Topological Phase Effects

J.M. Robbins
Basic Research Institute in the Mathematical Sciences
HP Laboratories Bristol
HPL-BRIMS-97-3
January, 1997

geometric phases;
Berry phase;
holonomy

Quantum eigenstates undergoing cyclic changes acquire a phase factor of geometric origin. This phase, known as the Berry phase, or the geometric phase, has found applications in a wide range of disciplines throughout physics, including atomic and molecular physics, condensed matter physics, optics and classical dynamics. In this article, the basic theory of the geometric phase is presented along with a number of representative applications.

The article begins with an account of the geometric phase for cyclic adiabatic evolutions. An elementary derivation is given along with a worked example of two-state systems. The implications of time-reversal are explained, as is the fundamental connection between the geometric phase and energy level degeneracies. We also discuss methods of experimental observation. A brief account is given of geometric magnetism; this is a Lorenz-like force of geometric origin which appears in the dynamics of slow systems coupled to fast ones.

A number of theoretical developments of the geometric phase are presented. These include an informal discussion of fibre bundles, and generalizations of the geometric phase to degenerate eigenstates (the nonabelian case) and to nonadiabatic evolution. There follows an account of applications. Manifestations in classical physics include the Hannay angle and kinematic geometric phases. Applications in optics concern polarization dynamics, including the theory and observation of Pancharatnam's phase. Applications in molecular physics include the molecular Aharonov-Bohm effect and nuclear magnetic resonance studies. In condensed matter physics, we discuss the role of the geometric phase in the theory of the quantum Hall effect.

TOPOLOGICAL PHASE EFFECTS

JM Robbins[†‡]

[†]Basic Research Institute in the Mathematical Sciences,
Hewlett Packard Laboratories Bristol, Filton Road, Stoke Gifford,
Bristol BS12 6QZ, UK

[‡]School of Mathematics, University Walk, Bristol BS8 1TW, UK

Abstract. Quantum eigenstates undergoing cyclic changes acquire a phase factor of geometric origin. This phase, known as the Berry phase, or the geometric phase, has found applications in a wide range of disciplines throughout physics, including atomic and molecular physics, condensed matter physics, optics, and classical dynamics. In this article, the basic theory of the geometric phase is presented along with a number of representative applications.

The article begins with an account of the geometric phase for cyclic adiabatic evolutions. An elementary derivation is given along with a worked example for two-state systems. The implications of time reversal are explained, as is the fundamental connection between the geometric phase and energy level degeneracies. We also discuss methods of experimental observation. A brief account is given of geometric magnetism; this is a Lorentz-like force of geometric origin which appears in the dynamics of slow systems coupled to fast ones.

A number of theoretical developments of the geometric phase are presented. These include an informal discussion of fibre bundles, and generalizations of the geometric phase to degenerate eigenstates (the nonabelian case) and to nonadiabatic evolution. There follows an account of applications. Manifestations in classical physics include the Hannay angle and kinematic geometric phases. Applications in optics concern polarization dynamics, including the theory and observation of Pancharatnam's phase. Applications in molecular physics include the molecular Aharonov-Bohm effect and nuclear magnetic resonance studies. In condensed matter physics, we discuss the role of the geometric phase in the theory of the quantum Hall effect.

Introduction
1 Basic Account of the Geometric Phase 4
1.1 *Quantum adiabatic theorem* 5
1.2 *The geometric phase* 5
1.3 *Two-state systems* 7
1.4 *Degeneracies* 9
1.5 *Time-reversal symmetry* 9
1.6 *Experimental observation* 10
1.7 *Geometric magnetism* 11
2 Theoretical Developments 13
2.1 *Geometric phase as holonomy* 13
2.1.1 *Fibre bundles.* 13
2.1.2 *Connections and curvature.* 14
2.1.3 *Geometric phase revisited.* 15
2.1.4 *Chern classes.* 15
2.2 *Degenerate eigenstates and nonabelian geometric phases* 16
2.3 *Nonadiabatic evolution* 19
2.4 *Geometric metric tensor* 21
2.5 *Beyond adiabaticity* 23
2.5.1 *Corrections to the geometric phase* 23
2.5.2 *The geometric amplitude* 25
3 Classical Systems 26
3.1 *The Hannay angle* 26
3.1.1 *The classical adiabatic theorem.* 26
3.1.2 *Basic account of the Hannay angle.* 27
3.1.3 *Examples.* 29
3.1.4 *Semiclassical limit.* 30
3.1.5 *Classical reaction forces.* 31
3.2 *Symmetry and kinematic geometric phases* 31
3.2.1 *Falling cats and swimming paramecia.* 31
3.2.2 *Geometric phases and symmetry reduction.* 33
3.2.3 *Optimization and control.* 35
4 Optics 36
4.1 *Geometric phase of coiled light.* 36
4.2 *Pancharatnam's phase.* 38
5 Molecular and Condensed Matter Physics 40
5.1 *Pseudorotation in triatomic molecules* 40
5.2 *Nuclear magnetic resonance* 42
5.3 *The quantum Hall effect* 45
Glossary 49
References 49
Further Reading 51

Introduction

In a paper published in 1984, Sir Michael Berry (Berry 1984) found that a quantum system undergoing adiabatic evolution acquires a phase factor of purely geometrical origin. This discovery, now called the Berry phase, or the geometric phase, initiated a tremendous surge of research in a range of disciplines across physics. Berry and subsequent workers quickly established connections to a number of apparently disparate phenomena, including amongst others the Aharonov Bohm effect (Aharonov and Bohm 1959), its molecular physics analogue (Mead and Truhlar 1979), polarization optics (Pancharatnam 1956), the quantum Hall effect (Thouless *et al* 1982), and the Foucault pendulum (Eco 1989). Beyond providing a unified description of these phenomena, the geometric phase led to predictions of new effects, subsequently observed experimentally, and a number of substantial extensions and generalizations. Among these, the discovery by Hannay (1985) of an analogous effect in classical mechanics, the Hannay angle, stimulated much new research in purely classical physics.

The mathematical phenomenon which underlies the geometric phase is holonomy. Holonomy describes transformations in a quantity induced by cyclic changes in the variables which control it. An example from geometry is the rotation of vectors parallel transported along closed paths on a curved surface. The mathematics of holonomy and curvature has long played a role in physics through general relativity and quantum field theory. With Berry's work came the appreciation of a broader scope of application. Whenever a physical system is divided into parts and one attempts to describe a subsystem in isolation, the influence of the excluded part of the system is manifested through geometric phase effects. The pervasiveness througout physics of this reductive procedure accounts for the ubiquity of the geometric phase. Introducing a collection of major papers in the field, Shapere and Wilczek (1989a) wrote, "We believe that the concept of a geometric phase, repeating the history of the group concept, will eventually find so may realizations and applications in physics that it will repay study for its own sake, and become part of the *lingua franca*."

An introductory account of the geometric phase is given in Section 1. This is followed in Section 2 by a discussion of several theoretical developements. Applications to classical systems, optics, and molecular and condensed matter physics are given in Sections 3 -5. The topics chosen for discussion are meant to be representative rather than comprehensive. Among several omissions, we mention the theory of fractional statistics, which is covered extensively in the volume by Shapere and Wilczek (1989a), and applications to elementary particle physics, which are reviewed by Aitchison (1987). The historical development of the subject is discussed by Berry (1990a).

1. Basic Account of the Geometric Phase

The geometric phase was discovered by Sir Michael Berry in 1983 in the course of a critical reexamination of the adiabatic theorem in quantum mechanics. This theorem says, in essence, that if a quantum system is changed sufficiently slowly, it does not undergo transitions. Thus a system initially in an eigenstate of a slowly changing Hamiltonian remains in the evolved eigenstate. Berry considered the overall phase accumulated by the changing eigenstate and found an unexpected contribution. This extra phase is of a geometrical nature, in that it does not depend on the rate at which the adiabatic process is carried out.

One way that a slowly varying Hamiltonian can arise physically is when the system under consideration interacts with another, much slower system. This leads to the consideration of the quite general problem of two coupled systems evolving on different time scales. In this context, the geometric phase describes an effect on the fast system produced by the slow one. As action in physics is accompanied by reaction, there is a reciprocal effect on the slow system generated by the fast. This takes the form of an effective vector potential in the slow Hamiltonian; the associated Lorentz-like force is called geometric magnetism. The same geometry underlies both geometric magnetism and the geometric phase.

This geometrical structure is intimately connected to degeneracies in the energy spectrum. A typical Hamiltonian has a nondegenerate spectrum (excepting systematic degeneracies due to symmetry, which are discussed in Section 2.2). Varying a single parameter (eg, time itself in the case of a time dependent Hamiltonian) is unlikely to produce any degeneracies; typically two or three parameters (the number depends on whether time-reversal symmetry is present or not) must be adjusted independently in order to find one. Nevertheless, the presence of degeneracies in Hamiltonians which are, in some sense, near the Hamiltonians of the adiabatic process, are largely and sometimes wholly responsible for the geometric phase. This connection underlies several applications, including to the quantum Hall effect (Section 5.3).

1.1. Quantum adiabatic theorem

Adiabatic invariants are quantities which are nearly conserved when the environment of an isolated system is slowly changed (slowly means in comparison with the system's internal dynamics). Adiabatic theorems provide estimates as to how accurately this conservation is respected. In quantum mechanics, the adiabatic invariants are the probabilities $P_n = |\langle n|\psi\rangle|^2$ to be in energy eigenstates. These are constant for stationary Hamiltonians, but vary if H is time-dependent. The quantum adiabatic theorem states that as dH/dt goes to zero, so do the probabilities of transitions between eigenstates (the rate depends on how smooth dH/dt is), so that the P_n are nearly constant. Therefore, within the adiabatic approximation, dynamics under a slowly varying Hamiltonian $H(t)$ is essentially no more complicated than under a stationary one. In particular, if $|\psi\rangle$ begins in an eigenstate, it remains in one (nonadiabatic

transitions are discussed in Section 2.5), and its evolution is determined up to a phase factor.

We shall now examine this phase. Suppose that at $t = 0$, $|\psi\rangle$ is in the nth eigenstate $|n\rangle$ (assumed to be nondegenerate) of H. For its subsequent evolution we may write

$$|\psi(t)\rangle \approx \exp\left(-i/\hbar \int_0^t E_n(\tau)d\tau\right) |n(t)\rangle. \tag{1}$$

In so doing we have explicitly identified part of the overall phase, the time integral of the frequency $E_n/\hbar$, which is called the *dynamical phase*. It is just the phase one would expect for a solution of the Schrödinger equation. Any remaining phase is contained implicitly in the eigenstate $|n(t)\rangle$. This phase is determined by requiring that $\langle n|(i\hbar\partial_t - H)\,\psi\rangle$ vanish (it would automatically if $|\psi(t)\rangle$ were an exact solution of the Schrödinger equation). This requirement leads to the condition

$$\langle n|\dot{n}\rangle = 0. \tag{2}$$

which determines the phase of $|n(t)\rangle$ given an initial choice at $t = 0$.

1.2. The geometric phase

What Berry discovered is that condition (2) yields a path dependent phase factor. This phase factor is particularly interesting when the Hamiltonian is cycled, ie made to return to itself after some time. Then it depends only on the cycle itself, and not, in contrast to the dynamical phase, on the rate at which the Hamiltonian is cycled. For this reason it is called, as well as the *Berry phase*, the *geometric phase*.

To discuss this further, it helps to formulate the problem slightly differently. Suppose the time dependence of the Hamiltonian is expressed through its dependence on some external parameters $\mathbf{R} = (X, Y, Z)$, which are made to vary slowly in time along a path $\mathbf{R}(t)$. Then the eigenstates $|n(\mathbf{R})\rangle$ depend on $\mathbf{R}$ as well, and (assuming no degeneracies) are determined for each $\mathbf{R}$ up to an overall phase. Let us fix this phase so that it varies smoothly but otherwise arbitrarily with $\mathbf{R}$. By fixing the phase independently of the dynamics, we may no longer satisfy the condition (2) along every path $\mathbf{R}(t)$. We therefore allow for an additional phase factor $\exp(i\gamma_n(t))$ in the adiabatic ansatz (1), which becomes

$$|\psi(t)\rangle \approx \exp\left(-i/\hbar \int_0^t E_n(\mathbf{R}(\tau))\,d\tau\right) \exp(i\gamma_n(t))\, |n(\mathbf{R}(t))\rangle. \tag{3}$$

The condition $\langle n|(i\hbar\partial_t - H)|\psi\rangle = 0$ now yields an evolution equation for γ_n. This may be written in the form $\dot{\gamma}_n(t) = -\mathbf{A}_n(\mathbf{R}(t)) \cdot \dot{\mathbf{R}}(t)$, where $\mathbf{A}_n(\mathbf{R})$ is a vector field in parameter space given by

$$\mathbf{A}_n(\mathbf{R}) = \mathrm{Im}\,\langle n(\mathbf{R})|\,\nabla n(\mathbf{R})\rangle. \tag{4}$$

or equivalently, in terms of eigenfunctions $\psi_n(\mathbf{r},\mathbf{R}) = \langle \mathbf{r}|n(\mathbf{R})\rangle$, by $\mathbf{A}_n(\mathbf{R}) = \mathrm{Im}\int \psi_n^*(\mathbf{r},\mathbf{R})\boldsymbol{\nabla}\psi_n(\mathbf{r},\mathbf{R})\,d^3r$ (here $\boldsymbol{\nabla}$ acts on the parameters $\mathbf{R}$).

If the Hamiltonian is cycled round a closed curve C in parameter space, then $|n(t)\rangle$ acquires a geometric phase

$$\gamma_n(C) = -\int_0^T \mathbf{A}_n \cdot \dot{\mathbf{R}}\,dt = -\oint_C \mathbf{A}_n \cdot d\mathbf{R} \tag{5}$$

Its expression as a line integral reveals its geometrical character: the geometric phase depends only on the cycle C, and not on the manner in which the cycle is traversed in time. Using Stokes' theorem, γ_n can be expressed as

$$\gamma_n(C) = -\int_S \mathbf{V}_n(\mathbf{R}) \cdot d\mathbf{S}, \tag{6}$$

the flux of the vector field $\mathbf{V}_n = \boldsymbol{\nabla} \times \mathbf{A}_n$ through a surface S in parameter space bounded by C. $\mathbf{V}_n$ is given by

$$\mathbf{V}_n = \mathrm{Im}\,\langle \boldsymbol{\nabla} n| \times |\boldsymbol{\nabla} n\rangle. \tag{7}$$

From (6) it is evident that only cycles which enclose nonzero area can have nontrivial geometric phases. For example, $\gamma_n(C)$ vanishes for cycles consisting of a path and its retrace.

It is easy to see that the geometric phase does not depend on the choice of phase for $|n(\mathbf{R})\rangle$. Indeed, under the change of phase

$$|n(\mathbf{R})\rangle \to \exp i\chi(\mathbf{R})|n(\mathbf{R})\rangle, \tag{8}$$

$\mathbf{A}_n$ transforms according to $\mathbf{A}_n(\mathbf{R}) \to \mathbf{A}_n(\mathbf{R}) + \boldsymbol{\nabla}\chi(\mathbf{R})$; this shift in $\mathbf{A}_n$ by a perfect gradient leaves its integral round a closed loop, $\gamma_n(C)$, and its curl, $\mathbf{V}_n$, unchanged. Thus, the geometric phase is an intrinsic property of the family of eigenstates $|n(\mathbf{R})\rangle$.

It is useful to draw an analogy with electromagnetism, in which $\mathbf{A}_n$ corresponds to the vector potential and $\mathbf{V}_n$ to the magnetic field. The geometric phase γ_n corresponds to a magnetic flux. The change of phase (8) corresponds to a gauge transformation, which shifts the vector potential by a gradient and leaves the magnetic field unchanged.

Substituting the first-order perturbation expansion

$$|\boldsymbol{\nabla} n\rangle = \sum_{m\neq n} \frac{\langle m|\boldsymbol{\nabla} H|n\rangle}{(E_n - E_m)}|m\rangle - \langle n|\boldsymbol{\nabla} n\rangle\,|n\rangle \tag{9}$$

into (7) leads to a useful formula,

$$\mathbf{V}_n = \mathrm{Im}\sum_{m\neq n} \frac{\langle n|\boldsymbol{\nabla} H|m\rangle \times \langle m|\boldsymbol{\nabla} H|n\rangle}{(E_n - E_m)^2}. \tag{10}$$

In this form, the independence of $\mathbf{V}_n$ on the eigenstate phases is manifest. Another formula with this property is

$$\mathbf{V}_n = \mathrm{Im}\,\mathrm{Tr}\,(\rho_n \boldsymbol{\nabla}\rho_n \times \boldsymbol{\nabla}\rho_n), \tag{11}$$

where $\rho_n = |n\rangle\langle n|$ is the density operator. It too is useful in applications.

One of the applications discussed by Berry (1984) is to the Aharonov-Bohm effect (Aharonov and Bohm 1959). The Aharonov-Bohm effect describes the dependence of quantum mechanics on electromagnetic fields in regions of space which are physically inaccessible. This dependence is mediated through the vector and scalar potential, whose relation to the electric and magnetic fields is nonlocal. The most celebrated example concerns an electron moving in the vicinity of an impenetrable perfect solenoid of infinite extent. Since outside the solenoid the magnetic field is zero, the classical mechanics of the electron is insensitive to any magnetic flux inside the solenoid. But the quantum mechanical behaviour of the electron does depend on the interior flux, as Aharonov and Bohm showed by calculating its scattering cross-section (a prediction subsequently experimentally confirmed by Chambers (1960)).

The Aharonov-Bohm effect can be understood in terms of interference between de Broglie waves passing on opposite sides of the solenoid. The phase difference between the two waves includes the integral

$$\gamma_{AB} = -\frac{e}{\hbar c}\oint_C \mathbf{A}^{(s)}(\mathbf{R})\cdot d\mathbf{R} \tag{12}$$

of the vector potential $\mathbf{A}^{(s)}$ of the solenoid round a loop C enclosing it. (In (12), c is the speed of light and $-e$ is the electron charge.) The Aharonov-Bohm phase γ_{AB} is just 2π times the magnetic flux in the solenoid measured in units of the flux quantum hc/e.

It may be regarded as a geometric phase in the following manner. We imagine the electron confined to a large box, centred at $\mathbf{R}$, which is not penetrated by the solenoid. The eigenstates of the electron $|n(\mathbf{R})\rangle$ depend parametrically on $\mathbf{R}$; specifically, changes in $\mathbf{R}$ induce gauge transformations in the eigenfunctions. One finds that $\mathrm{Im}\langle n(\mathbf{R})|\nabla n(\mathbf{R})\rangle = e/\hbar c\,\mathbf{A}^{(s)}(\mathbf{R})$, so that in this case the geometric vector potential (4) coincides with the physical vector potential of the solenoid. On being transported (not necessarily adiabatically in this case) round the solenoid, the electron eigenstates acquire a geometric phase which is precisely the Aharonov-Bohm phase.

1.3 Two-state systems

Two-state systems provide a canonical example of the geometric phase. The requisite calculations are easily carried out, and the main features of the general case are represented. The parameterized family of Hamiltonians can be written in the form

$$H(\mathbf{R}) = E(\mathbf{R})\mathbf{1} + \mathbf{F}(\mathbf{R})\cdot\boldsymbol{\sigma}, \tag{13}$$

where 1 is the two-by-two identity matrix and $\boldsymbol{\sigma}$ are the Pauli spin matrices. The term $E(\mathbf{R})\mathbf{1}$ shifts the energy by a constant and affects neither the eigenstates nor their geometric phases. Both of these are determined by the field $\mathbf{F}(\mathbf{R})$.

An illustrative example is a spin-1/2 particle in a uniform magnetic field $\mathbf{B}(\mathbf{R}) = \mathbf{R}$. The parameters $\mathbf{R}$ are just the components of the magnetic field, so that

$\mathbf{F}(\mathbf{R}) = \mu\mathbf{R}$, where μ is the magnetic moment. The eigenstates

$$|+(\mathbf{R})\rangle = \begin{pmatrix} \cos\theta/2 \\ e^{i\phi}\sin\theta/2 \end{pmatrix}, \quad |-(\mathbf{R})\rangle = \begin{pmatrix} -\sin\theta/2 \\ e^{i\phi}\cos\theta/2 \end{pmatrix} \tag{14}$$

depend on the polar angles (θ, ϕ) of $\mathbf{R}$, and the energy levels $E_\pm(\mathbf{R}) = \pm\mu R$ on its magnitude. Straightforward calculation using (4) and (7) yields

$$\mathbf{A}_\pm(\mathbf{R}) = \mp\tfrac{1}{2}\frac{1-\cos\theta}{R\sin\theta}\hat{\boldsymbol{\phi}}, \quad \mathbf{V}_\pm(\mathbf{R}) = \pm\tfrac{1}{2}\frac{\mathbf{R}}{R^3}. \tag{15}$$

Thus, $\mathbf{A}_\pm$ and $\mathbf{V}_\pm$ are respectively the vector potential and magnetic field of a magnetic monopole of strength $\pm 2\pi$ at the origin. Under the influence of a slowly varying magnetic field $\mathbf{R}(t)$ taken through a cycle C, a spin initially in an up state, say, returns to itself with a dynamical phase $-\mu/\hbar\int_0^T R dt$ and a geometric phase $\gamma_+(C) = -\int_S \mathbf{V}_+ \cdot d\mathbf{S}$. Gauss's law implies that

$$\gamma_+(C) = -\tfrac{1}{2} \times \text{solid angle subtended by } C. \tag{16}$$

This spin-1/2 geometric phase has been observed in neutron spin rotation experiments by Bitter and Dubbers (1987).

