A pedagogical note on geometric phases

A. Alexandradinata[1]

[1]*Department of Physics and Santa Cruz Materials Center, University of California Santa Cruz, Santa Cruz, CA 95064, USA*

CONTENTS

I. BASIC NOTIONS IN PARALLEL TRANSPORT

The geometric notion of parallel transport is well-defined where the space of states at each momentum is separated into two subspaces, such that transport is 'parallel to' one of the two subspaces. In electronic insulators, the Fermi energy distinguishes occupied from empty subspaces, and we parallel transport within the occupied subspace. Mathematically, parallel transport is defined for a sub-bundle of the full space of states, which forms a trivial bundle.

Parallel transport is a set of rules to gauge-fix a set of Hamiltonian eigenfunctions at k_2, given the eigenfunctions at $k_1 \neq k_2$. The gauge-fixing at k_2 can be determined by

$$A(k)_{mn} = \left\langle u_{m,k} \,|\, \partial u_{n,k} \right\rangle = 0 \quad \text{for} \quad k \in \quad \text{loop connecting} \quad k_1 \quad \text{and} \quad k_2. \tag{1}$$

Here, A is the (possibly non-abelian) Berry connection, though in principle this is not the only connection one can choose. Therefore, defining rules for parallel transport is equivalent to defining a connection. A is also known as a $U(N)$ gauge field.

II. BERRY AND ZAK PHASES

If the Hamiltonian is parametrized by $\alpha(t)$, we consider an arbitrary basis $|\, 0(\alpha)\,\rangle$ at each α, which is both single-valued in the domain of interest and differentiable. In the adiabatic approximation, the ground state evolves as

$$\Psi_0(t) = e^{-i\int_{t_0}^{t} E_0(t')dt'}\, e^{i\phi_B(t)}\, |\, 0(t)\,\rangle \tag{2}$$

where we include ϕ_B because our arbitrarily chosen basis need not coincide with the state at hand. By coincidence, I mean two states have the same $U(1)$ phase, though the distinction is mathematical and unmeasurable (all expectation values are bilinears). Schrodinger's equation

$$i\partial_t \Psi_0 \Big|_t = H(t)\, e^{-i\int_{t_0}^{t} E_0(t')dt'}\, e^{i\phi_B(t)}\, |\, 0(t)\,\rangle = E_0(t)\Psi_0(t) \tag{3}$$

determines $\phi_B(t)$. This is because Schrodinger evolution is determininistic. Given an initial state $\Psi_o(0)$, $\Psi_0(t)$ is determined in the adiabatic regime, and ϕ_B is defined by

$$\left\langle 0(t) \,|\, \Psi_0(t) \right\rangle = e^{-i\int_{t_0}^{t} E_0(t')dt'}\, e^{i\phi_B(t)}. \tag{4}$$

Now $\Psi_0(t)$ is not determined by the most general

$$\mathrm{T}\, e^{-i\int_0^t H(t')dt'}. \tag{5}$$

The adiabatic approximation lies in the projection to a certain subspace:

$$\Psi_0(t) = e^{-i\int_0^t E(t')dt'} \prod_{\alpha(t)\leftarrow\alpha(0)} P(\alpha)\,\Psi_0(0) = e^{-i\int_0^t E(t')dt'}\,|\,0(t)\,\rangle\langle\,0(t)\,|\prod_{\alpha(t)\leftarrow\alpha(0)} P(\alpha)\,|\,0(0)\,\rangle. \tag{6}$$

The Wilson-line phase is well-defined if we specify both initial and final bases. The Wilson-loop phase is defined such that initial and final bases are identical, but arbitrary. This arbitrariness allows us to formulate a basis-independent expression for the Abelian Berry phase: the unimodular eigenvalue of the operator

$$\hat{\mathcal{W}} = \prod_{l}^{R\leftarrow R} P(l). \tag{7}$$

More commonly, the Berry phase is expressed with a single-valued and differentiable basis $|\,\psi(R)\,\rangle$ at each parameter R:

$$\gamma = i\int_C d\,R_\mu \big\langle \psi(R)\,|\,\partial_\mu\psi(R)\big\rangle. \tag{8}$$

The Berry phase factor is invariant under transformations that preserve these two properties. However, the formal identity with the previous formula shows that differentiability may be relaxed in a computation where projections are used. The logic is $A = B$ for the same basis satisfying two conditions. $B = C$ for two bases of the same projection formula, but now the basis of C need not be differentiable.

