INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR: 718/5

COURSE ON GEOMETRIC PHASES

(6 - 17 September 1993)

"Some Geometric Phases"

presented by:

M. BERRY
University of Bristol
H.H. Wills Physics Laboratory
Royal Fort, Tyndall Avenue
BS8 1TL Bristol
United Kingdom

MAIN BUILDING STRADA COSTIERA, 11 TEL. 22401 TELEFAX 224163 TELEX 460392 ADRIATICO GUEST HOUSE VIA GRIGNANO, 9 TEL. 224241 TELEFAX 224531 TELEX 460449

Some geometric phases

transparencies for three lectures at CERN

February 1993

Michael Berry

H H Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK

Some references

Reviews

Shapere, A. and Wilczek, F (eds.) 1989 *Geometric phases in physics* (Singapore: World Scientific)

Markovski, B, and Vinitsky, S.I., 1989 *Topological phases in quantum theory* Singapore: World Scientific)

Vinitskii, S.I., Derbov, V.N., Dubovik, V.N., Markovski, B.L and Stepanovskii, Yu P. 1990 *Sov. Phys. Uspekhi* **33** 403-429 'Topological phases in quantum mechanics and polarization optics'

Zwanziger, J.W., Koenig, M. and Pines, A. 1990 *Annu. Rev. Phys. Chem.* **41** 601-46 'Berry's Phase'

Mead, C.A. 1992 *Revs. Mod. Phys.* 64 51-85 'The geometric phase in molecular systems'

Phase-related works by Michael Berry

Nye, J.F., & Berry, M.V., 1974, *Proc.Roy.Soc. Lond.*, **A336**, 165-90, 'Dislocations in wave trains'.

Berry, M.V., Chambers,R.G., Large,M.D., Upstill,C., & Walmsley, J.C., 1980, *Eur.J.Phys.* **1**, 154-162, 'Wavefront dislocations in the Aharonov-Bohm effect and its water-wave analogue'.

Berry, M.V., 1981, 'Singularities in Waves' in *Les Houches Lecture Series Session XXXV* , eds. R.Balian, M.Kléman and J-P. Poirier, North-Holland: Amsterdam, 453-543.

Berry, M.V., & Wilkinson, M., 1984, *Proc.Roy.Soc. Lond.*, **A392**, 15-43, 'Diabolical points in the spectra of triangles'.

Berry, M.V., 1984, *Proc.Roy.Soc. Lond.*, **A392**, 45-57, 'Quantal phase factors accompanying adiabatic changes'.

Berry,M.V., 1985, *J.Phys.A.* **18**, 15-27, 'Classical adiabatic angles and quantal adiabatic phase'.

Berry,M.V., 1985, 'Quantum, classical and semiclassical adiabaticity in *Theoretical and Applied Mechanics*, eds. F.I.Niordson and N.Olhoff, Elsevier, North-Holland, Amsterdam, pp 83-96.

Berry,M.V., & Mondragon,R.J. 1986, *J.Phys.A.* **19**, 873-885, 'Diabolical points in one-dimensional Hamiltonians quartic in the momentum'.

Berry, M.V., & Robnik,M., 1986, *J.Phys.A.* **19**, 1365-1372, 'Quantum states without time-reversal symmetry: wavefront dislocations in a nonintegrable Aharonov-Bohm billiard'.

Berry, M.V., 1986, 'Adiabatic phase shifts for neutrons and photons', in *Fundamental aspects of quantum theory*, eds. V.Gorini and A.Frigerio, Plenum, NATO ASI series vol.144, 267-278.

Berry, M.V., 1986, 'The Aharonov-Bohm effect is real physics not ideal physics' in *Fundamental aspects of quantum theory*, eds. V.Gorini and A.Frigerio, Plenum, NATO ASI series vol. 144, 319-320.

Berry, M.V., 1987, *Nature*, **326**, 277-278, 'Interpreting the anholonomy of coiled light'.

Berry, M.V., 1987, *Proc.Roy.Soc.* **A414**, 31-46, 'Quantum phase corrections from adiabatic iteration'.

Berry, M.V., 1987, *J.Modern Optics*, **34**, 1401-1407, 'The adiabatic phase and Panacharatnam's phase for polarized light'.

Berry, M.V., & Hannay, J., 1988, *J.Phys.A.* **21**, L325-331, 'Classical nonadiabatic angles'.

Berry, M.V., 1988, 'The geometric phase', *Scientific American*, 256 (12), 46-52.

Berry, M.V., 1989 'The quantum phase, five years after' in 'Geometric Phases in Physics', eds: A.Shapere, F.Wilczek. (World Scientific) 7-28.

Mondragon,R.J and Berry,M.V. 1989, 'The quantum phase 2-form near degeneracies: two numerical studies' *Proc.Roy.Soc.* **A424** 263-278.

Berry, M.V. 'Quantum adiabatic anholonomy' in *Anomalies, phases, defects'* eds U M Bregola, G.Marmo and G.Morandi (Naples: Bibliopolis) 125-181.

Berry, M.V. and Lim, R. 1990 *J. Phys. A.*, **23**, L655-L657, 'The Born-Oppenheimer electric gauge force is repulsive near degeneracies'.

Berry, M.V. 1990 *Proc.Roy.Soc.Lond.* **A430** 405-411, 'Geometric amplitude factors in adiabatic quantum transitions'.

Berry, M.V. 1990 *Proc. Roy. Soc. Lond*, **A431**, 531-537 'Budden and Smith's 'additional memory' and the geometric phase'.

Berry, M.V. 1990 *Physics Today*, **43** (no.12) 34-40, 'Anticipations of the geometric phase'.

Berry, M.V. 1991 In 'Sir Charles Frank OBE FRS, an eightieth birthday tribute' Eds. R.G. Chambers, J.E. Enderby, A. Keller, A.R. Lang and J.W. Steeds (Adam Hilger, Bristol), 'Bristol Anholonomy Calendar', pp.207-219.

Berry, M.V. 1991 in 'Quantum Coherence', ed. Jeeva S. Anandan (World Scientific) 'Wave Geometry: A plurality of singularities' pp.92-98.

Robbins, J. M. and Berry, M. V. 1992, *Proc. Roy. Soc. Lond.* **A436**, 631-661, 'The geometric phase for chaotic systems'

M. V. Berry and J. M. Robbins 'Classical geometric forces of reaction: an exactly solvable model', submitted to *Proc. Roy. Soc.* December 1992

M. V. Berry and J. M. Robbins 'Chaotic classical and half-classical adiabatic reations: geometric magnetism and deterministic friction', submitted to *Proc. Roy. Soc.* January 1993

Lecture 1

(Some geometric phases, Michael Berry Geneva 1993)

SOME GEOMETRIC PHASES

Three central ideas

1) Adiabaticity

Mechanical phenomena on the border between statics and dynamics.

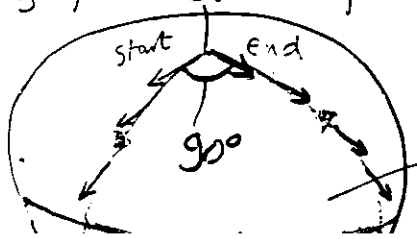
statics: study of things, i.e. states that do not change; in quantumspeak: eigenstates of Hamiltonian operators

dynamics: study of happenings: evolving states, changing external fields

Border (the adiabatic regime): study of slow changes

2. Anholonomy: global change without local change

In geometry: when some quantities are varied round a cycle, others, dependent on them*, fail to return. e.g. parallel transport of vectors on a sphere



solid angle $\Omega = \frac{1}{8}$ sphere $= \frac{1}{8} 4\pi$

*and have no local rate of change

Anholonomy in optics

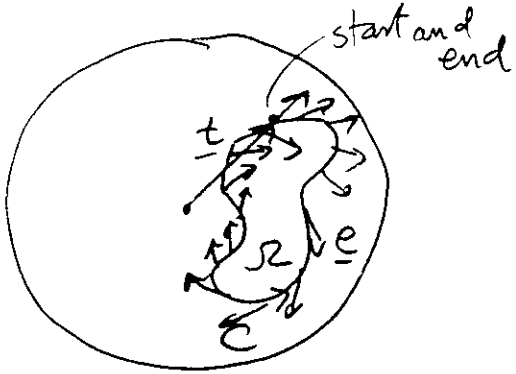
Vladimirsky 1941

Light rays curved in space:

Curving of $\underline{t}$ governed by Snell's law (1611)

What about polarisation $\underline{e}$?

This is parallel-transported: no rotation of $\underline{e}$ about $\underline{t}$. So $\underline{e}$ need not return with $\underline{t}$: it rotates by the solid angle Ω that the $\underline{t}$ circuit subtends on its sphere:

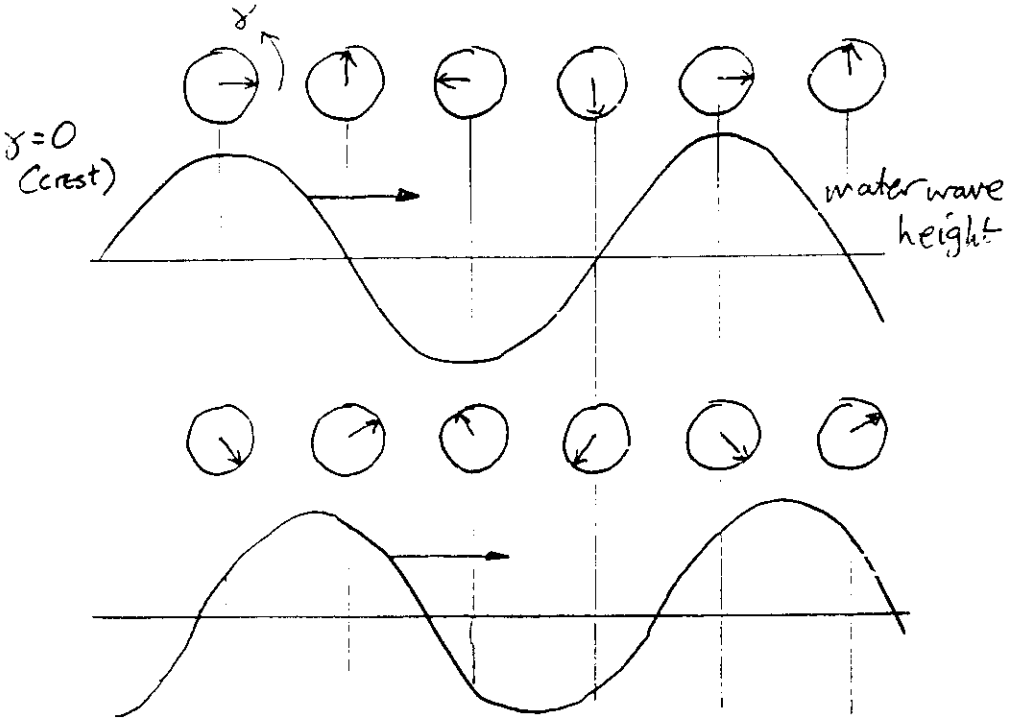


3. Phase

Oxford English Dictionary:

"A particular stage or point in a recurring sequence of movements or changes, e.g. a vibration or undulation, considered in relation to a reference position or time" (1861)

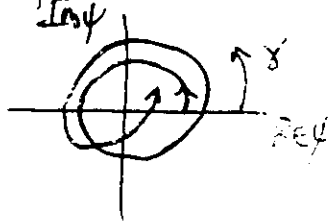
Naturally measured by an angle, γ.



Representation in terms of complex functions of time:

$$\psi(t) = \rho(t)\cos\gamma(t) + i\rho(t)\sin\gamma(t)$$

$$= \rho(t)\, e^{i\gamma(t)}$$



Solid-angle formula for parallel transport

Triad $\underline{r}$, $\underline{e}_1$, $\underline{e}_2$ $\underline{r}(t)$ radial, driving $\underline{e}_1$, $\underline{e}_2$

Let $\underline{e}$ be any vector $\perp \underline{r}$ (e.g. $\underline{e}_1$ or $\underline{e}_2$). Then

$\dot{\underline{e}} = \underline{\Omega} \wedge \underline{e}$ $\quad \underline{\Omega}$ = angular velocity

Most general

$$\underline{\Omega} = a\,\underline{r} + b\,\dot{\underline{r}} + c\,\underline{r}\wedge\dot{\underline{r}}$$

Parallel transport (no twist about $\underline{r}$) $\rightarrow a = 0$

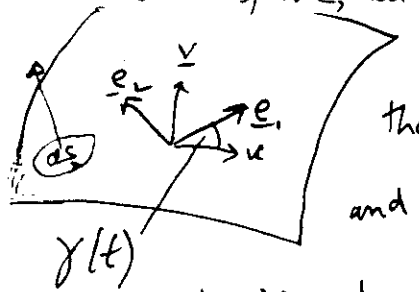
Then $\underline{e}\cdot\underline{r} = 0 \longrightarrow b = 0$, $c = 1$

$$\therefore \underline{\Omega} = \underline{r}\wedge\dot{\underline{r}} \quad : \quad \boxed{\dot{\underline{e}} = (\underline{r}\wedge\dot{\underline{r}})\wedge\underline{e} = -\,\underline{e}\cdot\dot{\underline{r}}\;\underline{r}}$$

Useful expression for equation of motion:

let $\underline{\phi} \equiv \underline{e}_1 + i\underline{e}_2$. Then $\boxed{\underline{\phi}^*\cdot\dot{\underline{\phi}} = 0}$

Chart passage of $\underline{e}_1, \underline{e}_2$ relative to given orthogonal coordinates on sphere, with unit directions $\underline{u}$, $\underline{v}$, angle γ.



let $\underline{n} \equiv \underline{u} + i\underline{v}$

then $\underline{\phi} = \underline{n}\, e^{-i\gamma}$

and $\underline{\phi}^*\cdot\dot{\underline{\phi}} = 0 \longrightarrow \dot{\gamma} = \mathrm{Im}\;\underline{n}^*\cdot\dot{\underline{n}}$

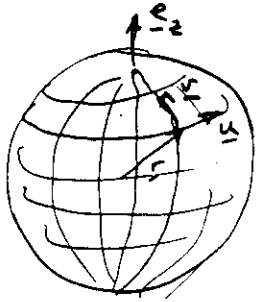
$$\therefore \gamma = \mathrm{Im}\oint \underline{n}^*\cdot\underline{n}\,dt = \mathrm{Im}\iint_{\partial S = C} \nabla\underline{n}^*\wedge\nabla\underline{n}\cdot d\underline{S}$$

$d\underline{S}$ on sphere

· scalar product

$$\boxed{\gamma(C) \equiv \mathrm{Im}\iint_{\partial S = C} d\underline{n}^*\wedge d\underline{n}}$$

For $\underline{u}$, $\underline{v}$ choose e.g. polar coordinate directions



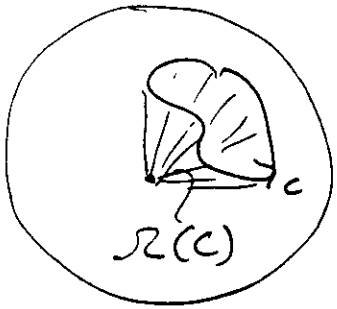
$$\underline{u} = \frac{\underline{r} \wedge \underline{e}_z}{|\underline{r} \wedge \underline{e}_z|} \quad , \quad \underline{v} = \frac{\underline{r} \wedge \underline{u}}{r}$$

Then $\boxed{\mathrm{Im}\, \nabla \underline{n}^* \wedge \cdot \nabla \underline{n} = \frac{\underline{r}}{r^3}}$

= monopole field

$$\therefore \boxed{\gamma(C) = \iint_{\partial S = C} \frac{\underline{r} \cdot d\underline{S}}{r^3} = \Omega(C)}$$

Solid angle subtended by C at $\underline{r} = 0$

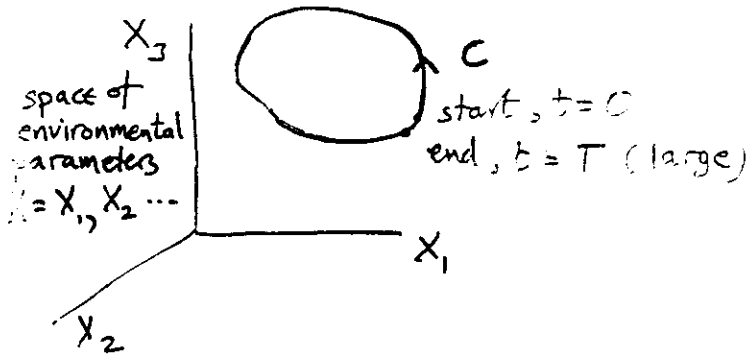


2-form notation: if X_1, X_2 are coordinates on sphere

$$\boxed{d\underline{n}^* \cdot \wedge d\underline{n} \equiv \left(\frac{\partial \underline{n}^*}{\partial X_1} \cdot \frac{\partial \underline{n}}{\partial X_2} - \frac{\partial \underline{n}^*}{\partial X_2} \cdot \frac{\partial \underline{n}}{\partial X_1} \right) dX_1 dX_2}$$

The geometric phase

Take any quantum system (atom, electron, molecul in one of its discrete energy states (labelled n) and slowly alter its environment (e.g. electric or magnetic forces) round a cycle.



Afterwards, environment has returned, and system is still in state n (quantum adiabatic theorem)
But its phase has changed, by $\gamma = \gamma(T) - \gamma(0)$.

First guess at γ:

Imagine the environment doesn't change. Then

$$\gamma = -\omega(X) T = -\frac{E_n(X)}{\hbar} T \qquad \text{(Planck)}$$

energy of state n

When X changes, generalize to

$$\boxed{\gamma = (?) = -\frac{1}{\hbar} \int_0^T dt\, E_n(X(t))}$$

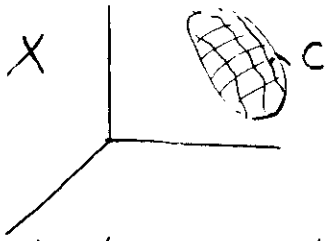
But this is wrong. The true phase is

$$\gamma = -\frac{1}{\hbar}\int_0^T dt\, E_n(X(t)) + \gamma_n(C)$$

"dynamical phase" "geometric phase"

dynamical phase: increases with T; system's answer to:
"How long did your trip take?"

geometric phase: independent of T, dependent only on n and the geometry of C; system's answer to:
"where have you been?"



$\gamma_n(C)$ is phase anholonomy

As with other circuit-dependent quantities in physics,
$\gamma_n(C)$ = flux of something through C
(e.g. emf = flux of rate of change of magnetic field)

Here (from Schrödinger equation)

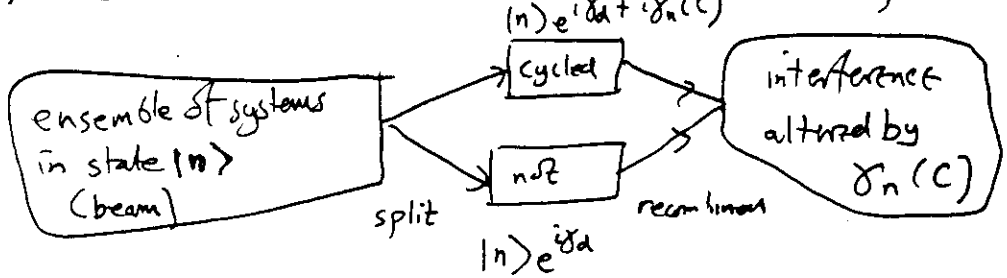
$$\gamma_n(C) = -\iint_{\partial S = C} \mathrm{Im}\langle dn|n\rangle|dn\rangle \quad \text{abstractly (2 form)}$$

$$= -\iint_{\partial S=C} dX_1 dX_2 \iiint d\underline{r}\, \mathrm{Im}\left[\frac{\partial \psi_n^*(\underline{r};X)}{\partial X_1}\frac{\partial \psi_n(\underline{r};X)}{\partial X_2} - \frac{\partial \psi_n^*(\underline{r};X)}{\partial X_2}\frac{\partial \psi_n(\underline{r};X)}{\partial X_1}\right]$$

coordinates spanning C — concretely

3 types of phase experiment

1) Two Hamiltonians, one state (interferometry)



2) One Hamiltonian, two states (superposition)

$$|\psi(0)\rangle = a_m|m\rangle + a_n|n\rangle$$

$$|\psi(T)\rangle = a_m|m\rangle e^{i\gamma_{dm} + i\gamma_m(C)} + a_n|n\rangle e^{i\gamma_{dn} + i\gamma_n(C)}$$

measure any operator $\hat{A}$ not commuting with $\hat{H}$:

$$\langle\psi(T)|\hat{A}|\psi(T)\rangle = |a_n|^2\langle n|\hat{A}|n\rangle + |a_m|^2\langle m|\hat{A}|m\rangle + 2\,\mathrm{Re}\, a_n^* a_m \langle n|\hat{A}|m\rangle e^{i\{\gamma_{dn} - \gamma_{dm} + \gamma_n(C) - \gamma_m(C)\}}$$

depends on $\gamma_n(C)$

3) Repeated cycling

$$|\psi(nT)\rangle = e^{-\frac{in}{\hbar}\int_0^T E_n(t)dt}\, e^{in\gamma_n(C)}|\psi(0)\rangle$$

$$= e^{-i\left(\bar{\omega} - \frac{\gamma_n(C)}{T}\right)nT}|\psi(0)\rangle \qquad \left(\bar{\omega} = \frac{1}{\hbar T}\int_0^T dt\, E_n(t)\right)$$

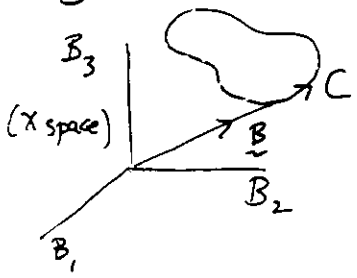
$$|\psi(nT)\rangle \equiv e^{-i\omega' nT}|\psi(0)\rangle \rightarrow \text{frequency shift}$$

Example TURNING SPINS

System: quantum spins

environment: whatever can change spin, e.g. a magnetic field $\underline{B}$ for particles with magnetic moment.

cycle: slow circuit in $\underline{B}$ space



Quantum states labelled by component $n\hbar$ of angular momentum along $\underline{B}$.

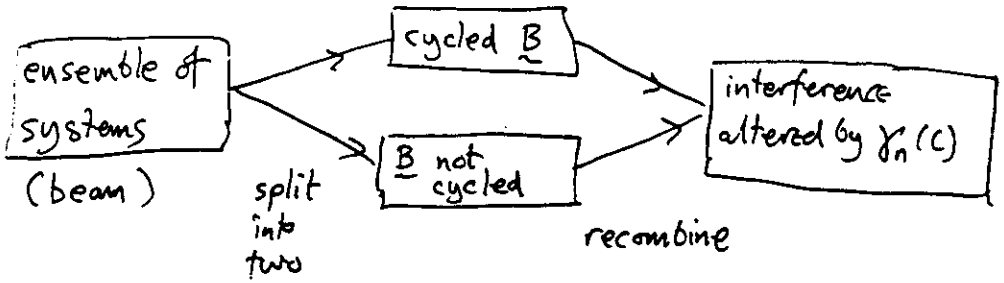
(n = integer for bosons (α particles, photons)
= half-integer for Fermions (electrons, neutrons))

Geometric phase is

$$\gamma_n(C) = -n\Omega(C)$$

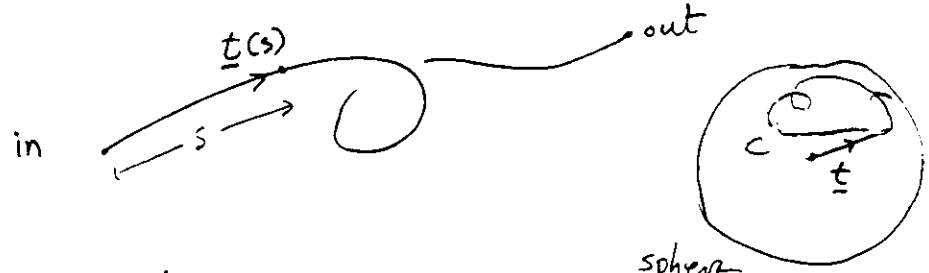
solid angle subtended by C at $\underline{B}=0$

Schematic experiment



PHOTONS spin $n = \pm 1$ (Chiao-Wu-Tomita)

Photons don't interact with electric or magnetic fields, but their spins are parallel to their direction of propagation $\underline{t}$, and so can be turned by changing $\underline{t}$, e.g. by coiling an optical fibre along which the light is travelling



phase shifts

$\gamma = -\Omega(C)$ (spin 1 : right circular polarization)

$+\Omega(C)$ (spin −1 : left circular polarization)

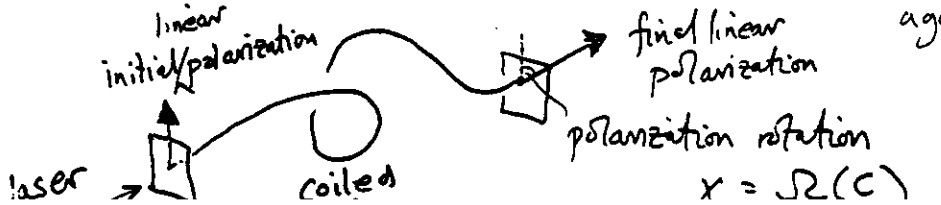
Instead of an interference experiment, it is easier to send in linearly polarized light:

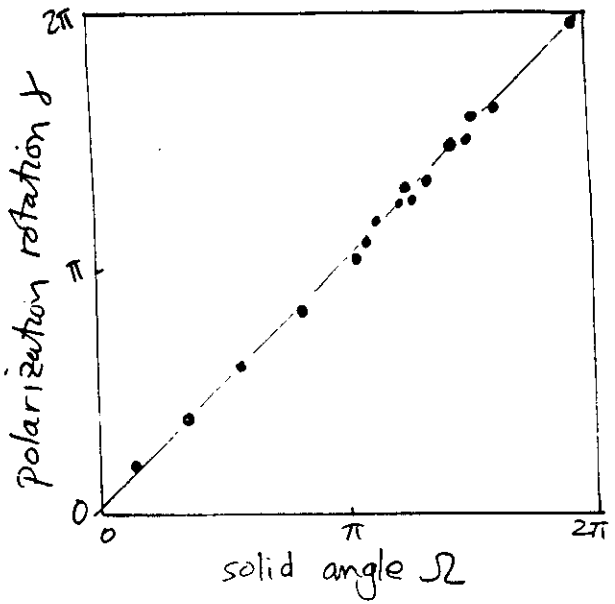
linear polarization = superposition of R circular + L circular

afterwards:

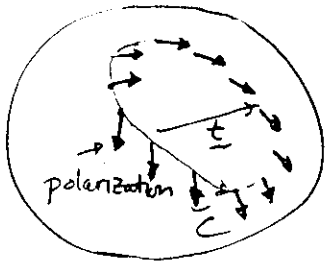
$e^{-i\Omega} \times R + e^{+i\Omega} \times L$ = linear polarization rotated by Ω

Vladimirski again!





This is 'geometric optical activity' the solid-angle dependence means that the polarization direction is parallel-transported over the $\underline{t}$ sphere



(Ross 1984)

A natural question is the effect quantum or classical?

Chiao and Wu:

"... would rather think of these effects as topological features of classical Maxwell theory which originate at the quantum level, but survive the correspondence-principle limit ($\hbar \to 0$) into the classical level"

(They also survive the short-wave limit ($\lambda \to 0$) into geometrical optics - Rytov 1938, Vladimirsky 1941)

Feynman "The photon equation ... is just the same as Maxwell's eqn"

ray arguments strictly not applicable in monomode fibres; need full Maxwell equations.

Quantum or classical?

Pseudoproblem because (Feynman)

'The photon equation is just the same as Maxwell's equations ...'

