band gap. For a weak disorder ($W = 0.1$) and open boundary

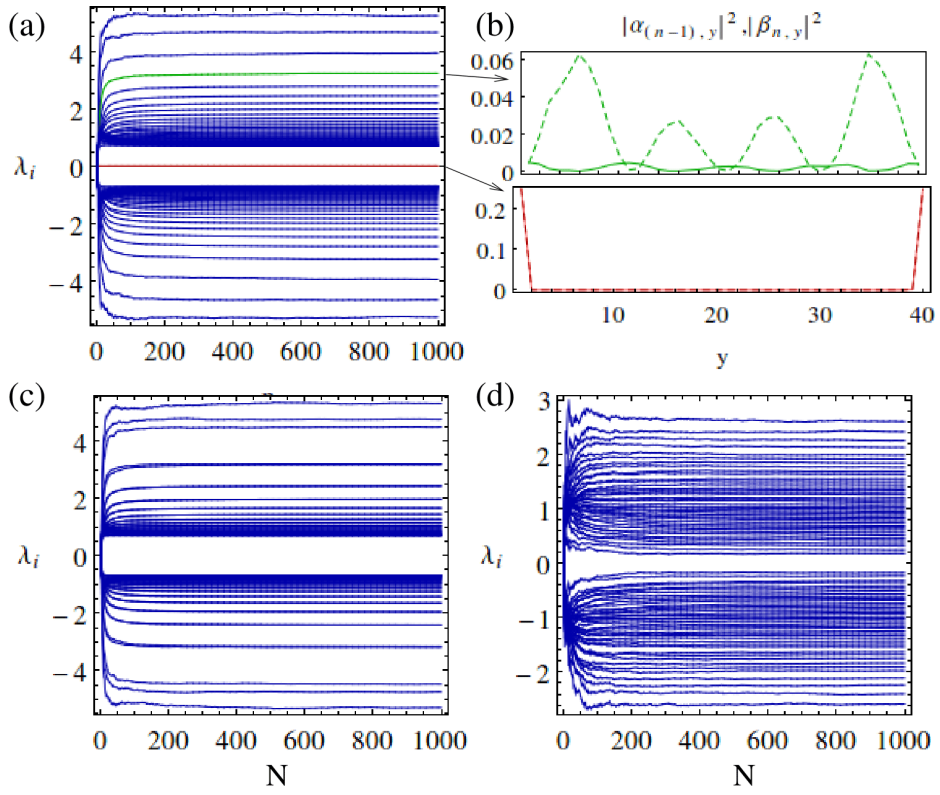


Figure 7.9: Numerically computed estimates of Lyapunov exponents as a function of system length N for Chern insulator on a strip geometry with width $L_y = 40$ and parameters $m = 1.0$, $\varepsilon = 0$. For large N, the estimates converge to the Lyapunov exponents $\{\lambda_i\}$. (a) $W = 0.1$ and OBC along y shows robust metallic edge modes (in red) with $\lambda_i = 0$ in this scale. Also highlighted is an insulating bulk mode (green trace) with $\lambda_i \approx 3$. (b) Spatial profile of eigenstates along y at $N = 10^3$ for modes with $\lambda \approx 3$(green) and $\lambda \approx 0$(red). The latter is strongly localized at $y = 0, L_y$. Arrows mark the position of these eigenmodes in the Lyapunov spectrum. (c) The same system as in fig (a) with PBC along y, which shows no metallic edge states. (d) Strongly disordered case ($W = 5.0$) with open boundaries and absent metallic states.

conditions along y, there are two quasi-1D metallic modes with $\lambda_i \approx 0$ at the center of the spectrum, highlighted in red in Fig 7.9(a). Numerically, the relevant exponents are never zero to machine precision, but are much smaller ($|\lambda_i| < 10^{-5}$ at $N = 10^4$) than the other Lyapunov exponents and systematically decrease (as a power law) with increasing N. To confirm the identification of these modes as topological edge states, we plot their spatial profile along y in Fig 7.9(b), which clearly shows localization at the edge, in contrast to an insulating localized mode with $\lambda_i \approx 3$. Furthermore, for the same parameters but with closed periodic boundaries, no metallic modes are observed, as shown in Fig 7.9(c). We note that tuning the mass parameter m to the topologically trivial range or moving ε into the center of the bulk band also removes these metallic modes.

Finally, for a strong enough disorder ($W = 5.0$), the metallic modes are also absent, as shown in Fig 7.9(d), which signals a *percolation transition*. We observe that the lifting of the metallic edge modes from

the asymptotic value $\lambda_i = 0$ occurs continuously with changes of tuning parameters, in agreement with the theory of continuous Anderson transitions[114]. However, further work is needed to verify that the scaling exponents $\{\nu_i\}$ corresponding to the divergence of the localization lengths $\{l_i\}$ at the metal-insulator transition agrees with the expectations for the Integer Quantum Hall transition[117, 118]. We leave such numerical investigations for future work.

8 Conclusions

In Part I of this thesis, we have generalized the chiral kinetic theory, originally proposed to compute U(1) anomaly in $3+1$ dimensions, to nonabelian anomalies in higher dimensions. There were two central ingredients in this generalization: the symplectic formulation of classical mechanics and the dequantization of nonabelian guage fields. The former is crucial to study the latter, since the coadjoint orbits obtained from the gauge group are compact and often support no global coordinates. The higher dimensional computation was also conceptually cleaner, once we have discarded the special features of the low dimensional dynamics.

The abelian chiral anomaly or the singlet anomaly for nonabelian gauge fields are a breakdown of an ordinary conservation law of the form $\partial_\mu J^\mu = 0$, which can be interpreted as a continuity equation. Thus, the anomaly can be interpreted as a breakdown of particle number conservation, so that chiral particles are being spontaneously created (out of a Dirac sea) under a *spectral flow*. This fact was used to motivate the computation of the breakdown of Liouville's theorem, a more general "continuity equation" in phase space. However, we also obtained the nonabelian gauge anomaly using the same formalism, which is a breakdown of a covariant conservation law $\mathcal{D}_\mu J^\mu_a = 0$ and thus does not conventionally have a spectral flow interpretation as an influx of discrete particles. It is interesting to ponder whether such a calculation points towards a possible spectral flow argument for the nonabelian anomalies.

Our computation of the anomalous contribution to the hydrodynamic currents from a classical kinetic theory approach can be thought of as a counterpoint to the usual anomalous hydrodynamic calculations. In deriving a low energy hydrodynamic description of a QFT with anomalous currents, one takes the anomalies as given and explores their consequences on the macroscopic dynamics, which one derives from thermodynamic constraints. Ideally, however, one would like to start with a microscopic field theory and systematically integrate out the high energy modes, thereby arriving at a macroscopic hydrodynamic description. Our computation, to some extent, does part of that for noninteracting Weyl fermions by a circuitous route, by "guessing" the semiclassical phase space which encodes the anomaly and describes the physics at microscopic scales, and then using classical kinetic theory to compute the macroscopic currents. As long as one can sensibly talk about discrete particle-like excitations in the underlying QFT, one can potentially improve this picture by including the collision terms, which would then lead to dissipative hydrodynamic effects.

To complete the picture, one would like to be able to also derive the gravitational contributions to the gauge anomaly using a chiral kinetic theory formalism, which would require dynamics on a curved spacetime manifold. Our work in that direction so far[3] (which did not make it to the final cut of this thesis) has been unsuccessful; however, we have discovered much interesting physics in the process. One basic issue is to derive a Lorentz covariant formulation of our theory, which can then be thought of as invariance under frame rotations when one puts the theory on a curved spacetime manifold. Since the Hamiltonian dynamics treats the space and time coordinates differently, the Lorentz invariance is nor manifest. Furthermore, one needs to define a generalization of the Berry curvature 2-form which depends on all components of the energy-momentum 4-vector instead of just the space parts. We have derived such a generalization in Ref [3].

A more serious problem is the frame-dependence of the "position" of the Weyl fermion! This sounds absurd for a classical particle, but we are essentially trying to define the dynamics of a wavepacket, whose centroid is the 'position' in question. In this picture, one way to think of the anomalous velocity, which leads to a sideways group velocity, is a result of the different Berry phases picked up by different momentum modes which are superimposed to define the wavepacket. This is, in essence, similar to the spreading of a localized mode in a dispersive medium.

The frame dependence of the 'position' can also be understood as an artifact of the nontrivial representation theory of the Lorentz group for massless spinning particles. To wit, a massive particle has a rest frame in which one can define the centroid, and the symmetry group in the rest frame (termed the *little group*) SO(3), which is compact and whose representations are used to label the *spin* of the particle, and the corresponding values in other frames can be obtained by Lorentz boosts. However, for massless particles, there is no rest frame and the little group is the noncompact group $E(2) \cong \mathbb{R}^2 \rtimes \mathrm{SO}(2)$, the symmetry group of $\mathbb{R}^2$, so that the *spin* (helicity, to be precise) of the particle is labeled by the representation of the compact subgroup, *viz*, SO(2). We are still left with a *gauge freedom* corresponding to $\mathbb{R}^2 \subset E(2)$, which leads to a sideways translation of vectors upon boosts, termed *Wigner translations*. This problem is worse in curved spacetime, when the gauge freedom from Wigner translations mix with the gauge freedom of the theory under local diffeomorphisms of the spacetime manifold.

Finally, our computation can be applied to the study of macroscopic effects driven by anomalies in condensed matter systems, for instance in Weyl semimetals[34]. The effect of the band structure can be encoded by using the band energy spectrum instead of $\varepsilon = |\mathbf{p}|$, while some of the interaction/scattering effects can be encoded in the collision integral of the Boltzmann equation. Like the conventional semiclassical dynamics used to study transport in metals, this provides a much more tractable approach to include real effects than using the complete quantum field theoretic description.

In Part II, we have essentially filled a gap in the conventional usage of transfer matrix for tight-binding models, by bringing the models with singular hopping matrices within its reach. By generalizing the transfer matrix construction, however, we have also discovered many other interesting algebraic and geometric aspects of the problem, some of which were originally discussed by Hatsugai for the special case of the Hofstadter model. The construction also formalizes (and essentially mechanizes) the application of transfer matrices to nonsquare lattice for translations along arbitrary directions.