Let's give an example of a basis which is differentiable but not single-valued. The simplest 1D topological insulator:

$$h(k) = \cos k\,\sigma_3 + \sin k\,\sigma_1. \tag{9}$$

Since the Hamiltonian is real, we can choose a smooth gauge in which all wavefunctions are real. However, because the insulator is nontrivial, we'll find that this real gauge is not periodic. The nontriviality arises because there exists an inversion symmetry

$$\sigma_3\,h(\text{-}k)\,\sigma_3 = h(k) \tag{10}$$

and the wavefunctions at 0 and π have opposite inversion eigenvalues. The two-level Hamiltonian can be written as $d\cdot\sigma$ so d can be thought of as a magnetic field, with spin-half eigenstates aligned parallel and anti-parallel to the field. Parametrizing $d = |d|(\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta)$, the spinor that is anti-parallel to the field is written

$$\psi(\theta,\phi) = \begin{pmatrix} e^{\text{-}i\phi/2}\ \sin\frac{\theta}{2} \\ \\ \text{-}\,e^{i\phi/2}\ \cos\frac{\theta}{2}\end{pmatrix}. \tag{11}$$

In this simple Hamiltonian we make the identification $k = \theta$ and $\phi = 0$, hence

$$|\,\psi_k\,\rangle = \begin{pmatrix} \sin\frac{k}{2} \\ \\ \text{-}\cos\frac{k}{2}\end{pmatrix}. \tag{12}$$

Since $A(k) = 0$ in this real gauge, naively the Berry phase is zero. Actually the right answer is π. There are two ways to get the right answer:

(i) To eat both cake, it is necessary to move off the real plane by the following phase rotation:

$$|\,\psi_k\,\rangle = e^{ik/2}\begin{pmatrix} \sin\frac{k}{2} \\ \\ \text{-}\cos\frac{k}{2}\end{pmatrix}. \tag{13}$$

In this periodic gauge, $A(k) = 1/2$. Thus we obtain the right answer for the Berry phase.

(ii) Choose the real basis only in the half-closed interval $[-\pi, \pi)$, and formally identify $\psi(\pi) = \psi(-\pi)$. Then the basis is single-valued but not differentiable. But this is ok if we use the projection formula. Since (12) is the parallel-transport gauge, by definition,

$$|\,\psi_\pi\,\rangle = \prod_k^{\pi \leftarrow -\pi} P(k)\,|\,\psi_{-\pi}\,\rangle. \tag{14}$$

Since we measure Berry phase relative to the original basis vector at the base point,

$$e^{i\phi} = \langle\,\psi_{-\pi}\,|\prod_k^{\pi \leftarrow -\pi} P(k)\,|\,\psi_{-\pi}\,\rangle = \langle\psi_{-\pi}\,|\,\psi_\pi\rangle = -1. \tag{15}$$

It is assumed that the Hamiltonian is identical at the base and end points of the loop C. If the parameter space is 1D, as for a circle, we cannot apply Stoke's theorem and transform this to an integral of a curvature. Thus, we must choose a differentiable and single-valued basis in a domain that encloses the loop C. Berry phase is defined as the extra phase acquired relative to this single-valued basis. To make this explicit, we note that a state adiabatically evolves as

$$|\,\bar\psi(T)\,\rangle = e^{-i\int E(t)dt}\prod_R^{T\leftarrow S} P(R)\,|\,\psi(S)\,\rangle. \tag{16}$$

We choose to express the final state in the original single-valued basis:

$$|\,\bar\psi(T)\,\rangle = e^{-i\int E(t)dt}\,|\,\psi(T)\,\rangle\langle\,\psi(T)\,|\prod_R^{T^-\leftarrow S} P(R)\,|\,\psi(S)\,\rangle = e^{-i\int E(t)dt}\,|\,\psi(T)\,\rangle\,e^{-\int_S^T dR_\mu\left\langle\psi(R)\,|\,\partial_\mu\psi(R)\right\rangle}. \tag{17}$$

Generally, such phases have no measurable consequence unless we make a loop in parameter space, i.e., $T \equiv S$.