A special case is when C is confined to a plane containing the origin. Then its solid angle is 2π times the number of times m the origin is enclosed, and the corresponding phase factor is $(-1)^m$. This corresponds to the well-known sign change of spin-1/2 particles under rotations through 2π.

It is straightforward to generalize these results to the case of arbitrary $\mathbf{F}(\mathbf{R})$ in (13). The flux field is then given by

$$\mathbf{V}_\pm(\mathbf{R}) = \mp\tfrac{1}{2}\epsilon_{ijk}\frac{F_i\nabla F_j \times \nabla F_k}{F^3} \tag{17}$$

(ϵ_{ijk} is the antisymmetric Levi Cevita symbol), and it has monopole sources at points where $\mathbf{F}(\mathbf{R})$ vanishes. The monopoles have strength $\pm 2\pi\sigma$, where σ is the sign of the (assumed nonvanishing) determinant $\det \partial F_i/\partial R_j$ evaluated where $\mathbf{F} = 0$.

As is familiar from Dirac's analysis of magnetic monopoles (Dirac 1931), attached to the monopole there is necessarily a string of singularities of the vector potential. The position of the string depends on the choice of gauge, and Dirac's quantization condition for electric charge follows from demanding that all gauges be physically indistinguishable. In the context of the two-state geometric phase, the monopole strings of $\mathbf{A}_\pm(\mathbf{R})$ lie along discontinuities in the phases of $|\pm(\mathbf{R})\rangle$. When the eigenstates are chosen as in (14), these discontinuities lie along the negative z axis. By varying the phase of the eigenstates (as in (8)), the string can be made to pass in and out of a given cycle C. As it does, the geometric phase $\gamma_\pm(C)$ changes by $\pm 2\pi$. Therefore, the geometric phase factor $\exp i\gamma_\pm(C)$ remains unchanged, and there are no observable consequences.

1.4. Degeneracies

In the preceding example of a spin-1/2 particle in a magnetic field, the flux fields $\mathbf{V}_\pm(\mathbf{R})$ are generated by point sources of strength $\mp 2\pi$ at the origin $\mathbf{R} = 0$. What makes the origin special is that it is the only point in parameter space where the energy levels are degenerate. This relationship between the geometric phase and energy level degeneracies turns out to be quite general.

In the absence of symmetries, a typical Hamiltonian in a parameterized family has a nondegenerate spectrum. However, at certain points in parameter space, belonging to the so-called *degenerate set*, two (or more) energy levels become degenerate. In the absence of time reversal symmetry (the time-reversal symmetric case is discussed in the following section), the dimension of the degenerate set is typically three less than the dimension of parameter space itself (von Neumann and Wigner 1929). Thus, for the three-parameter families $H(\mathbf{R})$ we have been considering, degeneracies occur at isolated points. We have already seen this to be the case in the spin-1/2 example. (Note that there, the magnetic field is responsible for breaking time-reversal symmetry.)

More generally, let us consider the behaviour of the flux field $\mathbf{V}_n(\mathbf{R})$ in the neighbourhood of a point of degeneracy $\mathbf{R}_*$ between two eigenstates, say the nth and $n-1$st, of an arbitrary family of Hamiltonians. As $E_n(\mathbf{R}) - E_{n-1}(\mathbf{R})$ is small near $\mathbf{R}_*$, perturbation theory (cf (9)) implies that the dominant component of $|\nabla n\rangle$ lies along the ket $|n-1\rangle$. Therefore, for the purposes of approximating $\mathbf{V}_n$ (which is defined in terms of $|\nabla n\rangle$), the Hamiltonian can be restricted to the nearly-degenerate subspace spanned by the states $|1\rangle \stackrel{\text{def}}{=} |n(\mathbf{R}_*)\rangle$ and $|2\rangle \stackrel{\text{def}}{=} |n-1(\mathbf{R}_*)\rangle$. One obtains in this way a two-state system of the form (13), where the field $\mathbf{F}(\mathbf{R})$ is given by

$$F_x = \mathrm{Re}\,\langle 1|H(\mathbf{R})|2\rangle,\quad F_y = -\mathrm{Im}\,\langle 1|H(\mathbf{R})|2\rangle,\quad F_z = \frac{\langle 1|H(\mathbf{R})|1\rangle - \langle 2|H(\mathbf{R})|2\rangle}{2}. \tag{18}$$

In the neighbourhood of $\mathbf{R}_*$, $\mathbf{V}_n(\mathbf{R})$ is given approximately by (17). There it has a single monopole source at the point of degeneracy $\mathbf{R}_*$ of strength $\mp 2\pi$. We may apply a similar analysis in the neighbourhood of each degeneracy of $n(\mathbf{R})$ to obtain

$$\nabla \cdot \mathbf{V}_n = \sum_{\mathbf{R}_*} \pm 2\pi \delta^3(\mathbf{R} - \mathbf{R}_*). \tag{19}$$

Thus, the flux field for the nth eigenstate has monopole sources of strength $\pm 2\pi$ at points where $n(\mathbf{R})$ is degenerate. (The sign of the monopole depends on whether the degeneracy is with the state above or below, and also on the sign of the determinant $\det \partial F_i(\mathbf{R}_*)/\partial R_j$.) $\mathbf{V}_n$ may in addition have a purely divergenceless contribution.

It follows that the flux of $\mathbf{V}_n$ through a closed surface in parameter space is 2π times the number of degeneracies, counted with their signs, contained within.

1.5. Time reversal symmetry

The presence of time-reversal symmetry imposes constraints on the geometric phase. These are strongest when the eigenstates themselves are time-reversal symmetric. We

consider such cases first. Then the wave functions $\psi_n(\mathbf{r};\mathbf{R}) = \langle \mathbf{r}|n(\mathbf{R})\rangle$ can be chosen to be real, and the geometric phase factor is necessarily a sign ± 1, so that γ_n is an integer multiple of π. In this case, the geometric phase depends only on the topology of the closed curve, not on its geometry. As C is deformed, unless $\gamma_n(C)$ changes discontinuously, it necessarily remains constant. It can change (discontinously) when C passes through degeneracies (points $\mathbf{R}_*$ where $|n\rangle$ is degenerate). For systems with time-reversal symmetry, the dimension of this degenerate set is typically two less than the dimension of parameter space (von Neumann and Wigner 1929), so that it consists of lines in three-dimensional parameter space, or points in two-dimensional parameter space. The geometric phase $\gamma_n(C)$ is π times the number of degeneracies enclosed by C, counted with the appropriate sign. An important consequence is that the flux field $\mathbf{V}_n$ has delta-function singularities at the energy level degeneracies and is zero elsewhere.

The example of a spin 1/2 in a magnetic field (cf Section 1.3) provides an illustration if we fix Y (the y component of the magnetic field) to be zero. Then the spin Hamiltonian $\frac{1}{2}\mu\mathbf{B}.\boldsymbol{\sigma}$ is real symmetric rather than complex Hermitian, and we can take its eigenvectors to be real. The parameter space is the XZ plane, and the solid angles subtended by curves in this plane are multiples of 2π. The geometric phase $\gamma_\pm(C)$ is just $\pm\pi$ times the number of times C encircles the origin. Another example concerns the electronic states of triatomic molecules X_3 (cf Section 5.1).

In cases where the eigenstates are themselves not time-reversal symmetric, and instead occur in pairs related by time-reversal, the constraints on the geometric phases are weaker. They are no longer necessarily multiples of π (ie, the phase factors are no longer necessarily real). However, geometric phases for states related by time-reversal necessarily have opposite signs.

1.6. Experimental observation

Since the physical properties of a quantum system does not depend on its overall phase, one could wonder whether the geometric phase is observable. Numerous experiments have shown that it is.

Most direct experimental studies of the geometric phase belong to one of three types. The first is the 'one state, two Hamiltonian' experiment. An ensemble of states initially in an energy eigentate is coherently divided into two streams. The first evolves under a constant Hamiltonian, while the Hamiltonian driving the second is cyclically varied. The two streams are then recombined and allowed to interfere, and their relative phase determined. The relative phase includes both dynamical and geometric contributions, and to extract the geometric component the dynamical phase must either be made to vanish, or else subtracted off explicitly. Examples include the optical polarization experiments discussed in Section 4.2.

The second type is the 'two state, one Hamiltonian' experiment. Here an initial state $|\psi_i\rangle$ is prepared in a superposition $\alpha_i|m\rangle + \beta_i|n\rangle$ of energy eigenstates. The Hamiltonian governing the evolution is taken through an adiabatic cycle. The final

state $|n;f\rangle = \alpha_f|m\rangle + \beta_f|n\rangle$ is physically distinguishable from the initial one, as the relative phase of the superposed eigenstates has changed. Part of this (observable) change in the relative phase involves the difference $\gamma_m - \gamma_n$ in geometric phases. The observation of optical activity in coiled fibres by Tomita and Chiao (1986) (see Section 4.1) provides an example of this second type.

In the third type of experiment, the adiabatic cycle is performed not once but repeatedly, with period T. With each cycle, $|n\rangle$ acquires a phase $\alpha_n = -\int_0^T E_n(\mathbf{R}(t))dt/\hbar + \gamma_n$, so that there appears in its mean angular frequency $\omega_n = \alpha_n/T$ a geometric shift $\Delta\omega_n = \gamma_n/T$. This frequency shift can be detected in spectroscopic measurements, and is the basis for nuclear magnetic resonance studies of the geometric phase (cf Section 5.2).

1.7. Geometric magnetism

So far we have been regarding the evolution of the parameters $\mathbf{R}$ as externally prescribed. Now let us suppose the parameters of the Hamiltonian $H(\mathbf{R})$ are dynamical variables in their own right. For example, they could be the coordinates of a heavier (and therefore slower) system to which the original one, the fast system, is coupled. For definiteness, we take the Hamiltonian of the coupled system to be of the form $P^2/2M - H(\mathbf{R})$, where $\mathbf{P}$ is the momentum of the slow system, and the dependence on the fast coordinates and momenta is left implicit in $H(\mathbf{R})$. Both systems are to be treated quantum mechanically. The geometric phase describes an effect on the fast system due to the slow. There is a corresponding influence on the slow system due to the fast.

This is the natural setting for the Born-Oppenheimer approximation, in which the fast and slow systems are the electrons and nuclei of a molecule. (For consistency with the preceding discussion, we denote the nuclear coordinates by a single vector $\mathbf{R}$; it is easy to adapt the notation to accomodate more than one nucleus.) By assuming that the electronic state adjusts rapidly to changes in the nuclear configuration, one is led to look for approximate eigenfunctions of the form $\Psi_n(\mathbf{r},\mathbf{R}) = \phi_n(\mathbf{R})\psi_n(\mathbf{r},\mathbf{R})$, where the electronic wavefunction $\psi_n(\mathbf{r},\mathbf{R})$ is an eigenstate with energy $E_n(\mathbf{R})$ of the electronic Hamiltonian $H(\mathbf{R})$, in which the nuclear coordinates appear as parameters. The nuclear wavefunction $\phi_n(\mathbf{R})$ is an eigenstate of the nuclear Hamiltonian $-\hbar^2\nabla^2/2M + E_n(\mathbf{R})$; the electronic energy plays the role of a potential. The Born-Oppenheimer approximation consists of neglecting the relatively slow dependence of the electronic wave function $\psi_n(\mathbf{r},\mathbf{R})$ with the nuclear coordinates. (It is essentially the adiabatic approximation (1) adapted to coupled systems.)

This was the context for the seminal work of Mead and Truhlar (1979). The primary objective of this work, which anticipated certain aspects of the geometric phase, was to derive corrections to the Born-Oppenheimer approximation. Such corrections appear when the momentum operator $\mathbf{P} = -i\hbar\nabla$ is applied to the putative eigenstate $\phi_n(\mathbf{R})\,n(\mathbf{R})\rangle$. The result, $-i\hbar(\nabla\phi_n|n\rangle + \phi_n|\nabla n\rangle)$, is no longer proportional to the electronic eigenstate $|n(\mathbf{R})\rangle$. However, if we neglect electronic

components orthogonal to $|n\rangle$, we obtain $-i\hbar(\nabla c_n - \langle n|\nabla n\rangle)|n\rangle$, which is. The effect on the nuclear Schrödinger equation is to redefine the momentum operator.

$$\mathbf{P}\phi \stackrel{\text{def}}{=} (-i\hbar\nabla - \mathbf{A}_n)\phi, \tag{20}$$

where $\mathbf{A}_n(\mathbf{R})$ is given by (4). Thus, the field whose line integral gives the electronic geometric phase appears as a vector potential in the nuclear Hamiltonian. There it corresponds to an effective magnetic field $\mathbf{V}_n = \nabla \times \mathbf{A}_n$ which modifies the nuclear dynamics. This is called *geometric magnetism*.

The situation described above is quite general. Whenever there are two coupled systems evolving on different time scales, we can obtain an approximate description of the slow system by averaging over the fast motion. There appears in the slow dynamics an effective vector potential and associated magnetic field. The vector potential and magnetic field are precisely the fields whose line integrals and fluxes describe the geometric phase. (In cases where the coupling involves not only the slow coordinates but the slow momenta as well, this description has to be modified somewhat.)

Geometric magnetism is responsible for a variety of phenomena. Mead and Truhlar (1979) analysed its effects in nuclear rotation vibration spectra, which have since been observed experimentally (cf Section 5.1). These effects can be nontrivial, even when time-reversal symmetry is present (cf Section 1.5), due to the change in sign of the electronic wavefunction $\psi_n(\mathbf{r}, \mathbf{R})$ when the nuclear coordinates $\mathbf{R}$ are carried round a degenerate configuration. Like the magnetic field of the Aharonov-Bohm effect (Section 1.2), the flux field $\mathbf{V}_n(\mathbf{R})$ in this case vanishes everywhere except at these degenerate configurations, where it is singular. Mead (1980) has called this the "molecular Aharonov-Bohm effect".

Geometric magnetism survives in the classical limit. When this limit is applied to the slow system alone, it leads to geometric phase modifications of the semiclassical (Bohr-Sommerfeld) quantization conditions. These are discussed by Kuratsuji and Iida (1985), Wilkinson (1984), and Littlejohn and Flynn (1991).

The situation where both the fast and slow systems are fully classical is discussed in Section 3.1.5. Here let us mention a simple but compelling illustration discussed by Li and Mead (1992). Let us consider the Born-Oppenheimer approximation applied to an atom rather than a molecule. The electronic energy levels $E_n(\mathbf{R})$ depend on the atom's centre-of mass $\mathbf{R}$, and appear as potentials in the equations of motion $M\ddot{\mathbf{R}} = -\nabla E_n + \mathbf{F}(\mathbf{R})$, along with any external force $\mathbf{F}(\mathbf{R})$. In the absence of external potentials, the energy levels are independent of $\mathbf{R}$, and $\mathbf{F} \equiv 0$: the centre-of-mass behaves as a free particle. Now suppose the atom moves in a uniform magnetic field $\mathbf{B}$. The energy levels are still $\mathbf{R}$ independent (by translation invariance), but there now appears in the centre-of-mass dynamics a Lorentz force, so that $M\ddot{\mathbf{R}} = \mathbf{F} = +Ze\dot{\mathbf{R}} \times \mathbf{B}$. The predictions of this analysis are spectacularly unphysical: walking through the earth's magnetic field, you would experience a vertical force sufficient to hurl you into the air. What is missing from this analysis is the geometric magnetism from the electrons: the field $\mathbf{V}_n(\mathbf{R})$ introduces into the centre-of-mass dynamics a second

Lorentz force which either exactly or nearly (in case the electronic eigenstate has a nonzero magnetic moment) cancels the first.

2. Theoretical Developments

A number of theoretical developments of the geometric phase are described. These include its mathematical setting in terms of holonomy under parallel transport, as well as several generalizations which remove the assumptions of nondegeneracy and adiabaticity.

2.1. Geometric phase as holonomy

Objects rigidly transported round a closed path in a curved space can return with a different orientation — this is *holonomy*. Here is a simple (and surely physical) example. Hold your right arm outstretched before you with your palm facing down. Keeping your wrist rigid, swing your arm into your chest, then raise it clockwise in front of you, and finally swing it away downwards until it is outstretched before you again. Your arm has returned to its original position, but your hand has turned clockwise through 90°.

The mathematical setting for this phenomenon is the theory of connections on fibre bundles. What follows in Sections 2.1.1– 2.1.4 is an informal account of the relation between this theory and the geometric phase. No attempt is made at mathematical precision: the intention is to convey a sense of the principal ideas involved. Two references on fibre bundles oriented towards physicists are (Eguchi et al 1980) and (Nash and Sen 1983).

2.1.1. Fibre bundles. A *fibre bundle* E is a space, or manifold, composed of two smaller ones, the *base manifold* B and the *fibre* F. The fibre bundle is constructed by attaching a copy of the fibre to each point of the base in a particular way. The simplest case, a *trivial* fibre bundle, is where E is just the cartesian product $B \times F$ of the base with the fibre. A general fibre bundle cannot be so expressed. While it is always possible to partition it into smaller pieces which are themselves cartesian products $M_\alpha \times F$ of subsets of the base with the fibre, these pieces may fit together in an interesting and nontrivial way, which reflects the topological properties of the bundle.

The configuration space E of an outstretched arm can be described as a fibre bundle. In this case, the base manifold M is the unit sphere, whose polar coordinates (θ, ϕ) determine the arm's direction. The fibre F is the unit circle, whose single angle coordinate χ describes the orientation of the hand relative to some reference position. In this case E is not simply the cartesian product $M \times F$. This is because it is not possible to assign, in a continuous way, a reference hand position for every arm direction. This is a consequence of the "hairy ball theorem", ie the fact that it

is impossible to comb flat a hairy ball without introducing singularities (eg crowns, cowlicks, bald patches).

As in the arm example, it is often the case in physical applications that the fibre describes internal degrees of freedom. Usually the fibre has some additional mathematical structure as well. If it is a real or complex vector space (describing spin degrees of freedom, for example), then E is called a *real* or *complex vector bundle*. If the vector space is one-dimensional, E is called a *line bundle* (real or complex). Or else, the fibre might be a group (for example, the two- or three-dimensional rotation group, describing the axial or spatial orientation of a rigid body). Then E is called a *principal bundle*.

2.1.2. Connections and curvature. To a given path P on the base there correspond innumerable paths in the fibre bundle which project down to it: a motion in the base leaves the fibre motion undetermined. A *connection* is a rule for associating to P a particular path $\tilde{P}$ through a given point in the bundle, the lifted path (see Fig. 1). A point moving along the lifted path $\tilde{P}$ is said to be *parallel transported*. The connection rule is expressed in differential form by making the fibre velocity a linear function of the base velocity; this function can also depend on the base and fibre positions. For vector bundles, the dependence on fibre position should be linear as well. There are analogous restrictions in the case of principal bundles. In physical applications, a particular connection is often suggested by kinematical or dynamical considerations, eg, moving your arm without unnecessarily twisting your wrist.

Let us consider the lift $\tilde{C}$ of a closed path C on the base (see Fig. 2). Its endpoints necessarily lie in the same fibre, but they need not coincide. If they do not, the difference between the endpoints is called the *holonomy* of C. In the outstretched arm example, C is the spherical right triangle described by your arm, and its holonomy is the $\pi/2$ turn of your hand which results from traversing it in the manner prescribed. In case there is holonomy for arbitrarily small closed paths, the connection is said to have *curvature*. Curvature describes the infinitesimal holonomy acquired round an infinitesimal closed paths, and may be expressed in terms of derivatives (a generalized curl) of the connection.

Let us consider a very simple example. We take the base M to be three dimensional Euclidean space $\{\mathbf{R} \in \mathbb{R}^3\}$, the fibre F to be one-dimensional complex vector space $\{z \in \mathbb{C}\}$, and the bundle E to be their cartesian product $\{(\mathbf{R}, z) \in \mathbb{R}^3 \times \mathbb{C}\}$. Thus E is a trivial complex line bundle. A connection specifies the fibre velocity $\dot{z}$ as a function of the fibre position z, the base position $\mathbf{R}$ and the base velocity $\dot{\mathbf{R}}$; the dependence on z and $\mathbf{R}$ is linear. The general form of such a relation is given by $\dot{z} = (\mathbf{W}(\mathbf{R}) \cdot \dot{\mathbf{R}})z$, where $\mathbf{W}(\mathbf{R})$ is a complex vector field. $\mathbf{W}$ is called the connection form, or simply the connection.

Given a motion $\mathbf{R}(t)$ in the base, its lifted motion in the fibre is given by $z(t) = \exp\left(\int_{\mathbf{R}(t)} \mathbf{W} \cdot d\mathbf{R}\right) z_0$, where z_0 denotes the initial position in the fibre of the lifted path. If $\mathbf{R}(t)$ describes a closed path, then its holonomy is the scalar factor

$\exp\Gamma$, where $\Gamma = \oint_{\mathbf{R}(t)} \mathbf{W}(\mathbf{R}) \cdot d\mathbf{R}$. Using Stokes' theorem, we may express Γ as the flux of a vector field $\mathbf{V} = \boldsymbol{\nabla} \times \mathbf{W}$ through a surface S bounded by C. $\mathbf{V}$ is the curvature of the connection $\mathbf{W}$.

Of particular relevance to the geometric phase is the case in which the connection preserves the length $|c(t)|$ of vectors along the lifted path. Then the connection is said to be *hermitian*. The connection form $\mathbf{W}$ is purely imaginary, and may be written as $i\mathbf{A}$, where $\mathbf{A}(\mathbf{R})$ is real.