The central philosophy behind transfer matrices is the encoding of all possible asymptotic behavior of the solution of the Schrödinger equation in one matrix, which can be computed by knowing a very small part (essentially, a single supercell and its connectivity to its nearest neighbors, in our picture) of the system. Thus, crudely speaking, even though the transfer matrix formalism is a real space setup, it "knows" everything about the momentum-space behavior of the system, which makes it ideally suited for studying bulk-boundary correspondence, where the bulk invariants are computed in a momentum space picture, while the boundary invariants are most naturally defined in real space. A concrete problem in this direction would be to compute the Berry curvature, and hence the Chern number, directly from a given transfer matrix. A similar situation is that of lattice defects, especially modes localized at dislocations, where one seeks to match various "momentum" modes for a system where the periodicity is broken by the lattice defect, so that the momentum eigenstates are, strictly speaking, not defined.

The rank of the hopping matrix which determines the size of the transfer matrix, alongwith the corresponding singular values, provide an interesting perspective into the system, as they can be interpreted as a decomposition of the system into channels, with their strengths given by the singular values. Such a perspective can be useful in studying the system as a set of quasi-1D chains, in terms of the *most important* degrees of freedom, with a generalization of the "unfolding" transform that we described for Chern insulator. A practical situation where such an analysis could be useful is systems with disorder in the hopping strengths (besides/alongwith the on-site/Anderson disorder), where the rank of the hopping matrix can potentially change for different layers, but constructing the transfer matrix using only a few largest singular values may provide the useful statistics for the spectrum.

We note that many of the convenient simplifications that arose for the $r = 1$ case are due to the fact that the transfer matrix is symplectic. There is a wealth of interesting properties associated with symplectic matrices that can be relevant to the study of topological phase. For instance, we used the fact that $\pi_1\left(\mathrm{Sp}(2,\mathbb{R})\right) \cong \mathbb{Z}$ to define a *Maslov index* associated with the edge states. However, the symplectic matrices have $\pi_1\left(\mathrm{Sp}(2n,\mathbb{R})\right) \cong \mathbb{Z}$ for all n, so that the corresponding Maslov indices can also be interpreted as topological invariants of the edge states. We leave such analyses for future work.

Appendix

A Lie groups and coadjoint orbits

A.1 Basic notions

We begin with two definitions:

Definition A.1. *A Lie group is a smooth manifold G endowed with a group multiplication law, i.e, a smooth map $G \times G \to G$ satisfying the group axioms.*

Definition A.2. *A Lie algebra is defined as a real vector space V equipped with a bilinear map $[,] : V \times V \to V$, which satisfies the following conditions*

1. *Antisymmetry:* $[X, X] = 0 \quad \forall\, X \in V$.
2. *Jacobi identity:* $[X, [Y, Z]] + [Y, [Z, X]] + [Z.[X, Y]] = 0 \quad \forall\, X, Y, Z \in V$.

From the antisymmetry and bilinearity, given $X, Y \in V$, it follows that $[X, X] = [Y, Y] = 0$, and

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

so that $[X, Y] = -[Y, X]$, the conventional definition of antisymmetry.

Given a Lie Group G, we can associate a unique (upto isomorphisms) Lie algebra, usually denoted by $\mathfrak{g}$. Formally, this association is functor from the category of Lie groups to the category of Lie algebras. The functor is surjective and many to one on objects, i.e, a given Lie algebra may correspond to more than one Lie group. For instance, $\mathrm{SO}(3) \ncong \mathrm{SU}(2)$ but $\mathfrak{so}(3) \cong \mathfrak{su}(2)$.

Define $\mathfrak{g} \equiv T_e G$, i.e, the tangent space of G at the identity element $e \in G$. Then, $\mathfrak{g}$ is a vector space isomorphic to $\mathbb{R}^n$, with an additional bilinear function $[,] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ arising from the Lie bracket of tangent vector. Thus, $\mathfrak{g}$ is an algebra, termed the *Lie algebra* of G. More formally, As a vector space, $\mathfrak{g} \cong \mathbb{R}^n$, where $n = \dim(G)$, i.e, an atlas on G maps maps open sets to open sets of $\mathbb{R}^n$.

We would primarily be interested in a *representation* of G, i.e,

Definition A.3. *A representation of a group G on a vector space V is a homomorphism $\pi : G \to \mathcal{O}(V)$,*

the set of linear operators on V. More explicitly, to each $g \in G$, the representation π associates a linear operator $\pi_g : V \to V$ such that $\pi_{g_1 g_2} = \pi_{g_1} \circ \pi_{g_2}$.

A.2 Adjoint and Coadjoint representations

A natural representation of G is induced on the vector space $\mathfrak{g}$, termed the *adjoint representation*. To see this, consider the set of inner automorphisms of G, $\varphi_g : G \to G$, defined by the conjugations $\varphi_g : h \mapsto ghg^{-1}$. We can lift this map to a map $\varphi_{g,*} : TG \to TG$ on the tangent space by push-forward, which acts by changing the tangent vector as well as the base point. Expliclty, we always have $\varphi_{g,*} : T_h G \to T_{ghg^{-1}} G$.

Clearly, $\varphi_g : e \mapsto e$, so that e is a fixed point of G, and $\varphi_{g,*} : T_e G \to T_e G$. Thus, we have a representation of G on $\mathfrak{g} = T_e G$, termed the *adjoint representation*.

Definition A.4. *The adjoint representation* $\mathrm{Ad}\colon G \to \mathcal{O}(\mathfrak{g})$ *of a Lie group G on its Lie algebra $\mathfrak{g}$ is defined as the derivative of the adjoint action $h \mapsto ghg^{-1}$ at the identity. Explicitly,*

$$\mathrm{Ad}(g)\colon X \mapsto$$

The derivation of the adjoint representation induces an action of $\mathfrak{g}$ by itself, $\mathrm{ad}\colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$*, explicitly defined as*

$$\mathrm{ad}(Y)\colon X \mapsto \frac{d}{dt}\Big(\mathrm{Ad}\big(e^{tY}\big)\, X\Big)\Big|_{t=0} = [Y, X].$$

Thus, geometrically, $\mathrm{ad}(Y)X \in T_X \mathfrak{g} \cong \mathfrak{g}$.

Given the vector space $\mathfrak{g}$, we can construct its dual space $\mathfrak{g}^*$ as the set of linear functions $f : \mathfrak{g} \to \mathbb{R}$. By definition, we have a bilinear map $(\ ,\ )\colon \mathfrak{g}^* \times \mathfrak{g} \to \mathbb{R}$, defined as $(f, x) = f(x)\ \forall\, x \in \mathfrak{g}, f \in \mathfrak{g}^*$. Any linear operator $L : \mathfrak{g} \to \mathfrak{g}$ has an adjoint operator $L^* : \mathfrak{g}^* \to \mathfrak{g}^*$, defined by $(f, Lx) = (L^* f, x)$.

Definition A.5. *The coadjoint representation* $\mathrm{K}\colon G \times \mathfrak{g}^* \to \mathfrak{g}^*$ *is defined as* $\mathrm{K}(g) = \mathrm{Ad}(g^{-1})^*\colon \mathfrak{g}^* \to \mathfrak{g}^*$. *This induces the coadjoint action of $\mathfrak{g}$ on $\mathfrak{g}^*$, defined as*

$$\mathrm{k}(Y)\colon F \mapsto \frac{d}{dt}\Big(\mathrm{K}\big(e^{tY}\big)\, X\Big)\Big|_{t=0} = -\mathrm{ad}(Y)^* F.$$

Thus, geometrically, $\mathrm{k}(Y)X \in T_F \mathfrak{g}^* \cong \mathfrak{g}^*$.

Here, $\mathrm{K}(g)$ forms a representation of G since

$$\begin{aligned}\Big(\mathrm{Ad}(g_1g_2)^*F, X\Big) &= \Big(F, \mathrm{Ad}(g_1g_2)X\Big) = \Big(F, \mathrm{Ad}(g_1)\mathrm{Ad}(g_2)X\Big)\\ &= \Big(\mathrm{Ad}(g_1)^*F, \mathrm{Ad}(g_2)X\Big) = \Big(\mathrm{Ad}(g_2)^*\mathrm{Ad}(g_1)^*F, X\Big),\end{aligned}$$

so that

$$\mathrm{K}(g_1g_2) = \mathrm{Ad}\left(g_2^{-1}g_1^{-1}\right)^* = \mathrm{Ad}\left(g_1^{-1}\right)^* \mathrm{Ad}\left(g_2^{-1}\right)^* = \mathrm{K}(g_1)\mathrm{K}(g_2). \tag{A.1}$$

A.3 Symplectic structure on codjoint orbits

We define the coadjoint orbits as:

Definition A.6. *The coadjoint orbit of a fixed $F \in \mathfrak{g}^*$ is defined as $\mathcal{O}_F = \{\mathrm{K}(g)F \,|\, g \in G) \subseteq \mathfrak{g}^*$.*

Let $\mathrm{Stab}(F) = \{g \in G \,|\, \mathrm{K}(g)F = F\} \subseteq G$ be the stabilizer of F, which forms a subgroup of G since $\mathrm{K}(g)$ forms a representation. Then, the coadjoint orbit is simply the quotient $\mathcal{O}_F \cong G/\mathrm{Stab}(F)$, and G can be thought of as a fiber bundle over the base space $\mathcal{O}_X$ with fibers $\mathrm{Stab}(F)$, and the projection $\pi_F \colon G \to \mathcal{O}_F$ defined as $\pi_F(g) = \mathrm{K}(g)F$. Next, Let $\mathrm{stab}(F) \subseteq \mathfrak{g}$ be the Lie algebra of $\mathrm{Stab}(F)$, which forms a subalgebra of $\mathfrak{g}^*$. The projection π_F induces $\pi_{F*} \colon \mathfrak{g} \to T_F\mathcal{O}_F$, defined as the derivative of π_F, i.e, $\pi_{F*} \colon X \mapsto \mathrm{k}(X)F$. Then, we have a short exact sequence

$$0 \longrightarrow \mathrm{stab}(F) \hookrightarrow \mathfrak{g} \to T_F\mathcal{O}_F \longrightarrow 0, \tag{A.2}$$

so that $T_F\mathcal{O}_F \cong \mathfrak{g}/\mathrm{stab}(F)$, i.e, all vectors on $\mathcal{O}_F$ can be defined as $\xi_X \equiv \mathrm{k}(X)$ for some $X \in \mathfrak{g}$. These vectors act as derivatives on the functions $\mathcal{F} \colon \mathcal{O}_F \to \mathbb{R}$ as

$$\xi_X\mathcal{F}(F) = \frac{d}{dt}\mathcal{F}\Big(\mathrm{K}\left(e^{tX}\right)F\Big)\Big|_{t=0} = \mathcal{F}\left(\mathrm{k}(X)F\right). \tag{A.3}$$