Principle: an absolute phase isn't physical, but a relative phase between two wavefunctions is observable, if they are made to interfere. Observation of Berry phase must rely on interference. The conceptually-simple thought experiment is to split a particle into two, parallel transport one around a loop, while keeping the other fixed. When the two states meet again, they interfere in a manner that depends on the Berry phase acquired by the moving particle. In experiments, beam splitting is usually done, so both particles are moving – position is our parameter space. We may write:

$$\psi(r,t) = \psi_1(r)\,e^{i\phi_1(t)} + \psi_2(r)\,e^{i\phi_2(t)} \tag{18}$$

where ψ_1 and ψ_2 are localized wavepackets travelling in two different tubes. Let us denote R as the position where the tubes recombine. If tubes 1 and 2 are identical, then $\psi_1(R) = \psi_2(R)$ by symmetry, and also $E_1(r) = E_2(r)$ within the loop. Thus, a measurement made at R probes the Berry phase through:

$$\psi(R,t) = \psi_1(R)\,e^{i\phi_1(t)+i\phi_2(t)}. \tag{19}$$

If we insist on single-valuedness, the basis is not fully specified. We are allowed:

$$|\,\psi(R)\,\rangle \to e^{i\phi(R)}|\,\psi(R)\,\rangle \tag{20}$$

where $\phi(R)$ is generally a sum of a periodic function and a function with winding. Basis changes that involve winding are called large gauge transformations. Under (20), $\gamma \to \gamma + 2\pi n$, so the Berry phase factor is invariant under single-value-preserving gauge transformations.

Zak's phase has the same form as in (8), but is defined for non-contractible loops in a BZ. Let us consider the Bloch Hamiltonian

$$h(k) = e^{-ik\cdot r}\,H\,e^{ik\cdot r} \tag{21}$$

and denote its eigenstates as $u_{n,k}(r)$. The problem is not well-defined unless we have a boundary condition. The one we pick is $u_k(r) = u_k(r+R)$, so that $\psi_k = e^{ik\cdot r}u_k(r)$ has Bloch form, and is a legitimate eigenstate of discrete translations. The eigenvalue problem is

$$(\,h(k) - E\,)\,u_k(r) = 0;\quad u_k(r) = u_k(r+R). \tag{22}$$

Just like in electromagnetism, this problem has a gauge redundancy:

$$u_k(r) \to e^{-iG\cdot r}\, u_k(r); \quad h(k) \to h(k+G) \tag{23}$$

since

$$(\, h(k+G) - E\,)\, e^{-iG\cdot r}\, u_k(r) = (\, e^{-iG\cdot r}\, h(k)\, e^{iG\cdot r} - E\,)\, e^{-iG\cdot r}\, u_k(r) = e^{-iG\cdot r}\,(\, h(k) - E\,)\, u_k(r) = 0. \tag{24}$$

Alternatively stated, $e^{-iG\cdot r}\, u_k(r)$ satisfies the periodic BC, and is an eigenstate of $h(k+G)$ with the same energy E. Note that $u_k(r) \to e^{-ik'\cdot r}\, u_k(r)$ for arbitrary k' does not preserve the boundary condition, and is NOT a legal gauge transformation. (One might compare this to Laughlin's cylinder argument, where in the presence of extended states, only gauge transformations that preserve periodicity are legal and don't change the energy spectrum.) I used 'gauge redundancy' instead of 'gauge symmetry' because these two states correspond to the same physical state. Though they are obviously different functions of position, one does not take the overlap $\langle u_1 \,|\, u_2 \rangle$. A physical state is an eigenfunction of H, not $h(k)$. In gauge 1,

$$(\, h(k) - E\,)\, u_k(r) = 0 \quad \Rightarrow \quad (\, H - E\,)\, e^{ik\cdot r}\, u_k(r) = 0 \quad \Rightarrow \quad \psi_1 = e^{ik\cdot r}\, u_k(r), \tag{25}$$

and in the second gauge

$$(\, h(k+G) - E\,)\, e^{-iG\cdot r}\, u_k(r) = 0 \quad \Rightarrow \quad (\, H - E\,)\, e^{i(k+G)\cdot r}\, e^{-iG\cdot r}\, u_k(r) = 0 \quad \Rightarrow \quad \psi_2 = e^{ik\cdot r}\, u_k(r). \tag{26}$$