2.1.3. The geometric phase revisited. Simon (1983) observed that the Berry phase finds its natural mathematical expression as the holonomy of a connection on a fibre bundle. The fibre bundles in question are the complex line bundles $E^{(n)}$ associated with nondegenerate eigenstates of a quantum Hamiltonian $H(\mathbf{R})$. The base manifold is the parameter space (we continue to assume this to be $\mathbf{R}^3$), and the fibres are vectors $|v\rangle$ in Hilbert space which satisfy the Schrödinger equation $(H(\mathbf{R}) - E_n(\mathbf{R}))|v\rangle = 0$: these are just the unnormalized nth eigenstates, related to each other by a complex scalar.

Given an eigenstate $|\psi\rangle$ of $H(\mathbf{R})$, a connection determines an eigenstate $|\psi\rangle + |d\psi\rangle$ of the nearby Hamiltonian $H(\mathbf{R} + d\mathbf{R})$ by fixing its norm and phase. On physical grounds it is sensible to require that the connection conserve probability, ie that $|\psi\rangle + |d\psi\rangle$ and $|\psi\rangle$ both have unit norm. This implies that $\mathrm{Re}\,\langle\psi|d\psi\rangle = 0$. The simplest rule for fixing the phase of $|\psi\rangle + |d\psi\rangle$ is to take $\mathrm{Im}\,\langle\psi|d\psi\rangle = 0$ as well. This determines completely a connection on $E^{(n)}$, which is defined by the relation

$$\langle v\, dv\rangle = 0. \tag{21}$$

Under parallel transport round a closed path C, an eigenstate $|\psi\rangle$ acquires a phase factor $\exp i\gamma_n$. This phase factor, the holonomy of C, is precisely the geometric phase factor. This follows by noting that the connection (21) is just the condition (2) derived from adiabatic evolution.

There is an exact mathematical analogy between the outstretched arm and a variant of the spin example of Section 1.3. If a spin-1 (instead of a spin-1/2) particle, initially in its down state, is driven by a magnetic field which is slowly cycled round a closed path C, then its geometric phase γ_n is precisely the turn in the hand of an arm describing the same circuit. Both are equal to the solid angle subtended by C.

2.1.4 Chern classes. Some fibre bundles are trivial cartesian products $M \times F$, and others are not. Topology lies at the root of this distinction, and *characteristic classes* make this distinction precise. They describe an intrinsic twistedness in fibre bundles which cannot be combed away. Bundles whose characteristic classes are the same can be smoothly transformed into one another: bundles whose characteristic classes are different cannot.

For complex vector bundles, the characteristic classes are called *Chern classes*. The number of Chern classes is equal to the dimension of the fibre. The rth Chern class

associates an integer $c_1(N)$, the *Chern number*, to every $2n$-dimensional, boundaryless subspace N of the base manifold M. Thus, the first Chern class assigns an integer $c_1(S)$ to every closed surface S in M. There are various ways to calculate Chern numbers. Among these, there are formulas in terms of integrals over the subspace N of the curvature and functions thereof.

An illustrative example is the celebrated Gauss-Bonnet theorem from classical differential geometry. The theorem states that the integral $\int_S K dS$ of the Gaussian curvature K over a closed one-sided surface S is equal to 2π times an integer χ, the Euler characteristic. The Euler characteristic is a topological feature of the surface. It is related by the formula $\chi = 2 - 2g$ to the genus g, the number of handles on S when it is regarded through a topologist's eyes as a many-handled sphere. The Gauss-Bonnet theorem can be verified immediately for the unit sphere ($K = 1$, $g = 0$). What is remarkable is the implication that $\int_S K dS = 4\pi$ for any surface smoothly deformable to a sphere, but $\int_S K dS \neq 4\pi$ for any surface (such as a torus) which is not.

The Gauss-Bonnet theorem can be formulated in terms of fibre bundles. The bundle consists of the surface S (the base) together with its tangent vectors at each of its points (the fibres). A connection is defined through the Riemannian rule for parallel transport on a curved surface. By identifying two-dimensional real tangent vectors with one-dimensional complex ones, the fibre bundle may be regarded as a complex line bundle, and its first Chern number $c_1(S)$ turns out to be $1/2\pi$ times the integral $\int_S K dS$ of the Gaussian curvature.

Returning to considerations of the geometric phase, let us consider the complex line bundles $E^{(n)}$ of quantum eigenstates, discussed in the preceding section. Each has a single Chern class $c_1^{[n]}$ which assigns an integer to every closed surface S in parameter space. This integer is given by

$$c_1^{(n)}(S) = \frac{1}{2\pi} \oint_S \mathbf{V}_n \cdot d\mathbf{S} = \frac{1}{2\pi} \int_B \boldsymbol{\nabla} \cdot \mathbf{V}_n \, d\mathbf{R}, \tag{22}$$

where in the second equality we have used the divergence theorem; B is the three-dimensional region bounded by S. From (19), the last expression is precisely the number of energy level degeneracies (points $\mathbf{R}_*$ where $E_n(\mathbf{R}_*) = E_{n\pm1}(\mathbf{R}_*)$) contained in S, counted with their appropriate sign (see Section 1.4). In terms of the electromagnetic analogy (Section 1.2), the Chern number is the number of magnetic monopoles in S counted with their charge. Thus, energy level degeneracies manifest themselves in the nontrivial topology of the eigenstate line bundles $E^{(n)}$.

2.2. Degenerate eigenstates and nonabelian geometric phases

If a system possesses a sufficient degree of symmetry, its energy levels will be degenerate. Two examples are the angular momentum multiplets of rotationally symmetric systems, and Kramers degeneracy in systems with time-reversal symmetry and an odd number of fermions. If the symmetry persists throughout the parameterized family $H(\mathbf{R})$, then so do these degeneracies. Under a slowly varying

Hamiltonian $H(\mathbf{R}(t))$, the adiabatic theorem implies that transitions to states outside a degenerate multiplet are small. However, transitions to states within the multiplet need not be, no matter how slow the variation (transitions between degenerate states cost no energy).

Wilczek and Zee (1984) generalized Berry's analysis to degenerate eigenstates. They showed that under adiabatic cycling round a closed circuit C, such states need not return to themselves up to a phase, but instead can evolve into new states within the multiplet. The transformation from initial to final states is generated by a unitary matrix $U(C)$, called the *nonabelian geometric phase*. Nonabelian geometric phases can be interpreted as holonomy, but of a more general sort than was considered in Section 2.1.3. There is a corresponding generalization of geometric magnetism.

The analysis, which runs parallel to the nondegenerate case, proceeds as follows. Let $E(\mathbf{R})$ denote an r-fold degenerate energy level whose degenerate eigenspace is spanned by an orthonormal basis $\{|j(\mathbf{R})\rangle, j = 1, 2, \ldots, r\}$. We consider the evolution under a slowly changing Hamiltonian $H(\mathbf{R}(t))$ of states $|\psi_j(t)\rangle$ initially equal to $|j(\mathbf{R}(0))\rangle$. Neglecting transitions to states outside the multiplet, we look for approximate solutions of the form

$$|\psi_j(t)\rangle = \exp\left(-i/\hbar \int_0^t E(\mathbf{R}(\tau))d\tau\right) \sum_{k=1}^{r} U_{jk}(t)|k(\mathbf{R}(t))\rangle, \quad U_{jk}(0) = \delta_{jk}, \tag{23}$$

ie, a superposition of evolving degenerate eigenstates multiplied by an overall dynamical phase factor. For the inner products $\langle\psi_k(t)|\psi_j(t)\rangle$ to be conserved in time, the $r \times r$ matrix $U(t)$ must be unitary. Its evolution is determined by requiring $\langle k(\mathbf{R}(t))|(i\hbar\partial_t - H)|\psi_j(t)\rangle$ to vanish for each k (it would automatically if $\psi_j(t)$ were an exact solution of the Schrödinger equation). This requirement leads to the equation of motion

$$\dot{U}(t) = -iU(t)\mathbf{A}(\mathbf{R}(t)) \cdot \dot{\mathbf{R}}(t) = -i\mathrm{A}(t)U(t), \tag{24}$$

where $\mathbf{A}(\mathbf{R}) = (\mathrm{A}^{(1)}, \mathrm{A}^{(2)}, \mathrm{A}^{(3)})(\mathbf{R})$ is a vector of $r \times r$ Hermitian matrices whose matrix elements are given by

$$A^{(i)}_{jk}(\mathbf{R}) = -i\langle k(\mathbf{R})|\nabla_i j(\mathbf{R})\rangle, \tag{25}$$

and $\mathrm{A}(t) = \mathbf{A}(\mathbf{R}(t)) \cdot \dot{\mathbf{R}}(t)$.

The solution of (24) is given by the time-ordered product

$$\begin{aligned} U(t) &= \mathrm{T}\exp\left(-i\int_0^t \mathrm{A}(\tau)\,d\tau\right) \\ &= \mathrm{I} - i\int_{0<\tau<t} \mathrm{A}(\tau)\,d\tau - \int\int_{0<\tau<\tau'<t} \mathrm{A}(\tau)\mathrm{A}(\tau')\,d\tau d\tau' + \cdots, \end{aligned} \tag{26}$$

where I is the $r \times r$ identity matrix. The time ordering indicates that in expanding the exponential of the integral, all matrices are to be ordered with later times to

the right. If the matrices $A(\tau)$ at different times commute with each other, then the ordering of the factors is irrelevant, and the right-hand side of (26) simplifies to $I - i\int_{0<\tau<t} A(\tau)\,d\tau - \frac{1}{2}\int_{0<\tau<t} A(\tau)\,d\tau)^2/2! + \cdots = \exp\left(-i\int_0^t A(\tau)\,d\tau\right)$, eg simply the exponentiated integral. In particular, if $r = 1$, then $A(\tau)$ is a scalar, and $U(t)$ is a phase factor.

Suppose the Hamiltonian is taken round a closed cycle C. Then the relation between initial and final states is described, apart from the dynamical phase factor, by the unitary matrix

$$U(C) = \mathrm{T}\exp\left(-i\oint_C \mathbf{A}(\mathbf{R})\cdot d\mathbf{R}\right). \tag{27}$$

$U(C)$ is the nonabelian geometric phase. The terminology derives from the fact that the factors $\mathbf{A}(\mathbf{R})\cdot d\mathbf{R}$ in the time-ordered product do not in general commute. They do, however, in the nondegenerate case $r = 1$; then $U(C)$ is just the ordinary (abelian) geometric phase factor $\exp i\gamma(C)$. Like the ordinary geometric phase, $U(C)$ depends only on the path C and not on the rate at which it is traversed.

For infinitesimal circuits, $U(C)$ is of the form $\mathbf{I} + i\mathbf{\Gamma}(C)$, where $\Gamma(C)$ is an infinitesimal Hermitian matrix given by (minus) the flux of the covariant curl of $\mathbf{A}$.

$$\mathbf{V} = \boldsymbol{\nabla}\times\mathbf{A} - i\mathbf{A}\times\mathbf{A}. \tag{28}$$

through the infinitesimal area bounded by C. (Note that because the components of $\mathbf{A}$ are matrices, the cross product $\mathbf{A}\times\mathbf{A}$ does not vanish, just as the cross product $\boldsymbol{\sigma}\times\boldsymbol{\sigma} = i\boldsymbol{\sigma}$ of Pauli matrices does not vanish.) This result does not extend to finite circuits, though, as there is no nonabelian version of Stokes' theorem.

Like the ordinary geometric phase, $U(C)$ may be interpreted as holonomy, but now on a complex r-dimensional vector bundle rather than a line bundle. The base manifold is parameter space ($\mathbf{R}^d$ in the examples considered here), and the fibers are the degenerate eigenspaces spanned by $|j(\mathbf{R})\rangle$. The matrix-valued vector field $\mathbf{A}(\mathbf{R})$ is the connection form, which describes how states in the fibres are parallel transported along paths $\mathbf{R}(t)$ in the base. $\mathbf{V}(\mathbf{R})$ is the associated curvature form.

The nonabelian geometric phase does not depend on the choice of basis $|j(\mathbf{R})\rangle$ in any intrinsic way. Under a unitary change of basis

$$|j(\mathbf{R})\rangle \to |\tilde{j}(\mathbf{R})\rangle = \sum_{k=1}^{r} W_{jk}(\mathbf{R})|k(\mathbf{R})\rangle, \tag{29}$$

the nonabelian version of the gauge transformation (8), $U(C)$ transforms according to $U(C) \to WU(C)W^\dagger$. This unitary conjugation represents simply the change of basis applied to $U(C)$ and leaves its observable properties unchanged. The fields $\mathbf{A}(\mathbf{R})$ and $\mathbf{V}(\mathbf{R})$ transform according to

$$\mathbf{A} \to W\mathbf{A}W^\dagger - i\boldsymbol{\nabla}WW^\dagger, \quad \mathbf{V} \to W\mathbf{V}W^\dagger \tag{30}$$

under (29). These transformation properties characterize the potential and field strength of Yang-Mills field theories, generalizing the electromagnetic analogy with the geometric phase (cf Section 1.2) to the nonabelian case.

Given a particular closed circuit C, one can choose the basis $|j(\mathbf{R})\rangle$ so as to diagonalize $U(C)$. Then under parallel transport round C, each basis state returns to itself up to a phase factor. Under certain conditions, one can find a basis which diagonalizes every nonabelian geometric phase, regardless of C. In such cases, the degenerate and nondegenerate cases are essentially the same. In general, though, there is no choice of basis which renders every $U(C)$ diagonal. What distinguishes the general case is that the generators of the symmetries responsible for the degeneracies depend on parameters themselves. A particular example, the Λ-doubling of electron doublets in diatomic molecules, is discussed by Moody *et al* (1986).

The discussion of geometric magnetism in the Born-Oppenheimer approximation (cf Section 1.7) can also be generalized to the degenerate case. As before, we regard the parameters $\mathbf{R}$ as the coordinates of a slow system to which the original one, the fast system, is coupled. The Hamiltonian is given by $P^2/2M + H(\mathbf{R})$, where $\mathbf{P}$ is the slow momentum and the dependence on the fast coordinates and momenta is left implicit in $H(\mathbf{R})$. We look for approximate eigenfunctions of the coupled system in which the fast system belongs to a degenerate multiplet of $H(\mathbf{R})$. These are of the form $\Psi(\mathbf{r},\mathbf{R}) = \sum_{j=1}^{r} \phi_j(\mathbf{R})\psi_j(\mathbf{r},\mathbf{R})$, where $\psi_j(\mathbf{r},\mathbf{R}) = \langle \mathbf{r}|j(\mathbf{R})\rangle$ are the fast eigenfunctions (which depend parametrically on the slow variables $\mathbf{R}$). Substituting this ansatz into the stationary Schrödinger equation and making approximations similar to those in the nondegenerate case (the details are omitted), one sees that the slow system is described by a multicomponent wavefunction $\boldsymbol{\phi}(\mathbf{R}) = (\phi_1(\mathbf{R}), \ldots, \phi_r(\mathbf{R}))$ which is an eigenfunction of the matrix-valued Hamiltonian

$$\mathrm{H} = (\mathbf{P}\mathrm{I} - \mathbf{A})^2/2M + E(\mathbf{R})\mathrm{I}. \tag{31}$$

Here $\mathbf{A}(\mathbf{R})$ is the matrix-valued vector potential of (25) and I is the $r \times r$ identity matrix. Thus, coupling to a degenerate fast eigenstate introduces a generalized form of geometric magnetism, in which the slow Hamiltonian acquires spin degrees of freedom which are coupled to a Yang-Mills-like field.

Finally, there is a connection between the nonabelian geometric phase and higher-order degeneracies, ie points in parameter space where two or more degenerate levels coalesce. These points act as generalized monopole sources for the Yang-Mills field $\mathbf{V}$.

2.3. Nonadiabatic evolution

Geometric phases appear in contexts other than adiabatic evolution. An important generalization was found by Aharonov and Anandan (1987), who showed that geometric phases accompany any cyclic evolution, not necessarily adiabatic, of a quantum state. Subsequently, Samuel and Bhandari (1988) showed that geometric phases can arise in evolutions which are nonunitary, not cyclic, and not even

continuous. What underlies these generalizations is a natural prescription for the parallel transport of quantum phase factors.

The setting for the Aharonov Anandan phase is quite general. A state $|\psi(t)\rangle$ evolves according to the Schrödinger equation $|i\hbar\partial_t - H(t)\rangle|\psi\rangle = 0$ under a (usually) time-dependent Hamiltonian $H(t)$. After a time T the state is presumed to return to itself up to a phase factor, so that $|\psi(T)\rangle = \exp(i\beta)\,\psi(0)\rangle$. In the trivial case where H is time-independent and $|\psi\rangle$ is an eigenstate, β is simply the dynamical phase $-ET/\hbar$. Guided by this simple case, we are led to subtract from β a generalized dynamical phase, namely the time integral of the expectation value of energy

$$\beta_d = -\frac{1}{\hbar}\int_0^T \langle\psi|H|\psi\rangle\,dt - \mathrm{Im}\int_0^T \langle\psi\,\dot\psi\rangle\,dt. \tag{32}$$

The phase which remains, $\beta_g \overset{\text{def}}{=} \beta - \beta_d$, is given by

$$\beta_g = \beta - \beta_d = \arg\langle\psi(0)|\psi(T)\rangle - \mathrm{Im}\int_0^T \langle\psi|\dot\psi\rangle\,dt \tag{33}$$

What makes the decomposition $\beta = \beta_d + \beta_g$ of the total phase interesting is that β_g is, in a sense to be explained, a geometric phase. First, β_g depends only on the path through Hilbert space described by $|\psi(t)\rangle$, and not on the rate this path is traversed. Indeed, the state $|\tilde\psi(t)\rangle = |\psi(\tau(t))\rangle$ traverses the same path but, because of the time reparameterization $t \to \tau(t)$, at a different rate. However, it is easy to verify that its geometric phase β_g, as given by (33), is the same as that of $|\psi(t)\rangle$, even though its dynamical phase is different. (Note that $|\tilde\psi(t)\rangle$ might not even satisfy the Schrödinger equation.) Second, β_g does not depend on the overall phase of $|\psi(t)\rangle$. Indeed, if in (33), $|\psi(t)\rangle$ is replaced by $|\tilde\psi(t)\rangle = \exp i\varphi(t)|\psi(t)\rangle$, a simple calculation shows that β_g is unchanged.

As $|\psi(t)\rangle$ is periodic up to a phase, the density matrix $\rho_\psi(t) = |\psi(t)\rangle\langle\psi(t)|$ is strictly periodic, and over the time interval $0 \le t \le T$ describes a closed curve C in the space of density matrices. What the preceding discussion shows is that β_g is a geometric property of C. It depends not on the rate at which C is traversed, and it can be computed with (33), substituting any state $|\tilde\psi(t)\rangle$ whose density matrix $|\tilde\psi(t)\rangle\langle\tilde\psi(t)|$ is equal to $\rho_\psi(t)$.

The geometric phase for adiabatic cycles can be obtained as a special case of (33). Within the adiabatic approximation, a solution to the Schrödinger equation for a slowly cycled Hamiltonian $H(\mathbf{R}(t))$ is given by $|\psi(t)\rangle = \exp(i\delta(t))|n(\mathbf{R}(t))\rangle$ (cf (3)). From (32), the dynamical phase β_d is given by $-1/\hbar\int_0^T E_n(\mathbf{R}(t))dt$. The geometric phase is easily computed from (33) if $|\tilde\psi(t)\rangle = |n(\mathbf{R}(t))\rangle$ is substituted for $|\psi(t)\rangle$ (this is permissible, since the two states differ by an overall phase). Since $\arg\langle n(\mathbf{R}(0))|n(\mathbf{R}(T))\rangle = 0$, β_g is given by $-\mathrm{Im}\int_0^T\langle\dot{\tilde\psi}|\tilde\psi\rangle dt = -\mathrm{Im}\oint_C\langle n(\mathbf{R})|\nabla n(\mathbf{R})\rangle\cdot d\mathbf{R}$, in agreement with (5).

What underlies this general setting for the geometric phase is once more the mathematics of connections and fibre bundles. However, now the entire Hilbert space is

to be regarded as a fibre bundle! The base manifold $\mathcal{H}$ consists of all pure state density matrices ρ, normalized so that $\mathrm{Tr}\,\rho = 1$. (This is sometimes called the *projective Hilbert space*.) The fibre attached to ρ consists of all state vectors $|\psi\rangle$ whose density matrices $|\psi\rangle\langle\psi|$ are equal to ρ. (Thus, states in the same fibre differ only by a phase factor.) To an infinitesimal displacement $d\rho$ in the base there corresponds displacements $|d\psi\rangle$ in the fibre satisfying $|\psi + d\psi\rangle\langle\psi + d\psi| - \rho = d\rho$. With a little algebra this reduces to $d\psi\rangle - \langle\psi|d\psi\rangle|\psi\rangle = d\rho|\psi\rangle$. This equation does not determine $d\psi\rangle$ uniquely: to any solution $|d\psi\rangle$ can be added an infinitesimal imaginary component $i\,d\alpha|\psi\rangle$ along $|\psi\rangle$. A connection fixes this component and determines a particular $|d\psi\rangle$ for a given $d\rho$. A natural choice is to require that

$$\langle\psi|d\psi\rangle = 0 \tag{34}$$

This connection is precisely the condition (2) found by Berry for adiabatic cycles, but now generalized to arbitrary cycles in Hilbert space. Its holonomy is precisely the geometric phase β_g given by (33).

This setting admits further generalizations (Samuel and Bhandari 1988). For example, one can calculate a geometric phase for a nonunitary evolution $|\phi(t)\rangle$, for which $\langle\phi|\phi\rangle$ is no longer constant, simply by evaluating (33) with $|\psi(t)\rangle \stackrel{\mathrm{def}}{=} |\phi(t)\rangle/\langle\phi|\phi\rangle^{1/2}$.