Maxwell

$$\partial_t (\text{Fields}) = (\text{matrix linear in } \nabla)(\text{fields})$$

$\times i\hbar$ $\times i\hbar$

$$i\hbar\, \partial_t (\text{Fields}) = (\text{matrix linear in } \underline{p} = -i\hbar\nabla)(\text{fields})$$

$\to$ Schrödinger equation if matrix is Hermitian

Connection with spin:

$$-i\hbar \nabla \wedge \underline{E} = \underline{p} \wedge \underline{E} = -i(\underline{p} \cdot \underline{\underline{S}})\underline{E}$$

$$-i(\underline{p} \cdot \underline{\underline{S}})\underline{E} = \begin{pmatrix} 0 & -p_z & p_y \\ p_z & 0 & -p_x \\ -p_y & p_x & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}$$

$$\underline{\underline{S}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

spin 1 matrices

Explicitly:

$$\partial_t \underline{D} = \nabla \wedge \underline{H} \; ; \; \partial_t \underline{B} = -\nabla \wedge \underline{E} \; ; \; \underline{B} = \mu(\underline{r})\,\underline{H} \; ; \; \underline{D} = \varepsilon(\underline{r})\,\underline{E}$$

$\varepsilon(\underline{r})$: fibre refractive index

$$i\hbar\,\partial_t |\underline{\Psi}\rangle = \hat{H}\,|\underline{\Psi}\rangle$$

$$|\underline{\Psi}\rangle = \begin{pmatrix} \varepsilon^{1/2}(\underline{r})\,\underline{E}(\underline{r},t) + i\mu^{1/2}(\underline{r})\,\underline{H}(\underline{r},t) \\ \varepsilon^{1/2}(\underline{r})\,\underline{E}(\underline{r},t) - i\mu^{1/2}(\underline{r})\,\underline{H}(\underline{r},t) \end{pmatrix}$$

$$\hat{H} = c \begin{pmatrix} \underline{\Pi}\cdot\underline{\sigma} & i\hbar\,\underline{\xi}\cdot\underline{\sigma} \\ -i\hbar\,\underline{\xi}\cdot\underline{\sigma} & -\underline{\Pi}\cdot\underline{\sigma} \end{pmatrix}$$

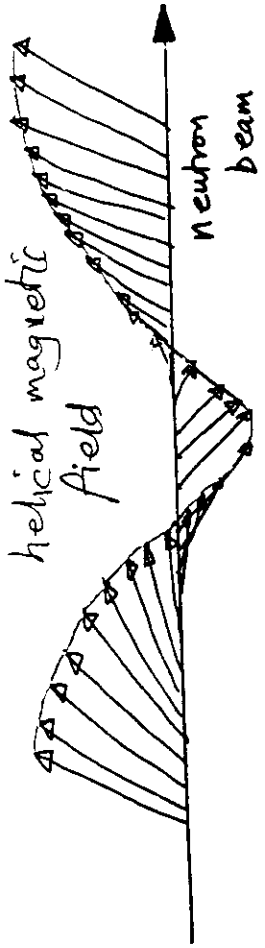
$$\underline{\Pi} = n^{+1/2}(\underline{r})\,\underline{p}\,n^{-1/2}(\underline{r}) \qquad n(\underline{r}) = \sqrt{\mu(\underline{r})\varepsilon(\underline{r})}$$

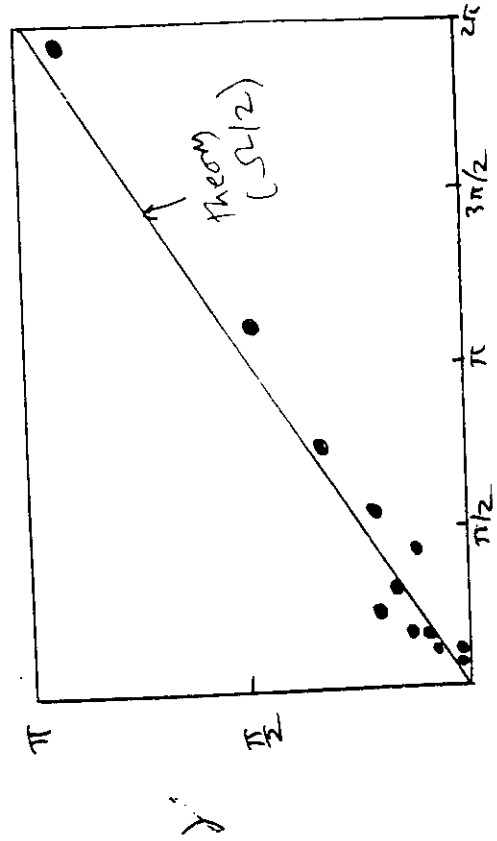
$$\underline{\xi}(\underline{r}) = \frac{1}{4n(\underline{r})}\,\nabla \log \frac{\varepsilon(\underline{r})}{\mu(\underline{r})}$$

So, Maxwell's equations (1865?), + $\hbar$ only relativistic but size

T. Bitter and D. Dubbers

INCLINONS, $n = \pm 1/2$

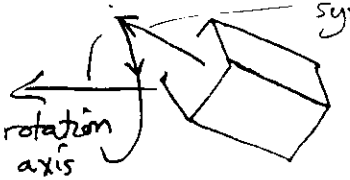




NUCLEI spin $n = \pm 3/2$

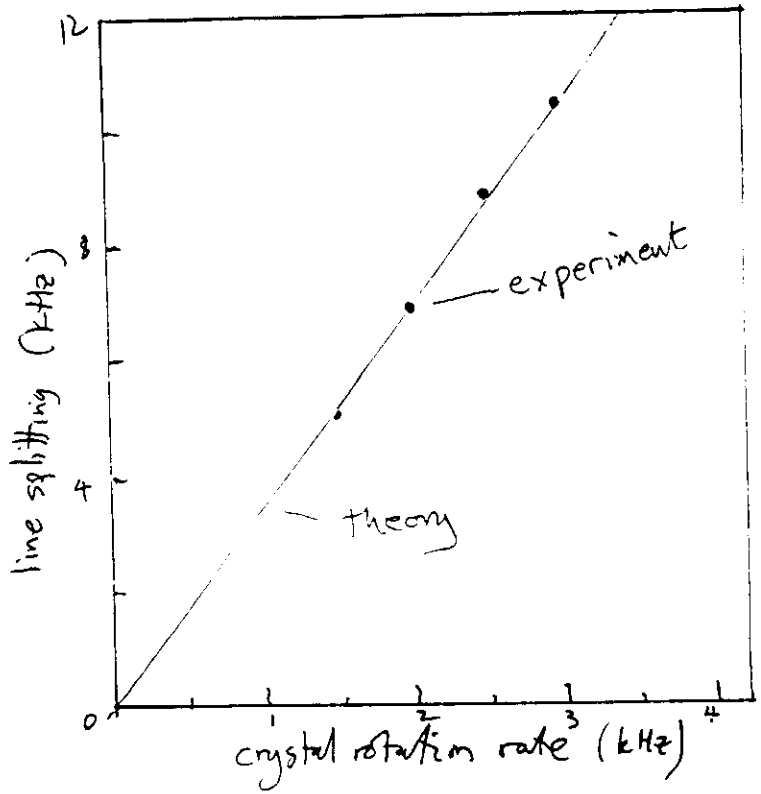
^{35}Cl nuclei in crystalline $NaClO_3$

Spin has electric (quadrupole) interaction with crystal electrons, and so can be turned by rotating a symmetry axis of the crystal



Phase of state increases by $\pm \frac{3}{2} \Omega$ per rotation. (Ω: solid angle)

Ever-changing phase $\longrightarrow$ frequency shift $\left(\frac{d\,\text{phase}}{dt}\right)$ measurable as line splitting in nuclear magnetic resonance



R. Tycko

Geometric contribution to adiabatic transition probability

'Adiabatic transitions' are the exponentially weak transitions between states, governed by

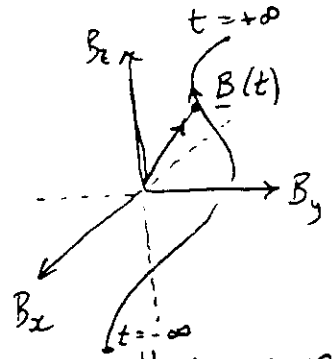
$$i\hbar \frac{\partial |\Psi\rangle}{\partial t} = \hat{H}(\delta t)\,|\Psi\rangle$$

when $\hat{H}(\delta t)$ is non-degenerate and δ is small

Example: Spin-½ in a $\underline{B}$ field

$$i\hbar \frac{\partial |\underline{\Psi}\rangle}{\partial t} = \mu\, \underline{B}(\delta t)\cdot \underline{\sigma}\, |\underline{\Psi}\rangle$$

($\underline{\sigma}$: Pauli)



Separation of adiabatic (instantaneous) eigenstates is $2\mu\,|\underline{B}(\delta t)|$ — never zero for real t, so transitions arise from complex degeneracies

ie $B_x^2(\tau_c) + B_y^2(\tau_c) + B_z^2(\tau_c) = 0$

Old theory: Transition probability

$$P_{+\to -} \sim e^{-\frac{4}{\hbar\delta}\,\mathrm{Im}\int_0^{\tau} d\tau\, B(\tau)} = e^{-\Gamma_d}$$

(τ: any real time)

Old theory wrong!

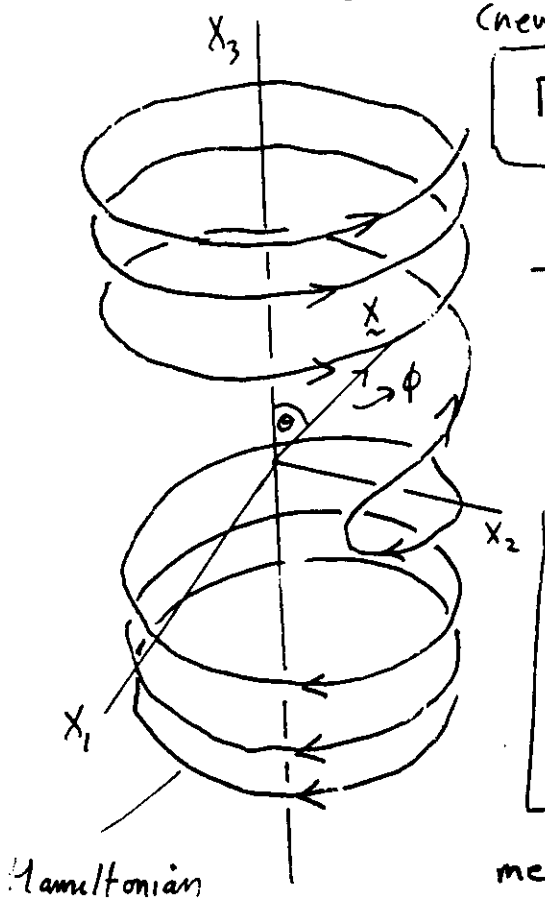
$$P_{+\to-}(\infty) = e^{-\Gamma_d} e^{\Gamma_g}$$

dynamical exponent (old)

$$\Gamma_d = \frac{4}{t\delta} \mathrm{Im} \int_{\tau_0}^{\tau} d\tau B(\tau)$$

geometric exponent (new) – from ∥ transport of adiabatic states

$$\Gamma_g = -2\,\mathrm{Im} \int_{\tau_0}^{\tau_2} d\tau \dot{\phi} \cos\theta$$



This is purely geometric – independent of δ and t_f.

Γ_g is a complex solid angle.

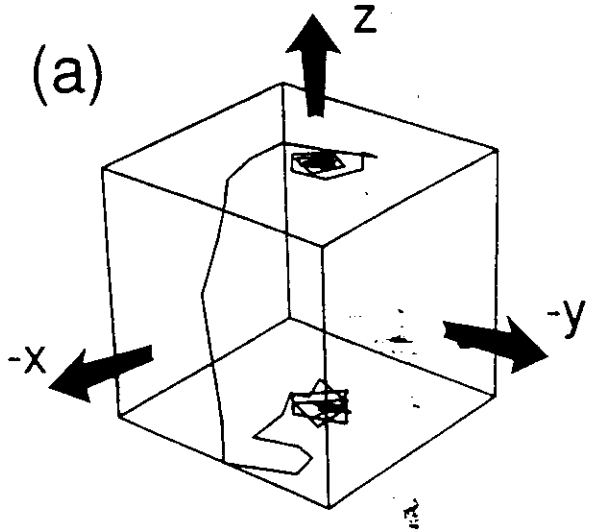
Example: Twisted Landau–Zener:

$$X_3 = A\tau$$
$$X_1 = \Delta \cos B\tau^2$$
$$X_2 = \Delta \sin B\tau^2$$

$$\Gamma_g = -\frac{\Delta^2}{A^2} B \,\mathrm{sgn}\, A$$

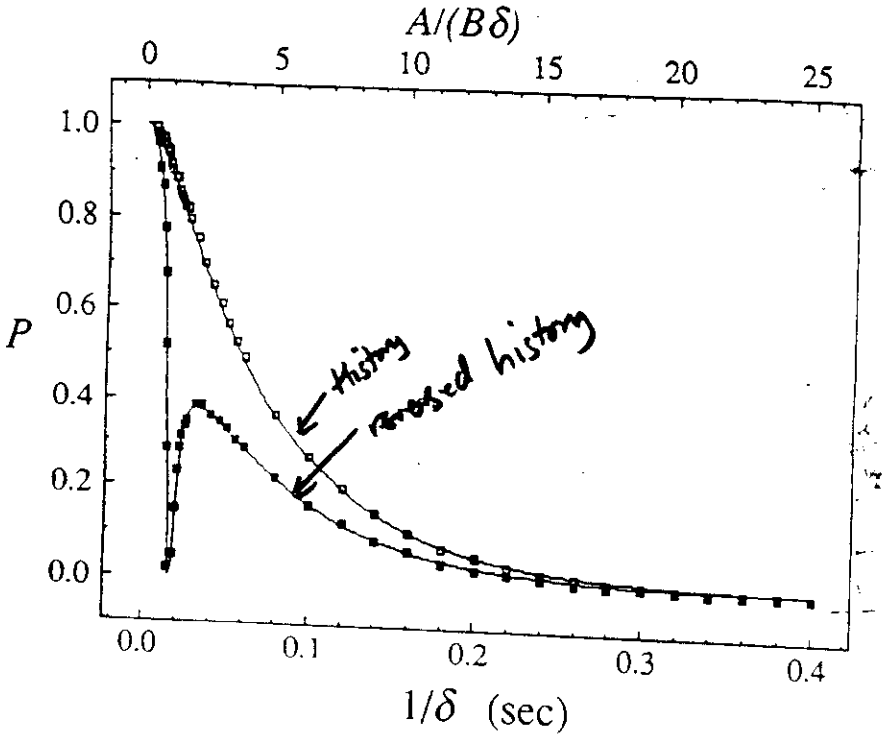
measurable in a spin experiment

Histories in Hamiltonian space



J.W. Zwanziger, S.P. Rucker and G.C. Chingas Phys Rev A 43 (1991) 3232-9

NMR in ^{13}C



Predict

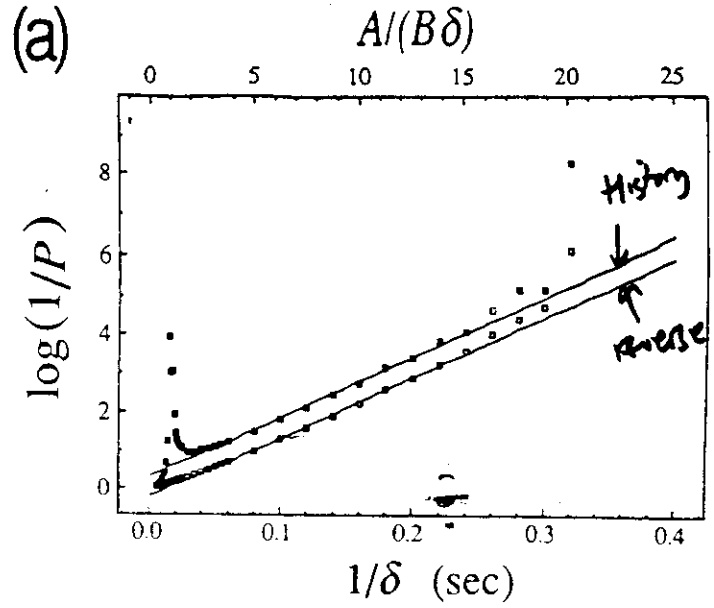
$$P_{+\to-} = e^{\Gamma_g - \Gamma_d}$$

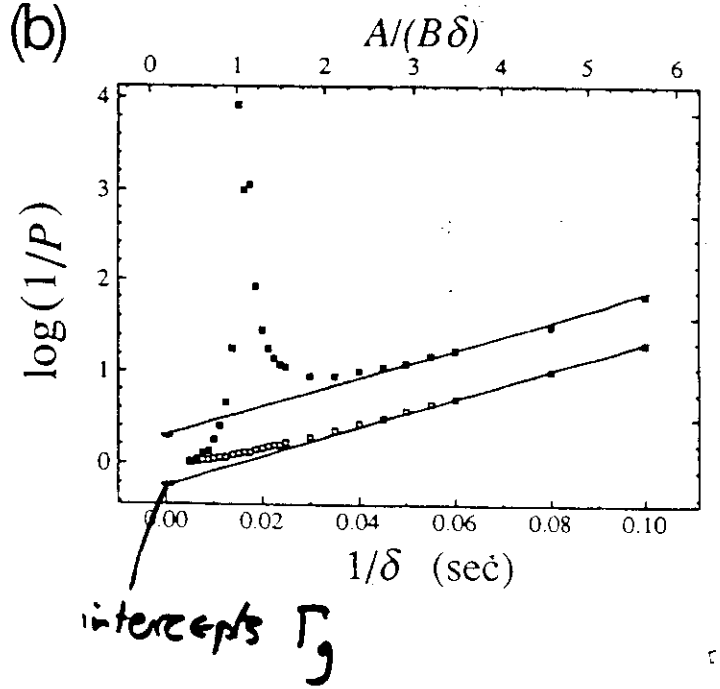
Γ_g: $O(\delta)$

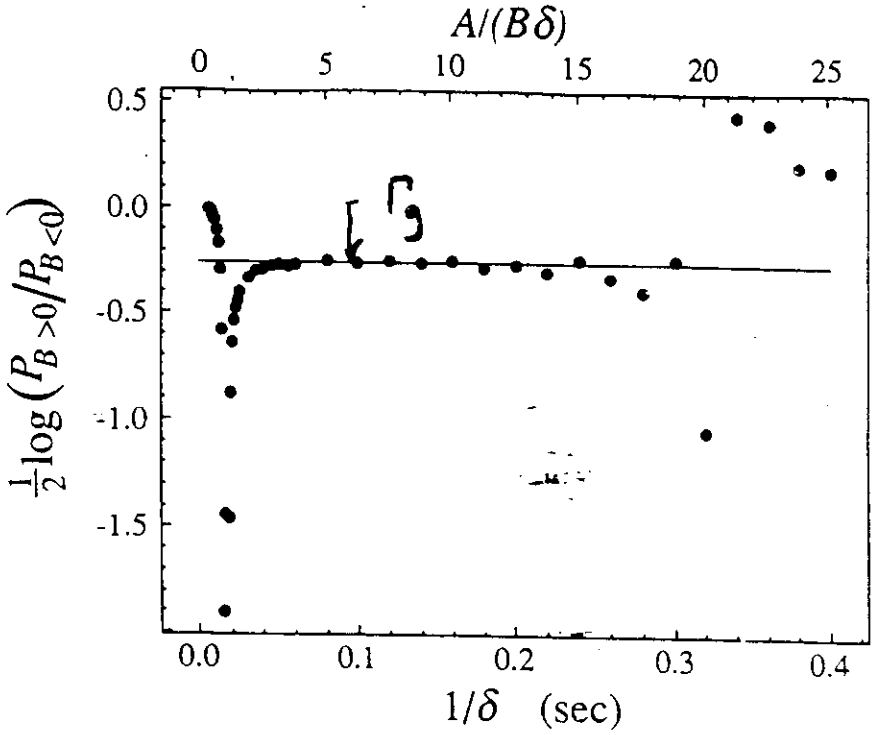
Γ_d: $\frac{1}{\delta\hbar}\int (E_1 - E_0)\,dt$

under reversal, Γ_g reverses

Γ_d does not







Theory $\Gamma_g = -0.243$

Expt $\Gamma_g = -0.26 \pm .01$

Figure 6

Lecture 2

(Some geometric phases, Michael Berry Geneva 1993)

BRISTOL ANHOLONOMY CALENDAR

1. Crystal dislocations as anholonomy (Frank 1951)

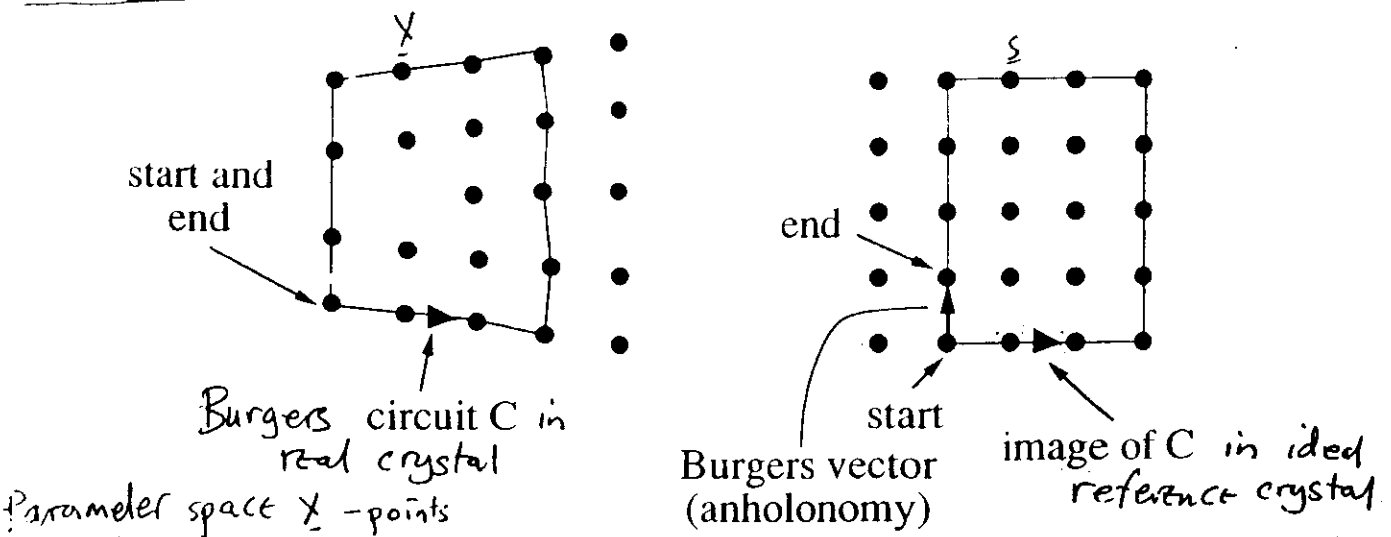


Figure 2. Crystal dislocation as anholonomy

2. Dislocations in waves (Nye and Berry 1974)

Wave $\psi(x) = \rho(x)\, e^{i\chi(x)}$

wavefronts $\chi(x) = \text{constant mod } 2\pi$

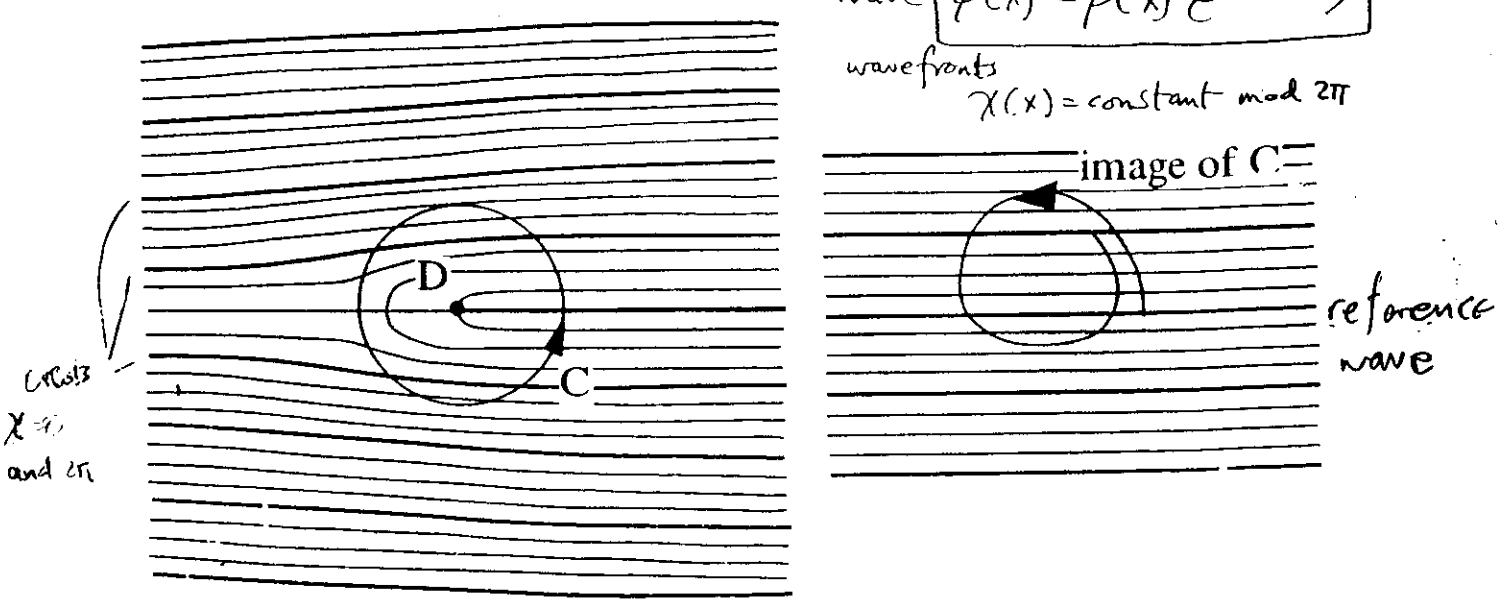


Figure 5. Wavefronts (at intervals of $\pi/4$ with crests bold) of a dislocated wave with a Burgers circuit surrounding the dislocation D, and its image in a plane wave

At dislocations, χ has a 2π anholonomy, and ρ is zero

$\frac{1}{2\pi}\oint_C d\chi = \text{integer}$

Dislocations are lines

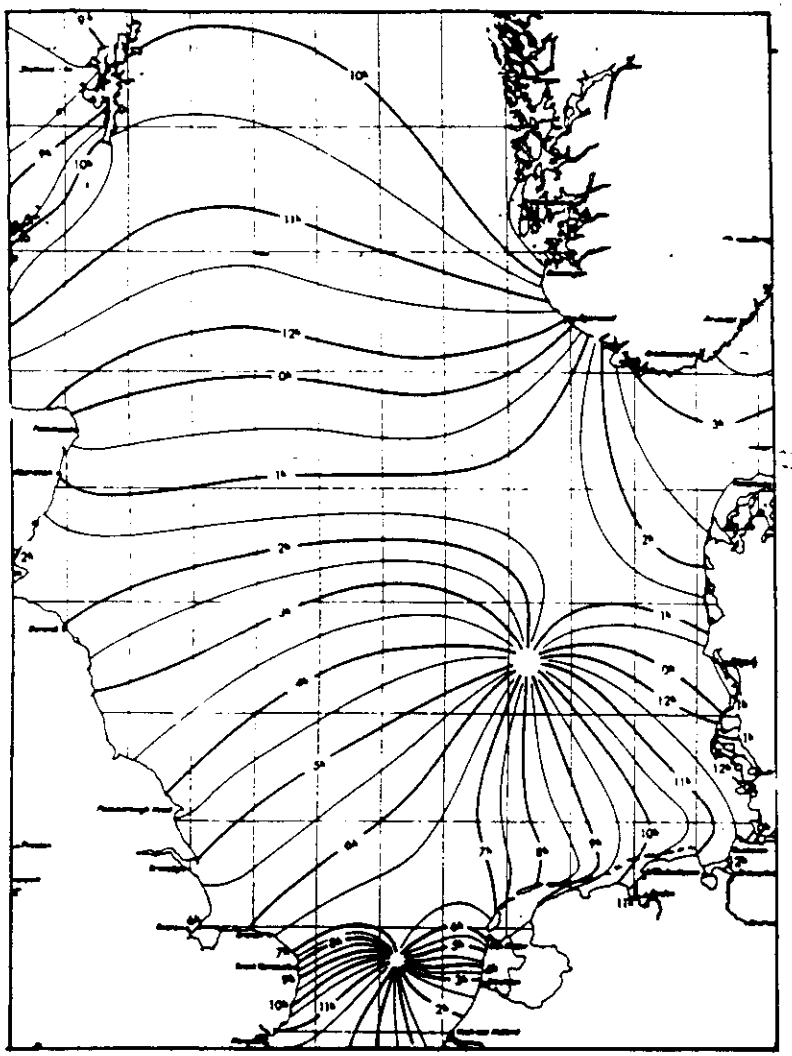
threads of silence

sound, microwaves, light, tides,

Dislocations in the tide wave (Whewell (1833)

Wavefronts : loci of instantaneous high tides at given times

Dislocations : 'amphidromic points' of no tide



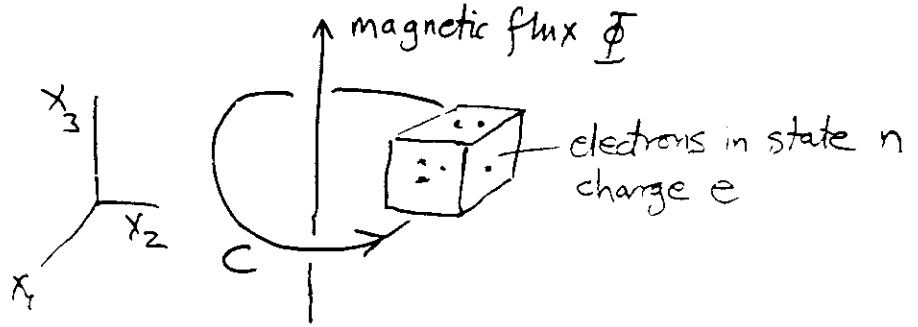
The tide wave (12h period) is a forced vibration of the whole earth. Rotating source → no time-reversal symmetry → complex waves (stationary but not standing) $|h(\theta,\phi,t) = Re[e^{-i\omega t}\psi(\theta,\phi)]|$

3. AHARONOV-BOHM EFFECT (1959)

System : electrons

environment: 3-D space of positions, containing a line of magnetic flux (e.g. long solenoid)

cycle : circuit around the flux line

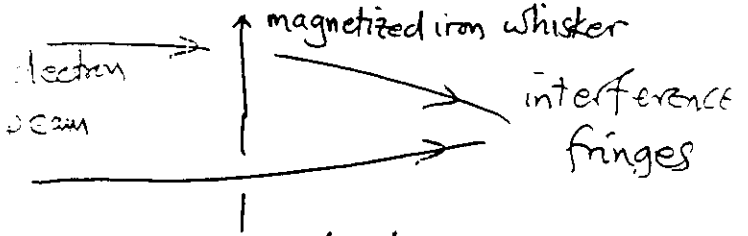


After the circuit, geometric phase is

$$\gamma = e\Phi/\hbar$$

(electrons influenced quantally by field elsewhere)

Chambers' experiment 1959



The flux Φ shifts the fringes, via an extra phase:

phase(→) − phase(↻)

= phase(→) + phase(↺)

= phase(⟲) = geometric phase

FOUR CLASSES OF AB EFFECT

effect	wave	flux	
AB	quantum	classical	electrons, mag flux
Tonomura	quantum	quantum	electrons on flux in superconduct
Steinberg	classical	quantum	sound waves on quantized vortex in He
Bathtub	classical	classical	ripples on bathtub vortex

But, can observe water-wave analogue of AB effect
For waves on a medium moving with speed $\underset{\sim}{v}(\underset{\sim}{r})$ ($\ll c$), $\underset{\sim}{v}(\underset{\sim}{r})$ acts like a vector potential, so for AB need

$\nabla \times \underset{\sim}{v} = A\delta(\underset{\sim}{r})$, ie. a vortex line in irrotational flow.

This can be realized with surface ripples scattered from a bathtub vortex

M.V.B., R.G. Chambers et. al, Eur. J. Phys 1 (1980) 154-162

Dislocations can be
1) unattached to any singularity of the medium (generic case)
2) coincident with line singularities (AB)
3) attached to a point singularity and otherwise movable (but not removable) by gauge transformation (Dirac string

4. Liquid crystal disclinations as anholonomy (Frank 1959)

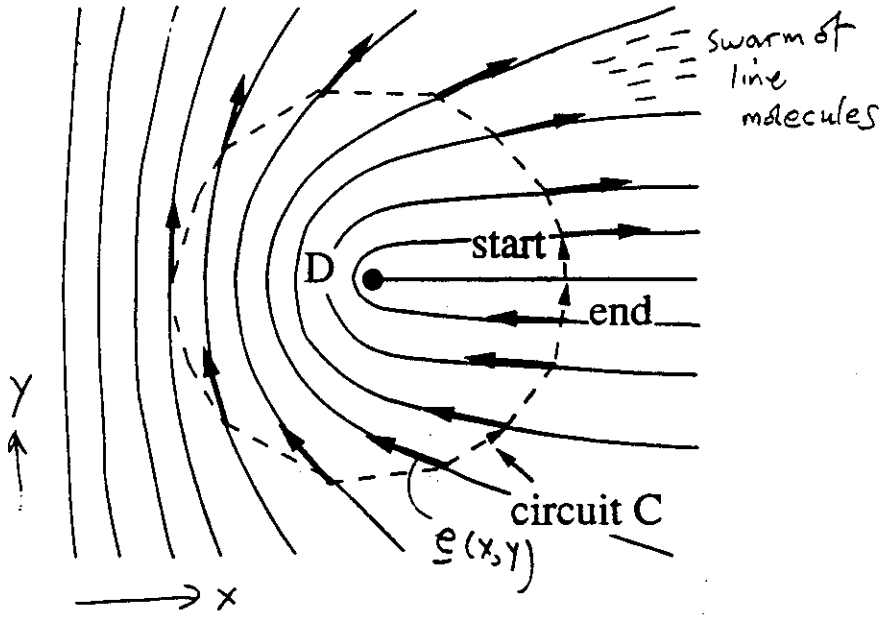


Figure 3. Reversal (π anholonomy) of liquid crystal direction around a disclination D

Can regard line field as field of eigendirections of a 2×2 real symmetric matrix

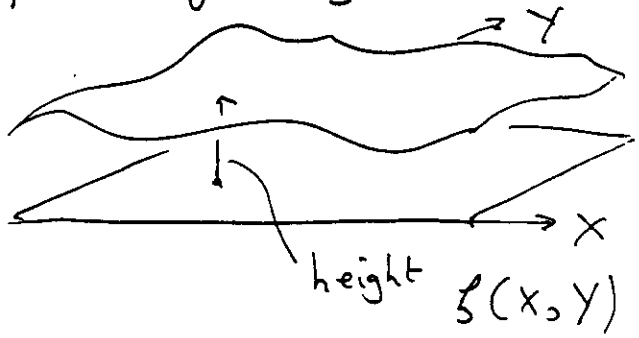
$$M(X,Y) = \begin{pmatrix} M_{11}(X,Y) & M_{12}(X,Y) \\ M_{12}(X,Y) & M_{22}(X,Y) \end{pmatrix}$$

$$\underline{M}\,\underline{e} = \lambda \underline{e}$$

Disclination D is a degeneracy $M_{11} = M_{22}$, $M_{12} = 0$

Darboux (1896)

Differential geometry of curved surfaces.



Curvature matrix

$$H(X,Y) = \begin{pmatrix} \partial^2\zeta/\partial X^2 & \partial^2\zeta/\partial X\partial Y \\ \partial^2\zeta/\partial Y\partial X & \partial^2\zeta/\partial Y^2 \end{pmatrix}$$

real, symmetric

parameters

Eigenvalues: principal curvatures at X, Y

Eigenvectors: (orthogonal) directions of principal curvatures at X, Y.