The symplectic structure on $\mathcal{O}_F$ is then defined by the following theorem:

Theorem A.1. *The coadjoint orbit $\mathcal{O}_F$ is a symplectic manifold, with the symplectic form defined as*

$$\rho(\xi_X, \xi_Y)\Big|_F = \Big(F, [X, Y]\Big). \tag{A.4}$$

Proof. In order to prove that ρ defines a symplectic structure on $\mathcal{O}_F$, we need to show that it is closed and

nondegenerate. We shall need

$$\xi_X \rho(\xi_Y, \xi_Z) = \xi_X \Big(F, [Y, Z]\Big) = \Big(\mathrm{k}(X)F, [Y, Z]\Big) = -\Big(F, \mathrm{ad}(X)[Y, Z]\Big) = -\Big(F, [X, [Y, Z]]\Big), \tag{A.5}$$

where we have used eq. (A.3). Thus, using Cartan's relations and $[\xi_X, \xi_Y] = -\xi_{[X,Y]}$), we can compute

$$\begin{aligned}
&d\rho(\xi_X, \xi_Y, \xi_Z) \\
&= \xi_X \rho(\xi_Y, \xi_Z) + \xi_Y \rho(\xi_Z, \xi_X) + \xi_Z \rho(\xi_X, \xi_Y) - \rho([\xi_X, \xi_Y], \xi_Z) - \rho([\xi_Y, \xi_Z], \xi_X) - \rho([\xi_Z, \xi_X], \xi_Y) \\
&= -\Big(F, [X, [Y, Z]]\Big) - \Big(F, [Y, [Z, X]]\Big) - \Big(F, [Z, [X, Y]]\Big) + \Big(F, [[X, Y], Z]\Big) + \Big(F, [[Y, Z], X]\Big) + \Big(F, [[Z, X], Y]\Big) \\
&= -\Big(F, [X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] - [[X, Y], Z] - [[Y, Z], X] - [[Z, X], Y]\Big) \\
&= -2\Big(F, [X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]]\Big) = 0,
\end{aligned} \tag{A.6}$$

where we have used Jacobi's identity in the last step. To show that ρ is nondegenerate, assume that $\rho(\xi_X, \xi_Y) = 0,\ \forall\, Y \in \mathfrak{g}$ for a fixed $X \in \mathfrak{g}$. Then, $\forall\, F \in \mathfrak{g}^*$,

$$0 = \rho(\xi_X, \xi_Y)(F) = \Big(F, [X, Y]\Big) = \Big(F, \mathrm{ad}(X)Y\Big) = \Big(\mathrm{k}(X)F, Y\Big), \tag{A.7}$$

so that $0 = \mathrm{k}(X) = \xi_X$, which completes our proof. □

Given this symplectic structure, we can define Hamiltonian flows on $\mathcal{O}_F$ corresponding to a Hamiltonian $\mathcal{H}\colon \mathcal{O}_F \to \mathbb{R}$. The symplectic form associates with $\mathcal{H}$ a Hamiltonian vector field $\mathfrak{d}\mathcal{H}$ via

$$i_{\mathfrak{d}\mathcal{H}}\rho = -d\mathcal{H} \implies \rho(\mathfrak{d}\mathcal{H}, \xi_X) = -\xi_X \mathcal{H}\ \ \forall\, X \in \mathfrak{g}. \tag{A.8}$$

We next prove the following lemma

Lemma A.1. *Given a Hamiltonian defined as $\mathcal{H}\colon F \mapsto (F, Z)$ for a fixed $Z \in \mathfrak{g}$, the corresponding Hamiltonian vector field is given simply by $\mathfrak{d}\mathcal{H} = \xi_Z$.*

Proof. We shall need to show that $\forall\, X \in \mathfrak{g}$,

$$\rho(\xi_Z, \xi_X)\Big|_{F'} = -\xi_X \mathcal{H}(F') \quad \forall\, F' \in \mathcal{O}_F. \tag{A.9}$$

The LHS is simply $\Big(F', [Z, X]\Big)$, while to evaluate the RHS, we use eq. (A.3) to compute the RHS as

$$-\xi_X \mathcal{H}(F') = -\mathcal{H}\left(\mathrm{k}(X)F'\right) = -\Big(\mathrm{k}(X)F', Z\Big) = -\Big(F', \mathrm{ad}(X)Z\Big) = \Big(F', [Z, X]\Big), \tag{A.10}$$

which completes the proof. □

A.4 Coadjoint orbits of SU(2)

The matrix group SU(2) is defined as

$$\mathrm{SU}(2) = \left\{M \in \mathrm{Mat}(2, \mathbb{C}) \,|\, M^\dagger = M^{-1},\ \det M = 1\right\}. \tag{A.11}$$

Parametrizing M, the constraints imply that

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SU}(2) \implies \delta = \alpha^*,\ \ \gamma = \beta^*,\ \ \alpha\delta - \beta\gamma = 1, \tag{A.12}$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{C}$. Thus, we set $\alpha = x_0 + ix_3$, $\beta = x_2 + ix_1$ to parametrize M using $x \in \mathbb{R}^4$'s using Pauli matrices as

$$M = \begin{pmatrix} x_0 + ix_3 & x_2 + ix_1 \\ -x_2 + ix_1 & x_0 - ix_3 \end{pmatrix} = x_0 \mathbb{1} + ix_i\sigma^i, \quad i = 1, 2, 3, \tag{A.13}$$

subject to the constraint $\sum_{\mu=0}^{3} x_\mu^2 = 1$. This is the equation of a 3-sphere in $\mathbb{R}^4$, so that as smooth manifolds, $\mathrm{SU}(2) \cong S^3$.

The Lie algebra is simply defined as $\mathfrak{su}(2) = \{X_i\sigma^i \,|\, X^i \in \mathbb{R}\}$. This can be derived either geometrically, by considering the tangent space of S^3 at the "north pole" corresponding to $x = (1, 0, 0, 0)$, or alternatively, by expanding $M(x)$ in the vicinity of $\mathbb{1}$. In the latter approach, we set $x_i = \epsilon X_i$, so that

$$M = \sqrt{1 - \epsilon^2|\mathbf{X}|^2}\mathbb{1} + \epsilon\, iX_i\sigma^i = \mathbb{1} + \epsilon\, iX_i\sigma^i + O(\epsilon^2), \tag{A.14}$$

and consider the terms at linear order. We define a basis for $\mathfrak{su}(2)$ as $\lambda_i = \frac{1}{2}\sigma_i$, so that

$$[\lambda^i, \lambda^j] = \frac{1}{4}[\sigma_i, \sigma_j] = i\epsilon^{ijk}\lambda^k, \qquad \langle\lambda^i, \lambda^j\rangle = \frac{1}{4}\mathrm{tr}\left\{\sigma_i\sigma_j\right\} = \frac{1}{2}\delta^{ij}. \tag{A.15}$$

We define

$$\mathrm{SU}(2) = \left\{ g_0 \mathbb{1} + i\, \mathbf{g}\cdot\boldsymbol{\sigma} \;\middle|\; g_0 \in \mathbb{R},\, \mathbf{g}\in\mathbb{R}^3,\, \sum_{\mu=0}^{3} g_\mu^2 = 1 \right\} \tag{A.16}$$

with its Lie algebra $\mathfrak{su}(2) = \{\mathbf{X}\cdot\boldsymbol{\sigma},\ \mathbf{X}\in\mathbb{R}^3\}$, and consider the coadjoint action of the Lie algebra on the Lie group, defined as $\mathrm{K}(g)\colon X \mapsto g^{-1}\, X\, g$. To egplicitly compute this product, we shall need some Pauli matrig identities. Recall that Pauli matrices are all Hermitian and traceless, and satisfy

$$[\sigma^i,\sigma^j]_+ = 2\delta^{ij}\mathbb{1}, \qquad [\sigma^i,\sigma^j]_- = 2i\,\epsilon^{ijk}\sigma^k. \tag{A.17}$$

Combining these two, we get

$$\sigma^i\sigma^j = \frac{1}{2}\left([\sigma^i,\sigma^j]_+ + [\sigma^i,\sigma^j]_-\right) = \delta^{ij}\,\mathbb{1} + i\epsilon^{ijk}\sigma^k. \tag{A.18}$$

Iterating and using the contraction for the Levi-Civita symbols, we also get

$$\begin{aligned}
\sigma^i\sigma^j\sigma^k &= \left(\delta^{ij}\,\mathbb{1} + i\epsilon^{ij\ell}\sigma^\ell\right)\sigma^k \\
&= \delta^{ij}\sigma^k + i\epsilon^{ij\ell}\left(\delta^{\ell k}\mathbb{1} + i\epsilon^{\ell km}\sigma^m\right) \\
&= \delta^{ij}\sigma^k + i\epsilon^{ijk}\mathbb{1} - \epsilon^{\ell ij}\epsilon^{\ell km}\sigma^m \\
&= \delta^{ij}\sigma^k + i\epsilon^{ijk}\mathbb{1} - \left(\delta^{ik}\delta^{jm} - \delta^{im}\delta^{jk}\right)\sigma^m \\
&= i\epsilon^{ijk}\mathbb{1} + \delta^{jk}\sigma^i - \delta^{ik}\sigma^j + \delta^{ij}\sigma^k.
\end{aligned} \tag{A.19}$$

The coadjoint action on $X = X_i\sigma^i$ can then be explicitly computed as

$$\begin{aligned}
g^{-1}Xg &= \left(g_0\mathbb{1} - ig_i\sigma^i\right)\left(-\frac{i}{2}X_i\sigma^i\right)\left(g_0\mathbb{1} + ig_i\sigma^i\right) \\
&= \frac{1}{2}\left[-ig_0^2X_i\sigma^i - \left(g_iX_jg_0 - g_0X_ig_j\right)\sigma^i\sigma^j - ig_iX_jg_k\sigma^i\sigma^j\sigma^k\right] \\
&= \frac{1}{2}\left[-ig_0^2X_i\sigma^i - g_0\left(g_iX_j - g_jX_i\right)\left(\delta^{ij}\mathbb{1} + i\epsilon^{ijk}\sigma^k\right) - ig_iX_jg_k\left(i\epsilon^{ijk}\mathbb{1} + \delta^{jk}\sigma^i - \delta^{ik}\sigma^j + \delta^{ij}\sigma^k\right)\right] \\
&= \frac{1}{2}\left[-ig_0^2X_i\sigma^i - 2ig_0\epsilon^{ijk}g_iX_j\sigma^k - 2i(g_jX^j)g_i\sigma^i + i(g_jg^j)X_i\sigma^i\right] \\
&= -\frac{i}{2}\left[(2g_0^2-1)X_i + 2(g_jX_j)g_i + 2g_0\epsilon^{ijk}g_jX_k\right]\sigma^i.
\end{aligned} \tag{A.20}$$