In fact, one can calculate a geometric phase for an open path $|\psi(t)\rangle$, where $|\psi(T)\rangle$ and $|\psi(0)\rangle$ are no longer proportional. The formula (33) for β_g can be computed for open paths, and the result does not depend on the parameterization in time of $|\psi(t)\rangle$, and remains unchanged by the change of phase $|\psi(t)\rangle \to \exp i\phi(t)\,|\psi(t)\rangle$. Further insight into the interpretation of the geometric phase for open paths is given in the following Section 2.4.

There is also a discrete version of the formula (33). Given a sequence of states $|\psi_1\rangle, |\psi_2\rangle, \ldots |\psi_N\rangle$, one can associate a phase according to the formula

$$\beta_g = \arg\langle\psi_1|\psi_2\rangle\langle\psi_2|\psi_3\rangle \ldots \langle\psi_N|\psi_1\rangle. \tag{35}$$

One sees easily that β_g does not depend on the overall phases of the states $|\psi_j\rangle$. If there are just two states, then $\beta_g = \arg|\langle\psi_1|\psi_2\rangle|^2 = 0$ identically. But if $N > 2$, the phase β_g is typically nonvanishing. Note that $|\psi_N\rangle$ need not be proportional to $|\psi_1\rangle$; the sequence of states need not close on itself.

The formula (35) is the discrete analogue of (33), and defines a geometric phase for discontinuous evolutions. Thus, a geometric phase can be defined for processes which include measurements and wavefunction collapse. A version of formula (35) appears in Pancharatnam's study of optical polarization dynamics, in which the $|\psi_j\rangle$ represent polarization states. Pancharatnam's phase is discussed in Section 4.2.

2.4. *Geometric metric tensor*

How close are two quantum states $|\psi\rangle$ and $|\phi\rangle$? The usual quantum mechanical measure of distance is the Hilbert space norm $\langle\psi - \phi|\psi - \phi\rangle^{1/2}$ of their difference.

There is another, introduced by Provost and Vallee (1980), namely the quantity $s^2(\psi,\phi) \stackrel{\text{def}}{=} 1 - |\langle\psi|\phi\rangle|^2/\langle\psi|\psi\rangle\langle\phi|\phi\rangle$. Clearly this vanishes if the states are the same. What distinguishes it from the Hilbert space norm is that it is unchanged if either $|\psi\rangle$ or $|\phi\rangle$ are multiplied by a complex scalar. In fact, what this quantity really describes is the distance between their normalized density matrices $\rho_\psi = |\psi\rangle\langle\psi|/\langle\psi|\psi\rangle$ and $\rho_\phi = |\phi\rangle\langle\phi|/\langle\phi|\phi\rangle$, and indeed can be alternatively expressed as $1 - \mathrm{Tr}\,\rho_\psi\rho_\phi$. For infinitesimally displaced states $|\phi\rangle = |\psi + d\psi\rangle$, a little algebra gives

$$ds^2(|d\psi\rangle) \stackrel{\text{def}}{=} s^2(|\psi\rangle, |\psi + d\psi\rangle) = 1 - \mathrm{Tr}(\rho_\psi\rho_{\psi+d\psi}) = \tfrac{1}{2}\,\mathrm{Tr}\,(d\rho_\psi{}^2) = \frac{\langle d\psi|I - \rho_\psi|d\psi\rangle}{\langle\psi|\psi\rangle}, \quad (36)$$

where $d\rho_\psi = \rho_{\psi+d\psi} - \rho_\psi$ and I is the identity operator. (36) defines a metric on the space of normalized density matrices $\tilde{\mathcal{H}}$ (or equivalently, the projective Hilbert space).

The quantum metric enters into the analysis of the geometric phase in a number of interesting ways. Below we describe two aspects: a deeper understanding of the geometric phase for open paths, and physically significant corrections to the Born-Oppenheimer approximation.

The quantum metric determines geodesics, ie curves $\rho(s)$ which minimize, or at least render stationary, the distance between two given density matrices in $\tilde{\mathcal{H}}$. These are most easily described in terms of their corresponding state vectors: it may be shown that if $|\psi(s)\rangle$ satisfies the differential equation $d^2|\psi\rangle/ds^2 + |\psi\rangle = 0$, and if the vectors $|\psi(0)\rangle$ and $|\psi'(0)\rangle$ are normalized and orthogonal to one another, then the curve $\rho_\psi(s) = |\psi(s)\rangle\langle\psi(s)|$ is a geodesic with respect to the metric (36), and the parameter s along the curve is the arclength. (The solutions $|\psi(s)\rangle$ are of the form $\cos s\,|\psi(0)\rangle + \sin s\,|\psi'(0)\rangle$.)

A simple illustration is provided by two-state systems. In this case, $\tilde{\mathcal{H}}$ may be identified with the unit sphere (in the context of polarization optics, this is the Poincaré sphere — see Section 4.2); each density matrix ρ may be expressed in the form $(I + \hat{\mathbf{e}}\cdot\boldsymbol{\sigma})/2$, where $\hat{\mathbf{e}}$ is a unit vector and $\boldsymbol{\sigma}$ are the Pauli matrices. The quantum metric (36) reduces to the usual metric on the sphere (up to a scalar factor), and the geodesics are the arcs of great circles.

As discussed in the previous Section 2.3, geometric phases can be defined via (33) for open as well as closed paths in $\tilde{\mathcal{H}}$. The two cases may be related as follows. An open path P may be closed by joining its endpoints with a geodesic. The resulting closed path, it turns out, has the same geometric phase as P. The geometric phase for continuous and discontinuous evolutions, (33) and (35), may be similarly related; given a closed sequence of density matrices $\rho_1, \rho_2, \ldots \rho_N, \rho_1$, we can construct a closed polygonal path $\rho(s)$ in $\tilde{\mathcal{H}}$ by connecting its points with geodesics. $\rho(s)$ turns out to have the same geometric phase as the discrete sequence ρ_j.

The quantum metric also plays a role in the Born-Oppenheimer description of the dynamics of a slow system coupled to a fast one (cf Section 1.7). There it appears in the slow Hamiltonian as an additional scalar potential $\Phi_n(\mathbf{R}) = \sum_a ds^2(|\nabla_a n\,\mathbf{R}\rangle)/2M$, which produces a 'geometric electric' force. The scalar potential appears at a higher

order of adiabatic expansion than the vector potential $\mathbf{A}_n(\mathbf{R})$: with appropriate scalings, the geometric magnetic force $\dot{\mathbf{R}} \times \mathbf{V}_n$ is of the order of the square root $\sqrt{m/M}$ of the mass ratio of the fast and slow systems, whereas the geometric electric force $-\nabla\Phi$ is of the order m/M. However, the electric force can become significant in the vicinity of degeneracies $\mathbf{R}_s$ (ie. points where the fast energy level $E_n(\mathbf{R})$ becomes degenerate). There the force is repulsive, and so tends to steer the slow dynamics away from points where the Born-Oppenheimer approximation would break down. A discussion of the scalar potential can be found in the reviews of Berry (1989) and Jackiw (1988). Additional terms in the slow dynamics of comparable order to geometric electricity have also been found by Weigert and Littlejohn (1993).

2.5. *Beyond adiabaticity*

2.5.1. Corrections to the geometric phase. For the adiabatically cycled eigenstates described by Eq. (3), the dynamical and geometric phases are but the first two terms in an asymptotic expansion for the total phase. An elegant iterative scheme for evaluating higher order corrections was developed by Berry (1987c). These corrections to the geometric phase turn out to be geometric phases themselves. The full asymptotic series diverges in a manner which reveals the breakdown of the adiabatic approximation.

The method is illustrated by the example of a spin-1/2 particle in a magnetic field (cf Section 1.3). We consider a time dependent Hamiltonian $H_0(\tau) = \mu\mathbf{B}_0(\tau)\cdot\boldsymbol{\sigma}$, cyclic over the infinite time interval $[-\infty, \infty]$ with $\mathbf{B}_0(\pm\infty) = B_z\hat{\mathbf{z}}$. $\mathbf{B}_0(\tau)$ is assumed to be analytic, and we assume as well that $B_0(\tau) \geq B_{\min}$ for all τ, so that the Hamiltonian does not pass through any degeneracies. $\tau = \epsilon t$ is a dimensionless time, and ϵ a small parameter (with dimensions of frequency) describing the slow rate of change of the Hamiltonian. (The corresponding dimensionless small parameter is $\lambda = \epsilon\hbar/\mu B_{\min}$.) Let us consider the evolution of a spin state $|w(t)\rangle$ with initial condition $|w(-\infty)\rangle = |-\rangle$. The analysis is facilitated by transforming to a rotating frame in which the direction $\hat{\mathbf{b}}_0$ of the magnetic field is fixed (say, for definiteness, along $\hat{\mathbf{z}}$). Such a frame is described by a rotation matrix $\mathcal{R}_1(\tau)$ satisfying

$$\hat{\mathbf{b}}_0(\tau) = \mathcal{R}_1(\tau)\cdot\hat{\mathbf{z}}. \tag{37}$$

This condition does not determine $\mathcal{R}_1(\tau)$ uniquely, but only up to a rotation (applied from the left) about $\hat{\mathbf{b}}_0$. Differentiating with respect to τ (this is denoted by a dot $\dot{}$), we obtain a condition

$$\dot{\hat{\mathbf{b}}}_0(\tau) = \boldsymbol{\omega}_1(\tau) \times \hat{\mathbf{b}}_0(\tau) \tag{38}$$

on the angular velocity $\boldsymbol{\omega}_1$ of the frame, where $\omega_{1i} = \frac{1}{2}\sum_{jk}\epsilon_{ijk}\dot{\mathcal{R}}_{1k}\mathcal{R}_{1j}$ (ϵ_{ijk} is the antisymmetric Levi-Civita symbol). (38) does not determine $\boldsymbol{\omega}_1$ uniquely, but only up to its component along $\hat{\mathbf{b}}_0$. We fix this indeterminacy (and the corresponding indeterminacy in $\mathcal{R}_1$) by requiring that

$$\boldsymbol{\omega}_1 \cdot \hat{\mathbf{b}}_0 = 0. \tag{39}$$

(39) is a condition for parallel transport, analogous to the requirement of moving your arm without unnecessarily twisting your wrist (cf Section 2.1).

In the rotating frame, the Hamiltonian $H_1(\tau)$ is given by $\mu \mathbf{B}_1(\tau)\cdot\boldsymbol{\sigma}$, where

$$\mathbf{B}_1(\tau) = B_0(\tau)\hat{\mathbf{z}} - \mathbf{B}_{L1}(\tau) \tag{40}$$

is the effective magnetic field. $\mathbf{B}_{L1}(\tau) = \epsilon\hbar/\mu\, \mathcal{R}_1^{-1}(\tau)\cdot\omega_1(\tau)$ is the Larmor field induced in the rotating frame (the magnetic analogue of the Coriolis force). As the rotation is slow, the Larmor field is small (of order ϵ), and in the lowest order adiabatic approximation (cf Sections 1.1, 1.2) it is simply neglected. Then $H_1(\tau)$ is diagonal, and the time-dependent Schrödinger equation is easily solved. The geometric phase $\gamma_-(C_0)$, half the solid angle of the cycle C_0 described by $\mathbf{B}_0(\tau)$, appears as a direct consequence of the parallel transport condition (39) when the solution is transformed back to the fixed frame.

If the Larmor field is not neglected, then $\mathbf{B}_1(\tau)$ may be regarded as a slowly-varying magnetic field to which the above procedure may be applied as before. That is, we transform to a second rotating frame $\mathcal{R}_2(\tau)$ in which $\hat{\mathbf{b}}_1(\tau)$ is fixed along z. The frame is uniquely determined by imposing the parallel transport condition $\omega_2\cdot\hat{\mathbf{b}}_1 = 0$, where ω_2 is the angular velocity of the frame $\mathcal{R}_2$. Note that because $\hat{\mathbf{b}}_1$ differs from $\hat{\mathbf{z}}$ by terms of order ϵ, the angular velocity ω_2 is of order ϵ. In this second frame, $H_2(\tau) = \mu\mathbf{B}_2(\tau)\cdot\boldsymbol{\sigma}$, where the effective magnetic field $\mathbf{B}_2(\tau)$ is given by $B_2(\tau)\hat{\mathbf{z}} - \mathbf{B}_{L2}(\tau)$, with $\mathbf{B}_{L2}(\tau) = \epsilon\hbar/\mu\, \mathcal{R}_2^{-1}(\tau)\cdot\omega_2(\tau)$. The new Larmor field $\mathbf{B}_{L2}$ is of order ϵ^2.

It is easy to see that this procedure can be iterated: we can transform through successive rotating frames $\mathcal{R}_{n+1}(\tau)$ to compensate for the Larmor fields $\mathbf{B}_{Ln}(\tau)$ induced in the preceding ones. The geometric phase and its corrections emerge when the sequence of rotations is unravelled and the solution transformed back to the fixed frame. One obtains a formal series

$$\gamma_-(C_0) + \sum_{k=1}^{\infty}\gamma_-(C_k), \tag{41}$$

in which the corrections to the geometric phase are themselves geometric phases: $\gamma_-(C_k)$ is half the solid angle of the cycle C_k described by $\mathbf{B}_k(\tau)$.

It turns out that the series (41) is divergent. Although successive cycles C_k initially decrease in extent by a factor of ϵ, asymptotically the terms $\gamma_-(C_k)$ behave as $\epsilon^k k!$, a form characteristic of divergent asymptotic series (Dingle 1973). An optimal approximation is obtained by truncating the series at its smallest term $k\sim 1/\epsilon$, for which $\epsilon^k k!\sim\exp(-1/\epsilon)$ is exponentially small.

That the series (41) diverges can be understood on general grounds. If, on the contrary, the series converged, along with the infinite sequence of rotating frames, then upon transforming back to the fixed frame, we would obtain an exact solution $|c(t)\rangle$ of the Schrödinger equation which would begin and end, at $t=-\infty$ and $t=+\infty$ respectively, in the spin down state $|\downarrow\rangle$. This would imply a vanishing transition

probability $\langle -|\psi(-\infty)\rangle$. But it is known that, with the exception of some special Hamiltonians, the transition probability is not zero, but instead (for $\mathbf{B}_0(\tau)$ analytic) is exponentially small, of order $\exp(-1/\epsilon)$ (Dykhne 1962). These exponentially small transitions (the subject of the following section) signal the ultimate failure of the adiabatic approximation, and cannot be captured by the iterative scheme. It follows that this scheme must inevitably diverge.

2.5.2. The geometric amplitude. Adiabatic evolution is an approximation. For typical slowly varying Hamiltonians $H(\tau)$, where $\tau = \epsilon t$, no matter how small the rate of change ϵ, there are transitions between evolving eigenstates. The transition probability vanishes as ϵ goes to zero, and if H depends analytically on τ it does so exponentially rapidly (faster than any power of ϵ).

In a study of adiabatic evolution in two-state systems, Berry (1990b) found that this transition probability may contain a geometric factor. The setting is similar to that of the preceding Section 2.5.1. We consider a two-state Hamiltonian $H(\tau) = \mathbf{R}(\tau)\cdot\boldsymbol{\sigma}$, where $\mathbf{R}(\tau)$ is analytic in the dimensionless time $\tau = \epsilon t$ and is nonvanishing for real τ. As $\tau \to \pm\infty$, $\mathbf{R}(\tau)$ approaches limiting values $\mathbf{R}_\pm$, which need not coincide, so that $\mathbf{R}(\tau)$ need not be closed. From (3), the adiabatic solution for the up-state is given by $|\psi(t)\rangle = \exp(i\delta_+(t) + i\gamma_+(t))\,|+(\mathbf{R}(\tau))\rangle$, where the eigenstates $|\pm(\mathbf{R})\rangle$ are given by (14) and the evolving dynamical and geometric phases $\delta_+(t)$ and $\gamma_+(t)$ are given explicitly by

$$\delta_-(t) = -\frac{1}{\epsilon\hbar}\int_0^\tau R(\tau')\,d\tau', \quad \gamma_+(t) = -\tfrac{1}{2}\int_0^\tau (1-\cos\theta)\dot{\phi}\,d\tau'. \tag{42}$$

Here (R,θ,ϕ) denote the polar coordinates of $\mathbf{R}(\tau)$, and the dot denotes the derivative with respect to τ. We have used (15) for the vector potential $\mathbf{A}_+(\mathbf{R})$.

Nonadiabatic transitions may be attributed to roots of $R(\tau)$ in the complex τ plane. The dominant contribution to the transition amplitude comes from the root τ_c closest to the real axis. (We assume τ_c is a simple zero of $R^2(\tau)$, so that it is a square-root branch point of $R(\tau)$.) Davis and Pechukas (1977) showed that the transition amplitude may be obtained by analytically continuing the adiabatic solution around the branch point τ_c. (Under this continuation, the up- and down-states $|+(\mathbf{R}(\tau))\rangle$ and $|-(\mathbf{R}(\tau))\rangle$ are interchanged.) The transition probability $P_{-\leftarrow+} = |\langle -(\mathbf{R}(\infty))\,|\,\psi(\infty)\rangle|^2$ is given by a product of factors $\exp(-2|\Delta_-|)\exp(2\Gamma_+)$, whose exponents are obtained from the analytical continuation of the dynamical and geometric phases. Explicitly, we have that

$$\Delta_+(t) = \frac{2}{\hbar\epsilon}\,\mathrm{Im}\int_0^{\tau_c} R(\tau')\,d\tau', \quad \Gamma_+ = -\tfrac{1}{2}\mathrm{Im}\oint \cos\theta\,d\phi. \tag{43}$$

In the expression for Γ_+, the contour is taken round the cut joining τ_c and its complex conjugate τ_c^*.

The factor $\exp(\Gamma_+)$ is the geometric amplitude. It depends only on the cycle followed by $\mathbf{R}(\tau)$, not on the slowness parameter ϵ nor on $\hbar$. It can be shown to vanish

if $\mathbf{R}(\tau)$ lies in a plane, or if $\mathbf{R}(\tau)$ can be rotated into itself about an axis through the origin. Thus it vanishes for the well-known Landau-Zener model $\mathbf{R}(\tau) = (a, a, A\tau)$ (Zener 1930), for which $\mathbf{R}(\tau)$ describes a vertical line. A simple example with nonzero geometric amplitude is $\mathbf{R}(\tau) = (a\cos(\omega\tau^2), a\sin(\omega\tau^2), A\tau)$. It describes a helix on the cylinder of radius a which winds for $\tau < 0$ and unwinds for $\tau > 0$. For this curve Γ_+ is given by $-a^2\omega \operatorname{sgn} A/A^2$.

The geometric amplitude was observed by Zwanziger et al (1991) in a nuclear magnetic resonance experiment (cf Section 5.2). A time-dependent spin Hamiltonian with the requisite properties was generated by applying a modulated rf magnetic field to a sample of spin-1/2 nuclei, and the induced magnetization was observed in a rotating frame. The logarithm of the transition probability P_+ was plotted as a function of the slowness parameter ϵ. Theory predicts that this plot should be linear (due to the dynamical amplitude) with a nonzero intercept (due to the geometric amplitude). This behaviour was confirmed in the experiment, with good quantitative agreement between the observed and predicted value of Γ_+.

3. Classical Systems

Following Berry's discovery of the geometric phase, Hannay (1985) found an analogous phenomenon in classical mechanics. Adiabatically cycled classical oscillators undergo a shift in their phase of oscillation. Part of this shift is simply the time integral of the frequency. But there appears in addition a purely geometrical contribution, now known as the Hannay angle. The Hannay angle includes and generalizes such well-known phenomena as the precession of the Foucault pendulum. In many cases it may be directly observed. The Hannay angle emerges as the classical ($\hbar \to 0$) limit of the Berry phase (in cases where chaos is absent from the classical dynamics). In the context of coupled systems, it leads to classical geometric magnetism in the slow dynamics.

Wilczek and Shapere (1989b-1989d) studied phenomena in classical kinematics similar to the Hannay angle. A simple example is the manner in which falling cats and platform divers change their spatial orientations through a cycle of internal motions. Provided their (conserved) angular momentum is zero, the change in their orientation does not depend on the rate at which the internal motions are performed. Classical geometric phases of this type have been investigated systematically by Marsden, Montgomery and coworkers (Marsden *et al* 1991, Marsden 1992) as part of a general theory of Hamiltonian dynamics with symmetry.

3.1 The Hannay Angle

3.1.1. The Classical Adiabatic Theorem. Imagine a swinging pendulum slowly shortened to half its original length. During this process, is anything conserved, either exactly or approximately? This question is addressed by the classical adiabatic

theorem. While energy is not conserved (work is being done), there is an approximately conserved quantity, namely the classical adiabatic invariant, or the *action*.

The classical adiabatic theorem is most simply described for systems of one degree of freedom, so we concentrate on this case. The dynamics in the two-dimensional phase plane is described by Hamilton's equations $\dot{q} = \partial_p H, \dot{p} = -\partial_q H$. The classical Hamiltonian $H(q, p, \mathbf{R})$ is taken to depend on a vector of parameters $\mathbf{R}$ (like the quantum Hamiltonians of Section 1). If these are held fixed, then energy is conserved, and orbits lie on contours of H. Let us assume the motion is bounded, so that the energy contours are closed. The orbits go round them periodically with oscillation frequency $\omega = (\partial_E I)^{-1}$, where

$$I(E, \mathbf{R}) = \frac{1}{2\pi} \int_{H(q,p,\mathbf{R}) \leq E} dq\, dp \tag{44}$$

is the action. Apart from the factor of 2π, the action is just the area enclosed by the orbit of energy E. If the parameters vary in time, then energy is no longer conserved. However, if the variation is slow compared to the oscillation frequency ω, then the action (44) is nearly conserved, and the variation of the energy is determined implicitly by

$$I(E(t), \mathbf{R}(t)) = I_0 \stackrel{\text{def}}{=} I(E(0), \mathbf{R}(0)). \tag{45}$$

This is the classical adiabatic theorem.