Degeneracy: umbilic point on surface (locally spherical)

Sign change → a line of curvature turns through π (reverses) in a circuit of an umbilic point

fig. 44
(disclinations of directors of liquid crystals)
(Berry and Hannay (1977))

(surface: $z(x,y)$; curvature matrix: $\begin{pmatrix} \partial^2 z/\partial x^2 & \partial^2 z/\partial x \partial y \\ \partial^2 z/\partial y \partial x & \partial^2 z/\partial y^2 \end{pmatrix}$; eigenvalues: principal curvatures; eigenvectors: directions of principal curvature; degeneracy = umbilic point (locally spherical))

Degeneracies These are singularities of the 2-form generating the phase

$$\gamma_n(C) = \iint_{\partial S = C} \mathrm{Im}\langle dn | \wedge | dn \rangle \equiv \iint_{\partial S = C} V_n(\underline{X})$$

Singularities of V_n at X^*, where $|n\rangle$ (transported state) degenerates with $|n+1\rangle$ or $|n-1\rangle$

locally a 2-state system (like spin $-\frac{1}{2}$), so can choose parameters in which Hamiltonian is, locally

$$H = \frac{1}{2}\begin{pmatrix} Z & X - iY \\ X + iY & -Z \end{pmatrix} \rightarrow V_n \approx \frac{\underline{R} \cdot d\underline{S}}{R^3}$$

monopole

Then $\gamma_n(C) \approx \pm \frac{1}{2}\Omega(C)$ for C near X^*

Important special case: H real (eg. time-reversal symmetry) i.e. $Y = 0$. Circuits in ZX plane

$\Omega = 2\pi$ (hemisphere)

$\therefore \gamma = \pm\pi$

$$e^{i\gamma} = -1$$

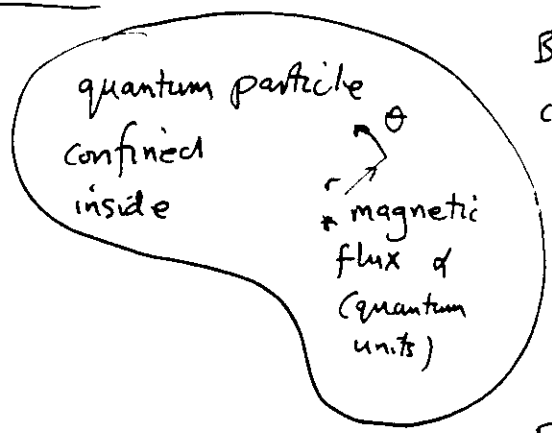
Sign change of real symmetric matrix round degeneracy

Exploring monopoles in parameter space

with Raul Mondragon

Need a 3-parameter family of quantum systems, uncomplicated by any symmetries (e.g. time reversal)

Such a family is Aharonov-Bohm chaotic billiards



Boundary is a cubic conformal transformation of unit disc, i.e

$$W(z) = z + Bz^2 + Ce^{i\chi}z^3$$

$|z| = 1$

Fix $\chi = \pi/3$

Hamiltonian

$$\hat{H} = -\left(\nabla - i\alpha \frac{\underline{e}_\theta}{r}\right)^2 \quad ; \quad \hat{H}|\psi\rangle = E|\psi\rangle$$

$\psi(\underline{r}) = \langle \underline{r}|\psi\rangle = 0$ on Boundary

Parameters $(B, C, \alpha) \equiv \underline{R}$

2 Form $V = \mathrm{Im}\langle \nabla_{\underline{R}}\psi_n | \wedge | \nabla_{\underline{R}}\psi_n\rangle$

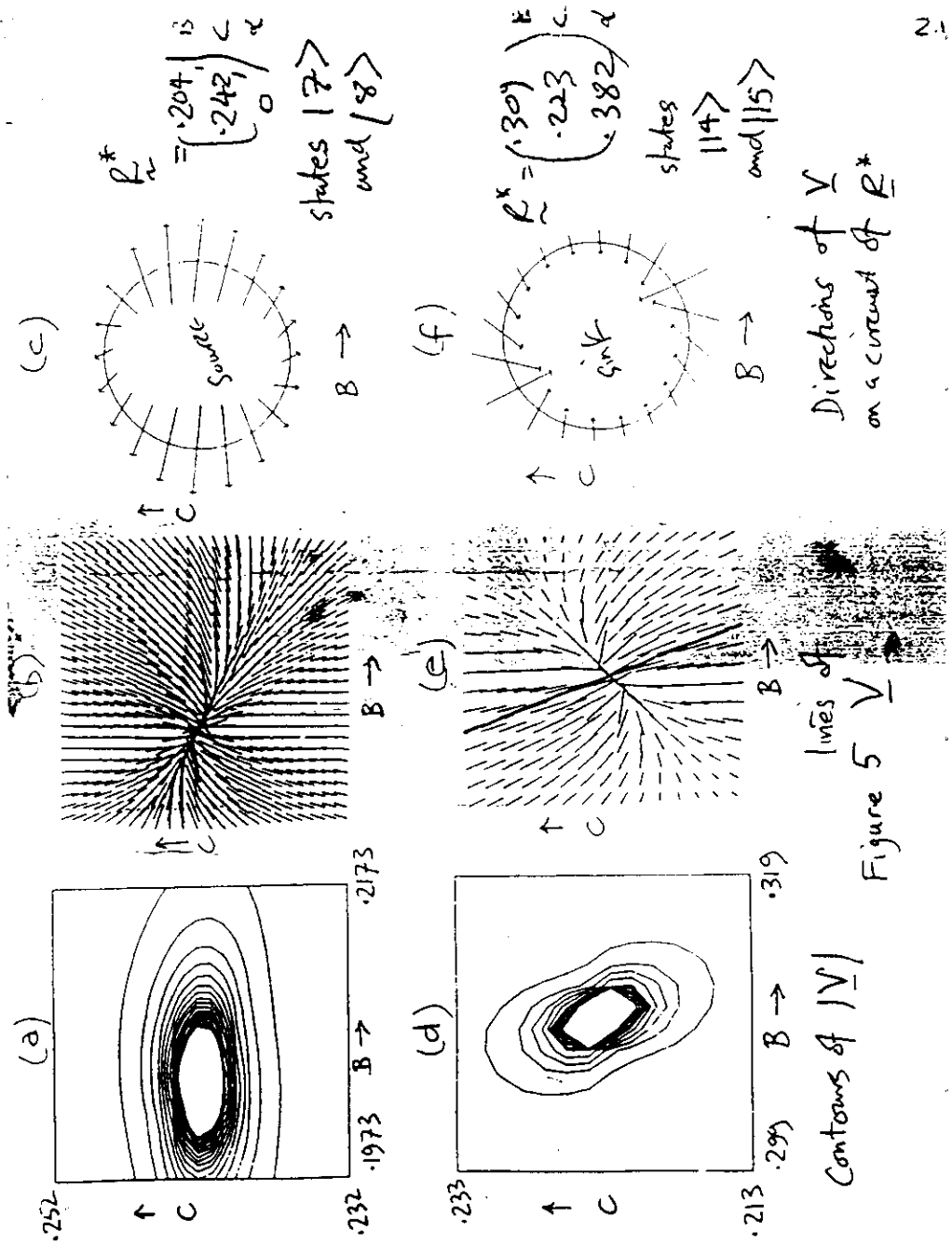


Figure 5

Monopole behaviour of $\underline{V}$:

$$|\underline{V}| \propto \frac{1}{r^2}$$

where $r = \sqrt{(B-B^*)^2 + (C-C^*)^2}$

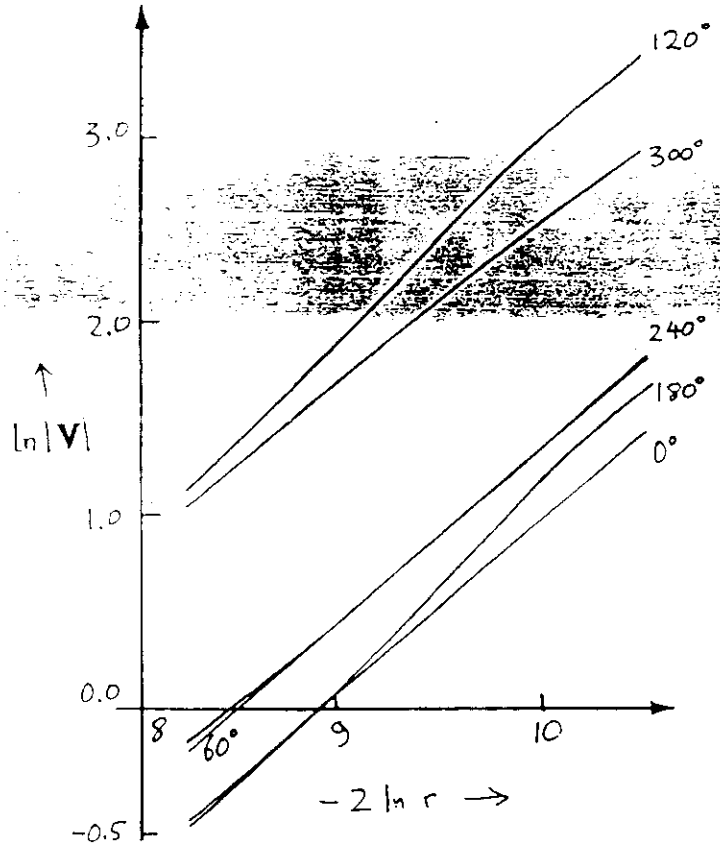
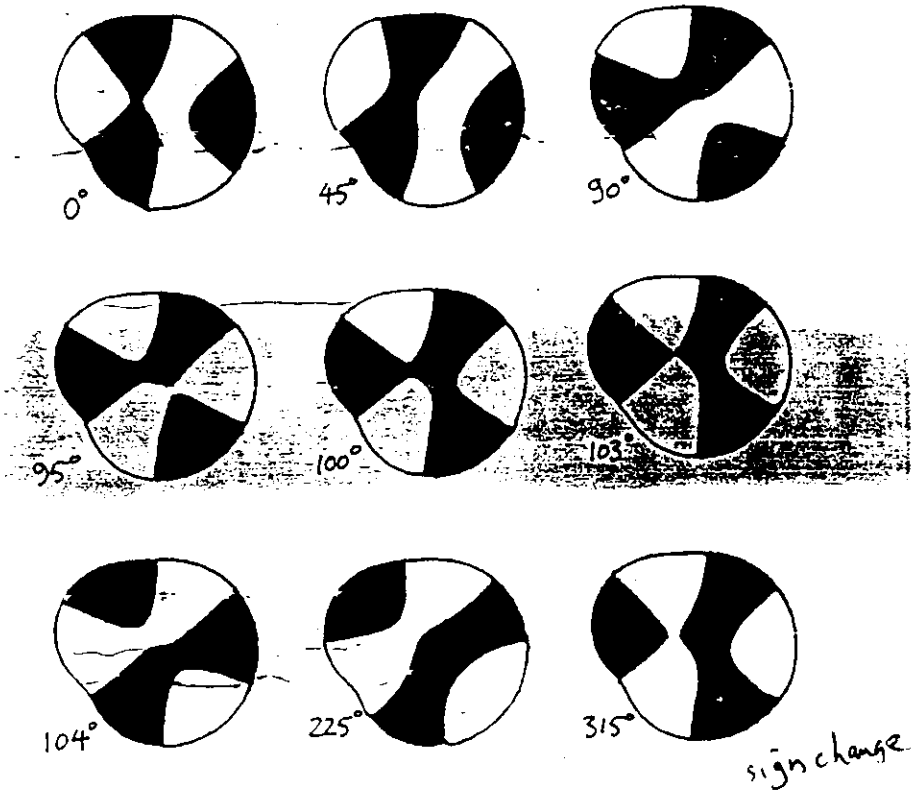


Figure 2

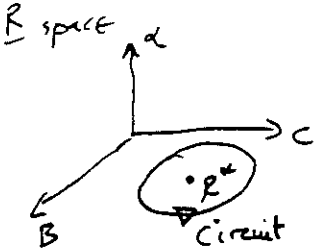
Rearrangement of nodal cells round degeneracy

at $\underset{\sim}{R}^* = (\underset{B}{.204}\ \underset{C}{.242}\ \underset{\alpha}{0})$ — ∴ states real

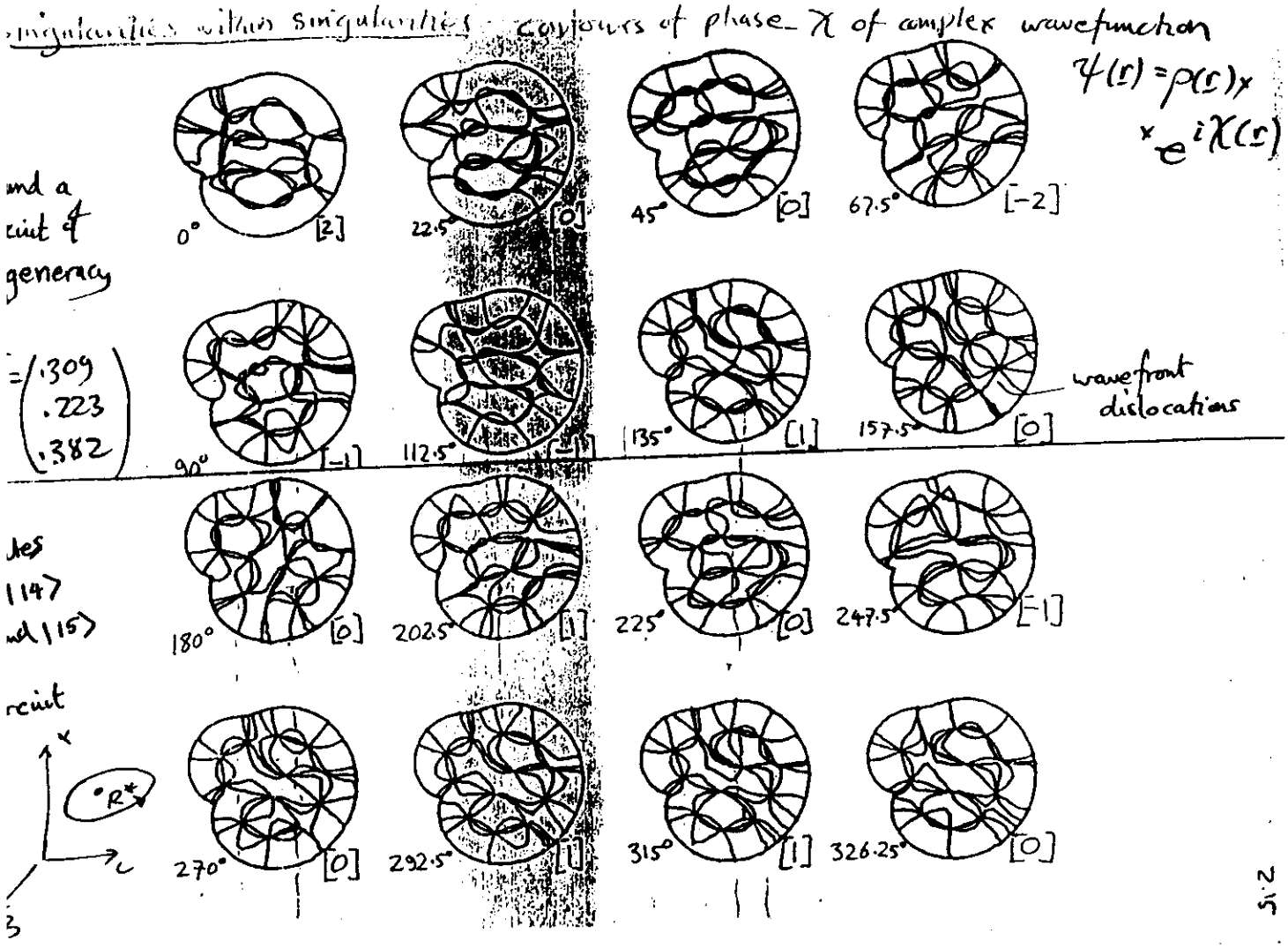


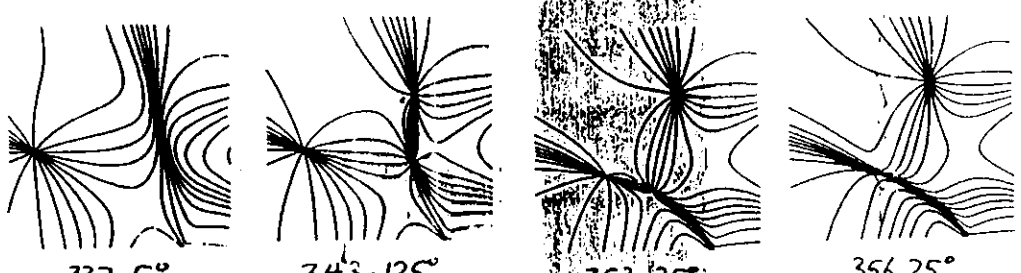
$$|\psi\rangle = \cos\tfrac{\Theta}{2}\,|7\rangle + \sin\tfrac{\Theta}{2}\,|8\rangle \qquad 0 \le \Theta \le 2\pi$$

Figure 7



two states at $\underset{\sim}{R}^*$





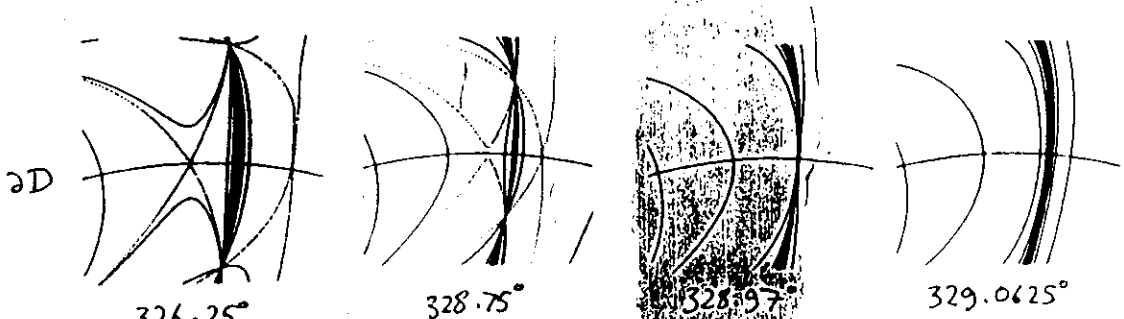


Figure 10
Singularities within singularities within singularities

5. Disclinations in waves (Nye 1983)

In general 3-D spatially varying vector waves, there are singular lines where $\underline{E}$ is linearly polarized or circularly polarized.

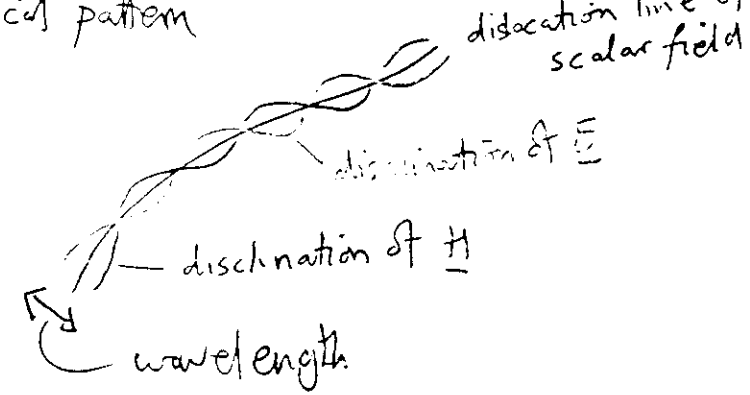
2π anholonomy — π anholonomy

These line singularities are 'a useful way of describing the spatial and temporal geometrical features of electromagnetic fields that would otherwise be far too complicated to visualize. The disclination carries with it a certain local field structure; therefore, a description of the arrangement and motion of the disclinations contains much of the essential geometrical information about the field itself. They constitute elements of structure in the field.'

NPL

Electromagnetic double helix:

Typical pattern



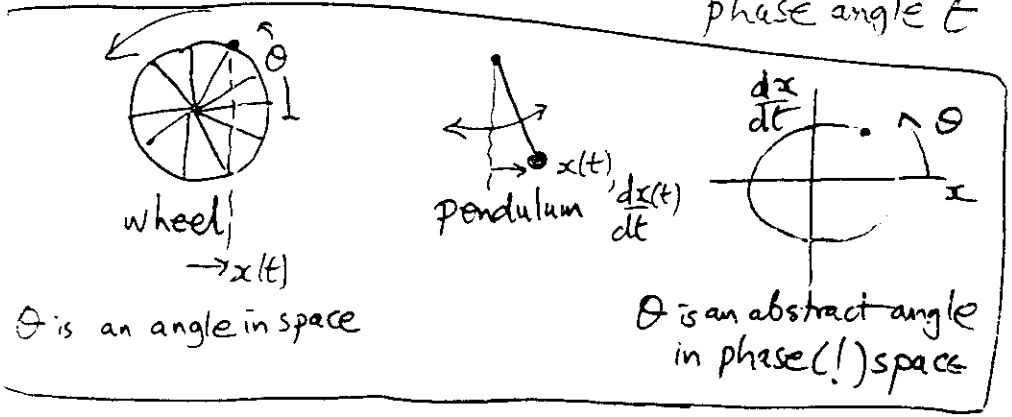
NEWTONIAN ANHOLONOMY

A quantum state with a fixed environment is just a complicated oscillator, and shows phase anholonomy as the environment is altered. In classical mechanics, oscillators abound — wheels, pendulums, in fact any periodic or multiply periodic motion — so these should show anholonomy too.

$$x(t;X) = A(X) \begin{matrix}\sin\\ \text{or}\cos\end{matrix}\{\omega(X)\,t\}$$

(t: time)

- oscillator coordinate
- parameters
- amplitude
- frequency
- generalized phase angle θ



over time T, θ changes by $\omega(X)T$ for fixed environment X. When X is slowly cycled, then, as before, θ acquires not only the expected dynamical change but a geometric one:

$$\theta_{end} - \theta_{start} = \underbrace{\int dt\,\omega(X(t))}_{\text{dynamical}} + \underbrace{\Delta\theta(C)}_{\text{Hannay's angle}}$$

($\omega = -\frac{\partial H}{\partial n}$)

chaos?

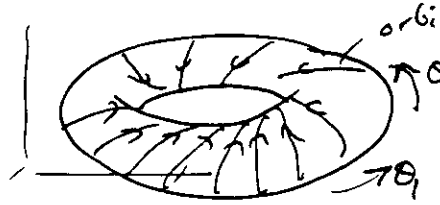
Hannay's angles in classical multiply periodic systems.

Classical Hamiltonian

$$H(\underline{q}, \underline{P}, \underline{X}(t)) \qquad D \text{ freedoms}$$

$q_1 \cdots q_D$ $P_1 \cdots P_D$ parameters $\{X_1, \ldots\}$

For fixed $\underline{X}$, H is integrable : D conserved quantities, motion on D-torus in $2D$ dim. phase space:



Tori labelled by $\underline{I} = \{I_1 \cdots I_D\}$

$$I_j = \frac{1}{2\pi} \oint_{\gamma_j} \underline{P} \cdot d\underline{q}$$

Points on tori labelled by q's

angles $\underline{\theta} = \{\theta_1 \cdots \theta_D\}$

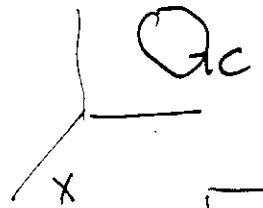
$1 \le j \le D$: D actions (conserved)

Can express H in terms of actions

$H = H(\underline{I})$; frequencies

$$\omega_j = \frac{\partial H(\underline{I})}{\partial I_j} = \dot{\theta}_j$$

Now let parameters $\underline{X}$ vary round a cycle C



At the end, the tori returns ($\underline{I}$ conserved: classical adiabatic theory)

But $\underline{\theta}$ does not.

Hannay:

$$\theta_{j\,\text{end}} - \theta_{j\,\text{start}} = \int_0^T dt\, \omega_j(\underline{X}(t)) + \Delta\theta_j(C)$$

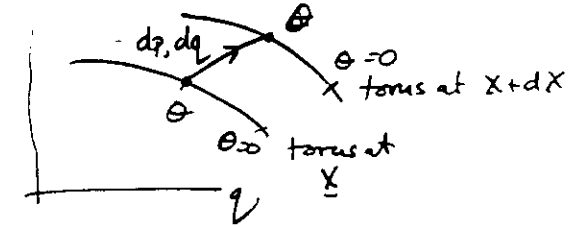
dynamical — Hannay

$$\Delta\theta_j = -\frac{\partial}{\partial I_j} \iint_{\partial S = C} \langle d\underline{P} \wedge \cdot\, d\underline{q} \rangle$$

From Hamilton's equations

$$\langle\ \rangle = \frac{1}{(2\pi)^D} \int_0^{2\pi} dq_1 \cdots dq_D$$

= torus average



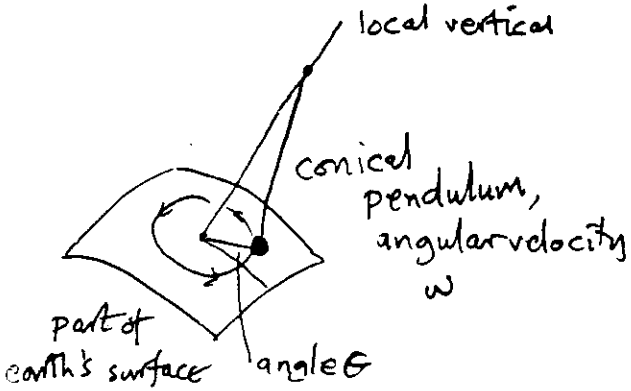
Semiclassical theory : from WKB association tori $\longleftrightarrow$ eigenstate

$$\gamma_n(C) \simeq \frac{1}{\hbar} \iint_{\partial S = C} \langle d\underline{P} \wedge \cdot\, d\underline{q} \rangle$$

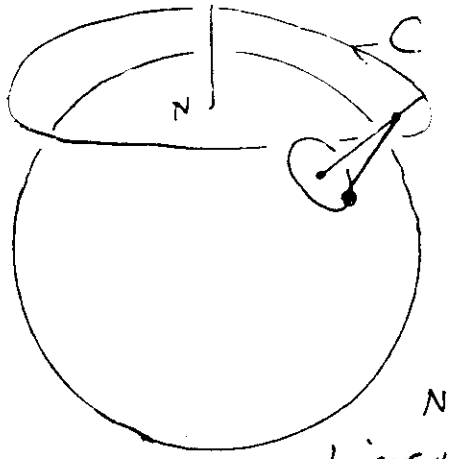
i.e. $$\Delta\theta_j = -\hbar \frac{\partial}{\partial I_j} \gamma_n(C) = -\frac{\partial}{\partial n_j} \gamma_n(C)$$

-- so $\gamma_n(C) = -n_j 2\pi$, so $\Delta\theta = 2\pi$

Example : Foucault pendulum



During 24 hours ($=T$), the local vertical (environment) turns:



angle turned through by bob of pendulum is

$$\theta_{end} - \theta_{start} = \omega T + \Omega$$

Ω : solid angle of C - Hannay

Now,

linear vibration = circular + circular

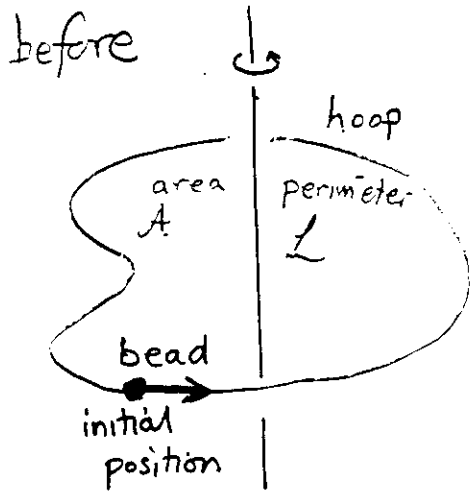
So, circular angle shift Ω $\longrightarrow$ rotation of linear vibration by $\Omega = 2\pi(1-\sin(\text{latitude}))$

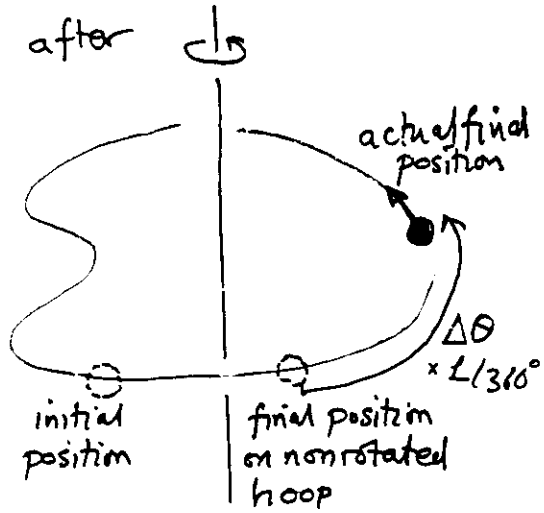
$\longrightarrow$ angular vel. of direction relative to earth's surface

$= (\Omega - 2\pi)/T = \frac{2\pi}{T}\sin(\text{latitude})$ (11.7°/h in Sicily)

= parallel transport of vibration (again!)

Example : Rotated rotator (Hannay's hoop)





Hannay angle:

$$\Delta\theta = 2\pi\left(1 - \frac{4\pi A}{L^2}\right)$$

purely geometric; zero for circle;
Euler force.

For chaotic classical dynamics, no tori no actions, no Hannay angles.

Does

$$\boxed{V = \mathrm{Im} \langle dn \wedge dn \rangle} \qquad \left(\gamma_n(C) = \iint_{\partial S = C} V \right)$$

have a classical limit? Yes. Main steps of derivation

1) $\boxed{\sum_m |m\rangle\langle m| = 1}$

2) $\boxed{\langle m\, dn \rangle = \frac{\langle m | d\hat{H} | n \rangle}{E_n - E_m}} \quad \text{(twice)}$

3) $\boxed{\frac{1}{(E_n - E_m)^2} = -\frac{1}{\hbar^2} \int_0^\infty dt\, e^{\frac{it}{\hbar}(E_n - E_m)}}$

4) quantum commutator → classical Poisson bracket

$$\boxed{[\hat{a}, \hat{b}] \longrightarrow i\hbar \{a(z), b(z)\}} \qquad z = \{q, p\}$$

5) $\boxed{\langle n | \hat{a} | n \rangle \to \langle a(z) \rangle}$ average over classical invariant manifolds corresponding to $|n\rangle$

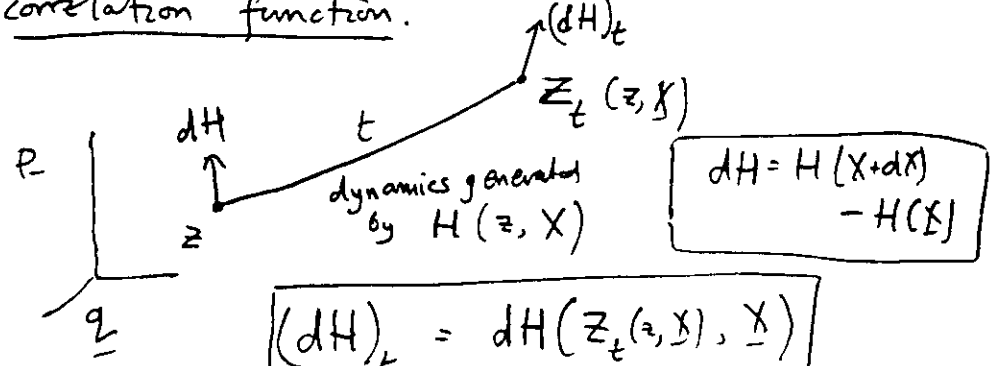
$$= \frac{1}{\partial_E \Omega} \int dz\, \delta(E - H(z, X))\, a(z) \quad \text{in ergodic case}$$

phase volume of energy surface $H = E$

Result

$$\boxed{V(X) \xrightarrow{\hbar \to 0} -\frac{1}{2\hbar\, \partial_E \Omega}\, d_E \left[\partial_E \Omega \int_0^\infty dt\, \langle (dH)_t \wedge dH \rangle \right]}$$

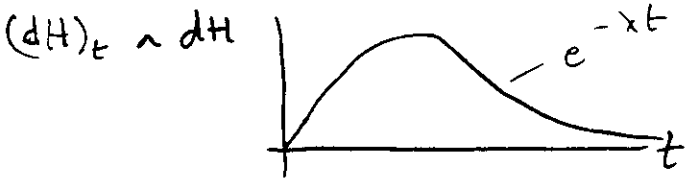
This is the time integral of a classical correlation function.



$$\boxed{dH = H(X + dX) - H(X)}$$

$$\boxed{(dH)_t = dH(Z_t(z, X), X)}$$

V involves in an essential way the history of the chaos estimate:



λ = entropy (rate of exponential separation)

Thus

$$\boxed{V(X) \approx -\frac{1}{2\hbar\, \partial_E \Omega}\, d_E \left[\frac{\partial_E \Omega}{\lambda^2} \langle \{H, dH\} \wedge dH \rangle \right]}$$

Lecture 3

(Some geometric phases, Michael Berry Geneva 1993)

DYNAMICS OF THE ENVIRONMENT

So far, have divided the world into

World = system + environment

(system: e.g. spin, electron; environment: acting through parameters $X(t)$)

But there is no unilateral action in physics, so the system must react back on the dynamics of the parameters $X(t)$, which are themselves subject to quantum mechanics. One reaction effect is on the phase of the quantum state of the environment: roughly, the world's state is single-valued, so the environment must acquire a phase $-\gamma_n(C)$ to compensate the system's $+\gamma_n(C)$.