Thus, on our coordinates, the coadjoint action sets

$$\mathrm{K}(g)\colon X_i \mapsto (2g_0^2-1)X_i + 2(g_jX_j)g_i + 2g_0\,\epsilon^{ijk}g_jX_k. \tag{A.21}$$

Defining vectors $\mathbf{g} = \{g_i\}$ and $\mathbf{X} = \{X_i\}$ on $\mathbb{R}^3$, this becomes

$$\mathrm{K}(g)\colon \mathbf{X} \mapsto \mathbf{X}^{\mathbf{g}} = \left(1 - 2|\mathbf{g}|^2\right)\mathbf{X} + 2\left(\mathbf{g}\cdot\mathbf{X}\right)\mathbf{g} \pm 2\,\mathbf{g}\times\mathbf{X}\sqrt{1-|\mathbf{g}|^2}. \tag{A.22}$$

For a fixed $\mathbf{X}$, the coadjoint orbit is simply the span of this map for $\mathbf{g} \in \mathbb{R}^3$, which can be shown to be a 2-sphere by simply computing

$$\begin{aligned}
|\mathbf{X}^{\mathbf{g}}|^2 =& \left(1-2|\mathbf{g}|^2\right)^2|\mathbf{X}|^2 + 4\left(\mathbf{g}\cdot\mathbf{X}\right)^2|\mathbf{g}|^2 + 4\left(1-2|\mathbf{g}|^2\right)\left(\mathbf{g}\cdot\mathbf{X}\right)^2 + 4\left(1-|\mathbf{g}|^2\right)|\mathbf{g}\times\mathbf{X}|^2 \\
=& \left(1-2|\mathbf{g}|^2\right)^2|\mathbf{X}|^2 + 4\left(1-|\mathbf{g}|^2\right)\left(\mathbf{g}\cdot\mathbf{X}\right)^2 + 4\left(1-|\mathbf{g}|^2\right)\left(|\mathbf{g}|^2|\mathbf{X}|^2 - \left(\mathbf{g}\cdot\mathbf{X}\right)^2\right) \\
=& \left[\left(1-2|\mathbf{g}|^2\right)^2 + 4|\mathbf{g}|^2\left(1-|\mathbf{g}|^2\right)\right]|\mathbf{X}|^2 = |\mathbf{X}|^2 .
\end{aligned} \tag{A.23}$$

where we have used $\mathbf{X}\cdot(\mathbf{g}\times\mathbf{X}) = \mathbf{g}\cdot(\mathbf{g}\times\mathbf{X}) = 0$, and

$$|\mathbf{g}\times\mathbf{X}|^2 = \epsilon_{ijk}g_jX_k\epsilon_{imn}g_mX_n = \left(\delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}\right)g_jX_kg_mX_n = |\mathbf{g}|^2|\mathbf{X}|^2 - \left(\mathbf{g}\cdot\mathbf{X}\right)^2 . \tag{A.24}$$

Thus, the orbit is simply given by the points with $\mathbb{R}^3$ with a fixed $|\mathbf{X}|$, i.e, a 2-sphere centered at the origin.

B Useful mathematical results

B.1 Block matrix manipulations

In this section, we describe a few well known results to do with operations on partitioned matrices with square blocks[119]. Consider a square matrix of dimensions $2n \times 2n$, consisting of blocks of dimensions $n \times n$:

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \tag{B.1}$$

We seek formulae relating the properties of M to those of A, B, C, D. The starting point is a decomposition of M in terms of triangular matrices,

$$M = \begin{pmatrix} A & 0 \\ C & \mathbb{1} \end{pmatrix} \begin{pmatrix} \mathbb{1} & A^{-1}B \\ 0 & D - CA^{-1}B \end{pmatrix}, \tag{B.2}$$

or, alternatively,

$$M = \begin{pmatrix} \mathbb{1} & B \\ 0 & D \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ D^{-1}C & \mathbb{1} \end{pmatrix}, \tag{B.3}$$

which can be verified by a direct computation.

Using this, we can compute the determinant of M as

$$\begin{aligned} \det M &= \det(A) \det(D - CA^{-1}B) \\ &= \det(D) \det(A - BD^{-1}C) \end{aligned} \tag{B.4}$$

The quantities of the form $A - BD^{-1}C$ that appear in these expressions are known as Schur complements, usually denoted by

$$M/D = A - BD^{-1}C, \tag{B.5}$$

where the order of the matrices in the second term is clockwise in M.

Now the inverse. For a lower triangular matrix with nonsingular A and D, the inverse can be computed

as

$$\begin{pmatrix} A & 0 \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{pmatrix}. \tag{B.6}$$

Similarly, for an upper triangular matrix,

$$\begin{pmatrix} A & B \\ 0 & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{pmatrix}. \tag{B.7}$$

An expression for inverse of M is

$$\begin{aligned} M^{-1} &= \begin{pmatrix} \mathbb{1} & A^{-1}B \\ 0 & M/A \end{pmatrix}^{-1} \begin{pmatrix} A & 0 \\ C & \mathbb{1} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} A^{-1} + A^{-1}B(M/A)^{-1}CA^{-1} & -A^{-1}B(M/A)^{-1} \\ (M/A)^{-1}CA^{-1} & (M/A)^{-1} \end{pmatrix} \end{aligned} \tag{B.8}$$

This expression is not very useful in general, but it illustrates the principle of decomposing a block matrix into a product of upper-triangular and lower-triangular matrices and computing the inverses individually, using the expressions above.

B.2 A parametrization of $\mathrm{Sp}(2, \mathbb{R})$

We seek to parametrize $Sp(2, \mathbb{R})$, and show that it is homeomorphic to a solid 2-torus[120, 121]. This can be shown using an Iwasawa decomposition[122]. Explicitly, let us consider a matrix $S \in Sp(2, \mathbb{R})$, parametrized as

$$S = \begin{pmatrix} a+b & c-d \\ c+d & a-b \end{pmatrix}, \tag{B.9}$$

with $(a, b, c, d) \in \mathbb{R}^4$. The determinant condition, $\det S = 1$, demands $(a^2 + d^2) - (b^2 + c^2) = 1$. Hence, $Sp(2, \mathbb{R})$ corresponds to a submanifold of $\mathbb{R}^4$ of codimension 1, which can be thought of as a 4-dimensional analogue of a hyperbola. We reparametrize

$$\begin{aligned} a &= \cosh\eta \cos\theta_1 \\ b &= \sinh\eta \cos\theta_2 \\ c &= \sinh\eta \sin\theta_2 \\ d &= \cosh\eta \sin\theta_1 \end{aligned}$$

where $\eta \in \mathbb{R}$ and $\theta_i \in [0, 2\pi)$. This makes $Sp(2, \mathbb{R})$ homeomorphic to $\mathbb{R} \times S^1 \times S^1 \cong \mathbb{R} \times T^2$. Define

$$\chi = \frac{1}{2}(1 + \tanh \eta) \in (0, 1), \tag{B.10}$$

so that $Sp(2, \mathbb{R}) \cong D \times S^1$. Finally, it is straightforward to embed the torus formed by $(\chi, \theta_1, \theta_2)$ in $\mathbb{R}^3$.

This parametrization also provides a particularly simple proof of the fact that $\pi_1(Sp(2n, \mathbb{R})) \cong \mathbb{Z}$ for the $n = 1$ case. Generally, the proof involves the fact[100, 71] that $U(n) \subset Sp(2n, \mathbb{R})$ is its maximally compact subgroup, so that $Sp(2n, \mathbb{R})$ has $U(n)$ as its strong deformation retract. Furthermore, $\pi_1(U(n)) \cong \mathbb{Z}$, which can be seen by the determinant map for $\mathcal{U} \in U(n)$ as $\mathcal{U} \mapsto \det \mathcal{U} \in S^1$, and $\pi_1(S^1) \cong \mathbb{Z}$.

For $Sp(2, \mathbb{R})$, consider the deformation retract

$$S_t = S(\eta t, \theta_1, \theta_2) : [0, 1] \to Sp(2n, \mathbb{R}). \tag{B.11}$$

For $t = 1$, we recover S, while for $\theta = 0$, we get

$$a_0 = \cos \theta_1, \quad d_0 = \sin \theta_1, \quad b_0 = c_0 = 0$$

so that S_0 is parametrized simply by $\theta_1 \in S^1$, which implies that S^1 is a deformation retract of $Sp(2n, \mathbb{R})$, which proves our result.