In higher dimensions, the classical adiabatic theorem generalizes provided the dynamics is either *integrable* or *ergodic*. Integrable systems are characterized by a high degree of symmetry. They possess N independent constants of the motion (where N is the number of degrees of freedom) which are in involution (that is, the Poisson bracket between any pair of them vanishes). Typical orbits in an integrable systems execute periodic or quasiperiodic motion on N dimensional tori embedded in the $2N$ dimensional phase space, and the equations of motion may be solved more or less explicitly. In sharp contrast stand ergodic systems, for which energy is the only constant of the motion. Typical orbits cover the $(2N-1)$-dimensional energy surface uniformly, so that time averages and microcanonical ensemble averages coincide. The adiabatic theorems in higher dimensions are rather more difficult to state and considerably harder to prove. The difficulty is related to the fact that the adiabatic invariants are not so well conserved as in the one-dimensional case (Lochak and Meunier 1988).

5.1.2. Basic account of the Hannay angle. The classical adiabatic theorem (45) determines the time evolution of the energy of a one-dimensional oscillator under a slowly changing Hamiltonian. Hannay (1985), motivated by Berry's analysis of the quantum geometric phase, asked a more refined question: what is the evolution of the phase of oscillator?

To frame this question precisely we introduce canonical coordinates, the *action angle variables* (θ, I). These are related to (q, p) by a canonical transformation

$q = q(\theta, I, \mathbf{R})$, $p = p(\theta, I, \mathbf{R})$ which depends on parameters. The action I, the generalized momentum, determines the energy of the oscillator according to (44). The angle θ, the generalized coordinate, describes the phase of the oscillator, and varies periodically around the orbit in the phase plane with period 2π. Expressed in terms of action-angle variables, the Hamiltonian $H(I, \mathbf{R})$ is independent of θ, so the canonical equations of motion, $\dot{\theta} = \partial_I H$, $\dot{I} = -\partial_\theta H$, imply that I is constant while θ evolves at a constant frequency $\omega \stackrel{\text{def}}{=} \partial_I H$.

Now suppose the parameters $\mathbf{R}$ of the Hamiltonian are taken slowly round a closed cycle C, returning to their original values $\mathbf{R}(0)$ after a time T. Since the action is nearly conserved, a trajectory returns approximately to the orbit, or contour of $H(I, \mathbf{R}(0))$, on which it started. Of course, it will be displaced by an angle $\Delta\theta$ from its initial position. Hannay obtained the two dominant contributions to $\Delta\theta$ (further corrections are of order $1/T$). The first is the *dynamical angle* $\theta_d = \int_0^T \omega(I, \mathbf{R}(t))dt$. It is simply the time integral of the frequency, taking into account its slow variation with parameters. The second, unexpected, contribution,

$$\theta_g(I, C) = -\partial_I \oint_C \mathbf{A}(I, \mathbf{R}) \cdot d\mathbf{R}, \tag{46}$$

is the *geometric angle*, or the *Hannay angle*. It is expressed in terms of the line integral around C of a vector field $\mathbf{A}(I, \mathbf{R})$ given by

$$\mathbf{A}(I, \mathbf{R}) = \frac{1}{2\pi} \int_0^{2\pi} p(\theta, I, \mathbf{R}) \nabla q(\theta, I, \mathbf{R})\, d\theta \tag{47}$$

The Hannay angle is geometric in the same sense as is the geometric phase: it depends only on the parameter cycle C, and not on the rate it is traversed. Using Stokes' theorem, the Hannay angle can also be expressed as the surface integral $-\partial_I \int_S \mathbf{V}(I, \mathbf{R}) \cdot d\mathbf{S}$, where S is a surface bounded by C and $\mathbf{V} = \nabla \times \mathbf{A}$. From (47),

$$\mathbf{V}(I, \mathbf{R}) = \frac{1}{2\pi} \int_0^{2\pi} \nabla p(\theta, I, \mathbf{R}) \times \nabla q(\theta, I, \mathbf{R})\, d\theta. \tag{48}$$

As with the geometric phase, the presence of time-reversal symmetry (cf Section 1.5) imposes constraints on the Hannay angle. In particular, if the orbits themselves are time-reversal symmetric (for all values of parameters), then the Hannay angle vanishes identically.

Like the geometric phase, the Hannay angle may be interpreted as the holonomy of a connection on a fibre bundle (cf Section 2.1.3). The parameters $\mathbf{R}$ constitute the base manifold, and the fibres attached to them are orbits of the Hamiltonian $H(q, p, \mathbf{R})$ with given action I. These orbits are parameterized by the angle θ, and the connection $\partial_I \mathbf{A}(I, \mathbf{R})$ describes how θ changes under parallel transport.

The Hannay angle can be detected by determining the initial and final angles of an adiabatically cycled oscillator, and subtracting from their difference the dynamical

angle θ_d. As θ_d depends only on the parameter cycle $\mathbf{R}(t)$, it can be computed without solving the equations of motion explicitly. In this way, Hannay's analysis, together with the adiabatic theorem (45), provides a complete asymptotic solution of the equations of motion for slowly varying Hamiltonians in one dimension. The analysis generalizes directly to integrable systems in higher dimensions, where there are as many Hannay angles as there are degrees of freedom. However, detecting them can be a more delicate matter, as the adiabatic theorem and corrections thereto become more complicated in higher dimensions (Golin *et al* 1989). For ergodic systems, there is no general analogue of the Hannay angle, though a related holonomy phenomenon emerges in some special cases (Robbins 1994, Jarzynski 1995, Schroer 1996).

3.1.3. Examples. A bead moving freely round a planar wire loop provides a simple example of the Hannay angle. Suppose the bead has unit mass, and the loop has circumference C, area A, and is otherwise arbitrary in shape. While the system as a whole has time-reversal symmetry, the individual orbits do not (they are either clockwise or counterclockwise). The action I is just the velocity v, and the angle $\theta = 2\pi s/C$ is proportional to the distance s along the wire. Suppose the loop is slowly rotated once around an interior point over a time T, and let Δs denote the total distance the bead has travelled during this time. The dynamical contribution s_d (which is $C/2\pi$ times the dynamical angle θ_n) is just vT (v is the adiabatic invariant). The geometric contribution s_g may be computed to be $-4\pi A/C^2$. It accounts for the distance the bead has slipped as the wire turned. For a fixed circumference C, $-s_g$ assumes its maximal value, C, for a circular loop; in this case, the bead simply slips a full revolution. For thin, elongated loops, s_g approaches zero; the only slippage occurs at the hairpin turns.

The second example is the family of harmonic oscillators $H = (Z - X)p^2/2 + Ypq + (Z + X)q^2/2$, where the term in qp breaks time-reversal symmetry. For $X^2 + Y^2 < Z^2$, the motion is bounded and oscillatory, and we consider adiabatic cycles confined to this region of parameter space. The flux field $\mathbf{V}$ of (48) may be computed to be

$$\mathbf{V}(I, \mathbf{R}) = -\frac{I}{2}\frac{\mathbf{R}}{(Z^2 - X^2 - Y^2)^{3/2}}. \tag{49}$$

Note that because $\mathbf{V}$ is linear in I, the Hannay angle θ_g is independent of I. θ_g may be described geometrically as follows. We regard the parameter space as three-dimensional Minkowski space, in which X and Y are space-like coordinates, Z is time-like, and the region $X^2 + Y^2 < Z^2$ is the interior of the light-cone. Then $\theta_g(C)$ is the area of the projection of the closed curve C onto the unit mass shell $Z^2 - X^2 - Y^2 = 1$.

The precession of a Foucault pendulum is yet another example of a Hannay angle. Consider a spherical pendulum initially set to oscillate in a plane. During the course of a day, the axis of gravity (whose direction represents the parameters of the pendulum) is carried once round a line of latitude α. At the day's end, the plane of oscillation has turned through an angle $2\pi(1 - \cos\alpha)$. This is just the solid angle described by the axis of gravity, and may be computed as a Hannay angle using (46). Indeed,

the phenomenon described by the Hannay angle is more general. If one turned the earth about its centre adiabatically (but otherwise arbitrarily) before returning it to its original orientation, the plane of oscillation of the spherical pendulum would still precess through an angle equal to the solid angle described by the earth's axis.

The spherical pendulum is a two degree of freedom system, integrable because of its axial symmetry. Therefore, it has two Hannay angles, one each for its polar and azimuthal degrees of freedom. For planar oscillations, however, only the azimuthal Hannay angle, which describes the Foucault precession, is nonzero.

Let us briefly mention two final examples drawn from plasma physics and astronomy, respectively. Littlejohn (1982) found a Hannay angle in the cyclotron oscillations of charged particles in strong magnetic fields. Berry and Morgan (1996) (see also Golin *et al* (1989)) compute "the Hannay angle of the world", an annual displacement of the earth along its orbit of approximately 150 meters (or 2×10^{-4} seconds of arc) caused by the adiabatic and periodic variation of Jupiter's gravitational field.

3.1.4. Semiclassical limit. Following Hannay's work, Berry (1985) showed that the Hannay angle emerges in the semiclassical limit of the geometric phase. The precise relation between the two is given by

$$\gamma_{n-1}(C) - \gamma_n(C) \overset{\hbar \to 0,\, n \to \infty}{\to} -\theta_J(I, C). \tag{50}$$

in which the classical action is related to the quantum number by the semiclassical quantization condition $I = n\hbar$. (50) conforms to the usual relation between quantum phases and classical angles, the most familiar instance being the relation $(E_{n-1} - E_n)/\hbar \to \omega$ between energy levels and frequencies, whose time integral connects the dynamical phase and the dynamical angle. The result (50) is derived by substituting the WBK approximation for the eigenstates $|n(\mathbf{R})\rangle$ into the formula (4) for the quantum vector potential $\mathbf{A}_n$. One obtains thereby the expression (47) for the classical vector potential. This analysis extends to integrable systems in more than one dimension, for which semiclassical eigenfunctions may be similarly determined.

On the other hand, for systems whose classical dynamics is chaotic, the quantum geometric phase has no classical limit in general. However, the vector field $\mathbf{V}_n$, when averaged over quantum number, does have a classical limit (Robbins and Berry 1992), which turns out to play a role in the classical Born-Oppenheimer approximation (cf Section 3.1.5). The classical limit

$$V(E, \mathbf{R}) = \frac{1}{2\partial_E \Omega(E, \mathbf{R})} \partial_E \left(\partial_E \Omega(E, \mathbf{R}) \int_0^\infty \langle \nabla H_t(\mathbf{q}, \mathbf{p}, \mathbf{R}) \times \nabla H(\mathbf{q}, \mathbf{p}, \mathbf{R}) \rangle_E \, dt \right) \tag{51}$$

is expressed in terms of the time integral of an antisymmetric correlation function of the parameter gradient of the Hamiltonian, $\nabla H(\mathbf{q}, \mathbf{p}, \mathbf{R})$. $\Omega(E, \mathbf{R}) = \int_{H(\mathbf{q},\mathbf{p},\mathbf{R}) < E} d^N q d^N p$ denotes the phase volume inside the energy shell $H = E$. $\langle \cdot \rangle_E$ denotes the normalized microcanonical average, and ∇H_t denotes the function ∇H evolved for a time t with the classical dynamics. In one dimension, (51) is equivalent to (48).

3.1.5. Classical reaction forces. As in Section 1.7, we suppose the parameters $\mathbf{R}$ are no longer externally prescribed, but instead are the coordinates of a slow system to which the original one, the one-dimensional oscillator of Section 3.1.1, is coupled. Both systems are governed by classical mechanics. We take their Hamiltonian to be $P^2/2M + H(q, p, \mathbf{R})$, where $\mathbf{P}$ is the slow momentum. The Hannay angle describes a holonomy in the fast dynamics induced by the slow dynamics. As in the quantum case, there is a reciprocal effect on the slow dynamics induced by the fast dynamics. This effect is found by averaging over the fast motion. According to the adiabatic theorem, the action of the fast system is nearly constant. In the lowest order approximation, we treat the action as strictly constant. Then the energy $H(I, \mathbf{R})$ acts as a potential in the slow dynamics, whose effective Hamiltonian is given by $P^2/2M + H(I, \mathbf{R})$. This is the classical analogue of the Born-Oppenheimer approximation.

First-order corrections account for the fluctuations of the action about its mean. These generate an effective vector potential in the slow Hamiltonian, so that $\mathbf{P} \to \mathbf{P} - \mathbf{A}(I, \mathbf{R})$. The vector potential is precisely the field (47) whose line integral gives the Hannay angle. This is classical geometric magnetism.

One example of geometric magnetism is the precession of a symmetric top (Berry and Robbins 1993a). In this case, the top's axis is the slow system, and the rotation about this axis is the fast system. Another example from plasma physics is the drift of the guiding centre of electrons in strong, slowly-varying magnetic fields (Littlejohn 1988). The guiding centre, ie the centre of the instantaneous Larmor orbit, constitutes the slow degrees of freedom (it remains fixed in constant fields), and the motion about the guiding centre, the fast degree of freedom.

Geometric magnetism generalizes to higher dimensional fast systems, both integrable and chaotic. In the latter case, geometric magnetism is partnered by an additional dissipative force, deterministic friction (Wilkinson 1990); the slow dynamics is no longer conservative, as the chaotic motion of the fast system acts as an effective heat bath. The magnetic and dissipative forces appear as the antisymmetric and symmetric parts of a linear response tensor (Berry and Robbins 1993b, Jarzynski 1993).

3.2. Symmetry and kinematic geometric phases

Kinematic geometric phases arise in systems with symmetry. They describe the evolution in degrees of freedom which the symmetry acts upon (eg, orientational degrees of freedom in case of rotational symmetry). Their structure depends only on how the symmetry is realized, and not on the details of the dynamics. The geometric character of their evolution is exactly enforced by the equations of motion, and does not depend on assumptions of adiabaticity. In the following sections, two different treatments of kinematic geometric phases are described.

3.2.1. Falling cats and swimming paramecia. Even with nothing to push against, bodies in free fall, such as cats and platform divers, can change their orientation

through internal motions. Perhaps counterintuitive at first, this phenomenon is a direct consequence of conservation of angular momentum, and can be calculated as such. The main difficulty in doing so lies in the definition of the body's orientation. This problem does not arise for rigid bodies, whose every configuration is related to a chosen reference by a unique rotation. But for a body which can bend and twist and stretch, ie which has internal degrees of freedom, it is not clear which configurations are 'upright' and which ones are 'tilted'.

There are various prescriptions for determining orientation. One is to refer to the orientation of the principal axes of inertia. Another is through the Eckart frame, a common convention in molecular physics. Both prescriptions have particular advantages: neither is well defined in all cases.

In fact, the structure of the solution emerges most clearly if the definition of orientation is made more or less arbitrarily. This is the point of view taken by Shapere and Wilczek (1989b). For simplicity, suppose the body consists of a finite number N of masses m_α at positions $\mathbf{r}_\alpha$. No assumption is made concerning their dynamics, except that they obey Newton's laws and that there are no external forces. By choosing a suitable reference frame, we may assume that the centre of mass $\sum_\alpha m_\alpha \mathbf{r}_\alpha / \sum_\alpha m_\alpha$ begins and remains at the origin.

By some means, certain configurations $(\mathbf{R}_1, \ldots, \mathbf{R}_N)$ are designated to be 'upright'. This designation must be consistent, in that if two configurations are related by a rotation, then only one of them can be upright. It must also be complete, so that every configuration can be made upright by applying a rotation. If both these conditions hold, then every configuration $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ can be obtained by applying a unique rotation $\mathcal{R}$ to a unique upright configuration $(\mathbf{R}_1, \ldots, \mathbf{R}_N)$. The body's motion,

$$\mathbf{r}_\alpha(t) = \mathcal{R}(t) \cdot \mathbf{R}_\alpha(t), \quad \alpha = 1, \ldots, N, \tag{52}$$

is then determined by the evolution of its orientation $\mathcal{R}(t)$ and its internal dynamics $\mathbf{R}_1(t), \ldots, \mathbf{R}_N(t)$. In the language of rigid body dynamics, $\mathbf{R}_\alpha(t)$ describes the motion in the body frame, and $\mathbf{r}_\alpha(t)$ the motion in the space frame.

An object fixed in the space frame appears to rotate with angular velocity $-\boldsymbol{\Omega}(t)$ in the body frame. $\boldsymbol{\Omega}(t)$ is defined by the relation

$$\dot{\mathcal{R}} \cdot \mathbf{X} = \mathcal{R} \cdot (\boldsymbol{\Omega} \times \mathbf{X}), \tag{53}$$

which can be made to hold for arbitrary vectors $\mathbf{X}$.

Regardless of the details of the dynamics, the total angular momentum $\mathbf{L} = \sum_\alpha m_\alpha \mathbf{r}_\alpha \times \mathbf{v}_\alpha$ in the space frame must be conserved. Conservation of angular momentum leads to the relation

$$\boldsymbol{\Omega}(t) = I^{-1} \cdot \left(\mathbf{L} - \sum_\alpha m_\alpha \mathbf{R}_\alpha \times \mathbf{V}_\alpha \right) \tag{54}$$

between the angular velocity $\boldsymbol{\Omega}$ and the angular momentum $\mathbf{L} = \mathcal{R}^{-1} \cdot \mathbf{l}$, velocities $\mathbf{V}_\alpha = \dot{\mathbf{R}}_\alpha$, and inertia tensor $I_{ij} = \sum_\alpha m_\alpha (R_\alpha^2 \delta_{ij} - R_{\alpha i} R_{\alpha j})$ in the body frame. Once

the internal dynamics $\mathbf{R}_a(t)$ is known, the relations (53) and (54) may be used to determine the orientation $\mathcal{R}(t)$.

The case of zero angular momentum ($\mathbf{I} = \mathbf{L} - 0$) is particularly interesting. Then one can show that the orientation $\mathcal{R}(t)$ does not depend on the rate of the internal dynamics. (One can, in effect, multiply through by dt in (53) and (54).) In particular, if the internal motion is periodic along a cycle C, then after one period it generates an overall rotation $\mathcal{R}_g = \mathcal{R}(T)\mathcal{R}(0)^{-1}$ of the body. The rotation $\mathcal{R}_g$ depends only on the cycle C, and not on the rate it is traversed. It is a purely geometrical quantity, a nonabelian analogue (rotations do not commute) of the Hannay angle. It is the means by which a cat dropped upside-down manages to land on its feet

Like the other geometric phases considered so far, this can be described in terms of holonomy on a fibre bundle. The fibre bundle is the configuration space of the masses m_a, constrained to have their centre of mass at the origin. The base manifold consists of the distinct internal configurations, no two of which can be made to coincide by a rotation. (Mathematically, this is the quotient of the configuration space by the rotation group.) It can be parameterized by the upright configurations, provided these are always well-defined. The fibre consists of the different possible orientations of a given internal configuration, and may be parameterized by the rotation group. A connection determines an infinitesimal rotation $\delta\mathcal{R}$ for an infinitesimal change of internal configuration $\delta\mathbf{R}_a$, and the physically relevant connection is determined by conservation of angular momentum. Parallel transport round a closed cycle C in the base yields the holonomy $\mathcal{R}_g$. The curvature of this connection, which describes the net rotation after an infinitesimal cyclic change of shape, may be computed using a formula analogous to (28).

The designation of upright configurations is arbitrary, and one can always choose them differently. It can be verified that the holonomies $\mathcal{R}_g$ do not depend intrinsically on this choice, and that the various quantities which describe the orientation transform appropriately when a different choice is made. In particular, the curvature and connection transform according to equations similar to (30).

A more complex example studied by Shapere and Wilczek (1989c) concerns swimming through viscous fluids at low Reynolds number. High viscosity and low velocity means that bodies remain at rest unless acted upon by a force – a realization of Aristotelian mechanics! Bodies locomote by executing a series of internal motions, ie swimming strokes. Conservation of linear and angular momentum determine the consequent changes in position and orientation, but now the motion of the fluid must be taken into account. Solving the fluid dynamical problem leads to a relation similar to (54) for both the linear and angular velocity in terms of rates of change of shape. Applications in biology include locomotion of microorganisms.

5.2.2. Geometric phases and reduction. Kinematic geometric phases can be framed in terms of the general theory of symmetry in Hamiltonian systems, called *reduction*. This formulation, extensively developed by Marsden, Montgomery and coworkers

(Marsden *et al* 1991, Marsden 1992), has led to a unified perspective on a variety of geometric phase phenomena as well as a number of interesting examples.

Dynamics is simplified by the presence of symmetry, and Hamiltonian dynamics especially so. A symmetry always reduces the dimension of a dynamical system by one, but in the Hamiltonian case, it reduces it by two. Indeed, when there is sufficient symmetry, the systems are integrable, and the dynamics can be solved analytically (cf Section 3.1.1).

The simplest symmetry is translation invariance in an ignorable coordinate q_N. Hamilton's equation $\dot{p}_N = -\partial H/\partial q_N = 0$ (or, more generally, Noether's theorem) implies that the conjugate momentum p_N is conserved. Fixing its value $p_N = p_{N0}$ leads to a Hamiltonian $H(q_1, \ldots, q_{N-1}, p_1, \ldots, p_{N-1}, p_{N0})$ in two fewer variables. Once their dynamics is obtained, the evolution of the ignorable coordinate may be determined by integrating Hamilton's equation $\dot{q}_N = \partial_{p_N} H(q_1, \ldots, q_{N-1}, p_1, \ldots, p_{N-1}, p_{N0})$. The theory of reduction extends this analysis to general symmetries. As in the simple example, the analysis consists of two stages. The first is reduction, wherein the conserved momenta are fixed, the ignorable coordinates ignored, and Hamilton's equations expressed in terms of a reduced set of coordinates. In all but the simplest cases, finding these coordinates is a nontrivial task (Jacobi's "elimination of the node" in the rotationally symmetric N-body problem is a representative example). The second state is reconstruction, wherein the evolution of the ignorable coordinates is determined from the reduced dynamics. It is in the reconstruction stage that kinematic geometric phases can appear. It turns out that the equations of motion for the ignorable coordinates can be separated into dynamical and geometric terms. The geometric term depends only on the path described by the reduced dynamics, and not on the rate it is traversed. The dynamical term, by contrast, does depend on the rate. Moreover, the form of the geometrical term depends only on the realization of the symmetry, and not on the details of the Hamiltonian.