Then it is clear $\langle \psi_1 \,|\, \psi_2 \rangle = 1$ up to a phase. In electromagnetism: we may solve a problem in two different gauges related by $A_1 = A_2 + \nabla\lambda$, and we know their wavefunctions are related by $\psi_1 \propto e^{i\lambda r}\, \psi_2$. Here: we may solve a problem in two different gauges (k and $k+G$), and we know their wavefunctions are related by $u_k \propto e^{iGr}\, u_{k+G}$.

Summary: the Hamiltonians at the base and end points are identical up to a legal gauge transformation. This implies that the basis chosen at k and $k+G$ may be arbitarily different. Suppose we chose a particular basis $\{\,|\, u_k \,\rangle\}$. We take an initial state $|\, u_{k_o} \,\rangle$ and performed parallel transport to $k_o + G$; Zak phase is defined as the extra phase acquired relative to $|\, u_{k_o+G} \,\rangle = e^{-iG\cdot r} u_{k_o}(r)$. The Abelian Zak phase is independent of the initial choice of basis $\{\,|\, u_k \,\rangle\}$, if we always calculate phase differences relative to $|\, u_{k_o+G} \,\rangle = e^{-iG\cdot r} u_{k_o}(r)$. In this sense we say the Zak phase is gauge-invariant.

Let us define a basis that satisfies

$$\psi_k = \psi_{k+G} \quad \Rightarrow \quad u_k(r) = e^{iGr}\, u_{k+G}(r) \tag{27}$$

as being in a periodic gauge. Since this is only a condition on the base and end points of a non-contractible loop, it is a rather weak constraint, and the basis within the loop is undetermined. A Bloch state transported around a non-contractible loop satisfies

$$|\, v_{k+G} \,\rangle = e^{-i\int E(t)dt}\, |\, u_{k+G} \,\rangle\langle\, u_{k+G} \,| \prod_{l}^{k+G\leftarrow k} P(l)\, |\, u_k \,\rangle = e^{-iGr}\, e^{-i\int E(t)dt}\, |\, u_k \,\rangle\, e^{-\int_k^{k+G} dk_\mu \left\langle u_k \,|\, \partial_\mu u_k \right\rangle}. \tag{28}$$

Sometimes the Zak phase factor is defined as

$$e^{-\int_k^{k+G} dk_\mu \left\langle u_k \,|\, \partial_\mu u_k \right\rangle}; \tag{29}$$

this lends to confusion, because the expression itself is gauge-covariant, depending on the relative phases at base and end points. (This is less confusing in the original context of Berry phase, where the Hamiltonians at base and end points are identical.) It is better to define the Zak phase factor as the unimodular eigenvalues of the operator

$$\hat{\mathcal{W}} = e^{iG\cdot r} \prod_{l}^{k+G\leftarrow k} P(l). \tag{30}$$

(i) The extra factor $e^{-iG\cdot r}$ is momentum-independent, thus the connection is single-valued: $A(k) = A(k+G)$.

(ii) In the tight-binding approximation, if the atoms are located only on Bravais lattice sites (i.e., $r = \sum_i n_i a_i$), then the factor is irrelevant. In a multi-atomic Bravais lattice, a gauge transformation matrix (analogous to e^{-iGr}) enters the tight-binding Wilson loop. A nontrivial gauge transformation tends to obscure the Berry phase in interference experiments. In an ideal experiment, the wavefunction is a coherent superposition of two wavepackets: $\psi = \psi_1 + \psi_2$. One wavepacket (ψ_1) stays put, and the other (ψ_2)

we send around the BZ. Assuming the dynamical phases are equal, the resultant wavefunction is an interference of two Bloch waves which differ by both a Berry phase and a gauge transformation:

$$\psi \sim \psi_1 + e^{i\phi_B}\, U_G^{-1}\, \psi_1. \tag{31}$$

(iii) The Zak phase is invariant under basis transformations that preserve the periodic gauge.