C Miscellaneous

C.1 Fermi-Dirac distribution and integrals

Consider a gas of fermions in the grand canonical ensemble. Owing to the Pauli exclusion principle, a given microstate can either be unoccupied or occupied by exactly one fermion. Thus, the 1-particle grand canonical partition function z is

$$z = \sum_{\text{states}} e^{-\beta(\varepsilon-\mu)} = 1 + e^{-\beta(\varepsilon-\mu)}, \tag{C.1}$$

where ε is the energy of the microstate, $\beta = T^{-1}$ is the inverse temperature and μ is the chemical potential. The corresponding grand potential g is

$$g = -\frac{1}{\beta}\ln z = -\frac{1}{\beta}\ln\left(1 + e^{-\beta(\varepsilon-\mu)}\right). \tag{C.2}$$

The grand potential is the generator for a variety of other relevant functions, such as the probability of occupation of a given state f (Fermi-Dirac distribution) or the 1-particle entropy h:

$$f = -\frac{\partial g}{\partial \mu} = \frac{1}{1 + e^{\beta(\varepsilon-\mu)}}, \quad h = -\frac{\partial g}{\partial T}. \tag{C.3}$$

These follow a highly nontrivial relation:

$$\begin{aligned}
h =& -\sum_{\text{states}} p_i \ln p_i = -f\ln f - (1-f)\ln(1-f) \\
=& \frac{1}{1+e^{\beta(\varepsilon-\mu)}}\ln\left(1+e^{\beta(\varepsilon-\mu)}\right) + \frac{e^{\beta(\varepsilon-\mu)}}{1+e^{\beta(\varepsilon-\mu)}}\ln\left(\frac{1+e^{\beta(\varepsilon-\mu)}}{e^{\beta(\varepsilon-\mu)}}\right) \\
=& \ln\left(1+e^{\beta(\varepsilon-\mu)}\right) - \frac{e^{\beta(\varepsilon-\mu)}}{1+e^{\beta(\varepsilon-\mu)}}\beta(\varepsilon-\mu) \\
=& \ln\left(1+e^{-\beta(\varepsilon-\mu)}\right) + \beta(\varepsilon-\mu)\left[1 - \frac{e^{\beta(\varepsilon-\mu)}}{1+e^{\beta(\varepsilon-\mu)}}\right] \\
=& \beta\left[-g + (\varepsilon-\mu)f\right].
\end{aligned}$$

Thus,

$$g = (\varepsilon - \mu)f - Th. \tag{C.4}$$

C.2 Symplectic forms in noninertial frames

Consider the generalized Liouville 1-form for the dynamics of a classical particle on $\mathbb{R}^{n,1}$ with an isotropic momentum-dependent Hamiltonian:

$$\eta_H = p_i dx^i - \mathcal{H}(|\mathbf{p}|)dt. \tag{C.5}$$

We seek a Hamiltonian formulation of this system as seen from a noninertial frame of reference. We switch frames by a time-dependent change of coordinate $x^i = O^i_{\ j}(w^j + \xi_j)$, where $w(t)$ corresponds to a Galilean boost and $O(t) \in \mathrm{SO}(n)$ to a time-dependent rotation, so that ξ_i is the position coordinate in the noninertial frame.

The derivation of a suitable symplectic form describing the dynamics in the noninertial frame then involves a choice of the definition of "momentum". The most straightforward choice is the *canonical* momentum, defined as $\pi_j = p_i O^i_{\ j}$, which preserves the canonical ($p_i dx^i$) form of η_H. We also define the velocity of the frame as $v^i = \partial_t w^i$ and its vorticity as $(O^{-1}\partial_t O)_{ij} = -\frac{1}{2}\omega_{ij}$, both of which may depend on time. The vorticity satisfies $\omega_{ij} = -\omega_{ji}$, which simply follows from the orthogonality of O.

The Liouville form becomes

$$\eta_H = \pi_i d\xi^i - \left[\mathcal{H} - \pi_i v^i + \frac{1}{2}\omega^{ij}\pi_i\xi_j\right] dt \equiv \pi_i d\xi^i - \mathcal{H}' dt, \tag{C.6}$$

where assuming a slowly accelerating and rotating frame, we have only retained the terms linear in ω and v. Thus, for the canonical momentum, the change of frame keeps the symplectic structure invariant, while changing the Hamiltonian. In other words, $\boldsymbol{\xi}$ and $\boldsymbol{\pi}$ are canonically conjugate.

An alternative choice of momentum is the *kinetic* momentum, which intends to keep the equation of motion for $\dot{\xi}^i$ invariant. To wit, consider the symplectic form in rotating coordinates

$$\rho_H \equiv d\rho_H = d\pi_i \wedge d\xi^i - \left[\frac{\partial \mathcal{H}}{\partial \pi_i} d\pi_i - v^i d\pi_i + \frac{1}{2}\omega^{ij}\left(\pi_i d\xi_j + \xi_j d\pi_i\right)\right] \wedge dt. \tag{C.7}$$

The equations of motion become

$$\dot{\xi}_i = \frac{\partial \mathcal{H}}{\partial \pi_i} - v^i + \frac{1}{2}\omega^{ij}\xi_j, \quad \dot{\pi}^i = -\frac{1}{2}\omega^{ij}\pi_j. \tag{C.8}$$

Then, one seeks the kinetic momentum ψ^i, in terms of which the equation of motion for ξ becomes $\dot{\xi}^i = \partial\mathcal{H}/\partial\psi_i$. We elucidate this by examples in the following.

Massive case: Consider a massive classical particle, so that

$$\mathcal{H} = \frac{|\mathbf{p}|^2}{2m} = \frac{|\boldsymbol{\pi}|^2}{2m} \implies \frac{\partial\mathcal{H}}{\partial\psi_i} = \frac{\pi_i}{m}.$$

Then, the kinetic momentum is defined by setting

$$\dot{\xi}_i = \frac{\psi_i}{m} \implies \pi^i = \psi_i + m\left(v_i - \frac{1}{2}\omega^{ij}\xi^j\right), \tag{C.9}$$

so that

$$\begin{aligned}
\mathcal{H}' &= \frac{|\boldsymbol{\pi}|^2}{2m} - \pi_i v^i + \frac{1}{2}\omega^{ij}\pi_i\xi_j \\
&= \frac{1}{2m}\left|\psi_i + mv_i - \frac{1}{2}m\,\omega^{ij}\xi^j\right|^2 - \left(\psi_i + mv_i - \frac{1}{2}m\,\omega^{ij}\xi^j\right)\left(v^i - \frac{1}{2}\omega^{ik}\right) \\
&= \frac{|\boldsymbol{\psi}|^2}{2m^2} + \psi_i v^i - \frac{1}{2}\omega^{ij}\psi_i\xi_j - \psi_i v^i + \frac{1}{2}\omega^{ij}\psi_i\xi_j + \text{second order terms} \\
&= \frac{|\boldsymbol{\psi}|^2}{2m^2} + \text{second order terms}.
\end{aligned} \tag{C.10}$$

It is precisely this cancellation that we seek in defining the kinetic momentum. Thus to linear order in v and ω, defining $a^i = \partial_t v^i$ and $\alpha_{ij} = \partial_t\omega_{ij}$, the symplectic form becomes

$$\rho_H = d\psi_i \wedge d\xi^i + \frac{1}{2}m\,\omega_{ij}d\xi^i \wedge d\xi^j + m\left(a_i + \alpha_{ij}\xi^j\right)dt \wedge d\xi^i - d\left(\frac{|\boldsymbol{\psi}|^2}{2m}\right) \wedge dt. \tag{C.11}$$

We combine the inertial terms as

$$\Omega = \frac{1}{2}\Omega_{\mu\nu}dx^\mu dx^\nu = \frac{1}{2}\omega_{ij}d\xi^i \wedge d\xi^j + \left(a_i + \alpha_{ij}\xi^j\right)dx^0 \wedge d\xi^i, \tag{C.12}$$

the symplectic form simply becomes

$$\rho_H = d\psi_i \wedge d\xi^i + m\,\Omega - d\mathcal{H} \wedge dt. \tag{C.13}$$

Here, $m\omega$ corresponds to the Coriolis force, ma to the inertial force and $m\alpha_{ij}\xi^j dtdx^i$ to the tangential acceleration due to a variable angular velocity. This does not capture the centrifugal force, as we have

ignored the terms at $O(\omega^2)$.

Massless case: For massless particles, $\mathcal{H} = c|\mathbf{p}| = c|\boldsymbol{\pi}|;\ c = \pm 1$, so that the equations of motion become

$$\dot{\xi}_i = c\hat{\pi}^i - v^i - \frac{1}{2}\omega^{ij}\xi_j, \quad \dot{\pi}^i = \frac{1}{2}\omega^{ij}\pi_j.$$

Taking a cue from the massive case, consider a definition of kinetic momentum as

$$\pi^i = \psi_i + c|\boldsymbol{\psi}|\left(v_i - \frac{1}{2}\omega^{ij}\xi^j\right), \tag{C.14}$$

where we have replaced m with $c|\boldsymbol{\psi}|$. Defining $\hat{\psi}^i = \psi^i/|\boldsymbol{\psi}|$, we again get a cancellation in $\mathcal{H}'$ at linear order:

$$\begin{aligned}
\mathcal{H}' &= c|\boldsymbol{\pi}| - \pi_i v^i + \frac{1}{2}\omega^{ij}\pi_i\xi_j \\
&= c|\boldsymbol{\psi}|\left\{\left|\hat{\psi}_i + v_i - \frac{1}{2}\omega^{ij}\xi^j\right| - \left(\hat{\psi}_i + v_i - \frac{1}{2}\omega^{ij}\xi^j\right)\left(v^i - \frac{1}{2}\omega^{ik}\right)\right\} \\
&= c|\boldsymbol{\psi}|\left\{\left[1 + 2c\hat{\psi}^i\left(v_i - \frac{1}{2}\omega^{ij}\xi^j\right)\right]^{1/2} - \hat{\psi}_i\left(v^i - \frac{1}{2}\omega^{ik}\right)\right\} + \text{second order terms} \\
&= c|\boldsymbol{\psi}|\left\{1 + c\hat{\psi}^i\left(v_i - \frac{1}{2}\omega^{ij}\xi^j\right) - \hat{\psi}_i\left(v^i - \frac{1}{2}\omega^{ik}\right)\right\} + \text{second order terms} \\
&= c|\boldsymbol{\psi}| + \text{second order terms}.
\end{aligned} \tag{C.15}$$

Thus, the symplectic form becomes

$$\rho_H = d\psi_i \wedge d\xi^i + c|\boldsymbol{\psi}|\,\Omega - c\,d|\boldsymbol{\psi}| \wedge dt. \tag{C.16}$$

In considering Galilean boosts (instead of Lorentz boosts), we are ignoring the effect of time dilation, including which will lead to corrections at the next order in Ω.

C.3 Transfer matrix for Chern insulator

The Chern insulator is a 2-dimensional lattice model described by the lattice Hamiltonian[40]

$$\mathcal{H} = a\sin k_x\sigma^x + a\sin k_y\sigma^y + b(2 - m - \cos k_x - \cos k_y)\sigma^z \tag{C.17}$$

The system is gapped in the bulk, except for $m = 0, 2, 4$, when the gap closes. It is topological for $0 < m < 2$ with edge states around $k = 0$ and and for $2 < m < 4$ with edge states around $k = \pi$.

Let us put the Chern insulator on a cylinder which is periodic along y and finite along x. Then, we need to inverse Fourier transform along x (as k_x is not well-defined for a finite system) and write the Hamiltonian as

$$\begin{aligned}\mathcal{H}(k_y) = \sum_{n=0}^{N} \Bigg[& \frac{a}{2i} \left(c_{n+1}^\dagger \sigma^x c_n - c_n^\dagger \sigma^x c_{n+1} \right) \\ & - \frac{b}{2} \left(c_{n+1}^\dagger \sigma^z c_n + c_n^\dagger \sigma^z c_{n+1} \right) \\ & + c_n^\dagger \left(\sin k_y \sigma^y + b\Lambda(k_y)\sigma^z \right) c_n \Bigg] \end{aligned} \tag{C.18}$$

where $c_n(k_y)$ is a row vector, corresponding to the annihilation operator for the two degrees of freedom on each lattice site and $\Lambda(k_y) = 2 - m - \cos k_y$. Here, k_y ($\equiv \mathbf{k}_\perp$) just acts as a parameter in the Hamiltonian. We are only concerned with a topological state for $n \geq 0$.