This affects **molecules** when divided into:

molecule = electrons + nuclei

("system" - light, swift) ("environment" - ponderous, slow)

Environmental parameters are nuclear position coordinates X, whose quantization gives the vibration-rotation spectrum of the molecule.

6. Molecular electronic degeneracies

(Longuet-Higgins, Öpik, Pryce, Sack 1958)

Head of Bristol Physics dept

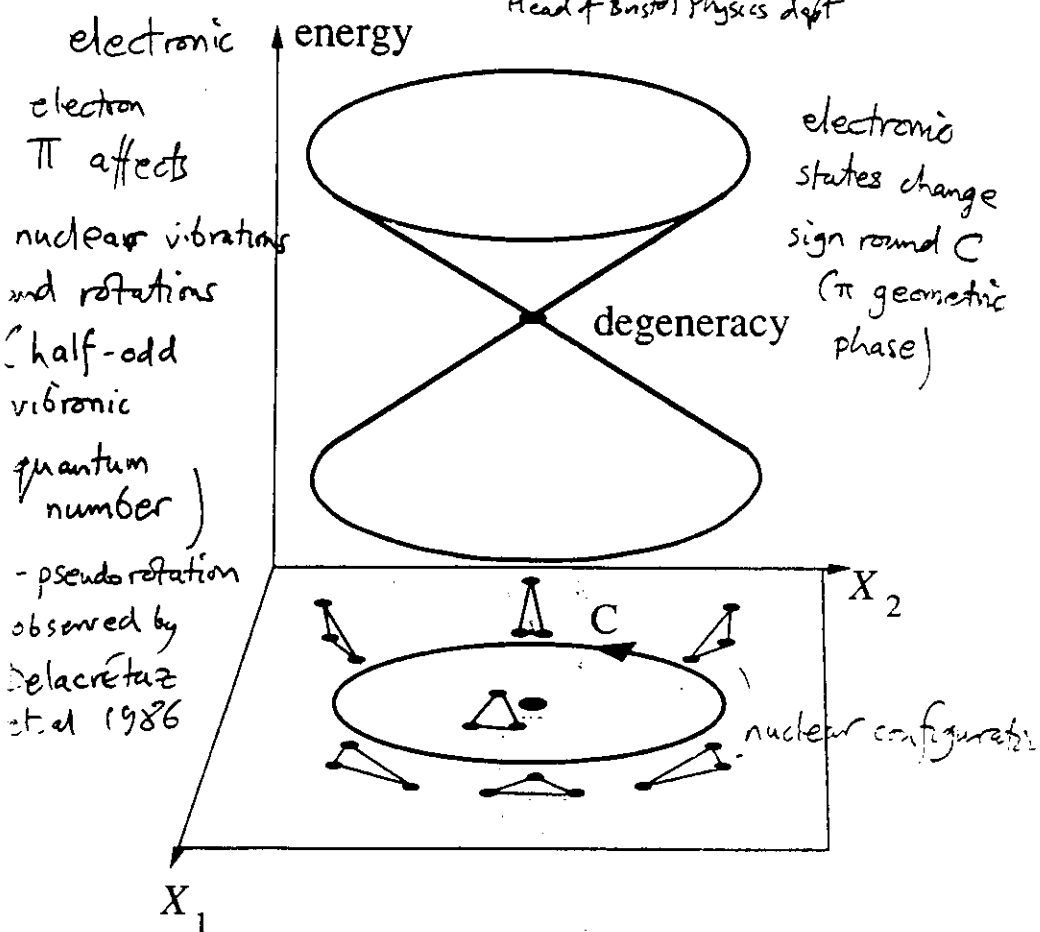


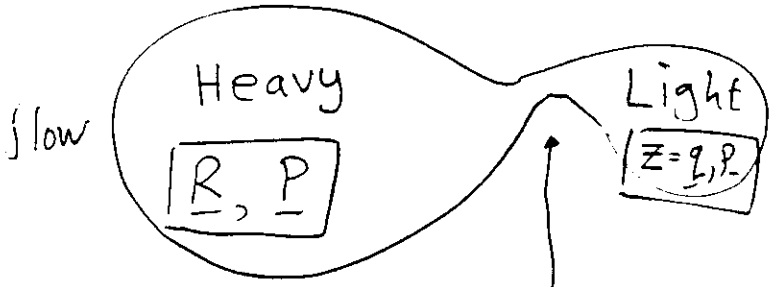
Figure 4. Circuit in shape space X of nuclei surrounding the equilateral molecule for which there is an electron energy degeneracy (conical intersection)

'This half oddness is at first sight strange, but may be understood by noting that [around a circuit] the electronic factor in the wave function will be multiplied by -1, so that the angular part of the nuclear factor must do likewise -1 th- [illegible]

Generalization:

CLASSICAL GEOMETRIC REACTION FORCES

Coupled systems (classical or quantum)



Here, concentrate on slow dynamics.

Simplest approximation (Born–Oppenheimer, Adiabatic averaging):

Averaged fast energy $E_{BO}(\underline{R})$ depends on slow coordinates, and its gradient acts as a reaction force on the slow system

$$\boxed{-\partial_i E_{BO}(\underline{R})} \qquad \partial_i \equiv \frac{\partial}{\partial R_i}$$

Born-Oppenheimer approximation depends on an adiabatic separation of time scales:

> Fast motion can be calculated with slow coordinates $\underline{R}$ acting as parameters, i.e. frozen.

An improved approximation, most consistent with the adiabatic assumption, includes **two extra reaction forces** on the slow motion: 'magnetic' and 'electric'

These first appeared in quantum mechanics (modifying vibration–rotation ('slow') energy levels in molecules) but have classical counterparts.

$$H(\underline{R},\underline{P},t) = \frac{1}{2}\sum\sum \underbrace{Q_{ij}}_{\text{inverse mass matrix (small)}} P_i P_j \quad + h(t,\underbrace{\underline{R}}_{\text{slow}},\underbrace{\underline{P}}_{\text{fast}})$$

Gauge approximation to slow motion based on effective Hamiltonian

$$H_g(\underline{R},\underline{P}) = \frac{1}{2}\sum\sum Q_{ij}(P_i - A_i(\underline{R}))(P_j - A_j(\underline{R})) + E_{Bo}(\underline{R}) + \Phi(\underline{R})$$

So 3 reaction forces act

Born–Oppenheimer	$-\partial_i E_{Bo}(\underline{R})$
Magnetic gauge	$B_{ij} = \partial_i A_j(\underline{R}) - \partial_j A_i(\underline{R})$
Electric gauge	$-\partial_i \Phi(\underline{R})$

Quantum formulae for geometric forces: let fast h have eigenstates $|n(\underline{R})\rangle$ and eigenvalues $E_n(\underline{R})$. Adiabatic assumption: no transitions, i.e. fast state remains fixed in one of the $|n\rangle$'s. Then effective slow Hamiltonian is

$$H_\zeta = \langle n|H|n\rangle = \frac{1}{2}\sum\sum Q_{ij}\langle n(\underline{R})| P_i P_j |n(\underline{R})\rangle + E_n(\underline{R})$$

But $P_i = -i\hbar\dfrac{\partial}{\partial R_i}$, so

$$P_j|n\rangle = -i\hbar|\partial_i n\rangle + |n\rangle P_j$$

leading to

$$E_{Bo}(\underline{R}) = E_n(\underline{R})$$

$$B_{ij}(\underline{R}) = i\hbar\left(\langle \partial_i n | \partial_j n\rangle - \langle \partial_j n | \partial_i n\rangle\right)$$

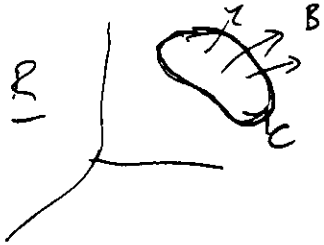
2form : $\langle dn | \wedge | dn \rangle$

$$\oint(\underline{R}) = \sum\sum Q_{ij} g_{ij}$$

$$g_{ij} = \frac{\hbar^2}{2}\langle \partial_i n | (1 - |n\rangle\langle n|) | \partial_j n\rangle$$

NB : B_{ij} and $\oint$ reflect (in the slow system) geometric aspects of the fast motion :

Geometric phase : in a cycle of $\underline{R}$, phase of fast state changes by



$$\gamma_n(C) = i\iint_{\partial S = C} \langle dn | \wedge | dn\rangle$$

Fast metric

$|n(\underline{R})\rangle$ $\quad$ $|n(\underline{R}+d\underline{R})\rangle$

$\underline{R}$ $\quad$ $\underline{R}+d\underline{R}$

$$(\text{distance})^2 \equiv 1 - |\langle n | n+dn\rangle|^2 = g_{ij}\, dR_i\, dR_j$$

(Provost & Vallée)

Remarkably, $\hbar$ drops out of the classical limit of these formulae, leaving geometric reaction forces in purely classical dynamics.

Main steps of classical limit : (with J.M. Robbins)

1) $$\sum_m |m\rangle\langle m| = 1$$

2) $$\langle m | \partial_i n\rangle = \frac{\langle m | \partial_i h | n\rangle}{E_n - E_m}$$ (twice)

3) $$\frac{1}{(E_n - E_m)^2} = -\frac{1}{\hbar^2}\int_0^\infty dt\, t\, e^{\frac{it}{\hbar}(E_n - E_m)}$$

4) quantum commutator $[a,b]$ $\longrightarrow$ $i\hbar\{a(z), b(z)\}$ classical Poisson bracket

5) $\langle n | a | n\rangle \to \langle a(z)\rangle$ average over classical invariant manifold corresponds to $|n\rangle$

$$= \frac{1}{\partial_E \Sigma}\int_{\text{phase volume}} dz\, \delta(E - h(z,R))\, a(z)$$ in ergodic case

These lead to classical formulae that apply also to chaotic fast systems :

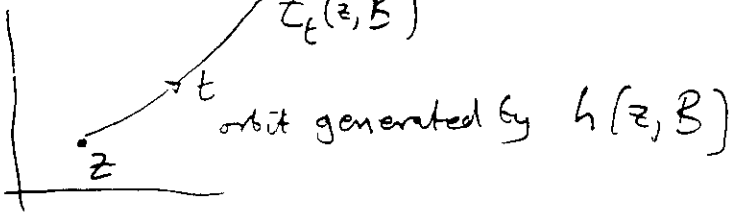
$E_{B_0}(\underline{B}) = E(\underline{B})$, determined by constancy of adiabatic invariant : phase volume

$$\Omega(E,\underline{B}) = \int dz\, \Theta\{E - h(z,\underline{B})\}$$

Θ : unit step

$$B_{ij}(\underline{B}) = -\frac{1}{2\partial_E \Omega}\,\partial_E\left[\partial_E \Omega \int_0^\infty dt \left\langle (\partial_i h)_t\, \partial_j h - (\partial_j h)_t\, \partial_i h \right\rangle_E\right]$$

where $(\partial_i h)_t = \partial_i h\left(Z_t(z,\underline{B},\underline{R}),\underline{B},\underline{R}\right)$



$$g_{ij}(\underline{B}) = -\frac{1}{2}\int_0^\infty dt\, t \left\langle (\partial_i h)_t - \partial_i E)(\partial_j h - \partial_j E)\right\rangle_E$$

fluctuation in $\partial_i h$ at t

These involve correlations

$$C_{ij}(t) = \left\langle (\partial_i h - \partial_i E)_t\,(\partial_j h - \partial_j E)\right\rangle_E$$

Connection with linear response theory :

Can derive geometric forces from classical mechanics, using shift of distribution of fast variables away from microcanonical.

In the chaotic case, get, in addition,

FRICTION – dissipation of slow motion by energy absorption by fast motion. Velocity dependant forces

$$\ddot{R}_i = -k_{ij}\dot{R}_i$$

antisymmetric part = magnetic force

symmetric part is friction

$$\frac{1}{2\partial_E \Omega}\,\partial_E\left[\partial_E \Omega \int_0^\infty dt\,\{C_{ij}(t) + C_{ji}(t)\}\right]$$

N.B. Rate of energy absorption by fast chaos is

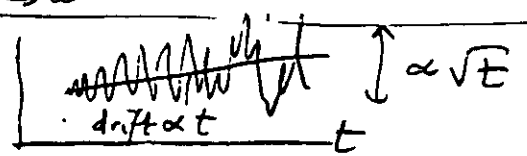
$$\dot{E}_a = \dot{R}_i\, k_{ij}\, \dot{R}_j$$

Drift in adiabatic invariant

$$\dot{\Omega} = \partial_E \Omega\; \dot{R}_i\, k_{ij}\, \dot{R}_j$$

– superimposed on fluctuations

$$(\Delta\Omega)^2_t \underset{t\to\infty}{\Rightarrow} (\partial_E \Omega)^2\, \dot{R}_i \dot{R}_j \left[\frac{t}{2}\int_0^\infty d\tau\, C_{ij}(\tau) - \int_0^\infty d\tau\, \tau\, C_{ij}(\tau)\right]$$

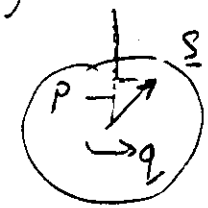


Spin model (with J. M. Robbins)

$$H = \frac{P^2}{2} + \underline{R}\cdot\underline{S}$$

Heavy particle in 3D, coupled with spin $\underline{S}$ via position $\underline{R}$ (sphere of monopolium)

$\underline{S}$ is fast system (1 freedom)



Adiabatic regime

$R = |\underline{R}|$ large (from eliminating mass)

Exact equations of motion (4 freedoms, 8D phase space)

fast: $\dot{\underline{S}} = \underline{R}\wedge\underline{S}$ — $\underline{S}$ precesses about instantaneous $\underline{R}$

slow $\dot{\underline{R}} = \underline{V} = \underline{P}$, $\dot{\underline{V}} = \dot{\underline{P}} = -\underline{S}$ — $\underline{S}$ forces $\underline{R}$

Motion probably nonintegrable, but can eliminate fast motion (unusual!) exactly, using conservation of total angular momentum:

$$\underline{J} = \underline{R}\wedge\underline{V} + \underline{S} = \text{constant}$$

Thus $$\dot{\underline{V}} = \underline{R}\wedge\underline{V} - \underline{J}$$

(measure-preserving but not Hamiltonian $\nabla\cdot\underline{R} \neq 0$)

(charged particle in monopolium, plus 'gravity')

conserved: $E = \frac{1}{2}V^2 + \underline{J}\cdot\underline{R}$ (energy)

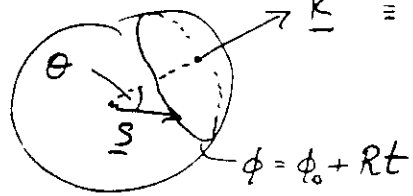
$|S| = |\underline{J} - \underline{R}\wedge\underline{V}|$ (magnitude of acceleration)

Geometric forces Fast motion generated by

$h = \underline{R}\cdot\underline{S}$, i.e. $\dot{\underline{S}} = \underline{R}\wedge\underline{S}$, $\underline{R}$ fixed

need $\nabla_{\underline{R}} h = \underline{S}$

motion integrable and ergodic



Invariant manifold: latitude circle θ

Adiabatic invariant

$$I = \underline{S}\cdot\underline{r} = S\cos\theta \quad \left(= \frac{1}{2\pi}\mathcal{L} = \frac{1}{2\pi}\oint p\,dq\right)$$

Thus $E_{Bo} = IR$

$$\underline{B} = -\frac{1}{2}\partial_E \int_0^\infty dt\, \langle \underline{S}_t \wedge \underline{S}\rangle_E$$

$$\Phi = -\frac{1}{2}\int_0^\infty dt + \langle (\underline{J} - I\underline{r}_1)\cdot(\underline{S} - I\underline{r})\rangle$$

Magnetic	$\underline{B} = -I\frac{\underline{r}}{R^2}$	monopole
electric	$\underline{\Phi} = \frac{S^2 - I^2}{2R^2}$	inverse-cube repulsive force

Thus adiabatic Theory predicts acceleration

$$\dot{\underline{V}} = -I\dot{\underline{r}} - I\frac{\underline{V}\wedge\underline{r}}{R^2} + \frac{(S^2-I^2)\underline{r}}{R^3}$$

Ahrunov Estem

Born-Oppenheimer — magnetic — electric

Study the geometric forces in isolation.

Magnetic only

Eliminate Born-Oppenheimer and electric (both radial) by confining motion to a sphere R = constant. Implement this constraint with

$$P^2 \longrightarrow P^2 - (\underline{P}\cdot\underline{r})^2$$

Then

$$\underline{V} = \underline{P} - \underline{P}\cdot\underline{r}\,\underline{r} \quad \text{i.e.} \quad \underline{V}\cdot\underline{r} = 0$$

Exact equations of motion for slow variables

$$\dot{\underline{V}} = \underline{R}\wedge\underline{V} - \underline{J} + \underline{J}\cdot\underline{r}\,\underline{r} - \frac{V^2}{R}\underline{r}$$

(conserves R, E, $S = |\underline{J} - \underline{R}\wedge\underline{V}|$) seek $\underline{r} = (\theta, \phi)$ about $\underline{J} = J\underline{\hat{z}}$

This motion is integrable:

$$K \equiv E + \frac{1}{2R^2}(J^2 - S^2) = \underline{R}\cdot\underline{J} + \frac{(\underline{r}\wedge\underline{V})\cdot\underline{J}}{R}$$

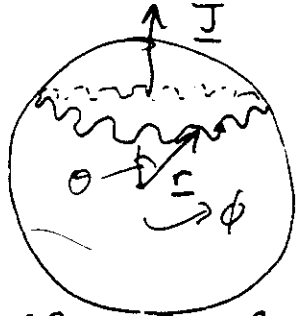
conserved

$$\therefore \quad 2(E-K) + \frac{|\underline{J}\wedge\underline{r}|^2}{R^2} = \left(\underline{V} - \frac{\underline{J}\wedge\underline{r}}{R}\right)^2$$

$$\therefore \quad \frac{2(E-K)}{R^2} + \frac{J^2\sin^2\theta}{R^4} = \dot{\theta}^2 + \sin^2\theta\left(\dot{\phi} - \frac{J}{R^2}\right)^2$$

$$K = J(\sin^2\theta\,\dot{\phi} + R\cos\theta)$$

Motion is precession about $\underline{J}$ ($\dot{\phi}$) with nutation (θ oscillations)



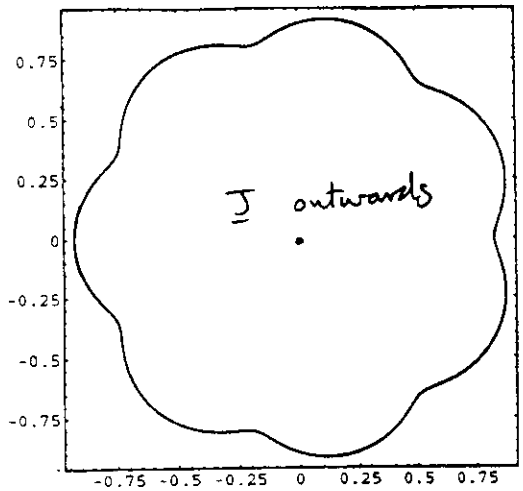
$(I = \underline{S}\cdot\underline{R} = \underline{J}\cdot\underline{R} = JR\cos\theta)$

Adiabatic limit

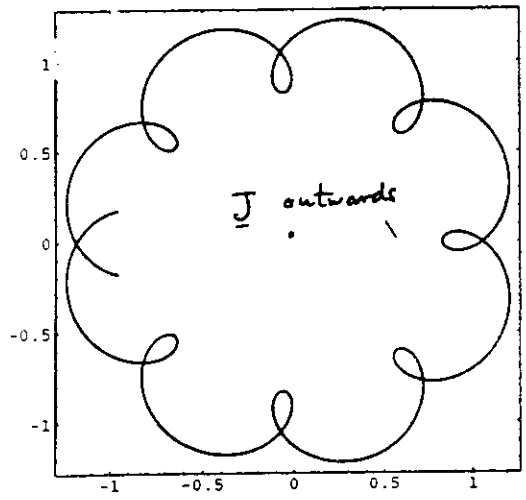
R large: $\dot{\theta} \to 0$ (no nutation)

$$\dot{\phi} = \frac{J}{R^2} \rightarrow \boxed{V = R\sin\theta\dot{\phi} = \frac{I}{R}\tan\theta}$$

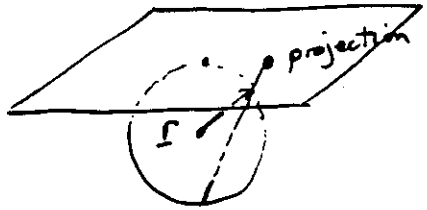
Stereographic projections from sphere of Σ sphere

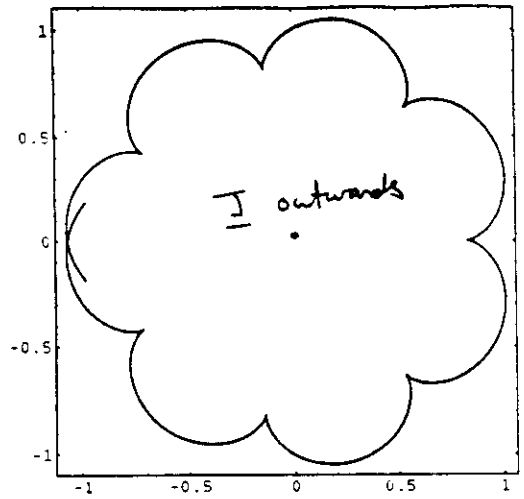


R = 3 J = 3

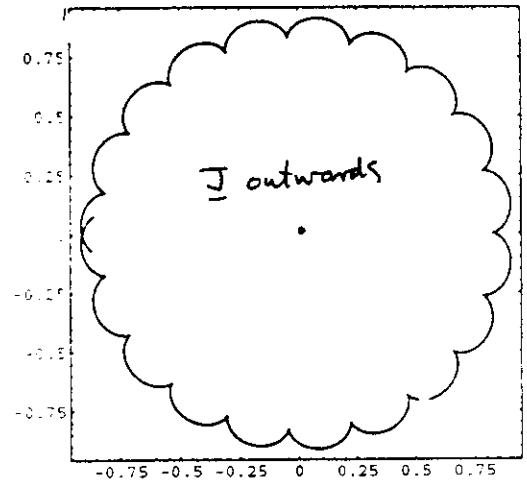


R = 3 J = 3





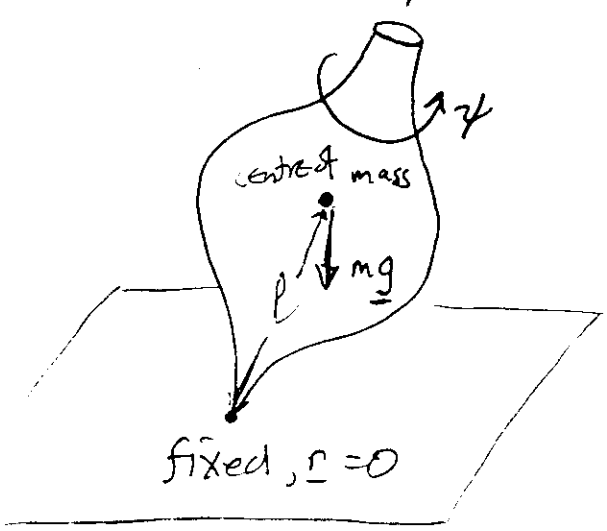
R = 3 J = 3



R = 4 J = 3

Adiabatically, amplitude of nutation diminishes

Analogy with heavy symmetrical (Lagrange) top



Moments of inertia A, B, B

Angular velocity

$$\underline{\Omega} = (\dot{\psi} + \dot{\phi}\cos\theta)\,\underline{r} + \underline{r} \wedge \dot{\underline{r}}$$

momentum

$$\underline{L} = A(\dot{\psi} + \dot{\phi}\cos\theta)\,\underline{r} + B\,\underline{r} \wedge \dot{\underline{r}}$$

$$L_\psi = \underline{L} \cdot \underline{r} = A(\dot{\psi} + \dot{\phi}\cos\theta) \quad \text{conserved}$$

$$L_\phi = \underline{L} \cdot \underline{z} = L_\psi\, \underline{r} \cdot \underline{e}_z + B\, \underline{r} \wedge \dot{\underline{r}} \cdot \underline{e}_z \quad \text{conserved}$$

torque equation $\dot{\underline{L}} = -mg\ell\, \underline{r} \wedge \underline{z}$

gives

$$\ddot{\underline{r}} = \frac{L_\psi}{B}\, \underline{r} \wedge \dot{\underline{r}} - \frac{mg\ell}{B}(\underline{z} - \underline{z}\cdot\underline{r}\,\underline{r}) - |\dot{\underline{r}}|^2 \underline{r}$$

Identical to equations of motion of spin model (constrained to a sphere) if

$$R = \frac{L_\psi}{B} \qquad \underline{J} = \frac{mg\ell L_\psi}{B^2}\underline{z}$$

$\therefore$ Adiabatic regime (R large $\longrightarrow$ L_ψ large = fast top)

$\left(K = \frac{J}{B} L_\phi\right)$

Puzzle: no obvious top analogy for spin

$$\underline{S} = \frac{L_\psi}{B^2}\left[mg\ell\,\underline{z} - L_\psi\, \underline{r} \wedge (\underline{R} \wedge \underline{r})\right]$$

Geometric force approximation:

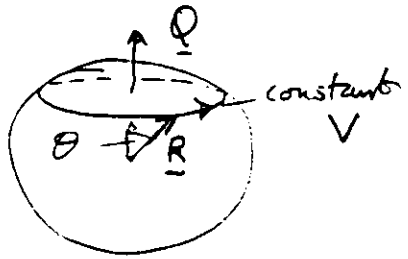
$$\dot{\underline{V}} = -I\,\frac{\underline{V} \wedge \underline{r}}{R^2} - \frac{V^2}{R}\,\underline{r}$$

monopole

conserves R, $E = \frac{1}{2}V^2$ and

$$\underline{Q} = \underline{R} \wedge \underline{V} + I\underline{r}$$

Hence $\underline{r} \cdot \underline{Q}$ constant $\longrightarrow$



Speed given by

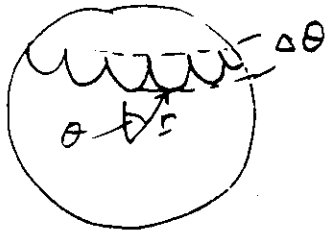
$$\tan\theta = \frac{|\underline{R} \wedge \underline{V}|}{I} \rightarrow \boxed{V = \frac{I\tan\theta}{R}}$$

Thus, magnetic gauge force describes exactly the adiabatic precession.
Born–Oppenheimer fails completely, predicting $\underline{R}$ at rest.

Accuracy of conservation of adiabatic invariant

$I = J\cos\theta$

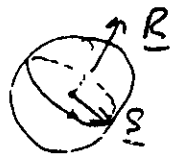
$$\frac{\Delta I}{I} = \frac{\Delta\cos\theta}{\cos\theta} \to \frac{+2J\sin^2\theta_0}{R^3\cos\theta_0} \quad \text{as } R\to\infty$$



So, precession of a top, with the axis (slow) forced by reaction from the spin (fast), is 'caused' by the monopole at the point that is held fixed.

Electric only

Eliminate Born–Oppenheimer and magnetic by choosing $I=0$



In particular, choose

$$\underline{J} = 0 \to I = \underline{S}\cdot\underline{r} = \underline{J}\cdot\underline{r} = 0 \text{ always}$$

Exact motion is

$$\dot{\underline{V}} = \underline{R}\wedge\underline{V}$$

('spin' $\underline{V}$, driven by $\underline{R}$, and coupled by $\dot{\underline{R}}=\underline{V}$, or charged particle in uniform sphere of monopolium)

This can be solved exactly.

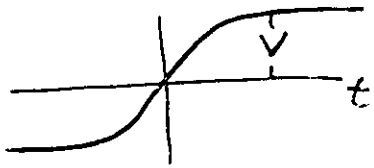
Speed $V = \text{constant}$.

$$\tfrac{1}{2}(\ddot{R^2}) = (\underline{R}\cdot\underline{V}) = V^2 \to R^2(t) = R_0^2 + V^2t^2$$

radial repulsion.

'Adiabatic invariant' for 'spin' $\underline{V}$ is not constant:

$$\underline{r}\cdot\underline{V} = \frac{\underline{R}\cdot\underline{V}}{R} = \frac{Vt}{\sqrt{t^2+R_0^2/V^2}}$$



So, asymptotically, orbits recede along radii

Can set $V = 1$, and parameterize orbit shapes just with R_0

Count of variables: 6 quantities to describe an orbit

eliminate: 1 by choosing $t=0$ at closest approach

3 by rigid rotation of orbit about $\underline{R} = 0$

1 by scaling: $\alpha \underline{R}(\alpha t)$ is an orbit if $\underline{R}(t)$ is

5 $\longrightarrow$ 1 left $= R_0$

Thus time t = arc length along orbit, so

$\underline{V}(t)$ is unit tangent. Orbit as space curve:

curvature

$$\boxed{\kappa = \sqrt{\dot{\underline{V}} \cdot \dot{\underline{V}}} = \text{constant} = R_0}$$

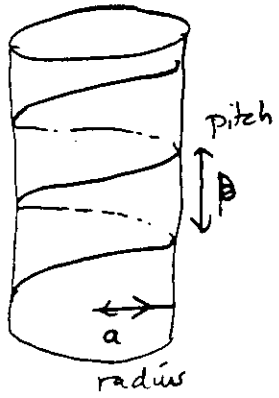
Normal and binormal

$$\underline{n} = \frac{\underline{R} \wedge \underline{V}'}{R_0} \qquad \underline{b} = \frac{\underline{R} - t\underline{V}}{R_0}$$

torsion $\boxed{\tau = -\dot{\underline{b}} \cdot \underline{n} = t}$

What curves have constant curvature R_0 and changing torsion t ?