In the example discussed in the preceding section, the flexible body composed of N point masses, the ignorable coordinates correspond to $\mathcal{R}$, the body's orientation, the conserved momentum to the space-fixed angular momentum $\mathbf{L}$, and the internal coordinates $\mathbf{R}_a(t)$ to the reduced system. The evolution of $\mathcal{R}(t)$ is obtained by integrating the linear system (53), in which angular velocity $\boldsymbol{\Omega}(t)$ is given in terms of the reduced dynamics by (54). Eq. (54) is an example of a reconstruction formula divided into its dynamical and geometric parts. The first term, $I^{-1}\cdot\mathbf{L}$, is the dynamical contribution. The second, $I^{-1}\cdot(\sum_a m_a \mathbf{R}_a \times \mathbf{V}_a)$, is the geometric contribution. As discussed above, its integral depends only on the path C described by the internal dynamics $\mathbf{R}_a(t)$.

Montgomery (1991) found an even simpler example in the dynamics of a free rigid body. Its configuration (with its centre of mass fixed at the origin) is described by a rotation $\mathcal{R}$, for which it will be convenient to use the Euler angle parameterization, $\mathcal{R} = \mathcal{R}_z(\varphi)\mathcal{R}_x(\theta)\mathcal{R}_z(\psi)$. The space-fixed angular momentum $\mathbf{L}$ is a constant of the motion, and without loss of generality we can assume it lies in the z direction. The

body-fixed angular momentum $\mathbf{L}(t) = R(t) \cdot \mathbf{l}$ has polar coordinates $(\theta(t), \phi(t))$ on the sphere of constant radius $L = |\mathbf{L}| = |\mathbf{l}|$. Its dynamics is determined by the Euler equations $\dot{\mathbf{L}} = \boldsymbol{\Omega} \times \mathbf{L}$, where $\boldsymbol{\Omega} = I^{-1} \cdot \mathbf{L}$ is the angular velocity and I is the inertia tensor. The Euler equations describe the reduced dynamics, in which $\mathbf{L}$, or equivalently θ and ϕ, are the reduced coordinates. The ignorable coordinate is the third Euler angle ψ.

Under the Euler equations, $\mathbf{L}$ executes a periodic motion with period T. The fact that $\mathbf{L}(T) = \mathbf{L}(0)$ implies that the rigid body returns after one period to its original orientation, up to a rotation about the direction of $\mathbf{L}(T)$. The amount of rotation is just $\Delta\psi = \psi(T) - \psi(0)$, the shift in the ignorable coordinate ψ.

With a little algebra, the equation of motion for ψ may be written in the form

$$\dot{\psi} = \frac{2E}{L} + \cos\theta\dot{\phi}, \tag{55}$$

where $E = \frac{1}{2}\mathbf{L} \cdot I^{-1} \cdot \mathbf{L}$ is the conserved rotational energy. (55) is another example of a reconstruction formula divided into dynamical and geometric contributions. In this case, the dynamical contribution integrates to $2ET/L$, whereas the geometric contribution gives (modulo 2π) the solid angle $\oint_C \cos\theta d\phi$ described by the body-fixed angular momentum $\mathbf{L}(t)$ over a single period.

Another example, due to Alber and Marsden (1992), concerns phase shifts in soliton collisions. A fundamental property of certain nonlinear partial differential equations, such as the Korteweg-de Vries equation $U_t + 6UU_x - U_{xxx} = 0$, is the existence of *solitons*. These are waveforms $U(x - vt)$ which propagate at a fixed velocity v whilst maintaining their shape, the nonlinearity in the wave dynamics just balancing the dispersion. When two solitons collide, they emerge after a complicated but transient period of interaction essentially intact. The only signature of their collision is a shift Δx in their position $x \pm \Delta x - vt$ (+ for the faster soliton, − for the slower one). This shift, which may be determined from an asymptotic analysis of the exact solutions, can be interpreted as a geometric phase. The same analysis applies to a variety of integrable nonlinear partial differential equations, including the nonlinear Schrödinger equation and the Klein-Gordan equation.

3.2.3. Optimization and control. The study of kinematic geometric phases suggests a problem in optimization. What is the most efficient internal motion for producing a given holonomy? As a specific example, consider a satellite orbiting the earth with a fixed orientation, so that its angular momentum about its centre of mass is zero. It is desired to change its orientation using as little energy as possible. The obvious method of applying external torques (supplied by jets attached to satellite, for example) has the drawback that any uncompensated force or torque leaves the satellite in a state of translation or rotation. The alternative is to execute a sequence of internal motions (e.g. flywheels driven by motors); in the manner described above, these generate a change in orientation without altering the total linear or angular momentum.

The problem of finding the most efficient cycle with a given holonomy leads to a constrained variational problem in the internal degrees of freedom, which can be formulated in terms of control theory. The solution depends very much on how efficiency is measured. In certain circumstances, it may be advantageous to make some sizeable internal motions initially in order to achieve a configuration for which the geometric phase curvature form (cf Section 2.1.2) is large; then much smaller subsequent motions can produce relatively large effects. In other situations, it may be necessary to restrict to small motions about some equilibrium configuration; in this case, the variational problem may be analyzed in terms of infinitesimal circuits. The solutions often have a physical interpretation in terms of motion in a magnetic, or more generally, a Yang Mills, field. Shapere and Wilczek (1989d) find efficient swimming strokes through viscous media in two- and three dimensions under various conditions. Marsden (1992) describes a variety of other examples.

4. Optics

Optics has provided fertile ground for studying the geometric phase. A comprehensive survey of experimental activity is given Bhandari (1996). Most investigations have concentrated on polarization effects. The polarization of a beam of monochromatic, coherent light behaves like a quantum mechanical spin, and under cyclic evolutions (and noncyclic ones as well) acquires a geometric phase similar to that of a spinning particle in a magnetic field (cf Section 1.3). There are various ways to change polarization. One is by varying the direction of propagation, as the polarization vector must remain transverse (Tomita and Chiao 1986); in this case the polarization has a spin-1, or vectorial character. Another is through optical activity induced by anisotropic media, in which case the polarization has a spin-1/2 character (Pancharatnam 1956, Bhandari and Samuel 1988, Bhandari 1988).

A beam of light passing through a cycle of polarization states may be recombined with a second beam whose polarization has remained unchanged. Their total phase difference (dynamical − geometric) can be determined, for example, from their interference pattern. In certain cases, the dynamical phase shift can be made to vanish. In others, it can be explicitly computed and subtracted from the total phase. In either case, one obtains a measurement of the geometric phase.

4.1. Geometric phase of coiled light

In a straight, single mode optical fibre of circular cross section, there is no preferred state of polarization: whatever the polarization state of the source, it is preserved along the fibre. This is no longer so if the fibre is made to bend, then the polarization must change in order to remain transverse to the direction of propagation $\hat{\mathbf{t}}$. Provided the radius of curvature $R = |d\hat{\mathbf{t}}/ds|^{-1}$ is large compared to the wavelength λ (here s measures distance along the fibre), the polarization vector $\vec{\mathbf{d}}(s)$ of the electric

displacement $\mathbf{D}(s,\rho) = \mathrm{Re}\, e^{i(2\pi s/\lambda - \omega t)} f(\rho)\hat{\mathbf{d}}(s)$ (here $f(\rho)$ is the radial modal profile) evolves along the fibre according to

$$\frac{d\hat{\mathbf{d}}}{ds} = -(\hat{\mathbf{t}}.\hat{\mathbf{d}})\hat{\mathbf{t}}. \tag{56}$$

The corrections to this approximate solution of Maxwell's equations are of order λ/R.

For a general elliptical polarization, $\hat{\mathbf{d}}$ is complex. Within the adiabatic approximation, both its real and imaginary parts $\hat{\mathbf{d}}_r$ and $\hat{\mathbf{d}}_i$ remain orthogonal to the direction of propagation $\hat{\mathbf{t}}$, and each satisfies (56) separately. As a function of s, $\hat{\mathbf{t}}(s)$ describes a curve C on the unit sphere, and the vectors $\hat{\mathbf{d}}_r(s)$ and $\hat{\mathbf{d}}_i(s)$ are tangent to the sphere at $\hat{\mathbf{t}}(s)$. Equation (56) is the condition for their parallel transport. (Infinitesimal parallel transport of a tangent vector $\hat{\mathbf{d}}$ from $\hat{\mathbf{t}}$ to $\hat{\mathbf{t}} - d\hat{\mathbf{t}}$ is accomplished by first rigidly transporting $\hat{\mathbf{d}}$ to $\hat{\mathbf{t}} + d\hat{\mathbf{t}}$ and then projecting the translated vector onto the plane tangent to $\hat{\mathbf{t}} + d\hat{\mathbf{t}}$.) If the curve C is made to close on itself, the vectors $\hat{\mathbf{d}}_r$ and $\hat{\mathbf{d}}_i$ return rotated through an angle θ_g equal to the solid angle subtended by C. θ_g is a geometric phase, depending only on the shape of C.

Tomita and Chiao (1986) observed this effect by introducing linearly polarized light from a He-Ne laser into a length of helically wound optical fibre whose initial and final directions coincided. The direction of linear polarization of the emergent beam was measured and found to have turned through an angle equal to the solid angle subtended by the cycle of directions. The experiment was performed with helices of both fixed and variable pitch, as well as for fibres constrained to lie in the xy plane. In this last case, the cycle of directions lies on the equator of the unit sphere, and the subtended solid angle is a multiple of 2π. As predicted, the direction of linear polarization remained unchanged.

The original analysis of the Tomita-Chiao experiment was based on quantum mechanical considerations, treating a photon propagating in the fibre as a spin-one particle whose helicity is constrained to be ± 1. (Classically this corresponds to left- and right-circularly polarized light.) Just as a spin with nonzero magnetic moment tracks the direction of a slowly changing magnetic field (cf Section 1.3), the helicity of the photon tracks the direction of propagation $\hat{\mathbf{t}}$, and acquires a geometric phase equal to the solid angle described by $\hat{\mathbf{t}}$. By decomposing a linear polarization into a superposition of left- and right circular polarizations, the observed rotation of linear polarization may be recovered.

A systematic derivation requires a more detailed analysis, most easily carried out within the context of classical (rather than quantum) electromagnetic theory. Indeed, the essential features of such an analysis are contained in work of Rytov (1938) and Vladimirsky (1941), who derived and interpreted the parallel transport law (56) in the short wavelength, or geometrical optics, limit (see (Berry 1990) for a discussion). Strictly speaking, geometrical optics is not appropriate for single-mode fibres, in which the transverse wavelength is comparable to the fibre diameter. Berry (1987a) has carried out a classical analysis using the full Maxwell equations, and it is the conclusions of this analysis which are described above.

Kitano *et al* (1987) performed an experiment related to Tomita and Chiao's, in which the cyclic change in the direction of propagation was accomplished nonadiabatically through a series of discrete reflections from mirrors. These reflections reverse helicity instead of conserving it, and the calculation of the geometric phase must be appropriately modified.

4.2. Pancharatnam's phase

When light passes through a homogeneous medium, its direction of propagation remains constant. However, if the medium is optically anisotropic, the polarization can change. Such changes in polarization are accompanied by a geometric phase, discovered by Pancharatnam in 1956. (See (Berry 1987b) for a contemporary discussion of Pancharatnam's work.) For definiteness, let us consider plane electromagnetic waves propagating in the z direction with fixed wavenumber k and frequency ω. The electric field $\mathbf{E}(\mathbf{r},t) = \mathrm{Re}\,\sqrt{I_A}e^{i(kz-\omega t)}|A\rangle$ is described by its intensity I_A and normalized polarization vector $|A\rangle = (A_x, A_y)$, whose complex components determine the phase and relative intensity along x and y. (Dirac notation is used for complex two-vectors to highlight an analogy with two-state systems, introduced below.)

Given two such waves, what is their relative phase? Unless their polarization states $|A\rangle$ and $|B\rangle$ happen to be proportional, it is not clear that the question is a sensible one. Pancharatnam found a natural and physically relevant prescription for their relative phase by considering the intensity I_{A+B} of the superposition, $\sqrt{I_A}|A\rangle + \sqrt{I_B}\,|B\rangle$. This intensity is given by

$$I_{A+B} = I_A + I_B + 2\sqrt{I_A I_B}\,\mathrm{Re}\,\langle A\,|B\rangle \tag{5.7}$$

If the constituent intensities I_A and I_B are fixed, the superposed intensity depends only on the interference term, and is maximized when $\langle A|B\rangle = 1$. This is Pancharatnam's criterion for the two polarization states $|A\rangle$ and $|B\rangle$ to be in phase. In general, their relative phase is defined to be $\arg\langle A|B\rangle$ (this is well defined unless $|A\rangle$ and $|B\rangle$ are orthogonal). It is the amount by which the phase of $|A\rangle$ must be retarded relative to the phase of $|B\rangle$ in order that the intensity of their superposition be a maximum.

Being in phase is not a transitive relation: if $|A\rangle$ is in phase with $|B\rangle$ and $|B\rangle$ is in phase with $|C\rangle$, then, in general, $|C\rangle$ is not in phase with $|A\rangle$. The relative phase $\arg\langle C|A\rangle$ has a simple geometrical interpretation, which we now describe. To every polarization state $|A\rangle$ may be associated a vector $\mathbf{e}_A$ on the unit sphere, namely the unit vector $\langle A|\boldsymbol{\sigma}|A\rangle$ (here $\boldsymbol{\sigma}$ denotes the Pauli matrices) rotated through $\pi/2$ about the y-axis. This is how polarization is represented, up to an overall phase factor, on the *Poincaré sphere*. The y rotation, a matter of convention, puts the left and right circular polarizations at the north and south pole of the Poincaré sphere, and the linear polarizations at the equator. (In more mathematical language, the Poincaré sphere is the projective Hilbert space for two-state systems—cf Section 2.3.)

One can show that the relative phase $\arg(C|A)$ is equal to minus half the solid angle subtended by the spherical triangle $\mathbf{e}_A\mathbf{e}_B\mathbf{e}_C$ on the Poincaré sphere, which is constructed by joining $\mathbf{e}_A, \mathbf{e}_B$ and $\mathbf{e}_C$ with arcs of great circles. A direct demonstration (Pancharatnam's) uses formulas from spherical trigonometry. Another way is to realize that a sequence of polarization states $|\psi(t)\rangle$ parallel-transported according to (34) along great circles on the Poincaré sphere are, according to Pancharatnam's criterion, in phase with each other. Thus $\arg(C|A)$ is precisely the geometric phase accumulated under parallel transport round $\mathbf{e}_A\mathbf{e}_B\mathbf{e}_C$ (for which the half-solid-angle formula is obtained in Section 1.3).

To establish a connection between these geometrical considerations and the actual physics, some analysis of the polarization dynamics is required. Pancharatnam considered systems of *ideal polarizers*, or *analyzers*, which completely transmit a polarization state $|A\rangle$ and absorb the state $|A_\perp\rangle$ orthogonal to it in such a way that the incident and transmitted states are always in phase. Ideal polarizers are very nearly approximated by very thin polarizing plates. (Mathematically, they are represented by the hermitian projection operator $|A\rangle\langle A|$.) Therefore, under a polarization cycle generated by ideal polarizers, the initial and final polarizations differ by Pancharatnam's phase.

Polarization cycles can also be generated by *ideal retarders*: these are perfectly transparent devices which introduce a relative phase difference between two given orthogonal polarization states $|A\rangle$ and $|A_\perp\rangle$. A particular example is a quarter wave plate, for which the orthogonal states are linearly polarized and the phase difference is $\pi/2$. (Mathematically, an ideal retarder is represented by $\exp(-i\pi\mathbf{e}_A\cdot\boldsymbol{\sigma}/4)$, a two-dimensional unitary matrix with determinant one.) Polarization cycles generated by ideal retarders produce a dynamical phase in addition to Pancharatnam's phase. The dynamical phase is easily calculated, however.

Polarization cycles can also arise through optical activity in a birefringent gyrotropic medium. Such a medium induces a relative phase between orthogonal polarization states whose orientation is determined by the local dielectric tensor. When the variation in the medium is small over a wavelength, the polarization of a monochromatic beam tracks the eigenstates of the dielectric tensor. Berry (1987b) applied an adiabatic analysis to Maxwell's equations and calculated the phase change over a polarization cycle generated by a spatially periodic medium. The analysis is formally the same as for a spin-1/2 particle in a slowly changing magnetic field (Section 1.3), and there appears a dynamical phase and a geometric phase: the geometric phase in this case is just Pancharatnam's.

Observations of Pancharatnam's phase are based on two beam interferometry. Such experiments were initiated by Pancharatnam himself, though current versions are greatly improved with the availability of laser sources. Schematically, a plane wave with polarization $|A\rangle$ is split into two coherent beams which propagate along the arms of an interferometer. Beam 1 propagates freely, while beam 2 passes through a sequence of optical elements which induce a polarization cycle. The beams are

then recombined, and their total relative phase determined from their interference. Differences in dynamical phases are determined and subtracted, and the geometric phase subsequently measured.

Bhandari (1988) used laser interferometry methods to measure changes in the geometric phase, rather than the geometric phase itself. These are readily generated (while keeping the relative dynamical phases fixed) by continuously rotating a half-wave plate in the polarization cycle.

Kwiat and Chiao (1991) observed Pancharatnam's phase in single-photon interference experiments using a source of entangled two-photon states. One photon, the idler, was passed through a narrow-band energy filter to a detector. Observation of this photon fixed the energy of its partner, which was made to pass through an interferometer. Detection rates varied in accordance with the intensity, which depends directly on Pancharatnam's phase. Various coincidence detectors were used to confirm that genuine single-photon interference was being observed.

Berry and Klein (1996) performed a modern version of Pancharatnam's experiment, determining the relative phase by direct measurement of photographed interference fringes. By increasing the number of optical components in the polarization cycle, they were also able to investigate the relation between discrete and continuous geometric phases.

5. Molecular and Condensed Matter Physics

Several applications of the geometric phase to molecular spectroscopy (Section 5.1), nuclear magnetic resonance (Section 5.2), and the quantum Hall effect (Section 5.3) are discussed. Amongst a number of applications which have been omitted, we mention Λ-doubling of rotational levels in diatomic molecules (Moody *et al* 1986), and atomic and reactive molecular scattering (Mead and Truhlar 1979, Zygelman 1987, Mead 1992, Kuppermann and Wu 1993).

5.1. Pseudorotation in triatomic molecules

Ordinarily, the angular degrees of freedom of a spinless system have integer quantum numbers. This is because, ordinarily, the angular momentum L_z conjugate to an angle ϕ is taken to be $-i\hbar\partial_\phi$: its eigenfunctions $\exp(i\nu\phi)$ are single-valued only if ν is an integer. The Aharonov-Bohm effect (cf Section 1.2) provides an exception. In the presence of a magnetic flux line on the z axis, with flux $\alpha hc/e$, the azimuthal angular momentum of an electron is given by $-i\hbar\partial_\phi + \alpha$, the constant term α coming from the magnetic vector potential $\mathbf{A}^{(\alpha)}(\rho,\phi,z) = \alpha/\rho\,\hat{\boldsymbol{\phi}}$. Therefore, the eigenvalues of L_z are no longer integral multiples of $\hbar$, but are given instead by $(m-\alpha)\hbar$, where m is integral.

As discussed in Section 1.7, there emerges in the Born-Oppenheimer approximation a molecular analogue of the Aharonov-Bohm effect. For a given electronic state $|n\rangle$, the nuclear configuration space is threaded by a magnetic flux concentrated at

points where $|n\rangle$ is degenerate. This flux appears in the nuclear Hamiltonian as a geometric vector potential (4). As discussed by Mead (1992), a triatomic molecule X_3 composed of identical atoms (eg, alkali metal trimers such as Na_3) provides a simple class of examples. A striking consequence of the molecular Aharonov-Bohm effect is the appearance of anomalous half integer quantum numbers. Their existence has been confirmed experimentally (Delacrétez *et al* 1986). First anticipated by Longuet-Higgens *et al* (1958), this phenomenon is most naturally understood in terms of the geometric phase.

The nuclear configuration space for a triatomic molecule has nine degrees of freedom. Six of these correspond to overall translational and rotational motion, which can be ignored for the purposes of this discussion. That leaves three internal degrees of freedom, which describe vibrations. The internal configuration space, which we denote by Q, may be parameterized by the distances R_{12}, R_{23} and R_{13} between the nuclei. It may be thought of as the space of noncongruent triangles (the R_{ij}'s determine the lengths of the triangle's sides) whose vertices are distinguishable.

In the Born-Oppenheimer approximation, the electronic energy levels and eigenstates are regarded as functions of the internuclear distances R_{ij}. Assuming the molecular Hamiltonian has time reversal symmetry, we expect degeneracies of the electronic states to lie along one-dimensional curves in Q (cf Section 1.5).