The corresponding recursion relation is

$$\begin{aligned} & \frac{1}{2i} \left(a\sigma^x - ib\sigma^z \right) \psi_{n+1} - \frac{1}{2i} \left(a\sigma^x + ib\sigma^z \right) \psi_{n-1} \\ & \qquad = \left(\varepsilon \mathbb{1} - a \sin k_y \sigma^y - b\Lambda(k_y)\sigma^z \right) \psi_n \end{aligned} \tag{C.19}$$

We identify the hopping matrix

$$J = \frac{1}{2i} \left(a\sigma^x - ib\sigma^z \right) \tag{C.20}$$

which has eigenvalues

$$\sigma(J) = \pm \frac{1}{2} \sqrt{b^2 - a^2}. \tag{C.21}$$

Hence, J becomes singular when $a = b$, which is precisely the case that we are interested in. For the subsequent calculations, we set $a = b = 1$. Hence,

$$J = \frac{1}{2i} \left(\sigma^x - i\sigma^z \right) = -\frac{1}{2} \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \tag{C.22}$$

and $\ker(J)$ is spanned by $\mathbf{v} = (1, i)^T$, while $\ker(J^\dagger)$ is spanned by $\mathbf{w} = (1, -i)^T$. The crucial fact, that helps us compute the transfer matrix, is that $\mathbf{v}$ and $\mathbf{w}$ are orthogonal, i.e, $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

To see that explicitly, we write out $\psi_n = (\psi_n^1, \psi_n^2)^T$, and the recursion relation as

$$\begin{aligned}\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}\begin{pmatrix} \psi_{n+1}^1 \\ \psi_{n+1}^2 \end{pmatrix} - \begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix}\begin{pmatrix} \psi_{n-1}^1 \\ \psi_{n-1}^2 \end{pmatrix} \\ = -2\begin{pmatrix} \varepsilon - \Lambda(k_y) & i\sin k_y \\ -i\sin k_y & \varepsilon + \Lambda(k_y) \end{pmatrix}\begin{pmatrix} \psi_n^1 \\ \psi_n^2 \end{pmatrix}\end{aligned} \tag{C.23}$$

We now premultiply the above expression by $(1, i)$ and $(1, -i)$ to get two recursion relations, one excluding ψ_{n+1} and one excluding ψ_{n-1}. We can simplify these expressions greatly by defining

$$\phi_n = \psi_n^2 + i\psi_n^2, \quad \bar{\phi}_n = \psi_n^2 - i\psi_n^2. \tag{C.24}$$

Notice that these are not complex conjugates, as ψ_n^i's are in general complex. In terms ϕ's, we get

$$\begin{aligned}(\varepsilon + \sin k_y)\phi_n - \Lambda(k_y)\bar{\phi}_n + \bar{\phi}_{n-1} = 0 \\ \phi_{n+1} - \Lambda(k_y)\phi_n + (\varepsilon - \sin k_y)\bar{\phi}_n = 0\end{aligned} \tag{C.25}$$

Replacing $n \to n+1$ in the former and reorganizing the terms, we get

$$\begin{pmatrix} \bar{\phi}_{n+1} \\ \phi_{n+1} \end{pmatrix} = \begin{pmatrix} \frac{1-\varepsilon^2+\sin^2 k_y}{\Lambda(k_y)} & \varepsilon + \sin k_y \\ -(\varepsilon - \sin k_y) & \Lambda(k_y) \end{pmatrix}\begin{pmatrix} \bar{\phi}_n \\ \phi_n \end{pmatrix}. \tag{C.26}$$

Hence, we have managed to compute the transfer matrix, acting as

$$\Phi_{n+1} = T\Phi_n, \quad \Phi_n = \begin{pmatrix} \bar{\phi}_n \\ \phi_n \end{pmatrix} \tag{C.27}$$

We can explicitly check that $\det(T) = 1$. The other useful quantity is the trace,

$$\Delta(\varepsilon, k_y) = \frac{1 - \varepsilon^2 + \Lambda^2(k_y) + \sin^2 k_y}{\Lambda(k_y)}. \tag{C.28}$$

This is equal to the trace obtained by using the formal construction in eq. (7.29). Finally, we can compute the band edges and edge states, as described in §6.4.

C.4 Closed form conditions for physical edge states

In this section, we construct a closed form expression combining the decay and the boundary conditions for the case when there are an equal number ($= r$) of eigenvalues are inside and outside the unit circle in the complex plane, which corresponds to an $(\varepsilon, \mathbf{k}_\perp) \in \mathscr{G}$, i.e, in the bulk gap. This implies that $\mathrm{Tr}\mathcal{P}_< = \mathrm{Tr}\mathcal{P}_> = r$, so that $\mathcal{P}_< + \mathcal{P}_> = \mathbb{1}$ and

$$\mathcal{P}_<\Phi_1 = \Phi_1 \implies \mathcal{P}_>\Phi_1 = 0. \tag{C.29}$$

We seek to represent $\mathcal{P}_>$ in terms of the (generalized) eigenvectors of T. Let $\rho_i \in \mathbb{C}$ be the generalized eigenvalues of T with corresponding left and right generalized eigenvectors being ϕ_i's and φ_i's. Furthermore, let us assume that that ρ_i lies outside the unit circle for $i = 1, \dots r$ while it lies inside the the unit circle for $i = r+1, \dots 2r$. Then, we define the left and right subspaces corresponding to $\mathcal{P}_>$ as

$$\mathcal{L}_> = (\phi_1, \dots, \phi_r), \quad \mathcal{R}_> = (\varphi_1, \dots, \varphi_r), \tag{C.30}$$

where $\mathcal{L}_>, \mathcal{R}_> \in \mathbb{C}^{2r\times r}$ span the co-kernel and range of $\mathcal{P}_>$, respectively.

If T were normal, i.e, diagonalizable by a unitary transform, then the right eigenvectors φ_i's form an orthonormal basis of $\mathbb{C}^{2r}$. As $\mathcal{P}_>$ projects along a subset of these eigenvectors, it is an orthogonal projection, which can be written as

$$\mathcal{P}_> = \varphi_1\varphi_1^\dagger + \varphi_2\varphi_2^\dagger + \cdots + \varphi_r\varphi_r^\dagger = \mathcal{R}_>\mathcal{R}_>^\dagger. \tag{C.31}$$

Alternatively, in terms of the left eigenvectors, $\mathcal{P}_> = \mathcal{L}_>\mathcal{L}_>^\dagger$. For general T, the analogous expression is the non-orthogonal representation[123] of $\mathcal{P}_>$

$$\mathcal{P}_> = \mathcal{R}_>(\mathcal{L}_>^\dagger\mathcal{R}_>)^{-1}\mathcal{L}_>^\dagger. \tag{C.32}$$

Hence, the decay condition $\mathcal{P}_>\Phi_1 = 0$ (eq. (C.29)) implies $\mathcal{L}_>^\dagger\Phi_1 = 0$, which, using eq. (C.30), can be written explicitly as

$$\sum_{j=1}^{r}(\phi_j^*)_i(\boldsymbol{\beta}_1)_j = 0, \quad i = 0, \dots, r, \tag{C.33}$$

which constitutes r linear equations for r variables $(\boldsymbol{\beta}_1)_j$. Note that $\boldsymbol{\beta}_1$ is unique up to a non-zero complex scalar since the right-hand sides are all zero. Thus the space of unique solutions really is the complex

projective $\mathbb{CP}^{r-1}$ valued. The equations (C.33) have a nontrivial solution if and only if

$$\det\left[\mathcal{L}_{>}^{\dagger}\mathcal{Q}_{\boldsymbol{\beta}}\right]=0. \tag{C.34}$$

which is essentially a Cramer's condition. The analogous right edge conditions reads as

$$\det\left[\mathcal{R}_{<}^{\dagger}\mathcal{Q}_{\alpha}\right]=0 \tag{C.35}$$

These conditions incorporate both the boundary and decay conditions and can be solved numerically to obtain ε as a function of $\mathbf{k}_{\perp}$ to obtain the edge spectrum, $\varepsilon_{edge}(\mathbf{k}_{\perp})$.

Equation (C.34) is very convenient for numerical computations, but we also present an alternative characterization which is more explicit in terms of T's projection. The general spectral decomposition of the resolvent of T[97] yields

$$\mathcal{P}_{>}=\oint_{|z|=1}\frac{dz}{2\pi i}\,(z-T^{-1})^{-1}=\oint_{|z|=1}\frac{dz}{2\pi i}\,T(zT-\mathbb{1})^{-1}. \tag{C.36}$$

Essentially, note that the integrand has poles whenever z equals an eigenvalue ρ_s of T so that $|\rho_s|>1$.

Now, in the simpler case of a normal T, we have $\mathcal{P}_{>}=\mathcal{L}_{>}\mathcal{L}_{>}^{\dagger}=\mathcal{R}_{>}\mathcal{R}_{>}^{\dagger}$, so that

$$\det\left[\mathcal{Q}_{\boldsymbol{\beta}}^{\dagger}\mathcal{P}_{>}\mathcal{Q}_{\boldsymbol{\beta}}\right]=\det\left(\mathcal{Q}_{\boldsymbol{\beta}}^{\dagger}\mathcal{L}_{>}\right)\det\left(\mathcal{L}_{>}^{\dagger}\mathcal{Q}_{\boldsymbol{\beta}}\right)=0. \tag{C.37}$$

Substituting the integral representation of $\mathcal{P}_{>}$ from eq. (C.36), we get

$$\det\left[\oint_{|z|=1}dz\left[T(\varepsilon)(zT(\varepsilon)-\mathbb{1})^{-1}\right]_{\boldsymbol{\beta}\boldsymbol{\beta}}\right]=0, \tag{C.38}$$

where $[*]_{\boldsymbol{\beta}\boldsymbol{\beta}}$ denotes the $r\times r$ sub-matrix of the argument and we have expressed the ε dependence of T explicitly. Such an equation, though impractical for numerical computations, make explicit the analytic properties of an edge dispersion $\varepsilon(\mathbf{k}_{\perp})$ in open neighborhoods where it exists as a solution.