Preliminary: uniform helix

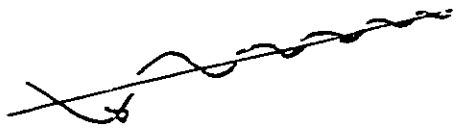


$$\kappa = \frac{4\pi^2 a}{p^2 + 4\pi^2 a^2} \qquad \tau = \frac{2\pi p}{p^2 + 4\pi^2 a^2}$$

$$\therefore \quad a = \frac{R_0}{R_0^2 + t^2} \longrightarrow \frac{R_0}{t^2}$$

$$p = \frac{2\pi t}{R_0^2 + t^2} \longrightarrow \frac{2\pi}{t}$$

So, orbits coil round asymptotes with shrinking radius and pitch



Winding reverses at $t=0$ (zero torsion), so predict **curly antelope horns** for $\underline{R}(t)$

and for $\underline{V}$ on its unit sphere predict a curve whose speed is the curvature R_0 of $\underline{R}(t)$ and whose curvature is the torsion t of $\underline{R}(t)$

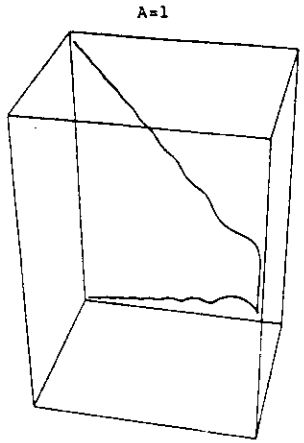
i.e. **an asymptotically infinitely coiling S**

Antelope horns: numerical solution of $\ddot{\underline{R}} = \underline{R}_n \ddot{\underline{R}}$

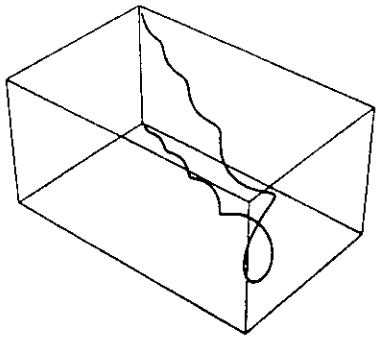
$V = |\dot{\underline{R}}| = 1$

closest approach $R = R_0$

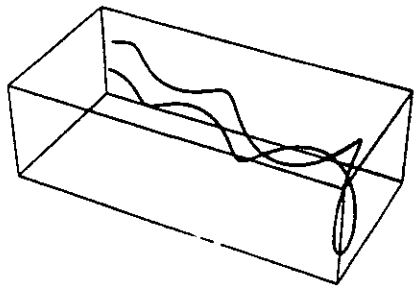
$R_0 = 1$



$R_0 = 1.5$



$R_0 = 2$

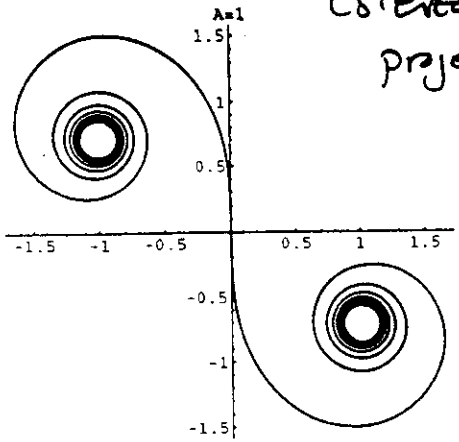


Horns close up as R_0 increases

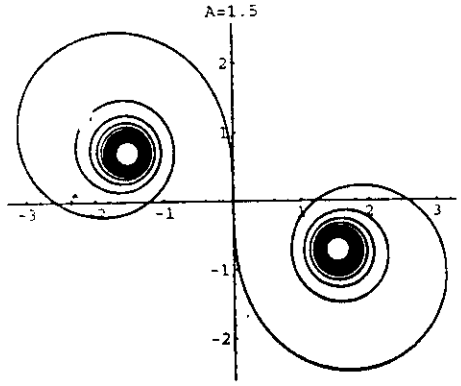
seek opening angle ϵ / ρ

3.23 Track of velocity $\underline{V}$ on unit sphere (Stereographic Spike projection)

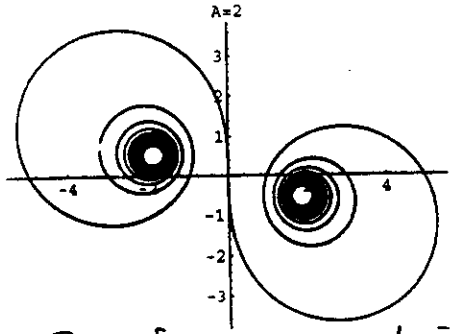
$R_0 = 1$



$R_0 = 1.5$



$R_0 = 2$

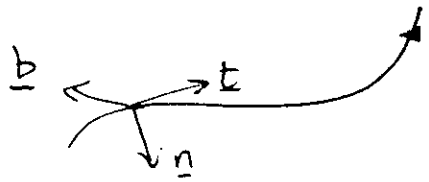


As R_0 increases, velocity asymptotes become antipodal

Intermezzo Constructing the curve

with given curvature $\kappa(s)$, torsion $\tau(s)$ functions of arc length s

Frenet equations



$$\underline{t}' = \kappa \underline{n}$$
$$\underline{n}' = -\kappa \underline{t} + \tau \underline{b}$$
$$\underline{b}' = -\tau \underline{n}$$

9 linear equations. Reduce to 2 by mapping onto evolution of a quantum spin-$\frac{1}{2}$ particle in a 'magnetic field'

$$\underline{B}(s) = (\kappa(s), 0, -\tau(s))$$

Schrödinger equation

$$i|\psi\rangle' = \underline{\sigma}\cdot\underline{B}(s)|\psi\rangle \qquad |\psi\rangle = \begin{pmatrix}\psi_+\\ \psi_-\end{pmatrix}$$

$$\underline{\sigma} = \tfrac{1}{2}\left[\begin{pmatrix}0&1\\1&0\end{pmatrix}_x, \begin{pmatrix}0&-i\\i&0\end{pmatrix}_y, \begin{pmatrix}1&0\\0&-1\end{pmatrix}_z\right] = \text{Pauli matrices}$$

take any 3 solutions $|\psi_x\rangle, |\psi_y\rangle, |\psi_z\rangle$

with $\langle\psi_i|\underline{\sigma}|\psi_i\rangle$ orthogonal at $t=0$.

Then

$$\underline{t} = 2\left[\langle\psi_x|\underline{\sigma}|\psi_x\rangle, \langle\psi_y|\underline{\sigma}|\psi_y\rangle, \langle\psi_z|\underline{\sigma}|\psi_z\rangle\right]$$

given $\underline{t}(s)$, the curve is $\underline{R}(s) = \int_0^s dt\, \underline{t}(s)$

In the spin model $\dot{\underline{V}} = \underline{R}\wedge\underline{V}$

The 'quantum Hamiltonian' is

$$\underline{\sigma}\cdot\underline{B}(t) = \tfrac{1}{2}\begin{pmatrix}-t & R_0\\ R_0 & t\end{pmatrix}$$

defining the Landau-Zener problem of quantum transit theory

Thus the velocity (unit tangent) is

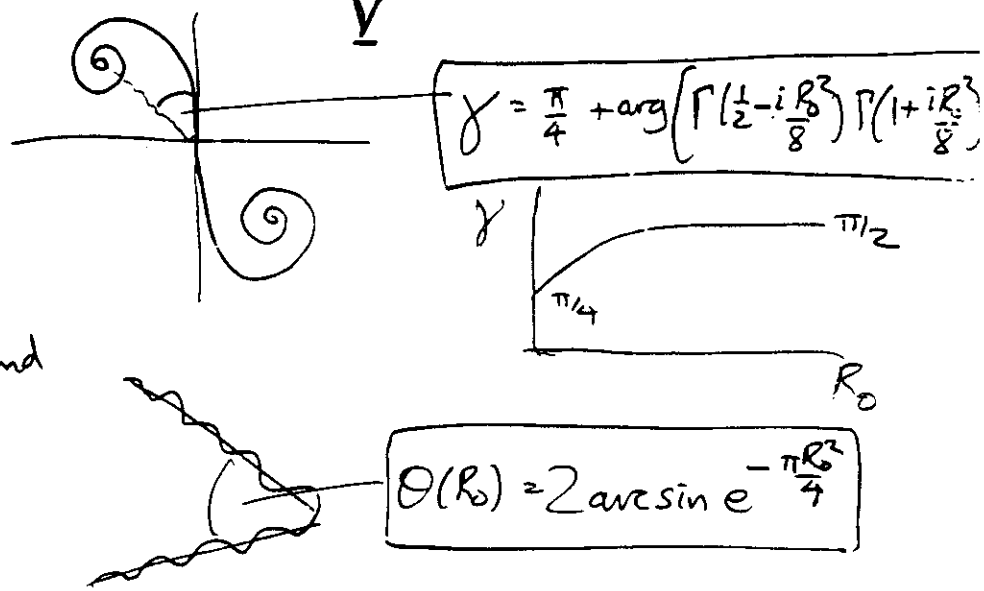
$$\underline{V}(t) = \left\{R_0 \operatorname{Im} y_1 y_2^*,\ R_0 \operatorname{Re} y_1 y_2^*,\ |y_1|^2 - \tfrac{1}{4}R_0^2|y_2|^2\right\}$$

where y_1, y_2 are the even and odd parabolic cylinder (Kummer, confluent hypergeometric) functions

$$y_1(t) = e^{-\frac{it^2}{4}} M\left(\tfrac{1}{8}iR_0^2, \tfrac{1}{2}, \tfrac{1}{2}it^2\right)$$

$$y_2(t) = t\, e^{-\frac{it^2}{4}} M\left(\tfrac{1}{8}iR_0^2 + \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{1}{2}it^2\right)$$

Asymptotics of these functions gives



and

Geometric force approximation

For large R_0, particle spirals in and back out along same direction, so 'guiding centre' motion is purely radial.

Gauge theory (electric force only) $I = 0$ so

$$\dot{\underline{V}} = \frac{S^2}{R^3}\underline{r} \quad , \quad S = |\underline{R}(0) \wedge \underline{V}(0)| = R_0$$

solution $\underline{R}(t) = \sqrt{R_0^2 + t^2}\ \underline{x}$

So electric gauge force describes exactly the repulsion of the guiding centre

Born-Oppenheimer fails completely

Conclusions

1. Geometric reaction forces of electric and magnetic type give a clear picture of average slow motion beyond Born-Oppenheimer and can capture the main features of this motion with little calculation.

2. New general formulae, giving B_{ij} and $\underline{\Phi}$ in terms of fast orbits with fixed $\underline{R}$, are easy to use and apply also to chaotic fast-motion.

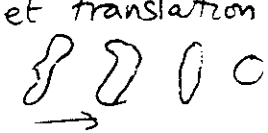
3. The particular model $\frac{P^2}{2} + \underline{S}\cdot\underline{R}$ much richer than the two special cases described, and includes chaos in exact slow motion, but geometric forces still produce distinctive effects (e.g. swerving out of plane) observed in numerical computations.

4. For chaotic fast motion, the magnetic force is the neglected antisymmetric cousin of friction.

MORE CLASSICAL ANHOLONOMY

SWIMMING BUGS (Shapere and Wilczek)

Swimming = cyclic alteration of a body's shape producing net translation through a fluid.



The shape variables return; the position variables do not. Therefore swimming is anholonomy.

For small creatures (insects, bacteria) the translation vector depends only on the sequence of body shapes, and not on how fast the movements are made, because the fluid reaction is dominated by viscous (rather than inertial) forces. So reciprocating motions get nowhere.

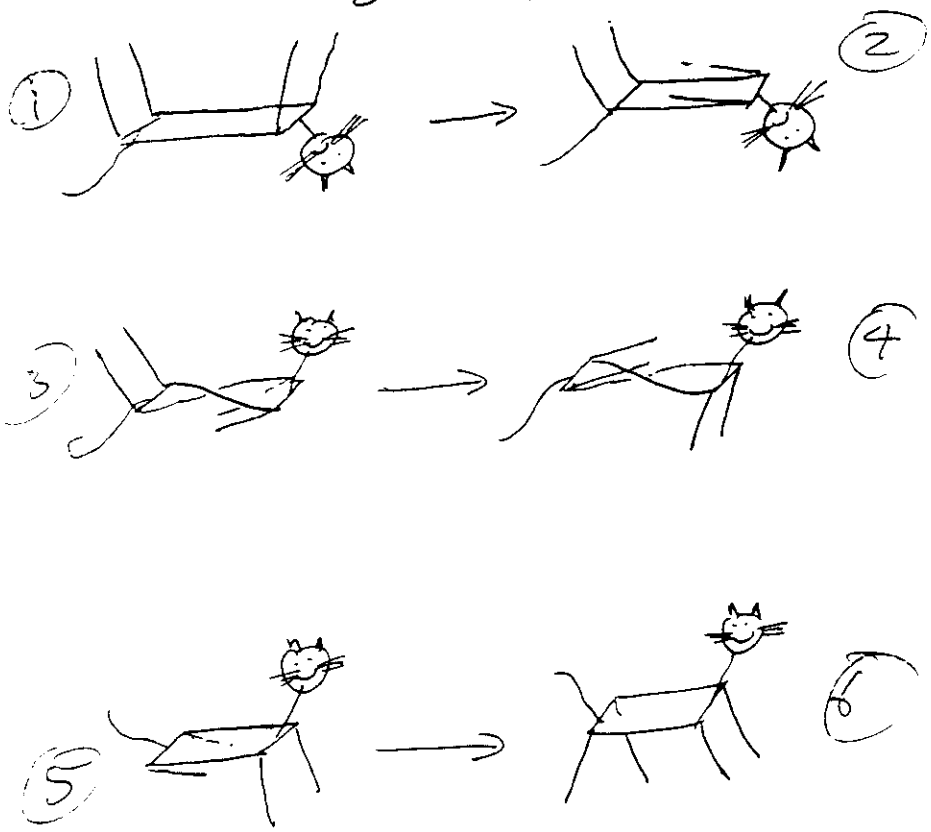
The body makes a circuit C in the abstract space of shapes. Wilczek et al show:

anholonomy (= translation vector)

= flux of something (complicated) through C.

CATS IN SPACE (Shapere and Wilczek)

With nothing to push against, the cat's angular momentum remains zero. Nevertheless it can rotate, by changing its shape:



The shape variables return, but the position variables do not, so the cat's turn is anholonomy.

A formula for the net rotation of the cat:

R_{ij} = rotation matrix turning initial into final object (with the same shape)

$$= P \exp\left[\oint A^{\mu}_{ij}(S)\, dS_{\mu} \right]$$

- P: path ordered product
- $\oint$: cycle in space of shapes S
- A^{μ}_{ij}: Infinitesimal rotation from infinitesimal μ deformation

This is geometry. For A, need dynamics, i.e. angular momentum zero:

$$A^{\mu}_{ij} \frac{dS_{\mu}}{dt} = -\varepsilon_{ijk} \left(I - 1 \mathrm{Tr} I \right)^{-1}_{kl} L_l$$

- $(I - 1 \mathrm{Tr} I)^{-1}_{kl}$: traceless inertia tensor of body in shape S in standard orientation
- L_l: apparent angular momentum of deforming body when reduced to standard orientation

A^{μ}_{ij} is a vector in shape space (with axes μ).
It is the (matrix-valued) potential of the gauge field

$$F^{\mu\nu} = \frac{\partial A^{\mu}}{\partial S_{\nu}} - \frac{\partial A^{\nu}}{\partial S_{\mu}} + [A^{\mu}, A^{\nu}]$$

Classical geometric forces of reaction: an exactly solvable model

M. V. Berry and J. M. Robbins

H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, U.K.

Abstract

We illustrate the effects of the classical 'magnetic' and 'electric' geometric forces that enter into the adiabatic description of the slow motion of a heavy system coupled to a light one, beyond the Born-Oppenheimer approximation of simple averaging. When the fast system is a spin S and the slow system is a massive particle whose spatial position $\boldsymbol{R}$ is coupled to S with energy (fast Hamiltonian) $S\cdot\boldsymbol{R}$, the magnetic force is that of a monopole of strength I (=adiabatic invariant $S\cdot\boldsymbol{R}/R$) centred at R=0, and the electric force is inverse-cube repulsion with strength S^2-I^2. Confining the slow particle to the surface of a sphere eliminates the Born-Oppenheimer and electric forces, and generates motion with precession and nutation exactly equivalent to that of a heavy symmetrical top. In the adiabatic limit the nutation is small and the averaged precession is precisely reproduced by the magnetic force. Alternatively, choosing the exactly conserved total angular momentum to vanish eliminates the Born-Oppenheimer and magnetic forces, and generates as exact orbits a one-parameter family of curly 'antelope horns' coiling in from infinity, reversing hand, and receding to infinity. In the adiabatic limit the repulsion of the 'guiding centre' of these coils is exactly reproduced by the electric force.

A by-product of the 'antelope horn' analysis is a determination of the shape of a curve with given curvature κ and torsion τ in terms of the evolution of a quantum 2-spinor driven by a planar 'magnetic field' with components κ and τ.

Short title: Solvable model of classical geometric forces

Submitted to: *Proc. Roy. Soc. Lond.*, December 1992

1. Introduction

When a light system, whose motion is fast, is coupled to a heavy system, that moves slowly, the dynamics can get quite complicated and hard to analyze. A useful approximation is to solve the fast motion for frozen values of the slow coordinates, and then consider the slow dynamics to be influenced by the average energy of the fast motion. This energy depends on the slow coordinates and so its gradient acts as a reaction force on the slow motion. In quantum mechanics the approximation technique is known as the Born-Oppenheimer method (Messiah 1962) and is commonly used to analyze molecules (where the light and heavy systems are the electrons and nuclei respectively). In classical mechanics it is the method of adiabatic averaging (Arnold et al. 1988, Lochak and Meunier 1988). We shall call this reaction force the Born-Oppenheimer force.

In recent years it has become clear (Mead and Truhlar 1979, Jackiw 1988, Berry 1989) that in quantum mechanics an improvement of the approximation, more consistent with the adiabatic assumption for the fast motion, should include in the slow dynamics two extra reaction forces, which have a gauge structure resembling that of electromagnetism, in addition to the Born-Oppenheimer force. These 'magnetic' and 'electric' reaction forces depend on the geometry of the fast dynamics in the space of frozen slow coordinates. They are related to the geometric phase (Shapere and Wilczek 1989). In the quantum mechanics of molecules, one effect of the magnetic gauge force (associated with degeneracies in the frozen electronic states) is to alter the energy levels of the nuclear motion (Longuet-Higgins et. al. 1959, Delacrétaz et. al. 1986).

Our purpose here is to study the analogous *classical* geometric forces in a simple Hamiltonian model where they have real and clearly identifiable effects. This is part of a larger programme in which we are exploring the origin and structure of the classical geometric forces. It is not our intention to give the general theory here, but it will be convenient to present some new formulae for the forces. The novelty is twofold. First, the formulae apply when the fast motion is ergodic, and so extend previous related studies (e.g. Hannay 1985, Berry 1985), which have been confined to integrable systems, to cover the case of chaotic motion,

thereby resolving a long-standing problem. (In the present example the fast motion is one-dimensional, and so the integrable and ergodic categories coincide.) Second, the forces are expressed as integrals along trajectories, and in the integrable case are manifestly independent of action-angle variables, making them easy to interpret and convenient to calculate.

In the model, introduced in §2, the fast system is a spin, and the slow system to whose coordinates it is coupled is a massive particle moving in three dimensions. The fast motion (with frozen slow coordinates as parameters) is integrable. An interesting feature is that the slow motion can be exactly decoupled from the fast motion - that is, independently of the adiabatic approximation - and the resulting equation is not Hamiltonian.

In §3 the gauge forces are calculated; the new formulae enable a speedy derivation, although the results are not new (see e.g. Berry 1989). The magnetic force is that of a monopole and the electric force an inverse-cube repulsion. These alter the slow motion in qualitatively different ways

§4 treats the special case where only the magnetic force acts. This leads to a surprising connection with the motion of an ordinary spinning-top, whose precession can be considered as caused by a magnetic gauge monopole describing the reaction of the (fast) spin on the (slow) dynamics of the symmetry axis, averaged over nutation. In a different special case, considered in §5, only the electric force acts; the repulsion describes precisely the slow motion after oscillations are averaged away. In an interesting technicality (appendix D) the determination of a curve whose curvature and torsion are given functions of arc length is shown to be equivalent to the 'Landau-Zener' evolution of a two-state quantum system, which in certain cases, including that of the orbit in §5, enables the exact analytic solution to be found.

Now we present the general formulae. Let the slow system have coordinates $\boldsymbol{R}=\{R_1, R_2...\}$ and momenta $\boldsymbol{P}=\{P_1, P_2,...\}$ and let the fast system have phase space variables $z=\{q,p\}=\{q_1,..p_1...\}$. For the Hamiltonian we take

$$H(\boldsymbol{R},\boldsymbol{P},z)=\tfrac{1}{2}\sum\sum Q_{ij}P_iP_j+h(z,\boldsymbol{R}) \qquad (1)$$

Here Q is an inverse mass matrix which is small in the adiabatic regime, and h is the fast Hamiltonian coupling the fast variables to the slow coordinates. This can describe a classical or a quantum system, depending on whether $\boldsymbol{R}$ and $\boldsymbol{P}$, and $\boldsymbol{q}$ and $\boldsymbol{p}$, commute. In the improved adiabatic approximation, the slow variables are governed by the effective Hamiltonian

$$H_g(\boldsymbol{R},\boldsymbol{P})=\tfrac{1}{2}\sum\sum Q_{ij}\big(P_i-A_i(\boldsymbol{R})\big)\big(P_j-A_j(\boldsymbol{R})\big)+E_{\mathrm{BO}}(\boldsymbol{R})+\Phi(\boldsymbol{R}) \qquad (2)$$

Corresponding to this are the three reaction forces

$$\begin{aligned}&\text{Born - Oppenheimer:} && -\partial_i E_{\mathrm{BO}}(\boldsymbol{R})\\ &\text{magnetic gauge:} && B_{ij}\equiv\partial_i A_j(\boldsymbol{R})-\partial_j A_i(\boldsymbol{R})\\ &\text{electric gauge:} && -\partial_i\Phi(\boldsymbol{R})\\ &\text{where} && \partial_i\equiv\frac{\partial}{\partial R_i}\end{aligned} \qquad (3)$$

In quantum mechanics, these forces depend on the eigenstate $|n(\boldsymbol{R})\rangle$ and energy $E_n(\boldsymbol{R})$ of the fast Hamiltonian h, corresponding to the adiabatically preserved state of the fast system. Berry (1989) obtained

$$\begin{aligned}E_{\mathrm{BO}}(\boldsymbol{R})&=E_n(\boldsymbol{R})\\ B_{ij}(\boldsymbol{R})&=i\hbar\Big[\big\langle\partial_i n\big|\partial_j n\big\rangle-\big\langle\partial_j n\big|\partial_i n\big\rangle\Big]\\ \Phi(\boldsymbol{R})&=\sum\sum Q_{ij}g_{ij}(\boldsymbol{R})\\ \text{where}\quad g_{ij}&=\frac{\hbar^2}{2}\Big\langle\partial_i n\Big|\big(1-|n\rangle\langle n|\big)\Big|\partial_j n\Big\rangle\end{aligned} \qquad (4)$$

Appendix A outlines a slightly different way of getting this well-known result. We remark that the quantities B_{ij} and g_{ij}, which here influence the slow motion, also have significance in the fast motion: B_{ij} is the 2-form whose flux through an $\boldsymbol{R}$ circuit gives the geometric phase, and g_{ij} is a metric governing the separation of quantum states in $\boldsymbol{R}$ space (Provost and Vallée 1989, Berry 1989).

It is remarkable that the quantum geometric forces have finite classical limits. We display the new classical formulae for the case where the fast motion is ergodic (this includes the example to follow, which is one-dimensional). The phase volume

$$\Omega(E,R)=\int \mathrm{d}z\,\Theta\{E-h(z,R)\} \tag{5}$$

where Θ denotes the unit step, is an adiabatic invariant (in one dimension it is the action times 2π). The fast energy $E_{\mathrm{BO}}(\boldsymbol{R})$ is determined by taking $\Omega(E_{\mathrm{BO}},\boldsymbol{R})$ as constant.

Let $Z_t(z,\boldsymbol{R})$ denote the orbit which starts at z at t=0 and evolves under h with parameters $\boldsymbol{R}$, and let

$$\left(\partial_i h\right)_t \equiv \partial_i h(Z_t,R) \tag{6}$$

denote the time-evolved parameter derivative of the fast Hamiltonian. Denote averaging over the energy surface in phase space by

$$\langle\cdots\rangle_E \equiv \frac{1}{\partial_E\Omega}\int \mathrm{d}z\,\delta(E-h(z,R))\cdots \tag{7}$$

Then the geometric forces are determined by

$$B_{ij}(R) = -\frac{1}{2\partial_E\Omega}\partial_E\left[\partial_E\Omega\int_0^\infty \mathrm{d}t\left\langle\left(\partial_i h\right)_t\partial_j h-\left(\partial_j h\right)_t\partial_i h\right\rangle_E\right]$$

$$g_{ij}(R) = -\tfrac{1}{2}\int_0^\infty \mathrm{d}t\,t\left\langle\left(\tilde{\partial}_i h\right)_t\left(\tilde{\partial}_j h\right)\right\rangle_E \tag{8}$$

where $\tilde{\partial}_i h=\partial_i h-\langle\partial_i h\rangle=\partial_i h-\partial_i E_{\mathrm{BO}}$

(The last equality is the classical analogue of the quantum Hellman-Feynman formula.)

The magnetic force B_{ij} has been derived by Robbins and Berry (1992) as the classical limit of the quantum expression in (4). The formula for the electric potential was found in 1988 by Berry and Wilkinson (unpublished) in the same way. In appendix B we outline this method. In one dimension the formula for B_{ij} is, despite appearances, equivalent to the classical 2-form discovered by Hannay (1985). Deriving and interpreting these formulae from classical mechanics in the multidimensional ergodic case involves subtle questions which we will discuss in later papers.

In a systematic study of the origin of quantum gauge forces, Weigert and Littlejohn (1992) have shown that an additional (LW) term arises in the effective Hamiltonian, which typically is of the same order as the electric geometric scalar potential. They have also considered its classical limit for spin problems (private communication) of which the one considered here is a special case. In §3 we show that in the particular case we study the LW force is smaller than the other forces, and we will not consider it further.

2. Spin model

The Hamiltonian (1) is

$$H=\tfrac{1}{2}P^2+R\cdot S \tag{9}$$

Here the heavy system is a particle with classical coordinates $\boldsymbol{R}=\{R_1,R_2,R_3\}$ and momenta $\boldsymbol{P}=\{P_1,P_2,P_3\}$. The light system $\boldsymbol{S}$ is a classical spin whose length is fixed and whose direction can be regarded as a Hamiltonian system with one freedom (the momentum is the projection of $\boldsymbol{S}$ onto any fixed axis, and the coordinate is the azimuth angle about that axis). The equations of motion are

$$\dot{\boldsymbol{R}}\equiv\boldsymbol{V}=\boldsymbol{P};\quad \dot{\boldsymbol{P}}=-\boldsymbol{S};\quad \dot{\boldsymbol{S}}=\boldsymbol{R}\wedge\boldsymbol{S} \tag{10}$$

Thus H has four freedoms: three slow and one fast. We shall denote the length of $\boldsymbol{R}$ by R, and its direction by the unit vector $\boldsymbol{r}$, i.e. $\boldsymbol{R}\equiv R\boldsymbol{r}$. $\boldsymbol{S}$ is slaved to the particle position $\boldsymbol{R}$ and precesses about its instantaneous direction with angular velocity R, and $\boldsymbol{S}$ influences $\boldsymbol{R}$ by providing the force on the particle.

The adiabatic regime corresponds to R large, so that the spin precession is fast and $\boldsymbol{S}$ turns many times while the direction $\boldsymbol{r}$ remains approximately constant. It is not necessary to introduce an explicit adiabatic parameter (for example a large mass) into the Hamiltonian (9), because in this case it can be removed by scaling - something not possible in general.

Classical or quantal aspects of the model (9) have been studied by many authors, for example Stone (1986), Gozzi and Thacker (1987),

Anandan and Aharonov (1989), Berry (1989), Aharonov and Stern (1992), Bulgac and Kusnezov (1992), and Littlejohn (private communication 1992). There are more general variants, in which for example $\boldsymbol{R}$ is replaced by any vector function of $\boldsymbol{R}$, but (9) suffices for the simple point we wish to make, namely that the gauge forces have real effects. It is possible to interpret (9) as describing the motion of a heavy uncharged spinning particle, with a magnetic moment, through a uniform sphere of monopolium (whose magnetic field is proportional to $\boldsymbol{R}$). In an alternative interpretation, the 'particle' could be a thin spinning molecule with an electric dipole moment along its axis, moving through a uniform sphere of charge.

The coupled motion (10) is apparently nonintegrable for most initial conditions. Nevertheless, it is a remarkable feature of this model that the fast motion can be eliminated without approximation (that is, independently of any adiabatic assumption), using the fact that because of rotational invariance (10) conserves the total angular momentum (orbital plus spin), namely

$$\boldsymbol{J} \equiv \boldsymbol{R} \wedge \boldsymbol{V} + \boldsymbol{S} = \text{constant} \tag{11}$$

Note that $\boldsymbol{J}\cdot\boldsymbol{R} = \boldsymbol{S}\cdot\boldsymbol{R}$. Thus (10) gives the acceleration of the slow variables as

$$\dot{\boldsymbol{V}} = \boldsymbol{R} \wedge \boldsymbol{V} - \boldsymbol{J} \tag{12}$$

Now the slow motion corresponds to that of a *charged* particle influenced by the Lorentz force inside a uniform sphere of monopolium, and also by a constant 'gravitational' force $-\boldsymbol{J}$. This Newtonian equation is measure-preserving in the phase space $\boldsymbol{R}$, $\boldsymbol{V}$. It is however not Hamiltonian, because the 'magnetic field' $\boldsymbol{R}$ has nonvanishing divergence and hence no vector potential exists. (It is possible to obtain (12) as the equation of motion generated by an energy and a 2-form, but the 2-form is not closed.) Two quantities conserved by (12) are the energy and length S of spin, namely

$$E = \tfrac{1}{2}V^2 + \boldsymbol{J}\cdot\boldsymbol{R}; \quad S = |\boldsymbol{J} - \boldsymbol{R} \wedge \boldsymbol{V}| \tag{13}$$

Note that conservation of S implies constancy of the magnitude of the acceleration in (12).

In §§4 and 5 we shall solve (3) exactly in special cases, in order to assess the effects of the gauge forces whose derivation follows now.

3. Geometric forces

In this model the slow coordinates form a vector in space, so that in (8) it is convenient to replace derivatives ∂_i by gradients ∇, and regard the magnetic gauge field as a vector $\boldsymbol{B}$. The gauge forces depend on the fast motion generated by

$$h = \boldsymbol{R}\cdot\boldsymbol{S}, \qquad \text{i.e.} \quad \dot{\boldsymbol{S}} = \boldsymbol{R} \wedge \boldsymbol{S} \tag{14}$$

with fixed $\boldsymbol{R}$, via the gradient $\nabla h = \boldsymbol{S}$. This describes precession about $\boldsymbol{r}$, which we take temporarily as the z axis, with angular velocity R. If the initial spin has polar coordinates θ, ϕ, its evolution is

$$\boldsymbol{S}_t = \big(S\sin\theta\cos(\phi + Rt),\ S\sin\theta\sin(\phi + Rt),\ S\cos\theta\big) \tag{15}$$

The constant-energy surface for this one-dimensional integrable and ergodic motion is a latitude circle on the spherical phase space S=constant, whose measure, appearing in (8), is

$$\partial_E \Omega = S\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \sin\theta\, \delta(E - SR\cos\theta) = \frac{2\pi}{R} \tag{16}$$

This is independent of E and so cancels from the magnetic field in (8). There is an adiabatic invariant (conserved under infinitely slow changes in $\boldsymbol{R}$), given by the component of spin along $\boldsymbol{R}$:

$$I = \boldsymbol{S}\cdot\boldsymbol{r} = S\cos\theta \quad \left(= \tfrac{1}{2\pi}\Omega = \tfrac{1}{2\pi}\oint p\mathrm{d}q\right) \tag{17}$$

Thus the Born-Oppenheimer, magnetic gauge and electric gauge forces are obtained from

$$E_{BO}(R) = IR$$

$$B(R) = -\tfrac{1}{2}\partial_E \int_0^\infty dt \langle S_t \wedge S \rangle_E \qquad (18)$$

$$\Phi(R) = -\tfrac{1}{2}\int_0^\infty dt t \langle (S_t - Ir)\cdot(S - Ir)\rangle_E$$

The averages $\langle ....\rangle_E$ are integrals round the latitude circle with $\cos\theta = I/S$, and a direct calculation using the orbit (15) gives the known results

$$B(R) = -I\frac{r}{R^2}$$

$$\Phi(R) = \frac{S^2 - I^2}{2R^2} \qquad (19)$$

As previously asserted, the magnetic force is that of a monopole at $R=0$, with strength $-I$, and the electric force is an inverse-cube radial repulsion with strength S^2-I^2. From (18) the Born-Oppenheimer force is radial and constant, and attractive or repulsive if $I>0$ or $I<0$. This calculation, based on the new formulae (8), is much simpler than the analogous calculation (Berry 1986) of B from the earlier formalism involving the shift of invariant tori with R.