It turns out that some of these degeneracies can be determined by symmetry considerations alone. For equilateral configurations $R_{12} = R_{23} = R_{13} = a$, the nuclear configuration is invariant under the dihedral group D_3 of symmetries of the equilateral triangle. Therefore, the electronic eigenstates $|n(a,a,a)\rangle$ must transform according to an irreducible representation, or irrep, of D_3 (Tinkham 1964). The group D_3 has two one-dimensional irreps, associated with states which are either invariant or else change sign under the symmetry operations, and a two-dimensional irrep, associated with a degenerate pair of states which transform into linear superpositions of each other under the symmetry operations. Therefore, if the state $|p(a,a,a)\rangle$ transforms according to the two-dimensional irrep, then the family of states $|p(R_{12},R_{23},R_{13})\rangle$ to which it belongs must be degenerate along the curve $R_{12} = R_{23} = R_{13} = a$ (with $a > 0$) in Q. Under parallel transport round this curve (cf (2)), $|p(R_{12},R_{23},R_{13})\rangle$ changes sign.

One might expect the energy $E_p(R_{12},R_{23},R_{13})$ of this state to assume its minimum in an equilateral configuration. The Jahn-Teller effect (Jahn and Teller 1937) shows this is not the case. It turns out to be energetically favourable for the molecule to assume an asymmetrical configuration, so that for the p-electronic state, the ground vibrational state of the molecule is distorted.

To analyze the vibrational motion, it is useful to introduce cylindrical-like coordinates, (ρ,ϕ,z), in Q. The vertical coordinate z is given by $R_{12}^2 + R_{23}^2 + R_{13}^2$ and determines the length scale; oscillations in z describe uniform dilations and contractions of the nuclei. The radial coordinate ρ is given by $\rho^2/4 = z^2 - 3(R_{23}^2R_{13}^2 + R_{12}^2R_{13}^2 + R_{12}^2R_{23}^2)$ and describes the degree of distortion. For collinear configurations, $\rho = z$; for equilateral configurations, $\rho = 0$. Thus, it is along the z

axis (where $\rho = 0$) that the electronic state $\psi(\rho, \phi, z)$ is degenerate. The angle ϕ is given by $\tan^{-1}[\sqrt{3}(R_{23}^2 - R_{13}^2)/(2R_{12}^2 - R_{23}^2 - R_{13}^2)]$, and describes the direction of the distortion. Displacements in ϕ are called *pseudorotations*. A 2π-pseudorotation returns the nuclei to their original positions, while a $2\pi/3$ rotation induces a cyclic permutation amongst them.

A systematic calculation of the vibrational Hamiltonian $H_{\rm vib}$, described in (Mead 1992), requires an explicit expression for the electronic energy $E_e(\rho, \phi, z)$ along with a careful derivation of the kinetic energy in terms of the (ρ, ϕ, z) coordinates. However, for this discussion, an approximate model treatment will suffice. The potential is taken to be harmonic in z and ρ and independent of ϕ, and the kinetic energy to be of the standard form. Thus

$$H_{\rm vib} = \frac{p_z^2}{2M} + \frac{L_z^2}{2M\rho^2} + \frac{p_\rho^2}{2M} + \tfrac{1}{2}M\omega_\rho^2(\rho - \rho_0)^2 + \tfrac{1}{2}M\omega_z^2(z - z_0)^2. \qquad (58)$$

where (ρ_0, z_0) denotes the asymmetrical configuration which minimizes E_e. The vibrational energy levels are characterized by three quantum numbers j, m and k, for the respective degrees of freedom ρ, ϕ and z, and are given approximately by

$$E(j, m, k) = (j + \tfrac{1}{2})\hbar\omega_\rho + \frac{(m\hbar)^2}{2I} + (k + \tfrac{1}{2})\hbar\omega_z. \qquad (59)$$

where the moment of inertia I about the z axis is taken to be $M\rho_0^2$. (The approximation consists of ignoring the dependence of I on the radial quantum number j).

The presence of the electronic degeneracy along the z axis introduces a geometric vector potential $\mathbf{A}_\rho = \alpha\hat{\boldsymbol{\phi}}/\rho$ into the nuclear Hamiltonian, which accounts for the sign change of the electronic eigenstate under parallel transport. (In the chosen coordinate system, $\alpha = \frac{1}{2}$.) Its effect is to replace the pseudoangular momentum operator $L_z = -i\hbar\partial_\phi$ by $-i\hbar\partial_\phi + \frac{1}{2}\hbar$, in analogy to what occurs in the magnetic Aharonov-Bohm Hamiltonian. Thus, the pseudorotational quantum number m in (59) is half-integral.

The half-integral quantization of pseudorotation was confirmed in a systematic spectral analysis of the vibrational spectrum of Na_3 by Delacrétaz *et al* (1986). A sequence of rovibrational levels corresponding to radial and pseudorotational motion were identified. It was found that the observed spectrum could be fitted to (59) only by assigning half-integer quantum numbers to the pseudoangular momentum. Such an assignment yields excellent agreement between observed and calculated energy levels.

5.2. Nuclear magnetic resonance

An rf magnetic field applied to polarized spins generates transitions between their stationary states. When the field is in resonance with the precession frequency of the spins, the transition probabilities can become large, even though the rf field may be weak. Transitions are manifested in oscillations of the spin magnetic moment, whose amplitude and frequency can be measured with great accuracy. This is the basis

for nuclear spin resonance experiments. In nuclear magnetic resonance (NMR), the nuclear spins are polarized by a static magnetic field. In nuclear quadrupole resonance (NQR), the spins are polarized by an electric field gradient (in the ambient crystalline potential, for example) which couples to the spin though the nuclear quadrupole moment.

Moody *et al* (1986) and Cina (1986) pointed out that cyclic variations in the rf field would generate geometric phases in the nuclear spins, which could be detected as shifts in the frequency spectrum of the nuclear magnetic moment (this is an example of the third type of experiment described in Section 1.6). Pursuing this suggestion, several groups have carried out NMR studies of the geometric phase, including its adiabatic (Tycho 1987, Suter *et al* 1987), nonadiabatic (Suter *et al* 1988), and nonabelian (Zwanziger *et al* 1990b) versions. These experiments have not only confirmed theoretical predictions, but have also brought to light interesting new phenomena in NMR spectroscopy.

In an NMR experiment, the nuclear spins (let $s = 1/2$, say) are initially polarized by a strong static magnetic field $B_z\hat{\mathbf{z}}$. The Hamiltonian H_0 is given by $\frac{1}{2}\hbar\gamma B_z\sigma_z$, where γ is the gyromagnetic ratio, and the eigenstates $|\pm\rangle$ have energies $\pm\hbar\omega_0/2$, where $\omega_0 = \gamma B_z$ is the Larmor frequency. For nonstationary states, the magnetic moment $\mathbf{M}(t) = \hbar\gamma\langle\psi(t)|\boldsymbol{\sigma}|\psi(t)\rangle/2$ precesses about $\hat{\mathbf{z}}$ at the Larmor frequency, so that its frequency spectrum has a single peak at ω_0.

A weak, circularly polarized rf field, $\mathbf{B}_{\rm rf}(t) = B_x\cos\omega t\hat{\mathbf{x}} + B_y\sin\omega t\hat{\mathbf{y}}$, is then applied. (Weak means that $B_{\rm rf} \ll B_z$). The subsequent dynamics is most easily analyzed in a rotating frame turning about $\hat{\mathbf{z}}$ with angular frequency ω. There the effective magnetic field $\mathbf{B}_{\rm rot}$ is static, and the Hamiltonian is given by

$$H_{\rm rot} = \tfrac{1}{2}\gamma\hbar\mathbf{B}_{\rm rot}\cdot\boldsymbol{\sigma}, \quad \mathbf{B}_{\rm rot} = B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}} + (B_z - \omega/\gamma)\hat{\mathbf{z}}. \tag{60}$$

The spin eigenstates $|\pm\rangle_{\rm rot}$ in the rotating frame are polarized along $\mathbf{B}_{\rm rot}$ with energies $\pm\hbar\omega_{\rm rot}/2$, and the magnetic moment $\mathbf{M}_{\rm rot}(t)$ precesses about $\mathbf{B}_{\rm rot}$ with frequency $\omega_{\rm rot}$, where $\omega_{\rm rot} = \gamma B_{\rm rot}$. Near resonance, where $\omega \approx \omega_0$, the axis of precession is nearly perpendicular to $\hat{\mathbf{z}}$ (even though $B_{\rm rf} \ll B_z$).

The magnetic moment $\mathbf{M}(t)$ in the laboratory frame is given by $\mathbf{M}_{\rm rot}(t)$ rotated back through an angle $-\omega t$ about $\hat{\mathbf{z}}$. Thus, the effect of the rf field is to superpose a slow longitudinal nutation (of frequency $\omega_{\rm rot}$) on the fast Larmor precession of $\mathbf{M}(t)$ about $\hat{\mathbf{z}}$ (of frequency ω_0). In consequence, the frequency spectrum of the magnetic moment has additional peaks at $\omega_0 \pm \omega_{\rm rot}$.

The Hamiltonian $H_{\rm rot}$ of (60) depends parametrically on the components of the rf and static magnetic fields, B_x, B_y and B_z, and vanishes for $B_x = B_y = 0$ and $B_z = \omega/\gamma$. If B_x, B_y and B_z are slowly modulated through a cycle C, with a period T much greater than the precession period $2\pi/\omega_{\rm rot}$ in the rotating frame, then the eigenstates $|\pm\rangle_{\rm rot}$ of $H_{\rm rot}$ evolve adiabatically, and acquire with each modulation cycle a geometric phase $\gamma_\pm(C)$ given by half the solid angle, α, described by C with respect to $(0, 0, \omega/\gamma)$ (cf Section 1.3). These geometric phases produce shifts $\mp\frac{1}{2}\hbar\alpha/T$ in the

energies $\pm\frac{1}{2}\hbar\omega_{\rm rot}$ of the unmodulated Hamiltonian $H_{\rm rot}$, and in turn shift the precession frequency of the magnetic moment from $\omega_{\rm rot}$ to $\omega_{\rm rot} - \alpha/T$. (This shift is the magnetic analogue of the daily precession of the Foucault pendulum discussed in Section 3.1.3.)

These frequency shifts were observed by Suter *et al* (1987) in the NMR response of proton spins in a water/acetone sample. The adiabatic cycle C was produced by placing the magnetization detector in a frame rotating with angular frequency $-\omega_{\rm rot}\hat{\mathbf{z}}$ about $\hat{\mathbf{z}}$. In the detector frame, the effective magnetic field turns about $\hat{\mathbf{z}}$ with angular frequency $\delta = \omega - \omega_{\rm rot}$; for small enough δ, the spins adiabatically track the turning field. A variation of this experiment was subsequently performed by Suter *et al* (1988), in which the geometric phase was observed as a rotation in the spin-echoed magnetic moment. This technique enabled observation of the individual geometric phases $\gamma_\pm$ rather than their difference $\gamma_- - \gamma_+$. The spins were also taken through nonadiabatic cycles, thereby incorporating the Aharonov-Anandan phase (cf Section 2.3).

In a nuclear quadrupole resonance experiment, the spin polarization is induced by an inhomogeneous electric field rather than a static magnetic field. Gradients in the electric field couple to spin through the nuclear electric quadrupole moment. The effective spin Hamiltonian is of the form $H_Q = \hbar\omega_Q(\mathbf{S}\cdot\hat{\mathbf{n}})^2$, where $\mathbf{S}$ denotes the dimensionless spin operators and $\hat{\mathbf{n}}$ is the symmetry axis of the electric field gradient tensor (assumed here to be axially symmetric). The eigenstates $|m;\hat{\mathbf{n}}\rangle$ are polarized along $\hat{\mathbf{n}}$ with energies $\hbar\omega_Q m^2$. Note that these are doubly degenerate for $m \neq 0$.

Geometric phase effects in the NQR spectrum of ^{35}Cl (for which $s = 3/2$) were studied by Tycko (1987). In the experiment, a coherent superposition of eigenstates of H_Q was created at $t = 0$ by an rf pulse; the induced magnetic moment $\mathbf{M}(t)$ was then observed and spectrally analyzed. The spectrum of H_Q consists of two degenerate doublets with energies $\frac{3}{4}\hbar\omega_Q$ and $\frac{1}{4}\hbar\omega_Q$ (corresponding to $m = \pm 3/2$ and $m = \pm 1/2$ respectively), so that the magnetic moment spectrum has a single peak at frequency $2\omega_Q$. Adiabatic cycling of H_Q was achieved by rotating a crystalline sample of $NaClO_3$ with angular frequency $\omega_R \ll \omega_Q$ about an axis inclined to the symmetry direction $\hat{\mathbf{n}}$. With each rotation, the spin eigenstates acquire geometric phases β_m (these may be computed explicitly), which shift their energies by $\hbar\beta_m\omega_R/2\pi$ and lift the double degeneracies. For the particular geometry realized in Tycko's experiment, theory predicts that the single peak in the magnetic moment spectrum is split into three peaks at frequencies $2\omega_Q$ and $2\omega_Q \pm \sqrt{3}\omega_R/\pi$. Precisely this splitting was observed in the experiment.

Since H_Q has degenerate energy levels, one might expect nonabelian geometric phases (*cf* Section 2.2) to appear. Tycko (1987) and Zee (1988) showed they do not; the experiment can be analyzed completely in terms of the ordinary (abelian) geometric phase. One can compute the nonabelian vector potentials $\mathbf{A}_{\pm 3/2}$ and $\mathbf{A}_{\pm 1/2}$ of (25) explicitly (there is one for each degenerate doublet of H_Q); they are matrix-valued, and are proportional to the matrix elements of the spin $\mathbf{S}$ taken between states in the degenerate doublets. For $m = \pm 3/2$, the off-diagonal matrix elements vanish identically ($\mathbf{S}$ has nonzero matrix elements only between states $|m\rangle$ and $|m\pm 1\rangle$), and the

diagonal matrix elements yield the ordinary geometric phases $\beta_{\pm 1/2}$. For $m = \pm 1/2$, the off diagonal elements $\langle \pm 1/2|\mathbf{S}| = 1/2\rangle$ do not vanish. However, for the adiabatic cycles used in the experiment, in which $\hat{\mathbf{n}}$ is turned about a fixed axis, a basis for the $m = \pm 1/2$ doublet can be found which diagonalizes $\mathbf{A}_{\pm 1/2}$. During the cycle there are no transitions between these basis vectors, and each acquires an ordinary geometric phase $\beta_{\pm 1/2}$.

Zee (1988) showed that more general adiabatic cycles of $\hat{\mathbf{n}}$ generate genuine nonabelian geometric phases in the NQR spectrum. Zwanziger *et al* (1990b) performed an experiment to observe these nonabelian effects. The more complicated adiabatic cycles were generated by mounting the $NaClO_4$ crystal on a double rotor. Careful analysis predicts that the NQR resonance should be split into five peaks instead of three. This is precisely what was observed, with excellent quantitative agreement between theory and experiment.

5.8. The quantum Hall effect

Charged particles in crossed uniform electric and magnetic fields experience a net drift along the direction $\mathbf{E} \times \mathbf{B}$ normal to both - this is the basis for the *Hall effect*. The Hall effect concerns specifically the behaviour of electrons in a planar conductor to which a perpendicular magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$ and a tangential electric field $\mathbf{E} = E\hat{\mathbf{x}}$ are applied. The resulting current density has a transverse component $j_y = \sigma_H E$. σ_H is the *Hall conductance*. Using classical theory, one obtains for large magnetic fields that

$$\sigma_H \approx \frac{nec}{B}, \tag{61}$$

where n is the two-dimensional electron density.

(Alternatively, one can consider conductors of finite transverse extent, eg long thin wires. Then the current density is necessarily longitudinal, whereas the electric field has a transverse component. In this case it is easier to compute the *Hall resistance* $\rho_H = E_y/j$. Classical theory gives $\rho_H = B/nec$ for all magnetic fields, not just large ones. The Hall conductance and resistance are the off-diagonal elements of the conductance and resistivity tensors σ and $\rho = \sigma^{-1}$, and the relation between them, $\sigma_H = \rho_H/(\rho_O^2 + \rho_H^2)$, depends on the longitudinal or Ohmic resistivity ρ_O.)

For large magnetic fields and at low temperatures, the classical prediction (61) fails: what is actually observed is a spectacular phenomenon, the quantum Hall effect. As B is increased, instead of falling off monotonically as $1/B$, the Hall conductance goes through a series of plateaus, where its value is constant, punctuated by intervals of monotonic decrease. The constant values assumed on the plateaus are sample-independent, and satisfy with extremely high accuracy (about one part in ten million) the quantization condition

$$\sigma_H = r\frac{e^2}{hc}. \tag{62}$$

As B increases, the coefficient r takes on integer values, and decreases with B in unit steps until it reaches one. This is the regime of the integer quantum Hall effect (von Klitzing *et al* 1980). As B is increased still further, one enters the regime of the fractional quantum Hall effect (Tsui *et al* 1982); r takes on rational values (less than one) with odd denominators.

The theory of the quantum Hall effect begins with the quantum mechanics of two-dimensional electrons in a magnetic field. The one-electron stationary states are the Landau levels $E_N = \hbar\omega_c(N + 1/2)$, $N = 0, 1, 2, \ldots$, where $\omega_c = eB/mc$ is the classical cyclotron frequency. The Landau levels are highly degenerate: each can accomodate B/Φ_0 electrons per unit area ($\Phi_0 = hc/e$ is the quantum of magnetic flux). In real materials, the crystalline potential, impurities, and electron-electron interactions all serve to break this degeneracy. However, for sufficiently large magnetic fields, there is negligible mixing between the Landau levels, so that each Landau level is split by the perturbations into B/Φ_0 sublevels per unit area.

If the electron density in the sample is n, then the ground state has the first $\nu = n\Phi_0/B$ Landau levels filled. (Thus, the larger the B-field, the fewer occupied Landau levels.) If ν has a fractional part, the uppermost level is only partially filled. The integer quantum Hall effect can now be explained by making two assumptions: i) only a small proportion of sublevels in each Landau level can carry current, and ii) each filled Landau level contributes precisely e^2/hc to the Hall conductance. Together, these assumptions imply that a nearly empty Landau level contributes nothing to the Hall conductance, as the occupied sublevels are unlikely to be the current-carrying ones, whereas, conversely, a nearly full Landau level contributes e^2/hc. From these qualitative considerations follows $\sigma_{xy} = [\nu]e^2/hc$ ($[\nu]$ denotes the integer nearest ν). This is precisely the relation (62), with $r = [\nu]$.

The first assumption can be justified on the basis of localization theory - most of the sublevels are localized, whereas only extended states can carry current. For the second assumption, Thouless, Kohmoto, Nightingale and den Nijs (TKN[2]) found, in 1982, a topological explanation (Thouless *et al* 1982). In retrospect, their analysis can be seen as an application of the theory of the geometric phase. Indeed, this connection was drawn as early as 1983 by Simon (1983).

In outline, the analysis of TKN[2] is as follows. We consider an electron in two dimensions in a uniform magnetic field $B\hat{\mathbf{z}}$ and periodic crystalline potential $V(x, y)$, with periods a and b in x and y. Because the vector potential $\mathbf{A} = (-By/2, Bx/2)$ is not periodic, neither is the Hamiltonian $mv^2/2 + V(x, y)$, in which $\mathbf{v} = (\mathbf{p} - e\mathbf{A}/c)/m$ is the velocity. However, the Hamiltonian is invariant under the magnetic translation operators $T_x = \exp(-imv_y a/\hbar)$ and $T_y = \exp(+imv_x b/\hbar)$, which shift x and y by a and b while leaving the velocity (rather than the momentum) unchanged.

While the magnetic translation operators commute with the Hamiltonian, they do not necessarily commute with each other. In general, $T_xT_y = \exp(2\pi iBab/\Phi_0)T_yT_x$. However, if we take the flux Bab through a unit cell to be a rational multiple p/q of the magnetic flux quantum Φ_0, then $(T_x)^q$ — a translation by q lattice spacings in x

- and T_y do commute, and Bloch's theorem applies. In this case, the energy levels form a band structure, whose bands $\epsilon_n(\mathbf{k})$ are periodic over the first Brillouin zone $[0, 2\pi/qa] \times [0, 2\pi/b]$. The eigenfunctions $\psi_{n,\mathbf{k}}(\mathbf{r})$ are Bloch waves $\exp(i\mathbf{k}.\mathbf{r})u_{n,\mathbf{k}}(\mathbf{r})$, where $u_{n,\mathbf{k}}(\mathbf{r})$ is an eigenfunction of the Hamiltonian

$$H(\mathbf{k}) = \frac{1}{2m}(-i\hbar\nabla + e\mathbf{A}/c + \hbar\mathbf{k})^2 + V(x, y), \tag{63}$$

with energy $\epsilon_n(\mathbf{k})$, which is invariant under the magnetic translations $(T_x)^q$ and T_y. For convenience we will sometimes denote the indices $n, \mathbf{k}$ by a single label α.

The Hall conductance can now be calculated perturbatively. An applied electric field $E\hat{\mathbf{x}}$ generates perturbed eigenstates $|u'_\alpha\rangle = |u_\alpha\rangle + |\delta u_\alpha\rangle$, which to lowest order are given by $\langle u_\beta|\delta u_\alpha\rangle = \langle u_\beta| - eEx\,u_\alpha\rangle/(\epsilon_\alpha - \epsilon_\beta)$, $\beta \neq \alpha$. The electric field induces a transverse current density $\langle u'_\alpha|j_y|u'_\alpha\rangle$, where $\mathbf{j} = e(\mathbf{v} + \hbar\mathbf{k}/m)/A_0$ and A_0 is the area of the sample. From (63), the current may also be written as $\mathbf{j} = e\nabla_{\mathbf{k}}H/A_0$. The Hall conductance σ_H is given by the ratio of the total induced current density $\sum_\alpha \langle u'_\alpha|j_y|u'_\alpha\rangle$ to the electric field strength E, where the sum is taken over the occupied levels, eg, for the ground state, the levels below the Fermi energy ϵ_F. Making use of the identity $[x, H] = i\nabla_{k_x}H$, we obtain the following explicit expression:

$$\sigma_H = \frac{e^2}{A_0\hbar c}\sum_{\epsilon_\alpha < \epsilon_F}\sum_{\beta \neq \alpha} \mathrm{Im}\,\frac{\langle u_\alpha|\nabla_{\mathbf{k}}H|u_\beta\rangle \times \langle u_\beta.\nabla_{\mathbf{k}}H|u_\alpha\rangle}{(\epsilon_\alpha - \epsilon_\beta)^2}. \tag{64}$$

(64) is the Kubo formula for the Hall conductance of the ground state.