In the most general case where $\mathcal{P}_{>}$ is oblique(non-orthogonal), the analogue of eq. (C.37) is

$$\det\left[[\mathcal{P}_{>}^{\dagger}\mathcal{P}_{>}]_{\boldsymbol{\beta}\boldsymbol{\beta}}\right]\equiv\det[\mathcal{Q}_{\boldsymbol{\beta}}^{\dagger}\mathcal{P}_{>}^{\dagger}\mathcal{P}_{>}\mathcal{Q}_{\boldsymbol{\beta}}]=0 \tag{C.39}$$

where $\mathcal{P}_{>}$ is still given by the integral equation (C.36).

References

[1] Michael Stone and Vatsal Dwivedi. Classical version of the non-abelian gauge anomaly. *Physical Review D*, 88(4):045012, 2013.

[2] Vatsal Dwivedi and Michael Stone. Classical chiral kinetic theory and anomalies in even space-time dimensions. *Journal of Physics A: Mathematical and Theoretical*, 47(2):025401, 2013.

[3] Michael Stone, Vatsal Dwivedi, and Tianci Zhou. Berry phase, Lorentz covariance, and anomalous velocity for Dirac and Weyl particles. *Physical Review D*, 91(2):025004, 2015.

[4] Vatsal Dwivedi and Michael Stone. Chiral kinetic theory and anomalous hydrodynamics in even spacetime dimensions. *arXiv preprint arXiv:1606.04945*, 2016.

[5] Vatsal Dwivedi and Victor Chua. Of bulk and boundaries: Generalized transfer matrices for tight-binding models. *Physical Review B*, 93(13):134304, 2016.

[6] Reinhold A Bertlmann. *Anomalies in quantum field theory*, volume 91. Oxford University Press, 2000.

[7] Stephen L. Adler. Axial-vector vertex in spinor electrodynamics. *Phys. Rev.*, 177:2426–2438, Jan 1969.

[8] John Stewart Bell and Roman Jackiw. A pcac puzzle: $\pi 0$? $\gamma\gamma$ in the σ-model. *Il Nuovo Cimento A*, 60(1):47–61, 1969.

[9] Michael Peskin and Dan Schroeder. *An introduction to quantum field theory*. Frontiers in physics. Westview Press, 1995.

[10] Carl Eckart. The thermodynamics of irreversible processes. iii. relativistic theory of the simple fluid. *Physical review*, 58(10):919, 1940.

[11] L D Landau and E M Lifshitz. *Fluid Mechanics*, volume 4 of *Courses of Theoretical Physics*. Butterworth-Heinemann, Oxford, 2 edition, 1 1987.

[12] Mukund Rangamani. Gravity and hydrodynamics: Lectures on the fluid-gravity correspondence. *Classical and quantum gravity*, 26(22):224003, 2009.

[13] Dam T Son and Piotr Surówka. Hydrodynamics with triangle anomalies. *Physical Review Letters*, 103(19):191601, 2009.

[14] Yasha Neiman and Yaron Oz. Relativistic hydrodynamics with general anomalous charges. *Journal of High Energy Physics*, 2011(3):1–12, 2011.

[15] R Loganayagam. Anomaly induced transport in arbitrary dimensions. *arXiv preprint arXiv:1106.0277*, 2011.

[16] R Loganayagam and Piotr Surówka. Anomaly/transport in an ideal Weyl gas. *Journal of High Energy Physics*, 2012(4):1–34, 2012.

[17] Tatsuo Azeyanagi, R Loganayagam, Gim Seng Ng, and Maria J Rodriguez. Holographic thermal helicity. *Journal of High Energy Physics*, 2014(8):1–66, 2014.

[18] Nabamita Banerjee, Suvankar Dutta, Sachin Jain, R Loganayagam, and Tarun Sharma. Constraints on anomalous fluid in arbitrary dimensions. *The Journal of high energy physics*, 2013(3):1–41, 2013.

[19] Kristan Jensen, R Loganayagam, and Amos Yarom. Thermodynamics, gravitational anomalies and cones. *Journal of High Energy Physics*, 2013(2):1–35, 2013.

[20] Kristan Jensen, R Loganayagam, and Amos Yarom. Anomaly inflow and thermal equilibrium. *Journal of High Energy Physics*, 2014(5):1–69, 2014.

[21] Kristan Jensen, R Loganayagam, and Amos Yarom. Chern-simons terms from thermal circles and anomalies. *Journal of High Energy Physics*, 2014(5):1–29, 2014.

[22] Di Xiao, Ming-Che Chang, and Qian Niu. Berry phase effects on electronic properties. *Rev. Mod. Phys.*, 82:1959–2007, Jul 2010.

[23] Robert Karplus and J. M. Luttinger. Hall effect in ferromagnetics. *Phys. Rev.*, 95:1154–1160, Sep 1954.

[24] J. M. Luttinger. Theory of the hall effect in ferromagnetic substances. *Phys. Rev.*, 112:739–751, Nov 1958.

[25] W. Kohn and J. M. Luttinger. Quantum theory of electrical transport phenomena. *Phys. Rev.*, 108:590–611, Nov 1957.

[26] MA Stephanov and Yi Yin. Chiral kinetic theory. *Physical Review Letters*, 109(16):162001, 2012.

[27] Ralph Abraham and Jerrold Marsden. *Foundations of mechanics*. Westview Press, 2 edition, 1994.

[28] Vladimir Igorevich Arnol'd. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer, 2010.

[29] Xiangang Wan, Ari M. Turner, Ashvin Vishwanath, and Sergey Y. Savrasov. Topological semimetal and fermi-arc surface states in the electronic structure of pyrochlore iridates. *Phys. Rev. B*, 83:205101, May 2011.

[30] Edward Witten. Three lectures on topological phases of matter. *arXiv preprint arXiv:1510.07698*, 2015.

[31] Sumathi Rao. Weyl semi-metals: a short review. *arXiv preprint arXiv:1603.02821*, 2016.

[32] A. A. Zyuzin and A. A. Burkov. Topological response in Weyl semimetals and the chiral anomaly. *Phys. Rev. B*, 86:115133, Sep 2012.

[33] Pavan Hosur and Xiaoliang Qi. Recent developments in transport phenomena in Weyl semimetals. *arXiv preprint arXiv:1309.4464*, 2013.

[34] D. T. Son and B. Z. Spivak. Chiral anomaly and classical negative magnetoresistance of Weyl metals. *Phys. Rev. B*, 88:104412, Sep 2013.

[35] Kenji Fukushima, Dmitri E Kharzeev, and Harmen J Warringa. Chiral magnetic effect. *Physical Review D*, 78(7):074033, 2008.

[36] Kenji Fukushima, Dmitri E Kharzeev, and Harmen J Warringa. Real-time dynamics of the chiral magnetic effect. *Physical review letters*, 104(21):212001, 2010.

[37] Alexander Vilenkin. Equilibrium parity-violating current in a magnetic field. *Physical Review D*, 22(12):3080, 1980.

[38] Nabamita Banerjee, Jyotirmoy Bhattacharya, Sayantani Bhattacharyya, Suvankar Dutta, R Loganayagam, and P Surowka. Hydrodynamics from charged black branes. *Journal of High Energy Physics*, 2011(1):1–37, 2011.

[39] Jing-Yuan Chen, Dam T Son, Mikhail A Stephanov, Ho-Ung Yee, and Yi Yin. Lorentz invariance in chiral kinetic theory. *Physical Review Letters*, 113(18):182302, 2014.

[40] B Andrei Bernevig. *Topological Insulators and Topological Superconductors*. Princeton University Press, 2013.

[41] Shun-Qing Shen. *Topological Insulators: Dirac Equation in Condensed Matters*, volume 174. Springer Science & Business Media, 2013.

[42] M. Z. Hasan and C. L. Kane. Colloquium: Topological insulators. *Rev. Mod. Phys.*, 82:3045, 2010.

[43] 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.

[44] Barry Simon. Holonomy, the quantum adiabatic theorem, and berry's phase. *Phys. Rev. Lett.*, 51:2167–2170, Dec 1983.

[45] Roger S. K. Mong and Vasudha Shivamoggi. Edge states and the bulk-boundary correspondence in Dirac Hamiltonians. *Phys. Rev. B*, 83:125109, Mar 2011.

[46] Srinidhi T. Ramamurthy and Taylor L. Hughes. Patterns of Electromagnetic Response in Topological Semimetals. *arXiv:1405.7377*, 2014.

[47] Yasuhiro Hatsugai. Edge states in the integer quantum Hall effect and the Riemann surface of the Bloch function. *Phys. Rev. B*, 48:11851–11862, Oct 1993.

[48] Yia-Chung Chang and J. N. Schulman. Complex band structures of crystalline solids: An eigenvalue method. *Phys. Rev. B*, 25:3975–3986, Mar 1982.

[49] D. H. Lee and J. D. Joannopoulos. Simple scheme for surface-band calculations. I. *Phys. Rev. B*, 23:4988–4996, May 1981.

[50] JulioCesar Avila, Hermann Schulz-Baldes, and Carlos Villegas-Blas. Topological invariants of edge states for periodic two-dimensional models. *Mathematical Physics, Analysis and Geometry*, 16(2):137–170, 2013.

[51] Antonis N. Andriotis, Madhu Menon, and Deepak Srivastava. Transfer matrix approach to quantum conductivity calculations in single-wall carbon nanotubes. *The Journal of Chemical Physics*, 117(6):2836–2843, 2002.

[52] Suraj Hegde, Vasudha Shivamoggi, Smitha Vishveshwara, and Diptiman Sen. Quench dynamics and parity blocking in majorana wires. *New Journal of Physics*, 17(5):053036, 2015.

[53] NG Stephen. Transfer matrix analysis of the elastostatics of one-dimensional repetitive structures. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 462, pages 2245–2270. The Royal Society, 2006.

[54] S. Eastham. *Spectral Theory of Periodic Differential Equations*. Hafner Press, 1974.

[55] B.M. Brown, M.S.P. Eastham, and K.M. Schmidt. *Periodic Differential Operators*. Operator Theory: Advances and Applications. Springer Basel, 2012.

[56] Jared C Bronski and Zoi Rapti. Counting defect modes in periodic eigenvalue problems. *SIAM Journal on Mathematical Analysis*, 43(2):803–827, 2011.