Adiabatic theory therefore predicts that the slow motion is determined by the equation of motion generated by the effective Hamiltonian (2), namely

$$\dot{V} = -Ir - I\frac{V \wedge r}{R^2} + \frac{\left(S^2 - I^2\right)}{R^3} r \qquad (20)$$

in which the acceleration is the sum of Born-Oppenheimer, magnetic and electric terms. This looks very different from the exact equation (12) for the slow motion, and yet we will find that it gives very good approximations to the average of the slow motion. Aharonov and Stern (1992) have derived the geometric forces in (20) by physical arguments involving careful averaging.

In this example the additional term discovered by Weigert and Littlejohn (1992) in the effective slow Hamiltonian is

$$H_{LW} = \frac{I|r \wedge P|^2}{2R^3} \qquad (21)$$

To estimate its importance, consider its effect when acting in conjunction with the Born-Oppenheimer and geometric electric forces, that is in the Hamiltonian

$$H_{\text{effective}} = \tfrac{1}{2}P^2 + IR + \frac{S^2 - I^2}{2R^2} + H_{LW} \qquad (22)$$

Conserved quantities are the angular momentum and energy:

$$L = R \wedge P = \frac{R \wedge V}{1 + I/R^3}$$

$$2E = \dot{R}^2 + 2IR + \frac{S^2 - I^2}{R^2} + \frac{L^2}{R^2} + I\frac{L^2}{R^5} \qquad (23)$$

Thus motion is planar, with the LW force contributing an inverse sixth-power repulsive force whose strength depends on L. Since L is proportional to R, this is smaller in the adiabatic regime of large R than the inverse-cube repulsion (inverse-square potential) of the electric gauge force, and from now on we neglect it. (The essential reason for the smallness of the LW force is that the Born-Oppenheimer force is radial.)

4. Magnetic force only

4.1. Exact solution

Both the Born-Oppenheimer and electric forces in (20) are radial. To eliminate their effects, and thereby study the case where the only adiabatic effective force is magnetic, we constrain the particle to the surface of a sphere. This can be achieved by replacing the kinetic energy in (9) by

$$P^2 \to P^2 - (P \cdot r)^2 \qquad (24)$$

because then Hamilton's equation gives the velocity as

$$V = P - P \cdot rr, \quad \text{i.e.} \quad V \cdot r = 0 \qquad (25)$$

The total angular momentum (11) is still conserved, and the fast variables S can be exactly eliminated, to give the slow acceleration as (12) together with forces incorporating the constraint:

$$\dot{V} = R \wedge V - J + J \cdot r r - \frac{V^2}{R} r \tag{26}$$

In addition to R, this conserves E and S given by (13), and hence the combination

$$K = E + \frac{1}{2R^2}\left(J^2 - S^2\right) = R \cdot J + \frac{(r \wedge V) \cdot J}{R} \tag{27}$$

To determine the motion of R on its sphere, we first introduce polar coordinates (θ,ϕ) for r, with J as axis, and thus obtain

$$\begin{aligned} E &= \tfrac{1}{2} R^2 \left[\dot{\theta}^2 + \sin^2\theta\, \dot{\phi}^2\right] + JR\cos\theta \\ K &= J\left(\sin^2\theta\, \dot{\phi} + R\cos\theta\right) \end{aligned} \tag{28}$$

Choosing the time origin when $\dot{\theta}=0$, with $\theta=\theta_0$ and azimuthal speed $\dot{\phi}_0$, we can eliminate E and K:

$$\tfrac{1}{2} R\dot{\theta}^2 = (c_0 - c)\left\{-\frac{R^3}{2s^2}(c_0 - c) + \frac{R^2 s_0^2}{s^2}\dot{\phi}_0 + \frac{R s_0^2}{2s^2}(c_0 + c)\dot{\phi}_0^2 + J\right\} \tag{29}$$

where $c \equiv \cos\theta$, $c_0 \equiv \cos\theta_0$, $s \equiv \sin\theta$, $s_0 \equiv \sin\theta_0$

This describes precession about the z axis (figure 1), with variable speed $\dot{\phi}$, accompanied by oscillations in θ (nutation) between limits θ_0 and θ_1 where $\dot{\theta}=0$. Exactly this motion is familiar in the heavy symmetrical top (Arnold 1978) and indeed as we show in Appendix C one form of the equations of motion for the top is precisely (26).

In the adiabatic limit of large R, the amplitude of nutation vanishes. To see this, we find the extreme θ_1 (i.e. $\dot{\theta}=0$) from the dominant terms in (29). Thus

$$\begin{aligned} \cos\theta_0 - \cos\theta_1 &\approx \frac{2\sin^2\theta_0\, \dot{\phi}_0^2}{R} \quad \text{if } \dot{\phi}_0 \neq 0 \\ &= \frac{2\sin^2\theta_0\, J}{R^3} \quad \text{if } \dot{\phi}_0 = 0 \end{aligned} \tag{30}$$

Both expressions vanish as $R\to\infty$ (the second because J is of order R - consistent with (13) for S constant), so that $\theta_1\to\theta_0$. This shows the accuracy to which the adiabatic invariant $I=S\cdot r=J\cdot r=J\cos\theta$ is conserved.

Taking the motion to be pure precession round a circle of latitude, we find the angular velocity $\dot{\phi}$ from (26):

$$\begin{aligned} \dot{\phi} &= \frac{R^2 - \sqrt{R^4 - 4JR\cos\theta}}{2R\cos\theta} \\ &\to \frac{J}{R^2} \quad \text{as } R \to \infty \end{aligned} \tag{31}$$

The speed of this adiabatic precession on the R sphere is

$$V = R\sin\theta\, \dot{\phi} = \frac{J\sin\theta}{R} = \frac{I\tan\theta}{R} \tag{32}$$

4.2. Motion under magnetic gauge force

In the adiabatic gauge model the equation of motion is (20), with the magnetic monopole force included and the two radial forces (Born-Oppenheimer and electric) omitted and replaced by the force constraining the particle to the sphere:

$$\dot{V} = -I\frac{V \wedge r}{R^2} - \frac{V^2}{R} r \tag{33}$$

In addition to R this conserves energy and a modified angular momentum incorporating the monopole field:

$$E = \tfrac{1}{2} V^2 = \text{constant}, \quad Q \equiv R \wedge V + Ir = \text{constant} \tag{34}$$

These imply that $r\cdot Q = I$ is constant, so that the particle moves with constant speed in a circle with axis Q. The angle made by r with Q is θ, where, from (34),

$$\tan\theta = \frac{|R \wedge V|}{I} = \frac{RV}{I} \tag{35}$$

Comparing with (32), we see that the magnetic gauge field, from the monopole at R=0, describes precisely the steady mean precession of the particle on the sphere in the adiabatic limit when the amplitude of nutation is negligible.

Thus the precession from the magnetic gauge force describes the main global feature of the adiabatic dynamics. Note that ordinary Born-Oppenheimer theory, in which this force is omitted, fails completely because it predicts that the particle remains at rest on the sphere.

To see that the result (35) is not trivial, consider the case where the particle is released from rest at polar angle θ_0. Then from (28) and (29) it follows that $E=K=JR\cos\theta_0$, and the motion is nutation in a series of loops (figure 1) between latitude circles θ_0 and θ_1, given by

$$\begin{aligned}\cos\theta_1 &= \frac{R^3}{4J}\left(1-\sqrt{1-16\left(\frac{J\cos\theta_0}{2R^3}-\frac{J^2}{R^6}\right)}\right) \\ &\approx \cos\theta_0 - \frac{2J}{R^3}\sin^2\theta_0 \quad \text{for large } R\end{aligned} \tag{36}$$

(cf. the second equation in (30). In the adiabatic regime, the loops are tiny and roughly semicircular. 'Microscopically', the motion is far from steady, because the precession stops - that is, $\dot\phi$=0 - every time θ returns to θ_0. However, a short calculation, which we do not give, shows that (31) still gives the *average* rate of precession over a nutation cycle.

Returning to the analogy with the heavy symmetrical top (appendix C) we arrive at an unexpected picture of its motion. The fast spin of the top can be regarded as adiabatically transported by the slow motion of the axis, which in turn is driven by the reaction of the spin, in the form of the magnetic force from a monopole centred on the point that is held fixed.

5. Electric force only

5.1. Exact solution

Now we remove the constraint that kept R constant, and specialise the model of §2 differently, in order to eliminate the Born-Oppenheimer and magnetic forces in the adiabatic approximate equations of motion (20) and retain only the electric gauge force. This can be achieved by setting I=0, which we do in a particular way that enables the motion to be solved exactly, namely by demanding that the conserved quantity $\boldsymbol{J}$ defined by (11) vanishes. Thus $I=\boldsymbol{S}\cdot\boldsymbol{r}=\boldsymbol{J}\cdot\boldsymbol{r}=0$, and the adiabatic invariant remains zero not only adiabatically but exactly. Thus from (12) the motion is determined by

$$\dot{\boldsymbol{V}} = \boldsymbol{R} \wedge \boldsymbol{V} \tag{37}$$

This could describe the motion of a 'spin' $\boldsymbol{V}$, driven, as is $\boldsymbol{S}$, by the position $\boldsymbol{R}$, and coupled to it by $\dot{\boldsymbol{R}}=\boldsymbol{V}$, or, alternatively, a charged particle in the magnetic field of a uniform sphere of monopolium.

The shape of the orbits is determined by a single parameter. This is because five of the six quantities required to specify a solution of (37) can be eliminated: one by choosing the origin of time as the instant of closest approach; three by rigid rotation about $\boldsymbol{R}$=0, possible because of the rotational invariance of (37); and one by the dilation law obeyed by (37), namely that if $\boldsymbol{R}(t)$ is a solution then so is $\alpha\boldsymbol{R}(\alpha t)$ for arbitrary α.

In solving (37) exactly, the first step is to note that the speed V is constant. Next, we use

$$\tfrac{1}{2}\partial_t^2\left(R^2\right) = \tfrac{1}{2}\partial_t^2(\boldsymbol{R}\cdot\boldsymbol{R}) = \partial_t(\boldsymbol{R}\cdot\boldsymbol{V}) = V^2 + \boldsymbol{R}\cdot\dot{\boldsymbol{V}} = V^2 = \text{constant} \tag{38}$$

so that the distance from the origin obeys the repulsion law

$$R^2(t) = V^2t^2 + R_0^2 \tag{39}$$

Corresponding to the 'spin' $\boldsymbol{V}$ is the 'adiabatic invariant' $\boldsymbol{r}\cdot\boldsymbol{V}$; unlike the analogous quantity I for the original spin $\boldsymbol{S}$, this is not constant, but from (38) and (39) varies as

$$r \cdot V = \frac{R \cdot V}{R} = V \frac{t}{\sqrt{t^2 + R_0^2 / V^2}} \tag{40}$$

This changes sign at t=0 but has the asymptotically constant values $\pm V$, indicating that the orbits ultimately recede along straight lines.

Without loss of generality we can set V=1, and regard R_0 as the single shape parameter. Adiabatic orbits are those with R_0>>1.

With V=1, time equals arc length along the orbit, the velocity V is the unit tangent vector, and the determination of the orbit can be reduced to a problem in the differential geometry of space curves. The curvature κ of the curve (see e.g. Struik 1950) is the magnitude of the acceleration, which from (37-39) is

$$\kappa = \sqrt{\dot{V} \cdot \dot{V}} = \sqrt{(R \wedge V) \cdot (R \wedge V)} = \sqrt{R^2 - (R \cdot V)^2} = R_0 = \text{constant} \tag{41}$$

The unit normal n and binormal b of the curve are the vectors

$$n = \frac{\dot{V}}{R_0} = \frac{R \wedge V}{R_0}, \quad b = V \wedge n = \frac{V \wedge (R \wedge V)}{R_0} = \frac{R - tV}{R_0} \tag{42}$$

The torsion is

$$\tau = -\dot{b} \cdot n = \frac{t\dot{V} \cdot (R \wedge V)}{R_0^2} = \frac{t\left(R^2 - (R \cdot V)^2\right)}{R_0^2} = t \tag{43}$$

Thus the orbit is a curve with constant curvature R_0 and changing torsion t. To get a preliminary idea of its shape, note that, for a helix uniformly wound on a cylinder, κ and t are given in terms of the radius a and pitch p (longitudinal distance between coils) by

$$\kappa = \frac{4\pi^2 a}{p^2 + 4\pi^2 a^2}, \quad \tau = \frac{2\pi p}{p^2 + 4\pi^2 a^2} \tag{44}$$

Thus for large |t|, when the torsion can be regarded as locally constant, we expect the orbit to wind about its asympotes in a coil with shrinking radius and pitch

$$a = \frac{R_0}{R_0^2 + t^2} \to \frac{R_0}{t^2}, \quad p = \frac{2\pi t}{R_0^2 + t^2} \to \frac{2\pi}{t} \tag{45}$$

Since the torsion changes sign at t=0, the negative-t and positive-t asymptotic windings have opposite senses.

Thus each orbit $R(t)$ resembles a pair of curly antelope horns. Figure 2 shows some of these shapes, for different values of R_0, obtained by solving (37) numerically. It will also be convenient to study the track of the velocity V on its unit sphere; this is a curve whose speed is the curvature R_0 of $R(t)$, and whose curvature is the torsion t of $R(t)$, and which therefore has the shape of a S coiling infinitely rapidly into asymptotic directions $V(\pm\infty)$; figure 3 shows some of these shapes.

To compare with adiabatic theory, we need to know the asymptotic opening angle $\theta(R_0)$ of the antelope horns, defined by

$$\cos\{\theta(R_0)\} = r(-\infty) \cdot r(+\infty) = -V(-\infty) \cdot V(+\infty) \tag{46}$$

Figure 3 indicates that the asymptotic directions $V(\pm\infty)$ become mutually antipodal as R_0 increases. Since the angle between these directions is π-$\theta(R_0)$, we conclude that θ=0 in the adiabatic limit , so that the horns spiral out in the same direction as they spiralled in, but with opposite torsion. This is confirmed by figure 2. Now we will determine the geometry of the curve exactly from its κ and τ and find the angle $\theta(R_0)$ analytically.

In appendix D we explain how the triad V, n, b is determined by the solution of a problem in quantum mechanics: the evolution of a complex 2-spinor, describing for example a spin-1/2 particle, driven by a 'magnetic field' whose x and z components are κ and -τ. As a special case, the velocity in our problem is

$$\begin{aligned} V &= \langle \psi | \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} | \psi \rangle \\ &= \left(|\psi_{x+}|^2 - |\psi_{x-}|^2, \ |\psi_{y+}|^2 - |\psi_{y-}|^2, \ |\psi_{z+}|^2 - |\psi_{z-}|^2 \right) \end{aligned} \tag{47}$$

where $|\psi\rangle$ denotes the 3-vector of 2-spinors

$$|\psi\rangle \equiv \left\{ \begin{pmatrix} \psi_{x+} \\ \psi_{x-} \end{pmatrix}, \begin{pmatrix} \psi_{y+} \\ \psi_{y-} \end{pmatrix}, \begin{pmatrix} \psi_{z+} \\ \psi_{z-} \end{pmatrix} \right\} \tag{48}$$

each of whose components satisfy the Schrödinger equation

$$i|\dot{\psi}\rangle = \tfrac{1}{2}\begin{pmatrix} -t & R_0 \\ R_0 & t \end{pmatrix}|\psi\rangle \tag{49}$$

with inital conditions

$$|\psi_x(0)\rangle = \tfrac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \quad |\psi_y(0)\rangle = \tfrac{1}{\sqrt{2}}\begin{pmatrix}1\\ \mathrm{i}\end{pmatrix}, \quad |\psi_z(0)\rangle = \begin{pmatrix}1\\0\end{pmatrix} \tag{50}$$

Equation (49) defines the Landau-Zener problem of quantum transition theory (Zener 1932). It can be exactly solved via the second-order equation that the components $\psi_\pm$ satisfy separately, namely

$$\ddot{\psi}_\pm + \left(\tfrac{1}{4}t^2 + \tfrac{1}{4}R_0^2 \mp \tfrac{1}{2}\mathrm{i}\right)\psi_\pm = 0 \tag{51}$$

These two equations are complex conjugates of each other. We express the $|\psi_j\rangle$ as linear combinations of the standard even and odd solutions y_1 and y_2 with initial conditions

$$y_1(0) = 1, \quad y_1'(0) = 0; \quad y_2(0) = 0, \quad y_2'(0) = 1 \tag{52}$$

In terms of these, (50) fixes the quantum states as

$$\begin{aligned}
\psi_x &= \tfrac{1}{\sqrt{2}}\begin{pmatrix} y_1^* - \tfrac{1}{2}\mathrm{i}R_0 y_2^* \\ y_1 - \tfrac{1}{2}\mathrm{i}R_0 y_2 \end{pmatrix} \\
\psi_y &= \tfrac{1}{\sqrt{2}}\begin{pmatrix} y_1^* + \tfrac{1}{2}R_0 y_2^* \\ \mathrm{i}y_1 - \tfrac{1}{2}\mathrm{i}R_0 y_2 \end{pmatrix} \\
\psi_z &= \begin{pmatrix} y_1^* \\ -\tfrac{1}{2}\mathrm{i}R_0 y_2 \end{pmatrix}.
\end{aligned} \tag{53}$$

The velocity of the particle is now seen to be, from (47),

$$V = \left(R_0\,\mathrm{Im}\,y_1 y_2^*, \quad R_0\,\mathrm{Re}\,y_1 y_2^*, \quad |y_1|^2 - \tfrac{1}{4}R_0^2|y_2|^2\right) \tag{54}$$

To determine the orbit $R(t)$, integration of $V(t)$ is unnecessary, since from (42)

$$R(t) = tV(t) + R_0 b(t) \tag{55}$$

gives the orbit which from (D12) has initial conditions

$$R(0) = (-R_0, 0, 0), \quad V(0) = (0, 0, 1) \tag{56}$$

The solutions of (51) are parabolic cylinder functions, and the standard solutions can be expressed in terms of confluent hypergeometric (Kummer) functions. From §§19.2, 19.16 and chapter 13 of Abramowitz and Stegun (1964) we identify

$$\begin{aligned}
y_1(t) &= \exp\left\{-\tfrac{1}{4}\mathrm{i}t^2\right\} M\left(\tfrac{1}{8}\mathrm{i}R_0^2, \tfrac{1}{2}, \tfrac{1}{2}\mathrm{i}t^2\right) \\
y_2(t) &= t\exp\left\{-\tfrac{1}{4}\mathrm{i}t^2\right\} M\left(\tfrac{1}{8}\mathrm{i}R_0^2 + \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{1}{2}\mathrm{i}t^2\right)
\end{aligned} \tag{57}$$

From the known asymptotic behaviour of these functions we obtain, for the dominant terms,

$$\begin{aligned}
y_1(t) &\to \sqrt{\pi}\,\frac{\exp\left\{-\tfrac{1}{16}\pi R_0^2\right\}}{\Gamma\left(\tfrac{1}{2} - \tfrac{1}{8}\mathrm{i}\pi R_0^2\right)} \exp\left\{-\tfrac{1}{4}\mathrm{i}\left[t^2 + \tfrac{1}{2}R_0^2 \log\left(\tfrac{1}{2}t^2\right)\right]\right\} \\
y_2(t) &\to \sqrt{\frac{\pi}{2}}\,\frac{\exp\left\{-\tfrac{1}{16}\pi R_0^2\right\}}{\Gamma\left(1 - \tfrac{1}{8}\mathrm{i}\pi R_0^2\right)} \exp\left\{-\tfrac{1}{4}\mathrm{i}\left[t^2 + \tfrac{1}{2}R_0^2 \log\left(\tfrac{1}{2}t^2\right) - \pi\right]\right\}
\end{aligned} \tag{58}$$

as $t \to +\infty$

Thus (54) gives the asymptotic velocity

$$\begin{aligned}
V \to \sqrt{2\sinh\left(\tfrac{1}{4}\pi R_0^2\right)} \exp\left\{-\tfrac{1}{8}\pi R_0^2\right\}(-\sin\gamma, \cos\gamma, 0) + \\
+ \exp\left\{-\tfrac{1}{4}\pi R_0^2\right\}(0, 0, 1) \\
\text{as } t \to +\infty
\end{aligned} \tag{59}$$

where

$$\begin{aligned}
\gamma &\equiv \arg\left\{\Gamma\left(\tfrac{1}{2} - \tfrac{1}{8}\mathrm{i}R_0^2\right)\Gamma\left(1 + \tfrac{1}{8}\mathrm{i}R_0^2\right)\right\} + \tfrac{1}{4}\pi \\
&\approx \tfrac{1}{4}\pi \quad \text{if} \quad R_0 << 1 \\
&\approx \tfrac{1}{2}\pi \quad \text{if} \quad R_0 >> 1
\end{aligned} \tag{60}$$

For $t \to -\infty$, V is given by (59) with the signs of the x and y components reversed. Thus from (46) we obtain the opening angle of the antelope horns as

$$\theta(R_0) = 2\arcsin\exp\left\{-\tfrac{1}{4}\pi R_0^2\right\} \qquad (61)$$

In the adiabatic limit $R_0 \to \infty$ this vanishes, confirming our previous conclusion. From (59), the asymptotic direction of the horns in the xy plane makes an angle $\gamma+\pi/2$ with the x axis, and from (60) this varies from $3\pi/4$ to π as R_0 varies from 0 to ∞.

5.2. Motion under electric gauge force

From this lengthy analysis of the exact solution of the model (9) with zero total angular momentum $\boldsymbol{J}$, we concluded that in the adiabatic limit the particle spirals in from infinity to a distance R_0, and then spirals back out along the same direction. Averaged over oscillations, this motion becomes that of a 'guiding centre' that is purely radial, with the history of R changing according to the exact formula (39) with V=1. Now we show that precisely this behaviour follows from the adiabatic equations of motion (20).

Since J=0 implies I=0, (20) reduces to

$$\dot{V} = \frac{S^2}{R^3}\boldsymbol{r} \qquad (62)$$

where, from (11) and the initial conditions,

$$S = |\boldsymbol{R}(0) \wedge \boldsymbol{V}(0)| = R_0 \qquad (63)$$

Equation (62) describes scattering. The shapes of orbits are determined by energy E and impact parameter b, which we identify from the asymptotic form of the exact orbits of (37). Since V=1, E=1/2; and since the orbital angular momentum oscillates about zero ($\boldsymbol{S}$=-$\boldsymbol{R}$x$\boldsymbol{V}$ precesses equatorially about $\boldsymbol{R}$), b=0. The corresponding solution of (62) is

$$\boldsymbol{R}(t) = \left(-\sqrt{R_0^2 + t^2}, 0, 0\right) \qquad (64)$$

This is identical with the exactly-calculated radial behaviour (39).

Thus the repulsion from the electric gauge force describes the main global feature of the adiabatic dynamics. Note that ordinary Born-Oppenheimer theory, in which this force is omitted, fails completely

because it predicts that the particle remains at rest or moves without acceleration.

6. Concluding remarks

Our main aim has been to show by exact calculation that in classical mechanics the gauge forces of reaction produce real effects. These forces are appealing because of the simple picture they give for the averaged slow motion, and the ease with which they reproduce features that are quite deeply buried in exact solutions (where these are available).

Woven into the exact analysis has been a tissue of spin analogies. In the basic model (9) there is the spin $\boldsymbol{S}$, driven by the coordinate $\boldsymbol{R}$. In the 'electric' special case (§5.1), the velocity $\boldsymbol{V}$ appears as a spin, again driven by $\boldsymbol{R}$. And in the determination of a curve from its curvature and torsion (appendix D) each vector in the Frenet frame acts as a spin, driven by the Darboux vector $\boldsymbol{\Omega}$ (equation D5); this is equivalent to the quantum spin states $|\psi_j\rangle$ being driven by the 'magnetic field' $\boldsymbol{\Omega}$.

We have deliberately restricted ourselves to exactly-solvable special cases of the model (9). On the basis of some numerical exploration of the exact equation (12) governing the slow motion, we anticipate that in the general case there will be very complicated behaviour. For example, according to (20) the Born-Oppenheimer force is radial, and when acting alone (or in conjunction with the much smaller electric gauge force) would generate multiply-periodic planar loops around the origin if the force is attractive (i.e. when the adiabatic invariant I is positive, and scattering if the force is repulsive (i.e. when I<0). The effect of the magnetic gauge force is to cause these orbits to swerve out of the plane, and we have observed this (along with superimposed small nonadiabatic oscillations from the reaction of the spin precession) in the numerical solution of (12). Over long times, the exact motion for 'attractive' initial conditions is observed to consist of non-planar loops with erratically varying radii and inclinations, probably indicating chaos (as also seen in a similar model by Bulgac and Kusnezov 1992), with substantial changes in the adiabatic invariant near the closest approaches where the precession frequency R is smallest.

Acknowledgments

We thank Dr J H Hannay for a helpful suggestion. JMR thanks the NSF and The Royal Society for financial support.

Appendix A. Quantum origin of gauge forces

The evolution operator for the Hamiltonian (1), describing the development of quantum states for a finite time t, can be decomposed into the product of infinitesimal operators:

$$U(t) = \exp\left\{-\mathrm{i}t\frac{H(\boldsymbol{R},\boldsymbol{P},z)}{\hbar}\right\} = \left[\exp\left\{-\mathrm{i}\frac{t}{N}\frac{H(\boldsymbol{R},\boldsymbol{P},z)}{\hbar}\right\}\right]^N \qquad \text{(A1)}$$

where N is large. Between each pair of factors we can insert the resolution of unity for the complete set of eigenstates $|n(\boldsymbol{R})\rangle$ in the fast subspace:

$$U(t) = \sum_{n_0}|n_0(\boldsymbol{R})\rangle\langle n_0(\boldsymbol{R})|\prod_{i=1}^{N}\left[\sum_{n_i}\exp\left\{-\mathrm{i}\frac{t}{N}\frac{H(\boldsymbol{R},\boldsymbol{P},z)}{\hbar}\right\}|n_i(\boldsymbol{R})\rangle\langle n_i(\boldsymbol{R})|\right] \qquad \text{(A2)}$$

(of course $\boldsymbol{R}$ is an operator in the full Hilbert space, but acts as a c-number in the fast subspace).

Now we make the adiabatic assumption that there are no transitions between fast states, so that all the sums can be replaced by the single term corresponding to the initial state, $|n\rangle$ say. Thus the effective evolution operator governing the slow motion becomes

$$\begin{aligned} U_g(t) &= \left[\langle n(\boldsymbol{R})|\exp\left\{-\mathrm{i}\frac{t}{N}\frac{H(\boldsymbol{R},\boldsymbol{P},z)}{\hbar}\right\}|n(\boldsymbol{R})\rangle\right]^N \\ &= \left[\langle n(\boldsymbol{R})|1-\mathrm{i}\frac{t}{N}\frac{H(\boldsymbol{R},\boldsymbol{P},z)}{\hbar}|n(\boldsymbol{R})\rangle\right]^N \\ &= \exp\left\{-\mathrm{i}t\frac{\langle n(\boldsymbol{R})|H(\boldsymbol{R},\boldsymbol{P},z)|n(\boldsymbol{R})\rangle}{\hbar}\right\} \qquad \text{(A3)} \\ &\equiv \exp\left\{-\mathrm{i}t\frac{H_g(\boldsymbol{R},\boldsymbol{P})}{\hbar}\right\} \end{aligned}$$

where

$$H_g(\boldsymbol{R},\boldsymbol{P}) = \tfrac{1}{2}\left\langle n(\boldsymbol{R})\middle|\sum\sum Q_{ij}P_iP_j\middle|n(\boldsymbol{R})\right\rangle + E_n(\boldsymbol{R}) \qquad \text{(A4)}$$

Now the result (4) follows from a double application of

$$P_i|n(\boldsymbol{R})\rangle = -\mathrm{i}\hbar\partial_i|n(\boldsymbol{R})\rangle + |n(\boldsymbol{R})\rangle P_i \qquad \text{(A5)}$$

We do not claim that this argument is rigorous, but it has the merit of suggesting that the gauge forces are part of the lowest-order adiabatic approximation, in the sense that they are a consequence of the assumption that there are no transitions.

Appendix B. Classical limit of the quantum gauge forces

Inserting the resolution of the identity into B_{ij} and Φ in (4), and using

$$\langle m|\partial_i n\rangle = \frac{\langle m|\partial_i h|n\rangle}{E_n - E_m} \qquad \text{(B1)}$$

(where here and hereafter we do not write the $\boldsymbol{R}$-dependences explicitly), we obtain the known formulae

$$B_{ij}(R) = \mathrm{i}\hbar \sum_{m\neq n} \frac{\langle m|\partial_i h|n\rangle\langle n|\partial_j h|m\rangle - \langle m|\partial_j h|n\rangle\langle n|\partial_i h|m\rangle}{(E_n - E_m)^2}$$
$$g_{ij} = \frac{\hbar^2}{2} \sum_{m\neq n} \frac{\langle m|\partial_i h|n\rangle\langle n|\partial_j h|m\rangle}{(E_n - E_m)^2} \qquad \text{(B2)}$$

Now we substitute

$$\frac{1}{E^2} = -\frac{1}{\hbar^2}\int_0^\infty \mathrm{d}t t \exp\left\{\pm \mathrm{i}\frac{Et}{\hbar}\right\} \qquad \text{(B3)}$$

and thereby express the gauge forces as integrals over time:

$$B_{ij} = -\frac{\mathrm{i}}{2\hbar}\int_0^\infty \mathrm{d}t t \left\langle n\left|\left[(\partial_i h)_t, \partial_j h\right] - \left[(\partial_j h)_t, \partial_i h\right]\right|n\right\rangle$$
$$g_{ij} = -\tfrac{1}{2}\int_0^\infty \mathrm{d}t t \left\langle n\left|(\partial_i h - \partial_i E_n)_t (\partial_j h - \partial_j E_n)\right|n\right\rangle \qquad \text{(B4)}$$

Here

$$(\partial_i h)_t \equiv \exp\left\{\frac{\mathrm{i}ht}{\hbar}\right\}\partial_i h \exp\left\{-\frac{\mathrm{i}ht}{\hbar}\right\} \qquad \text{(B5)}$$

denotes the time-evolved operator, and the diagonal terms in the m-sum, necessary to include in (B2) in order to get (B4), are zero.

Now we can take the classical limit directly from the correspondence principle: the expectation value is replaced by the phase-space average (7) over the classical invariant manifold (here the energy surface) corresponding to the state, and commutators by $\mathrm{i}\hbar$ times the Poisson brackets. Thus we immediately obtain the classical formula for g_{ij} in (8); the formula for B_{ij} follows after substituting the identity (Robbins and Berry 1992)

$$\left\langle \left\{(A)_t, B\right\}\right\rangle_E = \frac{1}{\partial_E \Omega}\partial_E\left(\partial_E \Omega \left\langle \partial_t (A)_t B\right\rangle_E\right) \qquad \text{(B6)}$$

and integrating by parts over t.