(iv) K/V connected the Berry-Zak phase to polarization:

$$\int d^d r\, r_i\, |W(r)|^2 \text{ `=' } \tfrac{i}{(2\pi)^d} \int d^d k \left\langle u_k \,|\, \partial_i u_k \right\rangle. \tag{32}$$

The equality is true only in the periodic gauge, as shown in Sec. A. This is problematic for a Chern insulator. K/V showed that the change in polarization due to an adiabatic change of the Hamiltonian ($\lambda = 0 \to 1$) is an integral of a curvature over a 2D manifold – it is a physical quantity. Assuming a differential gauge on this 2D manifold, we may apply Stoke's theorem to convert the change in polarization to a loop integral. In the periodic gauge, cancellations lead to

$$\Delta \boldsymbol{P} \text{ `=' } \boldsymbol{P}^{(1)} - \boldsymbol{P}^{(0)} \tag{33}$$

where the geometric-phase polarization is defined as

$$P_i^{(\lambda)} = \tfrac{1}{2\pi i} \int d^d k \left\langle u_k^{(\lambda)} \,|\, \partial_i u_k^{(\lambda)} \right\rangle. \tag{34}$$

Thus, the periodic gauge is implicit in the definition of this polarization.

(v) The fundamental difficulty of measuring the Zak phase in condensed matter systems is that we typically do not know, apriori, which state we begin with; the eigenstates of a Bloch Hamiltonian are dependent on the choice of origin. In spite of this ignorance, the difference between two Zak phases has been measured by Atala et. al. for cold atoms in an optical lattice. The assumption in this experiment: by rapidly changing the optical lattice, an eigenstate of Hamiltonian 1 with a certain choice of origin becomes an eigenstate of Hamiltonian 2 with the same choice of origin. The difference in Zak phases is then independent of the choice of origin.

III. THE DIFFERENCE BETWEEN BERRY PHASE AND BERRY PHASE FACTORS

Clearly the quantity

$$\Gamma_l = \int_l A \cdot dk \tag{35}$$

may be changed by integers using large gauge transformations. What physical meaning can it have? To motivate this problem, we might consider Dirac fermions in a 2D crystal, for which individually $\Gamma_l = \pm\pi$ if l wraps around a node. Is the sign of its half-vorticity well-defined?

No for an individual fermion, but the relative sign of two Dirac fermions is well defined in a globally smooth gauge.

I will take a topological definition of 'globally smooth' as the absence of vortices (or antivortices) in the wavefunction, where the determinantal phase winds (anti-winds). An example of a vortex is shown as a cross in Fig. 1. Such vortices are clearly singular, and are responsible for nontrivial Chern numbers in a gapped system.

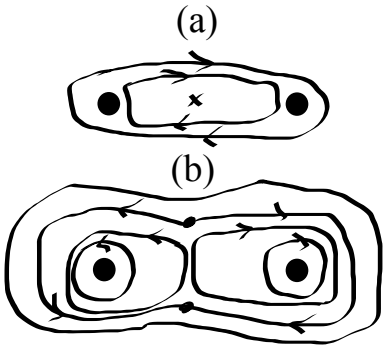

FIG. 1. (a) Illegal gauge transformation. (b) Legal gauge transformation.

Consider a single Dirac fermion. By a lauge gauge transformation around it, I can change its half-vorticity from $\pi \to -\pi$. Such a transformation is legal because I did not necessarily introduce a wavefunction vortex.