[57] W. Kohn. Analytic Properties of Bloch Waves and Wannier Functions. *Phys. Rev.*, 115:809–821, Aug 1959.

[58] V. A. Yakubovich and V. M. Starzhinskii. *Linear Differential Equations with Periodic Coefficients*, volume 1,2. John Wiley and Sons, New York-Toronto, Ontario, 1975. Translated from Russian by D. Louvish.

[59] Jared C Bronski and Zoi Rapti. Modulational instability for nonlinear Schrödinger equations with a periodic potential. *Dynamics of Partial Differential Equations*, 2(4), 2005.

[60] Yasuhiro Hatsugai. Chern number and edge states in the integer quantum Hall effect. *Phys. Rev. Lett.*, 71:3697–3700, Nov 1993.

[61] J. Bellissard, A. van Elst, and H. Schulz Baldes. The noncommutative geometry of the quantum hall effect. *Journal of Mathematical Physics*, 35(10):5373–5451, 1994.

[62] Johannes Kellendonk, Thomas Richter, and Hermann Schulz-Baldes. Edge current channels and chern numbers in the integer quantum hall effect. *Reviews in Mathematical Physics*, 14(01):87–119, 2002.

[63] Emil Prodan. Disordered topological insulators: a non-commutative geometry perspective. *Journal of Physics A: Mathematical and Theoretical*, 44(11):113001, 2011.

[64] Grigory E Volovik and GE Volovik. *The universe in a helium droplet*. Oxford University Press New York, 2009.

[65] Andrew M Essin and Victor Gurarie. Bulk-boundary correspondence of topological insulators from their respective greens functions. *Physical Review B*, 84(12):125132, 2011.

[66] I. C. Fulga, F. Hassler, and A. R. Akhmerov. Scattering theory of topological insulators and superconductors. *Phys. Rev. B*, 85:165409, Apr 2012.

[67] G. Brunlich, G.M. Graf, and G. Ortelli. Equivalence of topological and scattering approaches to quantum pumping. *Communications in Mathematical Physics*, 295(1):243–259, 2010.

[68] Xiao-Liang Qi, Yong-Shi Wu, and Shou-Cheng Zhang. General theorem relating the bulk topological number to edge states in two-dimensional insulators. *Phys. Rev. B*, 74:045125, Jul 2006.

[69] Herbert Goldstein, Charles P. Poole Jr, and John L. Safko. *Classical Mechanics*. Pearson, 3 edition, 2001.

[70] Michael Stone and Paul Golldbart. *Mathematics for Physics: A Guided Tour for Graduate Students*. Cambridge University Press, 1st edition, 2009.

[71] D MacDuff and D Salamon. *Introduction to symplectic topology*. Oxford Mathematical Monographs. Oxford University Press, 2nd edition, 1999.

[72] Abraham Aleksandrovic Kirillov. *Lectures on the orbit method*. AMS Bookstore, 2004.

[73] Mikio Nakahara. *Geometry, topology, and physics*. CRC Press, 2003.

[74] David Tong. Lectures on the quantum hall effect. *arXiv preprint arXiv:1606.06687*, 2016.

[75] David Tong. Lectures on kinetic theory. Online, 2012.

[76] Terence Tao. What is a gauge? Online, 2008.

[77] Kazuo Fujikawa and Hiroshi Suzuki. *Path integrals and quantum anomalies*. Number 122 in International Series of Monographs on Physics. Oxford University Press, 2004.

[78] William A Bardeen and Bruno Zumino. Consistent and covariant anomalies in gauge and gravitational theories. *Nuclear Physics B*, 244(2):421–453, 1984.

[79] Michael Stone. Gravitational anomalies and thermal hall effect in topological insulators. *Physical Review B*, 85(18):184503, 2012.

[80] Curtis G Callan and Jeffrey A Harvey. Anomalies and fermion zero modes on strings and domain walls. *Nuclear Physics B*, 250(1):427–436, 1985.

[81] Dan Friedan and Paul Windey. Supersymmetric derivation of the atiyah-singer index and the chiral anomaly. *Nuclear Physics B*, 235(3):395–416, 1984.

[82] Paul Windey. Supersymmetric quantum mechanics and the atiyah-singer index theorem. *Acta Physica Polonica. Series B*, 15(5):435–452, 1984.

[83] Nils Andersson and Gregory L Comer. Relativistic fluid dynamics: physics for many different scales. *Living Rev. Relativity*, 10(1), 2007.

[84] JW van Holten and NL Amsterdam. Relativistic fluid dynamics. Online, 2006.

[85] M. V. Berry. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 392(1802):45–57, 1984.

[86] Tosio Kato. On the adiabatic theorem of quantum mechanics. *Journal of the Physical Society of Japan*, 5(6):435–439, 1950.

[87] Shivaramakrishnan Pancharatnam. Generalized theory of interference and its applications. In *Proceedings of the Indian Academy of Sciences-Section A*, volume 44, pages 398–417. Springer, 1956.

[88] Ming-Che Chang and Qian Niu. Berry phase, hyperorbits, and the hofstadter spectrum: Semiclassical dynamics in magnetic bloch bands. *Phys. Rev. B*, 53:7010–7023, Mar 1996.

[89] Ganesh Sundaram and Qian Niu. Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and berry-phase effects. *Phys. Rev. B*, 59:14915–14925, Jun 1999.

[90] C Duval, Z Horváth, PA Horváthy, L Martina, and PC Stichel. Berry phase correction to electron density in solids and" exotic" dynamics. *Modern Physics Letters B*, 20(07):373–378, 2006.

[91] Frank Wilczek and A. Zee. Appearance of gauge structure in simple dynamical systems. *Phys. Rev. Lett.*, 52:2111–2114, Jun 1984.

[92] Yakir Aharonov and J Anandan. Phase change during a cyclic quantum evolution. *Physical Review Letters*, 58(16):1593, 1987.

[93] Di Xiao, Junren Shi, and Qian Niu. Berry phase correction to electron density of states in solids. *Phys. Rev. Lett.*, 95:137204, Sep 2005.

[94] David John Simms and Nicholas Michael John Woodhouse. Lectures on geometric quantization. In *Lectures on Geometric Quantization*, volume 53, 1976.

[95] Nicholas Michael John Woodhouse. *Geometric quantization*. Oxford University Press, 1997.

[96] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 4th edition, 2009.

[97] Tosio Kato. *Perturbation Theory for Linear Operators*. Springer Verlag Berlin-Heidelberg, 2 edition, 1995.

[98] Walther Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Education, 3rd edition, 1976.

[99] Hermann Schulz-Baldes. Signature and spectral flow of j-unitary $\mathbb{S}^1$ -fredholm operators. *Integral Equations and Operator Theory*, 78(3):323–374, 2014.

[100] Robert G Littlejohn. The semiclassical evolution of wave packets. *Physics Reports*, 138(4):193–291, 1986.

[101] L. L. Song, X. H. Zheng, R. L. Wang, and Z. Zeng. Dangling Bond States, Edge Magnetism, and Edge Reconstruction in Pristine and B/N-Terminated Zigzag Graphene Nanoribbons. *The Journal of Physical Chemistry C*, 114(28):12145–12150, 2010.

[102] Changwon Park, Jisoon Ihm, and Gunn Kim. Decay behavior of localized states at reconstructed armchair graphene edges. *Phys. Rev. B*, 88:045403, Jul 2013.

[103] Xiaoqian Dang, J. D. Burton, Alan Kalitsov, Julian P. Velev, and Evgeny Y. Tsymbal. Complex band structure of topologically protected edge states. *Phys. Rev. B*, 90:155307, Oct 2014.

[104] Shijun Mao, Yoshio Kuramoto, Ken-Ichiro Imura, and Ai Yamakage. Analytic Theory of Edge Modes in Topological Insulators. *Journal of the Physical Society of Japan*, 79(12):124709, 2010.

[105] Douglas R Hofstadter. Energy levels and wave functions of bloch electrons in rational and irrational magnetic fields. *Physical review B*, 14(6):2239, 1976.

[106] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3rd edition, 2007.

[107] Richard L. Burden, J. Douglas Faires, and Annette M. Burden. *Numerical Analysis*. Brooks Cole, 10th edition, 2015.

[108] W. P. Su, J. R. Schrieffer, and A. J. Heeger. Solitons in Polyacetylene. *Phys. Rev. Lett.*, 42:1698–1701, Jun 1979.

[109] H.-M. Guo and M. Franz. Topological insulator on the kagome lattice. *Phys. Rev. B*, 80:113102, Sep 2009.

[110] Evelyn Tang, Jia-Wei Mei, and Xiao-Gang Wen. High-temperature fractional quantum hall states. *Phys. Rev. Lett.*, 106:236802, Jun 2011.

[111] Robert G. Littlejohn and Jonathan M. Robbins. New way to compute Maslov indices. *Phys. Rev. A*, 36:2953–2961, Sep 1987.

[112] Liang Fu. Topological crystalline insulators. *Phys. Rev. Lett.*, 106:106802, Mar 2011.

[113] A MacKinnon and B Kramer. The scaling theory of electrons in disordered solids: additional numerical results. *Zeitschrift für Physik B Condensed Matter*, 53(1):1–13, 1983.

[114] Ferdinand Evers and Alexander D Mirlin. Anderson transitions. *Reviews of Modern Physics*, 80(4):1355, 2008.

[115] Bernhard Kramer and Angus MacKinnon. Localization: theory and experiment. *Reports on Progress in Physics*, 56(12):1469, 1993.

[116] J-P Eckmann and David Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of modern physics*, 57(3):617, 1985.

[117] Keith Slevin and Tomi Ohtsuki. Critical exponent for the quantum hall transition. *Physical Review B*, 80(4):041304, 2009.

[118] Hideaki Obuse, Arvind R Subramaniam, Akira Furusaki, Ilya A Gruzberg, and Andreas W W Ludwig. Conformal invariance, multifractality, and finite-size scaling at anderson localization transitions in two dimensions. *Physical Review B*, 82(3):035309, 2010.

[119] Dennis S Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2nd edition, 2009.

[120] S. Lang. $SL_2(R)$. Graduate texts in mathematics. Springer New York, 1985.

[121] Keith Conrad. Decomposing $SL(2, R)$. online.

[122] Kenkichi Iwasawa. On Some Types of Topological Groups. *Annals of Mathematics*, 50(3):pp. 507–558, 1949.

[123] Carl D Meyer. *Matrix analysis and applied linear algebra*. Siam, 2000.