Appendix C. Analogy with heavy symmetrical top

Let the top have moments of inertia A, B, B about the point of support R=0, mass m, and axis determined by the unit vector $\boldsymbol{r}$, and let gravity $\boldsymbol{g}$=-$g\boldsymbol{e}_z$ act through the centre of mass at $\boldsymbol{R}$=$l\boldsymbol{r}$. Let the axis $\boldsymbol{r}$, whose motion we seek, have polar angles θ, ϕ, and let the third Euler angle, describing the spin about $\boldsymbol{r}$, be ψ. In an inertial frame (i.e. not in the usual body frame), the angular velocity $\boldsymbol{\Omega}$ and angular momentum $\boldsymbol{L}$ are

$$\boldsymbol{\Omega} = \left(\dot\psi + \dot\phi\cos\theta\right)\boldsymbol{r} + \boldsymbol{r}\wedge\dot{\boldsymbol{r}}, \quad \boldsymbol{L} = A\left(\dot\psi + \dot\phi\cos\theta\right)\boldsymbol{r} + B\boldsymbol{r}\wedge\dot{\boldsymbol{r}} \qquad \text{(C1)}$$

Gravitation determines the motion of $\boldsymbol{L}$ through the torque

$$\dot{\boldsymbol{L}} = -mgl\boldsymbol{r}\wedge\boldsymbol{e}_z \qquad \text{(C2)}$$

By symmetry, the components of $\boldsymbol{L}$ along $\boldsymbol{r}$ and $\boldsymbol{e}_z$ are conserved:

$$L_\psi \equiv \boldsymbol{r}\cdot\boldsymbol{L} = A\left(\dot\psi + \dot\phi\cos\theta\right) = \text{constant}$$
$$L_\phi \equiv \boldsymbol{e}_z\cdot\boldsymbol{L} = L_\psi \boldsymbol{r}\cdot\boldsymbol{e}_z + B\boldsymbol{r}\wedge\dot{\boldsymbol{r}}\cdot\boldsymbol{e}_z = \text{constant} \qquad \text{(C3)}$$

Thus the equation of motion (C2) becomes

$$L_\psi\dot{\boldsymbol{r}} + B\boldsymbol{r}\wedge\ddot{\boldsymbol{r}} = -mgl\boldsymbol{r}\wedge\boldsymbol{e}_z \qquad \text{(C4)}$$

or, since $\boldsymbol{r}$ is a unit vector,

$$\ddot{\boldsymbol{r}} = \frac{L_\psi}{B}\boldsymbol{r}\wedge\dot{\boldsymbol{r}} - \frac{mgl}{B}\left(\boldsymbol{e}_z - \boldsymbol{e}_z\cdot\boldsymbol{r}\,\boldsymbol{r}\right) - |\dot{\boldsymbol{r}}|^2\boldsymbol{r} \qquad \text{(C5)}$$

Comparing this with (26) and recognising that V=$\mathcal{R}\dot{\boldsymbol{r}}$, we see that the top equation is identical with that for the spinning particle constrained to a sphere, provided we make the identifications

$$R = \frac{L_\psi}{B}, \quad \boldsymbol{J} = \frac{mglL_\psi}{B^2}\boldsymbol{e}_z \qquad \text{(C6)}$$

The conserved quantity K (equation 27) is essentially the vertical component of angular momentum, since $K=JL_\psi/B$.

Therefore the adiabatic limit regime of large R corresponds to large L_ψ, that is the top spinning fast. A puzzling feature of this analogy is that there is in the top no obvious physically-significant counterpart of the spin S of the 'particle' whose position corresponds to its axis, namely

$$S = J - R \wedge V = \frac{L_\psi}{B^2}\left[mgle_z - L_\psi r \wedge (\Omega \wedge r)\right] \tag{C7}$$

Appendix D. Quantum spinor method for finding a curve from its curvature and torsion

That the shape of a space curve is determined by its curvature κ and torsion τ, given as functions of arc length, is the content of the fundamental theorem of differential geometry (Struik 1950). We employ the conventional notation t, n, b for the orthogonal frame comprising the tangent, normal and binormal along the curve, which has arc length s, The curve $R(s)$ is obtained by finding t from the Frenet equations (Struik 1950)

$$t' = \kappa n, \quad n' = -\kappa t + \tau b, \quad b' = -\tau n \tag{D1}$$

and then integrating:

$$R(s) = \int_0^s \mathrm{d}s' t(s') \tag{D2}$$

To solve the nine coupled equations (C8) we first note that they can be written in the 3x3 matrix form

$$f_{ij}' = \varepsilon_{ikl}\Omega_k f_{lj} \tag{D3}$$

where f_{ij}, specifying the frame, is

$$f_{ij} \equiv \begin{pmatrix} -b \\ n \\ t \end{pmatrix} = \begin{pmatrix} -b_x & -b_y & -b_z \\ n_x & n_y & n_z \\ t_x & t_y & t_z \end{pmatrix} \tag{D4}$$

and

$$\{\Omega_k\} = (\kappa, 0, -\tau) \equiv \Omega \tag{D5}$$

is the angular velocity of the frame (Darboux vector). Note that the indices j are passive in (D3).

We claim that solutions of (D3) have the form

$$f_{ij} = 2\langle \psi_j | \sigma_i | \psi_j \rangle \tag{D6}$$

where σ_i denotes the three Pauli matrices

$$\sigma_i = \left\{ \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tfrac{1}{2}\begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\} \equiv \sigma \tag{D7}$$

and the spinors

$$|\psi_j\rangle = \begin{pmatrix} \psi_{j+} \\ \psi_{j-} \end{pmatrix} \tag{D8}$$

are any three solutions of the 'Schrödinger equation'

$$\mathrm{i}|\psi'\rangle = \Omega \cdot \sigma |\psi\rangle \tag{D9}$$

in which the 'Hamiltonian' is

$$\Omega \cdot \sigma = \tfrac{1}{2}\begin{pmatrix} -\tau(s) & \kappa(s) \\ \kappa(s) & \tau(s) \end{pmatrix} \tag{D10}$$

To prove this assertion, we differentiate (D6) and use, successively, (D9) and its conjugate, and the commutation relations for the Pauli matrices:

$$\begin{aligned} f_{ij}' &= 2\left(\langle \psi_j' | \sigma_i | \psi_j \rangle + \langle \psi_j | \sigma_i | \psi_j' \rangle\right) \\ &= 2\mathrm{i}\Omega_k \langle \psi_j | [\![\sigma_k, \sigma_i]\!] | \psi_j \rangle \\ &= -2\varepsilon_{kil}\Omega_k \langle \psi_j | \sigma_l | \psi_j \rangle \\ &= -2\varepsilon_{kil}\Omega_k f_{lj} \end{aligned} \tag{D11}$$

By antisymmetry, this is identical with (D3), which is what we wanted to show.

A convenient choice of initial conditions is that $|\psi_i(0)\rangle$ is the eigenstate of σ_i with eigenvalue +1/2, i.e. (50). Then (D6) fixes the orientation of the reconstructed curve (D2) by the following initial orientation of the triad:

$$f_{ij}(0) = \delta_{ij}, \quad \text{i.e.} \quad t(0) = (0,0,1), \quad n(0) = (0,1,0), \quad b(0) = (-1,0,0) \quad \text{(D12)}$$

The foregong argument is equivalent to the reduction of the Frenet equations, by Lie and Darboux (Struik 1950), to a single Riccati equation (this is satisfied by the logarithmic derivative of either component of any of the spinors $|\psi\rangle$).

An interesting duality, which follows from the symmetry of the Frenet equations (D1), or the 'Hamiltonian' (D10), is that once the curve specified by κ and τ is found, the curve with *torsion* κ and *curvature* τ is given by integrating $b(s)$ rather than $t(s)$.

When κ and τ are slowly-varying functions of s, the Schrödinger equation (D9) can be solved by the adiabatic approximation. The resulting 'adiabatic curves' form a class of helices of which the asymptotic coils of the antelope horns, discussed in the paragraph following (43), are an example.

References

Abramowitz M and Stegun I A 1964 'Handbook of mathematical functions' (Washington: National Bureau of Standards)

Anandan J and Aharonov Y 1989 *Phys. Rev. Lett.* **38** 1863-1870

Aharonov Y and Stern A 1992 Preprint, Department of Physics, Tel Aviv University

Arnold V I 1978 *Mathematical Methods of Classical Mechanics* (Springer: New York)

Arnold V I, Kozlov V V & Neishtadt A I 1988 'Dynamical systems III: Mathematical aspects of classical and celestial mechanics' *Encyclopedia of Mathematical Science*, vol. 3,(Springer: Berlin)

Berry M V 1985 *J.Phys.A.* **18**, 15-27 (reprinted in Shapere and Wilczek 1989)

Berry M V 1986, 'Adiabatic phase shifts for neutrons and photons', in *Fundamental aspects of quantum theory*, eds. V.Gorini and A.Frigerio, Plenum, NATO ASI series vol.144, 267-278.

Berry M V 1989 'The quantum phase, five years after' in pp 7-28 of Shapere and Wilczek 1989

Bulgac A and Kusnezov D 1992 *Nucl. Phys. A (Netherlands)* **545** 549c-560

Delacrétaz G, Grant E R, Whetten R L, Wöste L, & Zwanziger J W 1986 *Phys. Rev. Lett.* **56** 2598-2601 (reprinted in Shapere and Wilczek 1989)

Gozzi E and Thacker,W.D, 1987 *Phys.Rev.D.* **35**, 2398-2406

Jackiw R 1988 *Comm. At. Mol. Phys.* **21** 71-82

Hannay J H 1985 *J.Phys.A* **18**, 221-230 (reprinted in Shapere and Wilczek 1989)

Lochak P and Meunier C 1988 *Multiphase averaging for classical systems* (Springer: New York).

Longuet-Higgins H C, Öpik U, Pryce M H L & Sack R A 1959, *Proc. Roy. Soc. Lond.* **A244** 1-16

Mead C A and Truhlar D G 1979 *J Chem Phys* **70** 2284-2296 (reprinted in Shapere and Wilczek 1989)

Messiah A 1962 *Quantum Mechanics* (Wiley: New York) vol.2

Provost J P and Vallée G 1980 *Comm. Math. Phys.* **76** 289-301

Robbins J M and Berry M V 1992 *Proc. Roy. Soc. Lond.* **A436**, 631-661

Shapere A and Wilczek F (eds.) 1989 *Geometric Phases in Physics* (Singapore: World Scientific)

Stone M 1986 *Phys. Rev* **D33** 1191-4 (reprinted in Shapere and Wilczek 1989)

Struik D J 1950 *Lectures on Classical Differential Geometry* (Addison-Wesley: Reading, Mass)

Weigert S and Littlejohn R G 1992 Submitted to *Phys. Rev. A*

Zener C 1932 *Proc. Roy. Soc. Lond.* **A137**, 696-702

Figure captions

Figure 1. Stereographic (south pole) projection of the motion of the particle when confined to a sphere with radius R, calculated from (26) with (a-c): R=3, J=3; (d): R=4, J=3, for $\theta(0)$=45° and different initial azimuthal velocities. (d) shows the decrease in the amplitude of nutation as R increases and the adiabatic regime is approached.

Figure 2. 'Antelope horns' obtained by solving (37) (which corresponds to J=0) for the orbits $R(t)$, with speed V=1 and distances of closest approach (a): R_0=1; (b): R_0=1.5; (c): R_0=2.

Figure 3. Stereographic (south pole) projection of the velocity V (unit tangent vector) corresponding to the orbits of figure 2.

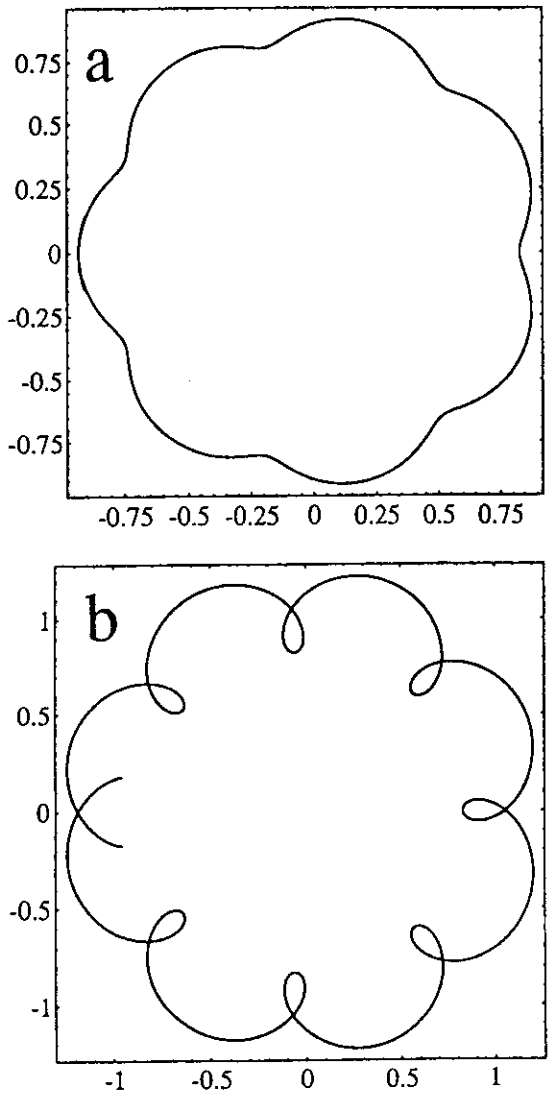


Figure 1

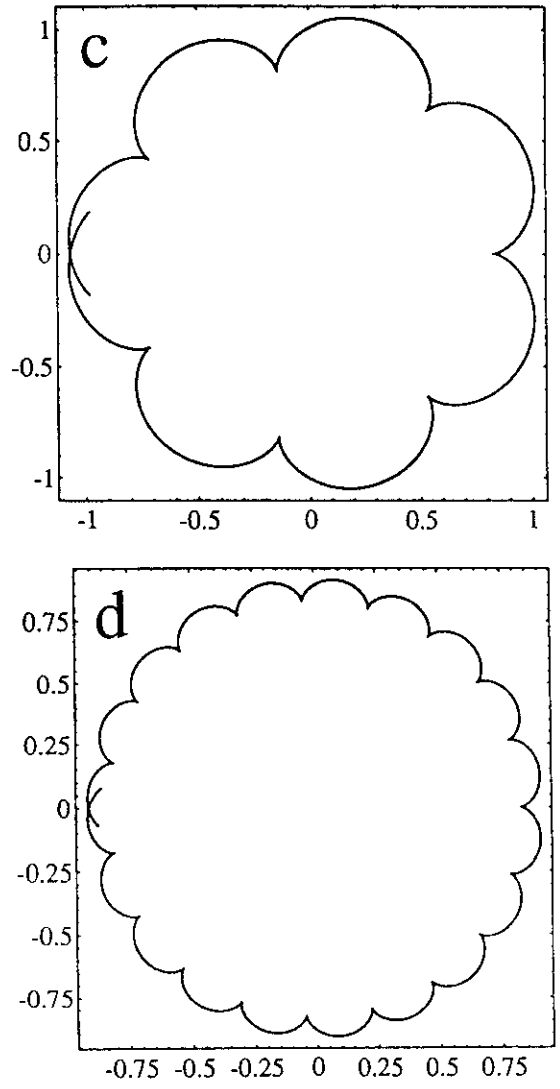


Figure 1

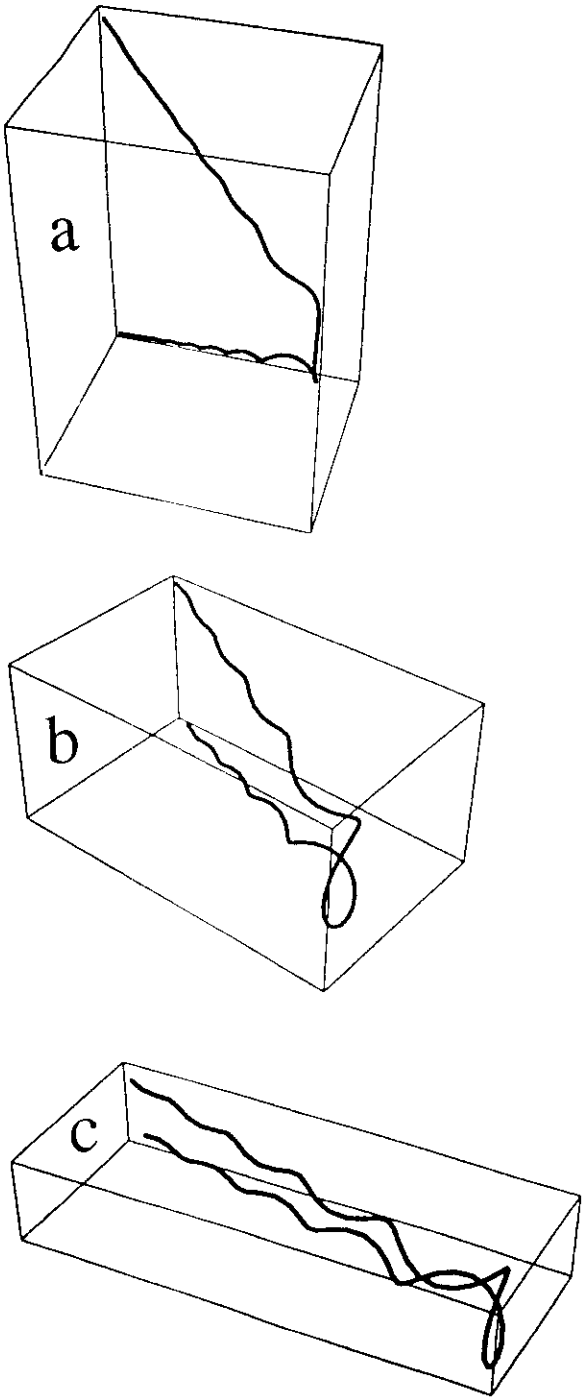


Figure 2

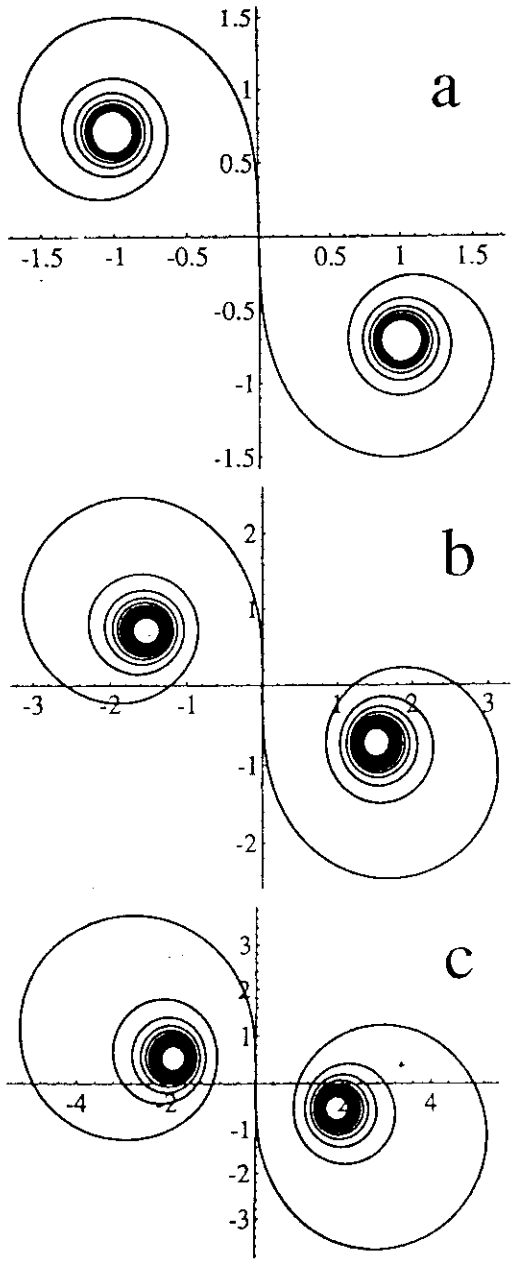


Figure 3

Chaotic classical and half-classical adiabatic reactions: geometric magnetism and deterministic friction

M. V. Berry and J. M. Robbins

H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, U.K.

Abstract

We study the dynamics of a heavy (slow) classical system coupled, through its position, to a classical or quantal light (fast) system, and derive the first-order velocity-dependent corrections to the lowest adiabatic approximation for the reaction force on the slow system. If the fast dynamics is classical and chaotic, there are two such first-order forces, corresponding to the antisymmetric and symmetric parts of a tensor given by the time integral of the force-force correlation function of the fast motion for frozen slow coordinates. The antisymmetric part is geometric magnetism, in which the 'magnetic field' is the classical limit of the 2-form generating the quantum geometric phase. The symmetric part is deterministic friction, dissipating slow energy into the fast chaos; previously found by Wilkinson, this involves the same correlation function as governs the fluctuations and drift of the adiabatic invariant. In the 'half-classical' case where the fast system is quantal with a discrete spectrum of adiabatic states, the only first-order slow force is geometric magnetism; there is no friction. This discordance between classical and quantal fast motion is explained in terms of the clash between the semiclassical and adiabatic limits. A generalization of the classical case is given, where the slow velocity, as well as position, is coupled to the fast motion; to first order, the symplectic form in the lowest-order Hamiltonian dynamics is modified.

Short title: Geometric magnetism and deterministic friction

Submitted to: *Proc. Roy. Soc. Lond.*, January 1993, revised May 1993

1. Introduction

Consider a light system, which may be classical or quantal, coupled to a heavy classical system. The evolution of the light system is rapid on the scale of the heavy motion, which can therefore be regarded as slow. Here we will concentrate on this slow classical motion, which is influenced by reactions from the fast system. In the usual 'Born-Oppenheimer' (Messiah 1962) or 'adiabatic averaging' (Arnold et al. 1988, Lochak and Meunier 1988) approximation, the energy of the fast motion, calculated for frozen slow variables, acts as a potential in which the slow system moves.

In the next approximation beyond this, there are two additional reaction forces linear in the slow velocities. The first is of magnetic type. The field which generates it (Mead and Truhlar 1979, Jackiw 1988, Berry 1989) is the 2-form that gives the geometric phase in the fast system (Berry 1984a, Shapere and Wilczek 1989). Because of the connection with the geometric phase, we call this force 'geometric magnetism'. For examples of the effects of geometric magnetism - and also of the higher-order electric reaction force - see Berry and Robbins (1993) and Aharonov and Stern (1992). The second force was calculated by Wilkinson (1990), and exists when the fast motion is classical and chaotic; it is deterministic friction. We use this term to emphasise that no thermodynamic limit is necessary for this kind of friction, which arises out of the deterministic chaos in low-dimensional fast systems.

Our aim here is to derive these first-order adiabatic reaction forces within a systematic framework. When the light system is classical (§2), we consider the fast motion to be ergodic and mixing. This case has long been problematic. Although the classical limit of the quantum phase 2-form has recently been calculated for chaotic motion (Robbins and Berry 1992a), it was not clear how to derive the reaction force it generates on the slow system using purely classical arguments. The reason is that the accuracy of conservation of the adiabatic invariant for chaotic motion (Ott 1979, Brown et al. 1987) is much poorer than in an integrable (or a quantum) system. To deal consistently with fluctuations, we employ the following physical model: the slow system is coupled to a (microcanonical) ensemble of fast systems (described by a phase-space distribution function), rather than to a single fast trajectory. Deterministic

friction emerges as the symmetric part of the tensor producing the velocity-dependent reaction force at first order. The antisymmetric part is geometric magnetism.

When the light system, coupled to the heavy classical one, is quantal (§3), we have a situation which (following a suggestion of Dr J.H. Hannay) we call 'half-classical mechanics' (there being no implication that the light system is semiclassical, i.e. near its classical limit). Then we couple the slow motion to a density matrix representing the fast system. From this formalism emerges geometric magnetism but not friction. This quantum/classical discordance - that friction vanishes in quantum mechanics, for all finite $\hbar$, but is not zero in the classical limit when the motion is chaotic - is discussed in §4 (for a related instance of it, see Robbins and Berry 1992b).

For simplicity of exposition, most of the discussion is restricted to a slow system with three freedoms, described by the motion of a vector

$$\boldsymbol{R}=\{R_i\}=\{X, Y, Z\} \tag{1}$$

Coupling to the fast motion is through this slow position $\boldsymbol{R}(t)$ (and not through the slow velocity $\boldsymbol{V}\equiv\dot{\boldsymbol{R}}$). The slow acceleration is determined by the reaction force, that is

$$\dot{\boldsymbol{V}} = \boldsymbol{F} \tag{2}$$

and the aim is to determine $\boldsymbol{F}$ in terms of averages over the fast motion, including terms linear in the velocity $\boldsymbol{V}$. It is however possible to give a much more general treatment, in which the slow system can have any number of freedoms and coupling can involve the slow velocities as well as coordinates. We show (§5) that in this case geometric magnetism and deterministic friction appear as modifications of the symplectic form generating the slow equations of motion.

2. Classical fast motion

Let the fast system be bound, with D freedoms and $2D$ phase-space variables

$$z=\{\boldsymbol{q},\boldsymbol{p}\}=\{q_1,\ldots,p_D\} \tag{3}$$

and let the ensemble of fast trajectories be represented by a phase-space distribution function $\rho(z,t)$ which is normalized, i.e.

$$\int \mathrm{d}z\, \rho = 1 \tag{4}$$

Let the time-dependent Hamiltonian h generating the fast motion be governed parametrically by the (changing) slow coordinates $\boldsymbol{R}$, i.e. $h=h(z,\boldsymbol{R}(t))$. Thus the evolution of the ensemble ρ is determined by its Poisson bracket with h:

$$\varepsilon\dot{\rho}(z,t) = \{h(z,\boldsymbol{R}(t)), \rho(z,t)\} \tag{5}$$

Here ε is an adiabatic parameter whose smallness guarantees that the z motion is indeed fast compared with the $\boldsymbol{R}$ motion; ε^2 could for example represent the light/heavy mass ratio, appearing in the form (5) after scaling the equations of motion. The force in (2) is defined as the following average over fast phase space:

$$\boldsymbol{F} = \int \mathrm{d}z\, \rho \nabla h \tag{6}$$

Here and hereafter gradients ∇ will act on the slow coordinates $\boldsymbol{R}$.

We shall find the first two terms of $\boldsymbol{F}$ in an expansion in powers of ε. This requires a formal expansion of the distribution function:

$$\rho = \sum_{r=0}^{\infty} \varepsilon^r \rho_r \tag{7}$$

From (5), the ρ_r are determined by

$$\{h,\rho_0\} = 0, \qquad \{h,\rho_r\} = \dot{\rho}_{r-1} \quad (r>0) \tag{8}$$

Thus ρ_0 is constant on an invariant manifold of the fast dynamics with frozen $\boldsymbol{R}$. For chaotic motion the invariant manfolds are the energy surfaces and, densely distributed over these, the periodic orbits. We choose the normalized microcanonical distribution

$$\rho_0(z,t) = \frac{\delta(E(\boldsymbol{R}) - h(z,\boldsymbol{R}))}{\partial_E \Omega(E(\boldsymbol{R}),\boldsymbol{R})} \tag{9}$$

Here $\boldsymbol{R}=\boldsymbol{R}(t)$ and Ω denotes the phase volume within the energy surface specified by E and $\boldsymbol{R}$, namely

$$\Omega(E,\boldsymbol{R}) = \int \mathrm{d}z \Theta(E - h(z,\boldsymbol{R})) \tag{10}$$

where Θ denotes the unit step. The function $E(\boldsymbol{R})$ determines the energy surfaces visited by ρ_0 as $\boldsymbol{R}$ changes with t. Its form is determined by the adiabatic theorem for ergodic systems (Lochak and Meunier 1988), namely Ω=constant for infinitely slow changes. (We invoke the adiabatic theorem at this point as a matter of convenience, as the adiabatic form of $E(\boldsymbol{R})$ actually emerges from the following analysis.)

Now we introduce two useful notations. Averaging over the energy surface E will be denoted by

$$\langle \cdots \rangle_E \equiv \int \mathrm{d}z\, \rho_0(z,t) \cdots = \frac{1}{\partial_E \Omega} \int \mathrm{d}z\, \delta(E - h(z,\boldsymbol{R})) \cdots \tag{11}$$

The fluctuation of the fast Hamiltonian relative to the adiabatically evolving energy $E(\boldsymbol{R})$ will be denoted by

$$\tilde{h} \equiv h(z,\boldsymbol{R}) - E(\boldsymbol{R}) \tag{12}$$

Adiabaticity then implies

$$\left\langle \nabla \tilde{h} \right\rangle_{E(\boldsymbol{R})} = 0 \tag{13}$$

We can write the force (6) as

$$\boldsymbol{F} = -\int \mathrm{d}z \rho \nabla h = -\left(\int \mathrm{d}z \rho \right) \nabla E - \int \mathrm{d}z \left(\rho_0 + \varepsilon \rho_1 + O\left(\varepsilon^2\right) \right) \nabla \tilde{h} \tag{14}$$

Since ρ is normalised and $\int \mathrm{d}z \rho_0 \nabla \tilde{h} = \left\langle \nabla \tilde{h} \right\rangle_{E(\boldsymbol{R})}$ vanishes (cf. 9 and 13), we have

$$\begin{aligned} &\boldsymbol{F} = -\nabla E(\boldsymbol{R}) + \varepsilon \boldsymbol{F}_1 + O\left(\varepsilon^2\right), \quad \text{where} \\ &\boldsymbol{F}_1 \equiv -\int \mathrm{d}z \rho_1 \nabla \tilde{h} \end{aligned} \tag{15}$$

The leading term $-\nabla E$ is the Born-Oppenheimer force. The focus of interest here is the next term $\varepsilon \boldsymbol{F}_1$, which is the desired first-order reaction force. To find it, we need the first correction ρ_1 in the distribution function, which we obtain using arguments of Ott (1979).

According to (8), the corrections to ρ_0 are determined by the solutions of equations of the form

$$\{h, f\} = g \tag{16}$$

where h and g are given functions of z, and f is to be determined. Because $\langle\{h, f\}\rangle_E = 0$ (micocanonical averages are constant in time), a necessary condition for (16) to have a solution is

$$\langle g \rangle_E = 0 \tag{17}$$

It is shown in appendix A that a particular solution of (16) is

$$f(z) = -\int_{-\infty}^{0} d\tau\, g(z_\tau(z)) \tag{18}$$

where $z_\tau(z)$ is the trajectory generated by the Hamiltonian h in time τ starting from z. As a function of τ, $g(z_\tau)$ is oscillatory with vanishing time average (as implied by (17) and ergodicity). Thus (18), like the integral of a random function, need not converge, but instead gives f as a disrtribution, whose averages over phase-space functions do converge. Equation (18) is the causal solution, depending only on the fast motion in the past, rather than the future. We may add to it any solution f_1 of the homogeneous equation $\{h,f\}=0$, which is necessarily a function of h because energy is the only constant of motion.

Using (8), the first correction can be determined from (18) as

$$\rho_1(z,t) = -\int_{-\infty}^{0} d\tau\, \partial_t \rho_0\left[z_\tau(z, \boldsymbol{R}(t)), t\right] + f_1\left[h(z, \boldsymbol{R}(t)), t\right] \tag{19}$$

where $z_\tau(z,\boldsymbol{R})$ is the orbit under the frozen dynamics, and the partial derivative ∂_t acts only on the second argument of ρ_0. Applied to (8) ,the condition (17) requires that $\langle \partial_t \rho_0 \rangle_E = 0$. A short calculation shows that

this is true if the average of the fluctuation $\nabla(h\text{-}E(\boldsymbol{R}))$ vanishes as in (13), that is if the adiabatic theorem holds (indeed this is one way to derive the adiabatic theorem) .

Substituting (9) into (19), we find

$$\rho_1 = \boldsymbol{V} \cdot \frac{1}{\partial_E \Omega} \int_{-\infty}^{0} d\tau \left[\partial_E \delta(E-h)\right]\left(\nabla \tilde{h}\right)_\tau + (\cdots) + f_1(h,t) \tag{20}$$

where

$$\left(\nabla \tilde{h}\right)_\tau = \nabla \tilde{h}\left(z_\tau(z, \boldsymbol{R}), \boldsymbol{R}\right) \tag{21}$$

and the term $(\cdots)$ is proportional to $\delta(E\text{-}h)$ and will not contribute to the first-order force.

The force produced by the term f_1 is velocity-independent. As shown by Jarzynski (1993), it can be determined by imposing the condition(17) on the second-order (r=2) equation in (8). This force is a correction to the Born-Oppenheimer force (and indeed vanishes when the latter does), because it may be expressed as the gradient of a potential. This potential is however time-dependent and involves the history of the slow motion; therefore Jarzynski's force is a memory effect. From now on we neglect the contribution from f_1, but it is certainly worth exploring.

From (15) and (20) $\boldsymbol{F}_1$ can be written as the time integral of the correlation function, over the fast dynamics with frozen $\boldsymbol{R}$, of the gradients of the fast Hamiltonian, that is the force-force correlation function. Thus

$$\boldsymbol{F}_1 = -\boldsymbol{V} \cdot \frac{1}{\partial_E \Omega} \partial_E \left[\partial_E \Omega \int_{-\infty}^{0} d\tau \left\langle \left(\nabla \tilde{h}\right)_\tau \nabla \tilde{h} \right\rangle_E \right] \tag{22}$$

with $E=E(\boldsymbol{R})$. Changing the sign of τ, and using the invariance of correlation functions to a time shift, we obtain

$$F_1 = -V \cdot \frac{1}{\partial_E \Omega} \partial_E \left[\partial_E \Omega \int_0^\infty d\tau \left\langle \nabla \tilde{h} \left(\nabla \tilde{h} \right)_\tau \right\rangle_E \right] = -\mathsf{K} \cdot V, \text{ where}$$

$$K_{ij} \equiv -\frac{1}{\partial_E \Omega} \partial_E \left[\partial_E \Omega \int_0^\infty d\tau C_{ij}(\tau) \right], \quad \text{and} \tag{23}$$

$$C_{ij}(\tau) \equiv \left\langle \left(\partial_i \tilde{h} \right)_\tau \partial_j \tilde{h} \right\rangle_E$$

For chaotic motion, the correlation function $C_{ij}(\tau)$ vanishes as $\tau \to \infty$, because of mixing, and we asume it decays fast enough for the integral in K_{ij} to converge.