This is no longer true when I perform a large gauge transformation around two Dirac fermions, which is illustrated as a phase winding of the outer loop in Fig. 1(a). I argue that this transformation necessarily introduces a vortex. Suppose I tried to impose a smooth gauge everywhere within the outer loop. Let me do this in steps. In each step, I define a successively smaller loop whose phase winding is not very different the adjacent bigger loop. Clearly, each smaller loop must also wind. Eventually, the smallest loop encloses both Dirac nodes like a super-stretched rubber band. This rubber band is composed of two horizontal strips and two crescents. In the limit that the crescent shrinks to zero, smoothness dictates that the phase winds mostly along the horizontal strips. If we consider a loop l_0 within this rubber band, but not enclosing the two nodes, smoothness dictates it has the same winding as the horizontal strip – l_0 must enclose a vortex.

Denoting Γ_{12} as the Berry phase around two nodes, and Γ_1 as the Berry phase around the first node, a smooth gauge implies

$$\Gamma_{12} = \Gamma_1 + \Gamma_2, \quad \text{which is invariant under smooth gauge transformations!} \tag{36}$$

For example, a legal gauge tranformation such as in Fig. 1(b) changes both Γ_1 and Γ_2 while keeping their sum invariant. This is useful if in addition we have a reality constraint that quantizes Γ_{12} to be integer multiples of π. In 2D, this can be imposed by C_2T, and in 3D by $\mathcal{I}T$. Quantization implies that Γ_{12} cannot change under smooth deformations of the loop.

In a smooth gauge, $\Gamma_{12} \neq 0$ with a reality constraint implies an obstruction to gapping the system within l_{12}.

Smoothness dictates that Γ_{12} must be zero if we shrink l_{12} to nothing. The reality constraint (e.g. imposed by C_2T in graphene) quantizes Γ_{12}. There are only two ways to change Γ_{12}: (a) if l_{12} crosses a wavefunction vortex, which are absent by assumption, and (b) if l_{12} crosses a gapless point.

In a 2D crystal, the above relation implies that the relative sign of any pair of Dirac nodes is an invariant. Thus choosing the sign of one node as a convention, we would determine the signs of all other nodes within this convention.

If we had no reality constraint, there is no obstruction to Γ_{12} smoothly shrinking to zero. This is consistent with the fact that any Dirac node in the plane can just wander away from the plane, by a codimension argument. Consider an unreal Hamiltonian with an accidental Dirac node in a plane. An infinitesimally small loop around the node will pick up a real Berry phase of $\pm\pi$, because the effective Hamiltonian is linear in k. Alternatively stated, for such a small loop, half the flux is guaranteed to exit the northern hemisphere. For a larger loop, nonlinear terms enter and the Berry phase factor will deviate from reality, i.e., there is generically an uneven flux distribution between northern and southern hemispheres.

Appendix A: Geometric-phase proof

$$
\begin{aligned}
&\int d^d r\, r_i\, |W(r)|^2 \\
&= \frac{1}{N} \int d^d r \sum_k e^{-ikr}\, u_k(r)^*\, r_i \sum_q e^{iqr}\, u_q(r) \\
&= \frac{1}{Ni} \int d^d r \sum_k e^{-ikr}\, u_k(r)^* \sum_q (\,\partial_i\, e^{iqr}\,)\, u_q(r) \\
&= \frac{1}{Ni} \int d^d r \sum_k e^{-ikr}\, u_k(r)^* \sum_q \left[\,\partial_i\,(\,e^{iqr}\, u_q(r)\,) - e^{iqr}\,\partial_i\, u_q(r)\right] \\
&= \frac{i}{N} \int d^d r \sum_{k,q} e^{i(q-k)r}\, u_k(r)^*\,\partial_i\, u_q(r) \\
&= \frac{i}{N} \sum_R \int_{unit} d^d \bar{r} \sum_{k,q} e^{i(q-k)(R+\bar{r})}\, u_k(r)^*\,\partial_i\, u_q(r) \\
&= i \int_{unit} d^d \bar{r} \sum_k u_k(\bar{r})^*\,\partial_i\, u_k(\bar{r}) \\
&= i \int \frac{d^d k}{(2\pi)^d} \int_{unit} d^d \bar{r}\, u_k(\bar{r})^*\,\partial_i\, u_k(\bar{r}).
\end{aligned}
\tag{A1}
$$

The integral over $\bar{r}$ is over one unit cell. The 4th equality is true $\underline{\text{if}}$ $\psi_{k_i+2\pi} = \psi_{k_i}$.