Suppose we regard $H(\mathbf{k})$ as a family of Hamiltonians parameterized by the Bloch vector $\mathbf{k}$. Under parallel transport round a closed cycle C, the eigenstate $|u_{n,\mathbf{k}}\rangle$ acquires a geometric phase $\gamma_{n,\mathbf{k}}(C)$, which may be expressed (cf Section 1.2) as the integral of the flux field $V_n(\mathbf{k}) = \mathrm{Im}\,\langle\nabla_{\mathbf{k}}u_{n,\mathbf{k}}| \times |\nabla_{\mathbf{k}}u_{n,\mathbf{k}}\rangle$ over the interior of C. Amongst the various expressions for $V_n(\mathbf{k})$ given in Section 1.2, we note that (10), which involves matrix elements of the gradient of the Hamiltonian with respect to the parameters $\mathbf{k}$, is contained precisely in the formula (64) for the Hall conductance. This allows us to rewrite (64) as $\sigma_H = e^2/(A_0\hbar c)\sum_n \sum_{\mathbf{k}} V_n(\mathbf{k})$. For samples large compared with the unit cell, the sum $\sum_{\mathbf{k}}$ over Bloch vectors may be replaced by an integral $A_0/4\pi^2 \int d^2k$ over the first Brillouin zone T_k. Thus we obtain

$$\sigma_H = \frac{e^2}{hc}\sum_n \sigma_n, \tag{65}$$

where

$$\sigma_n = \frac{1}{2\pi}\int_{T_k} V_n(\mathbf{k})\,d^2k \tag{66}$$

is the Hall conductance (in units of e^2/hc) of the nth band.

The central result of TKN[2] is that σ_n is an integer. The argument is as follows. Since the energy levels $\epsilon_n(\mathbf{k})$ are periodic over the Brioullin zone, the Bloch states

$|u_{n,\mathbf{k}}\rangle$ are periodic up to an overall phase factor. In particular, the Bloch states on the corners of the first Brioullin zone, ie $\mathbf{k} = 0$, $\mathbf{A}$, $\mathbf{A}+\mathbf{B}$, and $\mathbf{B}$ (here $\mathbf{A} = (2\pi/qa, 0)$ and $\mathbf{B} = (0, 2\pi/b)$ denote the reciprocal lattice vectors), differ only by a phase. Thus $|u_{n,\mathbf{A}}\rangle = e^{i\alpha}|u_{n,0}\rangle$, $|u_{n,\mathbf{A}+\mathbf{B}}\rangle = e^{i\beta}|u_{n,\mathbf{A}}\rangle$, $|u_{n,\mathbf{B}}\rangle = e^{i\gamma}|u_{n,\mathbf{A}+\mathbf{B}}\rangle$, and $|u_{n,0}\rangle = e^{i\delta}|u_{n,\mathbf{B}}\rangle$. Together these relations imply that $|u_{n,0}\rangle = e^{i(\alpha+\beta+\gamma+\delta)}|u_{n,0}\rangle$, and it follows that $\alpha - \beta + \gamma + \delta$ is necessarily an integer multiple m of 2π. Applying Stokes' theorem to (66), we obtain $\sigma_n = 1/(2\pi)\oint \mathrm{Im}\langle u_{n,\mathbf{k}}|\nabla_{\mathbf{k}} u_{n,\mathbf{k}}\rangle \cdot d\mathbf{k}$, where the integral is taken round the boundary of the first Brioullin zone. A little calculation shows that this integral gives precisely the accumulated phase change $(\alpha+\beta-\gamma-\delta)/2\pi$ round the first Brioullin zone T_*, ie the integer m. In the language of fibre bundles (cf Section 2.1.4), the Bloch states $|u_{n,\mathbf{k}}\rangle$ constitute a complex line bundle over T_* with curvature $V_n(\mathbf{k})$, and σ_n is the first Chern number of this line bundle.

For large magnetic fields, the bands $\epsilon_n(\mathbf{k})$ coalesce into clusters of p nearly degenerate sublevels, each cluster associated with one of the Landau levels of the pure magnetic Hamiltonian. It can be shown (the argument is omitted here) that sum of the conductances σ_n over states in a given Landau level is equal to one. This completes the topological explanation for the quantization of Hall conductance. TKN[4] also calculate the individual σ_n in the limit of a weak periodic potential: these are found to satisfy certain Diophantine equations in p and q.

TKN[4] was the first of a number of important applications of geometric phase methods to the study of the quantum Hall effect. Several subsequent works address the fractional effect, where electron-electron interactions come into play. Niu, Thouless and Wu (1984) show how the results of TKN[4] can be generalized to accomodate them. Arovas *et al* (1984) calculate properties of the elementary excitations of Laughlin's many body fractional-quantum-Hall ground state (Laughlin 1983) using geometric phase arguments. Concerning the integer case, Arovas et al (1988) establish a topological criterion for localization in terms of Chern numbers. Wilkinson (1984) incorporates the geometric phase into semiclassical calculations of the single-particle energy spectrum.

Finally, we mention studies by Avron and coworkers (Avron *et al* 1988, Avron 1995) of electron transport in quantum systems. While not specifically concerned with the quantum Hall effect, they elucidate essentially similar topological mechanisms. One class of models consists of planar networks of one-dimensional wires whose closed loops C_i are threaded by magnetic fluxes ϕ_i normal to the plane. The physical properties of the network are Φ_0-periodic in the fluxes (fluxes differing by an integral number of flux quanta are related by a gauge transformation). Fixing all but two of the fluxes, the stationary states $\psi_n(\phi_i, \phi_j)$ constitute a complex line bundle over the two-torus of fluxes (the analogue of the first Brillouin zone). The associated Chern number, the integral of the curvature, has the following physical interpretation: it represents the average with respect to ϕ_i of the number of electrons transported round C_j as ϕ_j is increased adiabatically by one flux quantum.

Glossary

dynamical phase The time integral of the energy of an evolving quantum state divided by Planck's constant h.

geometric phase A phase factor accumulated by a cycled quantum eigenstate which depends only on the geometry of its evolution.

geometric magnetism Effective Lorentz-like force which appears in the dynamics of a slow system coupled to a fast one after the fast degrees of freedom have been averaged away.

Hannay angle Displacement in the angle variables of adiabatically cycled integrable systems of purely geometric origin.

holonomy Change in phase induced by parallel transport of a quantum state around a closed curve.

integrable system A Hamiltonian system with as many constants of the motion in involution as there are degrees of freedom.

parallel transport A prescription to determine the overall phase of an evolving quantum state.

reduction General theory of Hamiltonian dynamics in the presence of symmetry.

References

Aharonov Y and Anandan J 1987 *Phys. Rev. Lett* **58** 145 148

Aharonov Y and Bohm D 1959 *Phys. Rev.* **115** 484–491

Aitchison IJR 1987 *Acta. Phys. Polonica* **B18** 207 235

Alber MS and Marsden JE 1992 *Comm. Maths Phys* **149** 217–240

Arovas DP, Schrieffer JR and Wilczek F 1984 *Phys. Rev. Lett.* **53** 722-723

Arovas DP, Bhatt RN, Haldane FDM, Littlewood PB and Rammal R 1988 *Phys. Rev. Lett.* **60** 619 622

Avron JE, Raveh A, Zur B 1988 *Rev. Mod. Phys.* **60** 873 915

Avron JE 1995 *Mesoscopic Quantum Physics, Les Houches, Session LXI* E. Akkermans et al, eds. Elsevier Science B.V . p 741–791

Berry MV 1984 *Proc. Roy. Soc. Lond A* **392**, 45 57

Berry MV 1985 *J. Phys. A* **18**, 15–27

Berry MV 1987a *Nature, Lond.* **326** 277 278

Berry MV 1987b *J. Mod. Optics* **34** 1401–1407

Berry MV 1987c *Proc. R. Soc.A* **414** 31 46

Berry MV 1989 *Geometric Phases in Physics* A Shapere and F Wilczek, eds . Singapore: World Scientific. p. 7–28

Berry MV 1990a *Physics Today* December, 34–40 Berry MV 1990b *Proc. R. Soc A* **430** 405–411

Berry MV and Robbins JM 1993a *Proc Roy Soc. Lond A* **442**, 641 658

Berry MV and Robbins JM 1993b *Proc. Roy. Soc Lond A* **442**, 659–672

Berry MV and Klein S 1996 *Journal of Modern Optics* **43** 165-180

Berry MV and Morgan MA *Nonlinearity* **9** 787-799

Bhandari R 1988 *Phys. Lett.A* **133** 1 3

Bhandari R "Polarization of light and topological phases", submitted to Physics Reports 1990

Bhandari R and Samuel J 1988 *Phys. Rev. Lett.* **60** 1211–1214

Bitter T and Dubbers D 1987 *Phys Rev. Lett.* **59** 251 254

Chambers RG 1960 *Phys Rev. Lett.* **5** 3 5

Cina JA *Chem. Phys. Lett.* **132** 393-395

Delacrétaz G, Grant ER, Whetten RL, Wöste L and Zwanziger JW 1986 *Phys. Rev. Lett.* **56** 161–164
Dingle RB 1973 *Asymptotic expansions: their derivation and interpretation* New York and London: Academic Press
Dirac PAM 1931 *Proc. R. Soc. A* **133** 60
Dykhne AM 1962 *Sov. Phys. JETP* **14** 941–943
Eco U 1989 *Foucault's pendulum* New York: Harcourt Brace Jovanovich
Eguchi T, Gilkey PB, and Hanson AJ 1980 *Phys. Rep.* **66**, 213–393
Ghim S, Knauf A and Marzuoli S 1989 *Comm. Math. Phys.* **110** 33–49
Hannay JH 1985 *J. Phys. A* **18**, 221–230
Jackiw R 1988 *Int. J. Mod. Phys.* **A3** 285–297
Jahn, HA and Teller, E 1937 *Proc. R. Soc. A* **161** 220–235
Jarzynski C 1993 *Phys. Rev. Lett.* **71** 839–842
Jarzynski C 1995 *Phys. Rev. Lett.* **74** 1732–1735
Kuppermann A and Wu YSM 1993 *Chem. Phys. Lett.* **205** 577–586
Kuratsuji H and Iida S 1985 *Prog. Theo. Phys.* **74** 439–445
Kwiat PG and Chiao RY 1991 *Phys. Rev. Lett.* **66** 588–91
Littlejohn RG 1988 *Phys. Rev. A* **38** 6034–6045
Littlejohn RG and Flynn WG 1992 *Phys. Rev. A* **44** 5239–5256
Laughlin RB 1983 *Phys. Rev. Lett.* **50** 1395–1398
Li Yin and Mead CA 1992 *Theor. Chim. Acta* **82** 397–406
Lochak P and Meunier C 1988 *Multiphase averaging for classical systems.* New York: Springer
Longuet-Higgins HC, Öpik U, Pryce MHL and Sack RA 1958 *PRS S* **244** 1–
Marsden JE, O'Reilly OM, Wicklin FJ and Zombro BW 1991 *Nonlinear Science Today* **1** 4–11, 14–21
Marsden JE 1992 *Lectures on Mechanics* Cambridge University Press, Cambridge
Mead CA 1980 *Chem. Phys.* **49** 23–40
Mead CA 1992 *Rev. Mod. Phys.* **64** 51–85
Mead CA and Truhlar DG 1979 *J. Chem. Phys.* **70(05)** 2284–2296
Montgomery R 1991 *Am. J. Phys* **59** 394–398
Moody J, Shapere A and Wilczek F 1986 *Phys. Rev. Lett.* **56** 893–896
Nash C and Sen S 1983 *Topology and geometry for physicists* New York: Academic
Niu Q, Thouless DJ, and Wu YS 1985 *Phys. Rev. B* **31** 3372–3377
Pancharatnam S 1956 *Proc. Ind. Acad. Sci. A* **44** 247
Provost J and Vallee G 1980 *Comm. Math. Phys.* **76**, 289–301
Robbins JM and Berry MV 1992 *Proc. Roy. Soc. Lond. A* **436** 631–661
Robbins JM 1994 *J. Phys. A: Math. Gen.* **27** 1179–1189
Rytov SM 1938 *Dokl. Akad. Nauk. USSR* **18**, 263; reprinted in B Markovsky and VI Vinitsky, *Topological Phases in Quantum Theory*, World Scientific, Singapore (1989), p. 6
Samuel J and Bhandari R 1988 *Phys. Rev. Lett.* **60** 2339 - 2342
Schroer CG 1996 *J. Phys. A: Math. Gen.* **29** 3289–3297
Shapere A and Wilczek F (eds) 1989a *Geometric phases in physics* Singapore: World Scientific
Shapere A and Wilczek F 1989b *Am. J. Phys.* **57** 514–518
Shapere A and Wilczek F 1989c *J. Fluid Mech.* **198** 557–585
Shapere A and Wilczek F 1989d *J. Fluid Mech.* **198** 587–599
Simon B 1983 *Phys. Rev. Lett* **51** 2167–2170
Suter D, Chingas GC, Harris RA and Pines A 1987 *Mol. Phys.* **61** 1327–1340
Suter D, Mueller KT and Pines A 1987 *Phys. Rev. Lett.* **60** 221–223
Thouless DJ, Kohmoto M, Nightingale MP and den Nijs M 1982 *Phys. Rev. Lett.* **49** 405–408
Tinkham M 1964 *Group theory and quantum mechanics* New York: McGraw Hill
Tsui DC, Stormer HL, and Gossard AC 1982 *Phys. Rev. Lett.* **48** 1559
Tomita A and Chiao R 1986 *Phys. Rev. Lett.* **57** 937–940
Tycko R 1987 *Phys. Rev. Lett* **58** 2281–2284
Vladimirskii VV 1941 *Dokl. Akad. Nauk. USSR* **21**, 222; reprinted in B Markovsky and VI Vinitsky *Topological Phases in Quantum Theory* World Scientific, Singapore (1989), p. 11
von Klitzing K, Dorda G and Pepper M 1980 *Phys. Rev. Lett.* **45** 494
von Neumann J and Wigner EP 1929 *Phys. Z* **30** 467–470
Weigert S and Littlejohn RG 1993 *Phys. Rev. A* **47** 3506–3512

Wilczek F and Zee A 1984 *Phys. Rev. Lett* **52** 141–144
Wilkinson M 1984 *J. Phys. A: Math. Gen.* **17** 3459–3476
Wilkinson. M 1990 *J. Phys. A. Math. Gen.* **23** 3603-3611
Zee A 1984 *Phys. Rev A* **38** 1–6
Zener C 1932 *Proc. R. Soc.A* **137** 696–702
Zwanziger JW, Koenig M, and Pines A 1990a *Annu. Rev. Phys. Chem.* **41** 601–646
Zwanziger JW, Koenig M, and Pines A 1990b *Phys. Rev.A* **42** 3107–3110
Zwanziger JW, Rucker SP and Chingas GC 1991 *Phys. Rev.A* **43** 3232-3240
Zygelman B 1987 *Phys. Lett. A* **125** 476–481

Further Reading

Aitchison IJR 1987 *Acta Phys. Polonica* **B18** 207–235
Avron JE 1995 *Mesoscopic Quantum Physics, Les Houches, Session LXI* E. Akkermans et al. eds Elsevier Science B.V., p. 741–791
Berry MV 1989 *Geometric Phases in Physics* A Shapere and F Wilczek, eds., Singapore: World Scientific, p 7–28
Berry MV 1989 *Scientific American* 256 no 12, 34–40
Berry MV 1990 *Physics Today* **D e**
cember, 34–40
Berry MV 1990 *Sir Charles Frank 80th Birthday Festschrift* RG Chambers et al. eds Adam Hilger, Bristol
Berry MV 1990 *Anomalies phases, defects* M Bregual, G Marmo and G Morandi, eds., Naples. Bibliopolis, pp 123-181
Bhandari, R "Polarization of light and topological phases", submitted to Physics Reports, 1996
Bohm A 1996 *Quantum Mechanics: Foundations and Applications, 3rd ed* Springer Verlag, New York
Jackiw R 1988 *Int. J. Mod. Phys.* **A3**, 285–297
Markovski B and Vinitsky SI (eds) 1989 *Topological Phases in Quantum Theory* World Scientific: Singapore
Marsden JE, O'Reilly OM, Wicklin FJ, and Zombro BW 1991 *Nonlinear Science Today* 1 4–11, 14–21
Marsden JE 1992 *Lectures on Mechanics* Cambridge University Press, Cambridge
Mead CA 1992 *Rev. Mod. Phys.* **64** 51–85
Shapere A and Wilczek F (eds) 1989 *Geometric phases in physics* Singapore: World Scientific
Shapere A and Wilczek F 1989 *Am. J. Phys.* **57** 514–518
Vinitskii SI, Derbov VN, Dubovik VN, Markovski BL and Stepanovskii YP *Sov. Phys. Usp.* **33** (6) 403–429
Zwanziger JW, Koenig M, and Pines A 1990 *Annu. Rev. Phys. Chem.* **41** 601–646

Figure captions

Figure 1. A path P on the base manifold and its lift $\tilde{P}$ in the fibre bundle

Figure 2. A closed path C on the base manifold and its lift $\tilde{C}$ in the fibre bundle. The difference between the endpoints of $\tilde{C}$, a displacement within a single fibre, is the holonomy of C.

Fig. 1

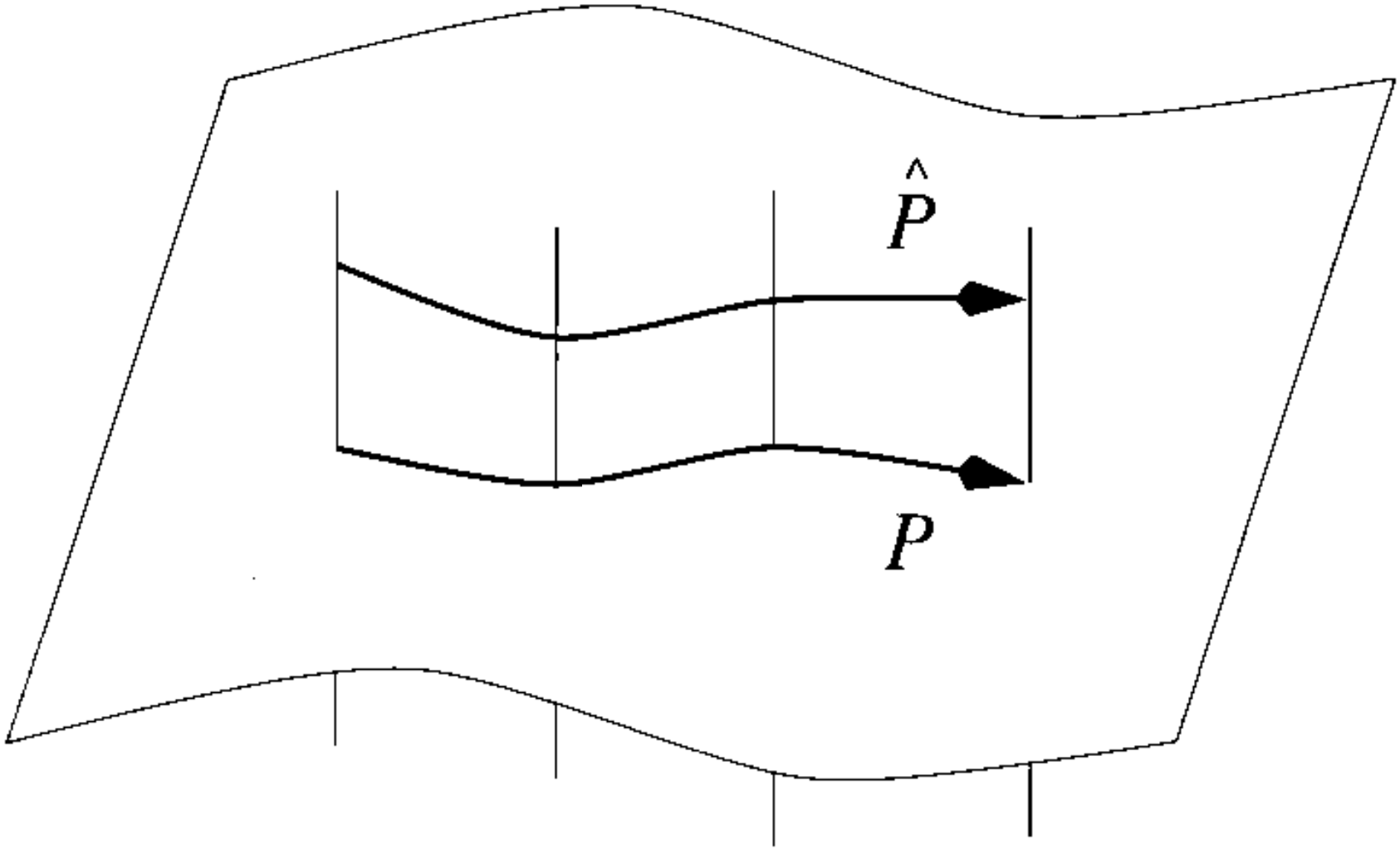

Fig 2