To interpret this result, we separate K_{ij} into symmetric and antisymmetric parts:

$$K_{ij}^{(s)} \equiv \tfrac{1}{2}\left(K_{ij} + K_{ji}\right), \quad K_{ij}^{(a)} \equiv \tfrac{1}{2}\left(K_{ij} - K_{ji}\right) \tag{24}$$

Consider first the antisymmetric part. The corresponding force $F_1^{(a)}$ is geometric magnetism, since

$$F_1^{(a)} = V \wedge B(R) \tag{25}$$

where the 'magnetic field' is given by

$$B(R) = -\frac{1}{2\partial_E \Omega} \partial_E \left[\partial_E \Omega \int_0^\infty d\tau \left\langle \left(\nabla \tilde{h} \right)_\tau \wedge \nabla \tilde{h} \right\rangle_E \right] \tag{26}$$

This is precisely the expression we previously found (Robbins and Berry 1992a) as the classical limit of the geometric phase 2-form. We showed that $\boldsymbol{B}$ vanishes when the fast dynamics possesses time-reversal symmetry (the two essential steps of the argument are invariance of the correlation function under time shift, and the existence of pairs of initial phase points z and Tz such that the forward evolution of ∇h from z is the same as its backward evolution from Tz). We also gave formal arguments suggesting that $\boldsymbol{B}$ is divergenceless.

The symmetric part of the first-order force is deterministic friction, that is irreversible viscous dissipation of slow energy by the fast chaos (Wilkinson 1990). We emphasise that this phenomenon requires only low-dimensional deterministic chaos: no heat bath has been introduced. To lowest order in ε, the dissipated energy is

$$\varepsilon V \cdot F_1 = -\varepsilon V_i V_j K_{ij} \tag{27}$$

In calling this dissipation, we are assuming that the tensor K_{ij} in (23) must be positive definite. Wilkinson (1987) shows that

$$\int_0^\infty d\tau C_{ii}(\tau) > 0 \quad (i \text{ not summed}) \tag{28}$$

but dissipation requires more: this quantity times $\partial_E \Omega$ must increase with E. This is true for scaling systems, where the fast Hamiltonian h is the sum of kinetic energy and a potential U satisfying

$$U(\alpha q, R) = \alpha^\mu U(q, R) \qquad (\mu > 0) \tag{29}$$

since then the quantity in square brackets in K_{ij} in (23) scales as

$$E^{\frac{(D+1)(\mu+2)}{2\mu}} \tag{30}$$

which increases with E. We cannot give a proof of the positivity of K_{ij} in the general case. Indeed, for certain low-dimensional systems we have found what appear to be counter-examples, although these can all be eliminated by increasing D. However, the inequality (28) is sufficient to guarantee dissipation if the unperturbed distribution ρ_0 is a decreasing function of h, e.g. the Boltzmann distribution or the Fermi-Dirac distribution (Wilkinson 1990), in contrast to the microcanonical ensemble considered here.

Unlike geometric magnetism, this frictional force exists whether or not the fast dynamics possesses time-reversal symmetry. But it does require the dynamics to be chaotic. To see this, consider the integrable case D=1, for which the correlation function is periodic:

$$C_{ij}(\tau) = \sum_{n=1}^{\infty}\left(S_{ij}(n)\cos n\omega\tau + A_{ij}(n)\sin n\omega\tau\right), \quad \omega > 0 \tag{31}$$

where S_{ij} is symmetric and A_{ij} is antisymmetric. Only S_{ij} could contribute to the dissipation (cf. 24 and 27), but its contribution is proportional to

$$\int_0^{\infty} d\tau \cos n\omega\tau = \pi\delta(n\omega) = 0 \tag{32}$$

(This argument can be generalised to higher-dimensional integrable systems.) Thus the irreversibility of deterministic friction arises from chaos, together with causality as embodied in the solution (18) of (16).

By assuming that correlations decay exponentially, as in systems with homogeneous chaos, we can obtain useful approximate expressions for the tensors in (23) and (24). We write

$$\begin{aligned} C_{ij}(\tau) &= \left(C_{ij}(0) + \tau\dot{C}_{ij}(0)\right)\exp\{-\lambda\tau\} \\ &= \left(\left\langle \partial_i\tilde{h}\,\partial_j\tilde{h}\right\rangle_E + \tau\left\langle \left\{h,\partial_i\tilde{h}\right\}\partial_j\tilde{h}\right\rangle_E\right)\exp\{-\lambda\tau\} \end{aligned} \tag{33}$$

where λ is the entropy characterising the chaos. Thus the tensor K_{ij} is

$$K_{ij} = \frac{1}{\partial_I\Omega}\partial_L\left[\partial_I\Omega\left(\frac{\left\langle \partial_i\tilde{h}\,\partial_j\tilde{h}\right\rangle_E}{\lambda} + \frac{\left\langle\left\{h,\partial_i\tilde{h}\right\}\partial_j\tilde{h} - \left\{h,\partial_j\tilde{h}\right\}\partial_i\tilde{h}\right\rangle_E}{2\lambda^2}\right)\right] \tag{34}$$

where in writing the second term we have used the invariance of the average under time shift. The first term is symmetric, and represents deterministic friction, while the second is antisymmetric and represents geometric magnetism. This expression is simple because it requires only the evaluation of fixed (that is, not evolved) phase-space quantities, in addition to λ.

The structure (23) of these first-order forces exemplifies linear response theory (Landau and Lifshitz 1980, Kubo et al. 1985), with the force εF_1 and the slow velocity V ('cause' and 'effect') being related by the tensor K_{ij}. Note that in the tensor K_{ij} describing 'nonequilibrium'

response (changing $\boldsymbol{R}$) there occurs the correlation function C_{ij} describing 'equilibrium' (constant $\boldsymbol{R}$) fluctuations (in ∇h). This is an example of the fluctuation-dissipation theorem. And indeed as shown by Ott (1979) and Brown et al. (1987) the same C_{ij} appears in expressions for the secular drift in the adiabatic invariant Ω (which is directly related to dissipation) and the growth rate of fluctuations in Ω.

When the underlying dynamics has time-reversal symmetry, the linear response tensor K_{ij} is symmetric. This is an example of Onsager's relation. When there is no time-reversal symmetry, K_{ij} has an antisymmetric part, which as we have seen corresponds to geometric magnetism. Thus the geometric phase 2-form appears in a new light, as the antisymmetric cousin of friction.

3. Half-classical mechanics

Now let the fast motion be quantum-mechanical, described by a density matrix $\rho(t)$ driven by a Hamiltonian $h(\boldsymbol{R})$, which is time-dependent because the (classical) slow position $\boldsymbol{R}$ is changing. h is a Hermitian operator, whose spectrum we assume to be discrete and non-degenerate for all $\boldsymbol{R}$. It is well known (Mead and Truhlar 1979, Wilkinson 1984, Berry 1989) that there appear magnetic (and electric) reaction forces at first (and second) order, associated with the geometric phase. Our purpose here is to give a derivation of geometric magnetism which parallels the discussion of §2, to facilitate the comparison of classical and half-classical results in §4.

The evolution of ρ is governed by the commutator

$$i\hbar\varepsilon\dot{\rho}(t) = \left[h(\boldsymbol{R}(t)),\rho(t)\right], \qquad \mathrm{Tr}\rho = 1 \tag{35}$$

where again ε is the adiabatic parameter. Equation (35) is the analogue of (4) and (5). The desired force is given by the analogue of (6), namely

$$\boldsymbol{F} = -\mathrm{Tr}\rho\nabla h \tag{36}$$

Consider first the case where ρ is an evolving pure state $|\psi(t)\rangle$:

$$\rho(t) = |\psi(t)\rangle\langle\psi(t)| = \rho(t)^2 \tag{37}$$

As in the classical case, we write ρ as the series (7) in powers of ε. The terms ρ_r are determined by the following equations, analogous to (8):

$$[h,\rho_0]=0, \qquad [h,\rho_r]=i\hbar\dot{\rho}_{r-1} \quad (r>0) \tag{38}$$

Thus ρ_0 must commute with the frozen fast Hamiltonian $h(\boldsymbol{R})$. If we define the adiabatic eigenstates and energy levels by

$$h(\boldsymbol{R})|m(\boldsymbol{R})\rangle = E_m(\boldsymbol{R})|m(\boldsymbol{R})\rangle \tag{39}$$

we can choose ρ_0 as one of these states, say the nth. Thus

$$\rho_0(t) = |n(\boldsymbol{R}(t))\rangle\langle n(\boldsymbol{R}(t))| \tag{40}$$

This is the natural analogue of the microcanonical distribution (9). It depends on time through the changing slow position $\boldsymbol{R}(t)$.

Now we can write the force (36) in a form analogous to (15):

$$\begin{aligned} \boldsymbol{F} &= -\mathrm{Tr}\rho_0\nabla h - \varepsilon\mathrm{Tr}\rho_1\nabla h + O(\varepsilon^2) \\ &= -\nabla E_n(\boldsymbol{R}) + \varepsilon\boldsymbol{F}_1 + O(\varepsilon^2) \end{aligned} \tag{41}$$

where

$$\boldsymbol{F}_1 = -\mathrm{Tr}\rho_1\nabla h = -\sum_{k,l}\langle k|\rho_1|l\rangle\langle l|\nabla h|k\rangle \tag{42}$$

As with (15), the leading term $-\nabla E_n$ (equal to $-\langle n|\nabla h|n\rangle$) is the Born-Oppenheimer force, and the next term is the desired first-order reaction. To find it, we need the first correction ρ_1 in the density matrix.

In the adiabatic basis the off-diagonal elements of the corrections ρ_r are determined by the commutator equations (38) as

$$\langle k|\rho_r|l\rangle = i\hbar\frac{\langle k|\dot{\rho}_{r-1}|l\rangle}{E_k-E_l} \qquad (k\neq l) \tag{43}$$

The diagonal elements are determined by the pure-state condition (37). A simple calculation using (43) shows that the first-order correction ρ_1 is

$$\begin{aligned} \langle k|\rho_1|l\rangle &= i\hbar\frac{\boldsymbol{V}\cdot\langle k|\nabla l\rangle(\delta_{nl}-\delta_{nk})}{E_k-E_l} \qquad (k\neq l) \\ &= 0 \qquad\qquad\qquad\qquad\quad (k=l) \end{aligned} \tag{44}$$

Using

$$\frac{\langle l|\nabla h|k\rangle}{E_k-E_l} = \langle l|\nabla k\rangle \qquad (k\neq l) \tag{45}$$

we now find the first-order force (42) as

$$\begin{aligned} \boldsymbol{F}_1 &= -i\hbar\boldsymbol{V}\cdot\sum_{k,l}\langle k|\nabla l\rangle(\delta_{nl}-\delta_{nk})\langle l|\nabla k\rangle \\ &= i\hbar\boldsymbol{V}\cdot\sum_k\left(\langle k|\nabla n\rangle\langle\nabla n|k\rangle-\langle\nabla n|k\rangle\langle k|\nabla n\rangle\right) \\ &= i\hbar\boldsymbol{V}\wedge\sum_k\langle\nabla n|k\rangle\wedge\langle k|\nabla n\rangle \end{aligned} \tag{46}$$

This has the form

$$\boldsymbol{F}_1 = \boldsymbol{V}\wedge\boldsymbol{B}(\boldsymbol{R}) \tag{47}$$

where the 'magnetic field' is

$$\boldsymbol{B}(\boldsymbol{R}) = -\hbar\,\mathrm{Im}\langle\nabla n(\boldsymbol{R})|\wedge|\nabla n(\boldsymbol{R})\rangle \tag{48}$$

(cf. the classical (25) and (26)).

Now we allow ρ to be a mixed state. From (38), the lowest-order approximation must still commute with h, and it follows that ρ_0 must have the form

$$\rho_0(t) = \sum_m c_m(t)|m(\boldsymbol{R}(t))\rangle\langle m(\boldsymbol{R}(t))|, \qquad \sum_m c_m(t)=1 \tag{49}$$

In fact the coefficients c_m must be constants; this is a consequence of the second equation in the hierarchy (38), which implies that $\langle m|\dot{\rho}_0|m\rangle=0$ for all adiabatic states $|m\rangle$. Now the calculation proceeds as for pure states, except that we cannot determine the diagonal elements of the correction ρ_1 (cf. 44) by using the pure-state condition. However, to order ε any such diagonal elements would give a contribution to $\boldsymbol{F}_1$ of the same form as the Born-Oppenheimer force - effectively a renormalization of the

coefficients c_m. The result is very similar to that for pure states: geometric magnetism (47), with the magnetic field given by the sum

$$\boldsymbol{B} = -\hbar \operatorname{Im} \sum_{m} c_m \langle \nabla m | \wedge | \nabla m \rangle \tag{50}$$

rather than (48).

We have therefore found that when the fast dynamics is quantum-mechanical the first-order force on the classical slow system consists entirely of geometric magnetism. The force is entirely antisymmetric, so there is no friction as when the fast system is classical and chaotic. Therefore quantum and classical mechanics are discordant, an interesting situation which we will discuss in the next section. The expression (48) for the magnetic field is familiar as the 2-form generating the geometric phase in the fast system when $\boldsymbol{R}$ is cycled.

As previously noted in connection with the geometric phase (Berry 1984a) the magnetic force $\boldsymbol{B}(\boldsymbol{R})$ is divergenceless except for monopole singularities at the degeneracies of the adiabatic spectrum. For typical h, without symmetry, degeneracies have codimension 3 and so correspond to points in the space of slow coordinates $\boldsymbol{R}$. The loss of global divergencelessness of $\boldsymbol{B}$ has an interesting consequence for the classical slow dynamics: this is measure-preserving but not globally Hamiltonian (although it is locally Hamiltonian). When the slow dynamics brings $\boldsymbol{R}$ near one of these monopoles, the orbits will, locally, be conical spirals (Goddard and Olive 1978). These classical effects will however be weak, since the monopole strength is $\pm\hbar/2$, which vanishes in the classical limit. Moreover, the breakdown of the adiabatic approximation will be greatest at the degeneracies, because of transitions between adiabatic states.

4. Discordance: classical but not quantal friction

In the classical treatment, deterministic friction originates in the symmetric part of the tensor K_{ij} in (23). For quantal fast motion, no symmetric part appeared. The quantal counterpart of the symmetric part of the correlation integral in (23) is

$$I_{ij} = \tfrac{1}{2}\int_0^\infty \mathrm{d}\tau \big(C_{ij}(\tau) + C_{ji}(\tau)\big), \quad \text{where}$$

$$C_{ij}(\tau) = \tfrac{1}{2}\Big[\langle n | \big(\partial_i \tilde{h}_n\big)_\tau \partial_j \tilde{h}_n | n\rangle + \langle n | \partial_j \tilde{h}_n \big(\partial_i \tilde{h}_n\big)_\tau | n\rangle\Big] \text{ and} \tag{51}$$

$$\tilde{h}_n \equiv h - E_n$$

Here we make use of time-evolved operators:

$$(A)_\tau \equiv \exp\Big\{\mathrm{i}\frac{\tau h}{\hbar}\Big\} A \exp\Big\{-\mathrm{i}\frac{\tau h}{\hbar}\Big\} \tag{52}$$

However, this quantum equivalent I_{ij} is zero. A short calculation shows that

$$\tfrac{1}{2}\big(C_{ij}(\tau) + C_{ji}(\tau)\big) = \sum_{m \neq n} \operatorname{Re}\big(\langle n|\partial_i h|m\rangle\langle m|\partial_j h|n\rangle\big)\cos\Big\{\frac{\tau}{\hbar}\big(E_n - E_m\big)\Big\} \tag{53}$$

This has the same form as (31), so that the time integral I_{ij} vanishes (cf. 32).

Therefore there really is no quantum friction in this theory, and the discordance with classical mechanics persists. Its origin is the same as that studied in a related, mathematically inspired, example by Robbins and Berry (1992b): a clash between the essence of quantization, namely the discrete spectrum of frequencies, and the essence of chaos, namely mixing and a continuous spectrum extending to zero frequency (which make the integrals in (23) converge to finite values).

At first this discordance appears paradoxical, and a violation of the correspondence principle: the I_{ij} vanish for all finite $\hbar$ but are finite for zero $\hbar$. The resolution lies in a careful consideration of the time scales involved in the correlation integrals. To see the quantal I_{ij} converge to zero, the τ integration must include times much longer than the reciprocal of the smallest frequency in the integrand (53). This is proportional to the level spacing, and scales as $\hbar^{D-1}$, so the integration time diverges semiclassically. Therefore the finite values attained by the classical integrals in a finite time (inversely proportional to the entropy characterising the chaos - cf. (33)) must be cancelled by quantal contributions over times large compared with $\hbar^{-(D-1)}$.

A model 'quantum' correlation function which shows this behaviour, and has the same structure as (53), is

$$C_{\rm qu}(\tau) = \hbar \sum_{m=1}^{\infty} \exp\{-\hbar^2 m^2\} \cos\{m\hbar\tau\} \tag{54}$$

Because of the discrete spectrum, the integral of this function from 0 to ∞ is zero. In the classical limit, summation can be replaced by integration, and

$$C_{\rm qu}(\tau) \to C_{\rm cl}(\tau) = \int_0^{\infty} {\rm d}x \exp\{-x^2\} \cos\{x\tau\} = \tfrac{1}{2}\sqrt{\pi} \exp\{-\tfrac{1}{4}\tau^2\} \tag{55}$$

whose integral is not zero but

$$I_{\rm cl} = \tfrac{1}{2}\pi \tag{56}$$

To resolve the discordance, we write (54) exactly in an alternative form, obtained by applying the Poisson sum formula, as follows:

$$\begin{aligned} C_{\rm qu}(\tau) &= \tfrac{1}{2}\hbar \sum_{m=-\infty}^{\infty} \exp\{-\hbar^2 m^2\} \cos\{m\hbar\tau\} - \tfrac{1}{2}\hbar \\ &= \tfrac{1}{2}\hbar \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} {\rm d}m \exp\{-\hbar^2 m^2\} \cos\{m\hbar\tau\} \exp\{2\pi i m n\} - \tfrac{1}{2}\hbar \\ &= \tfrac{1}{2}\sqrt{\pi} \sum_{n=-\infty}^{\infty} \exp\left\{-\tfrac{1}{4}\left(\tau - \frac{2\pi n}{\hbar}\right)^2\right\} - \tfrac{1}{2}\hbar \\ &= \sum_{n=-\infty}^{\infty} C_{\rm cl}\left(\tau - \frac{2\pi n}{\hbar}\right) - \tfrac{1}{2}\hbar \end{aligned} \tag{57}$$

Thus the 'quantum' correlation function is a sequence of copies of the classical function (55), centred on $\tau = 2\pi n/\hbar$, together with a classically vanishing negative offset $-\hbar/2$. In the integral over τ, each copy contributes π, except that centred on the origin, which contributes $\pi/2$. On the average, these are cancelled by the offset, since, for large T,

$$\begin{aligned} I_{\rm qu}(T) &\equiv \int_0^T {\rm d}\tau C_{\rm qu}(\tau) = \tfrac{1}{2}\pi + \pi {\rm Int}\left[\frac{\hbar T}{2\pi}\right] - \tfrac{1}{2}\hbar T \\ &= \pi\left({\rm Int}[x] + \tfrac{1}{2} - x\right) \quad \text{where } x = \frac{\hbar T}{2\pi} \end{aligned} \tag{58}$$

The mean value of the staircase function Int[x] is x-1/2, so that $I_{\rm qu}$ vanishes, as expected.

In emphasising the discordance between the classical model with friction and the corresponding quantum model without friction, we are of course not asserting that friction cannot be described within a quantum framework. This must be possible, since friction exists and the world is quantum-mechanical. Indeed, there is the well-known quantum formula of Kubo (see Kubo et al. 1985) and Greenwood (1958) for the linear transport coefficients of irreversible thermodynamics. (Moreover, these have an antisymmetric part representing, for example, the (nondissipative) quantum Hall effect (Thouless et al. 1982).) Our assertion is simply that friction does not appear as a first-order force when the fast motion is quantal and with a discrete spectrum. The argument for its nonexistence breaks down in the thermodynamic limit where the fast system is regarded as infinite, since then the spectrum becomes continuous and the correlation integrals need not vanish.

Even within the adiabatic framework for finite systems, there is dissipation at higher than the first order in ε, caused by nonadiabatic transitions at 'avoided crossings', where $\boldsymbol{R}$ passes near degeneracies of the adiabatic spectrum. For a given quantum system, that is with fixed $\hbar$, these transitions are exponentially small for small ε, that is 'beyond all orders' in ε. If now $\hbar$ is decreased, keeping ε fixed, the density of avoided crossings increases, and they can be treated statistically using random-matrix theory. In this regime, envisaged by Hill and Wheeler (1952) and studied in detail by Wilkinson (1988), dissipation arises as the collective effect of those occasional avoided crossings when levels approach closer than $O(\sqrt{\varepsilon})$, for which the transitions are not exponentially small. Making $\hbar$ even smaller leads to levels so close and rapidly changing that there are frequent transitions between many levels, and in this quantally nonadiabatic regime Wilkinson obtains the classical result (23) by taking the limit of the Kubo formula. That very different behaviour can occur

during classically and quantally adiabatic changes has been pointed out elsewhere (Berry 1984b): the adiabatic and semiclassical limits are both singular, and their combination is much more so.

5. Generalisation

Here we consider the case where both slow and fast dynamics are classical, and where the fast motion can be coupled to all the slow variables, that is velocities as well as coordinates. Let the slow coordinates and velocities be denoted by $Z=\{Z_\mu\}$, and let the dynamics of the combined system be governed by a Hamiltonian $\mathcal{H}(z,Z)$. As in §2, we consider the slow motion to be coupled to an ensemble of fast trajectories, described by a normalized distribution function $\rho(z,t)$ in fast phase space. From Hamilton's equations, the desired generalized force (rate of change of slow phase-space variables) is

$$\omega_{\mu\nu}\dot{Z}_\nu = \int \mathrm{d}z\,\rho\partial_\mu\mathcal{H} \tag{59}$$

This generalization of (6) involves the unit symplectic matrix

$$\omega_{\mu\nu} \equiv \begin{pmatrix} 0 & -\mathbf{I} \\ \mathbf{I} & 0 \end{pmatrix} \tag{60}$$

where $\mathbf{I}$ is the identity matrix with dimension equal to the number of slow freedoms. The evolution of the distribution function is slaved to $Z(t)$ by the generalization of (5), namely

$$\varepsilon\dot{\rho}(z,t) = \left\{\mathcal{H}\left(z,Z(t)\right),\rho(z,t)\right\} \tag{61}$$

(where of course the Poisson bracket is with respect to z, not Z).

Again we expand ρ in powers of ε. The leading term is the microcanonical distribution (cf. 9)

$$\rho_0(z,t) = \frac{\delta(\mathcal{E}(Z)-\mathcal{H}(z,Z))}{\partial_\mathcal{E}\Omega(\mathcal{E}(Z),Z)} \tag{62}$$

where again Ω denotes the phase volume, now defined as

$$\Omega(\mathcal{E},Z) = \int \mathrm{d}z\Theta(\mathcal{E}-\mathcal{H}(z,Z)) \tag{63}$$

and whose constancy defines the adiabatic energy $\mathcal{E}(Z)$. The generalized force is obtained as in (15)

$$\omega_{\mu\nu}\dot{Z}_\nu = \partial_\mu\mathcal{E}(Z) + \varepsilon\int \mathrm{d}z\rho_1\partial_\mu\tilde{\mathcal{H}} + O\left(\varepsilon^2\right) \tag{64}$$

where (cf. 12)

$$\tilde{\mathcal{H}} \equiv \mathcal{H} - \mathcal{E}(Z) \tag{65}$$

The first term in (64) is the Born-Oppenheimer energy, and the second term, involving the first correction ρ_1 to the distribution function, is the first-order force we are interested in.

The solution of (61) for ρ_1 is obtained by a procedure precisely analogous to that in §2, with the result (cf. 23 and 24)

$$\omega_{\mu\nu}\dot{Z}_\nu = \partial_\mu\mathcal{E} - \varepsilon\mathcal{K}_{\mu\nu}\dot{Z}_\nu + O\left(\varepsilon^2\right) \tag{66}$$

where

$$\mathcal{K}_{\mu\nu} = -\frac{1}{\partial_\mathcal{E}\Omega}\partial_\mathcal{E}\left[\partial_\mathcal{E}\Omega\int_0^\infty \mathrm{d}\tau\left\langle\left(\partial_\mu\tilde{\mathcal{H}}\right)_\tau\partial_\nu\tilde{\mathcal{H}}\right\rangle_\mathcal{E}\right] \tag{67}$$

(Here as above we are neglecting Jarzynski's force, associated with f_1 in equation (20).) Equation (66) can be written

$$\begin{aligned} &\omega'_{\mu\nu}(Z)\dot{Z}_\mu = \partial_\nu\mathcal{E}(Z), \quad \text{where} \\ &\omega'_{\mu\nu}(Z) \equiv \omega_{\mu\nu} + \varepsilon\mathcal{K}_{\mu\nu}(Z) \end{aligned} \tag{68}$$

In this effective slow equation of motion, accurate to order ε, the effect of the Born-Oppenhemer (ε^0) Hamiltonian $\mathcal{E}(Z)$ is modified by altering the symplectic form from ω to ω'. In the case where the fast motion in integrable, this phenomenon was noted by Gozzi and Thacker (1987) and (in the half-classical case) by Littlejohn and Flynn (1991). Because $\mathcal{K}_{\mu\nu}$ incorporates generalized friction through its symmetric part,

ω' is no longer symplectic, and the modified equations of motion are no longer Hamiltonian. The antisymmetric part of $\mathcal{K}_{\mu\nu}$ embodies a generalization of geometric magnetism.

This more general theory can of course reproduce the results of §2. This follows from the formulae, appropriate to that special case

$$\begin{aligned} &Z = (\boldsymbol{R}, V); \quad \mathcal{H}(Z) = \tfrac{1}{2}|V|^2 + h(z, \boldsymbol{R}); \\ &\mathcal{T}(Z) = \tfrac{1}{2}|V|^2 + E(\boldsymbol{R}); \quad \mathcal{K}_{\mu\nu} = \begin{pmatrix} K_{ij} & 0 \\ 0 & 0 \end{pmatrix} \end{aligned} \tag{69}$$

where K_{ij} is defined in (23).

Acknowledgment

We thank Professor V.I. Arnold for a suggestion leading to the generalization in §5, and Dr C. Jarzynski for sending us his paper before publication and thereby enabling us to correct an omission. JMR acknowledges support from the NSF (grant no. Int-9203313).

Appendix A. Solution of the Poisson-bracket equation (16)

To show that the expression (18) is a solution, we substitute it into (16):

$$\begin{aligned} \{h, f\} &= \int_{-\infty}^{0} \mathrm{d}\tau \{h(z), g(z_\tau(z))\} = -\int_{-\infty}^{0} \mathrm{d}\tau \{h(z_\tau(z)), g(z_\tau(z))\} \\ &= -\int_{-\infty}^{0} \mathrm{d}\tau \left(\partial_{q_\tau} h \cdot \partial_{p_\tau} g - \partial_{p_\tau} h \cdot \partial_{q_\tau} g\right) = \int_{-\infty}^{0} \mathrm{d}\tau \frac{\mathrm{d}}{\mathrm{d}\tau} g(z_\tau(z)) = g(z) \end{aligned} \tag{A1}$$

The second equality follows from the invariance of h under its own flow, and the third from the invariance of the Poisson bracket under τ translation, because this is a canonical transformation.

A similar argument shows that the 'anticausal' expression

$$f(z) = \int_{0}^{\infty} \mathrm{d}\tau g(z_\tau(z)) \tag{A2}$$

is also a solution of (16). If employed instead of (18), this correctly gives geometric magnetism according to (25) and (26), but generates antifriction (feeding energy from chaos into the slow motion) instead of friction.

References

Aharonov Y and Stern A 1992 *Phys. Rev. Lett.* **69** 3593-3597

Arnold V I, Kozlov V V & Neishtadt A I 1988 'Dynamical systems III: Mathematical aspects of classical and celestial mechanics' *Encyclopedia of Mathematical Science*, vol. 3,(Springer: Berlin)

Berry M V 1984a *Proc.Roy.Soc. Lond.* **A392** 45-57

Berry M V 1984b *J.Phys.A.* **17** 1225-1233

Berry M V 1989 'The quantum phase, five years after' in pp 7-28 of Shapere and Wilczek 1989

Berry M V and Robbins J M 1993 Submitted to *Proc. Roy. Soc. Lond.*

Goddard P and Olive D I 1978 *Rep. Prog. Phys.* **41** 1357-1437

Gozzi E and Thacker W D 1987 *Phys.Rev.D* **35** 2388-2397

Greenwood D A 1958 *Proc. Phys. Soc.* **71** 585-596

Hill and Wheeler 1952 *Phys. Rev.* **89** 1102-1045

Jackiw R 1988 *Comm. At. Mol. Phys.* **21** 71-82

Jarzynski C 1993 *Phys. Rev. Lett.* (in press)

Kubo R, Toda M and Hashitsume N 1985 *Statistical physics II: Nonequilibrium Statistical mechanics* (Springer: Berlin)

Landau L D and Lifshitz E M 1980 *Statistical Physics, Part 1* (3rd ed. Pergamon Press: Oxford)

Littlejohn R G and Flynn W G 1991 *Phys. Rev. A* **44** 5239-5256

Lochak P and Meunier C 1988 *Multiphase averaging for classical systems* (Springer: New York).

Mead C A and Truhlar D G 1979 *J Chem Phys* **70** 2284-2296 (reprinted in Shapere and Wilczek 1989)

Messiah A 1962 *Quantum Mechanics* (Wiley: New York) vol.2

Ott E 1979 *Phys. Rev. Lett.* **42** 1628-1631

Brown R, Ott E & Grebogi C 1987 *J. Stat. Phys.* **49** 511-550

Robbins J M and Berry M V 1992a *Proc. Roy. Soc. Lond.* **A436**, 631-661

Robbins J M and Berry M V 1992b *J. Phys. A* **25** L961-965

Shapere A and Wilczek F (eds.) 1989 *Geometric Phases in Physics* (Singapore: World Scientific)

Thouless D, Kohmoto M, Nightingale P & den Nijs M 1982 *Phys. Rev. Lett.* **49** 405-408

Wilkinson 1984, PhD Thesis, Department of Physics, University of Bristol

Wilkinson 1987 *J. Phys. A* **20** 2415-2423

Wilkinson 1988 *J. Phys. A* **21** 4021-4037

Wilkinson 1990 *J. Phys. A* **23** 3603-3611

FASTER THAN FOURIER

Michael Berry
H.H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, U.K.

Written to celebrate the 60th Birthday of Yakir Aharonov: deep, quick, subtle.

ABSTRACT

Band-limited functions $f(x)$ can oscillate for arbitrarily long intervals arbitrarily faster than the highest frequency they contain. A class of integral representations exhibiting these 'superoscillations' is described, and by asymptotic analysis the origin of the phenomenon is shown to be complex saddles in frequency space. Computations confirm the existence of superoscillations. The price paid for superoscillations is that in the infinitely longer range where $f(x)$ oscillates conventionally its value is exponentially larger. For example, to reproduce Beethoveen's ninth symphony as superoscillations with a 1Hz bandwidth requires a signal $\exp\{10^{19}\}$ times stronger than with conventional oscillations.

1. Model for superoscillations

My purpose is to decribe some mathematics inspired by Yakir Aharonov during a visit to Bristol several years ago. He told me that it is possible for functions to oscillate faster than any of their Fourier components. This seemed unbelievable, even paradoxical; I had heard nothing like it before, and learned only recently of just one related paper[1] in the literature on Fourier analysis (see §4). Nevertheless, Aharonov and his colleagues had constructed such 'superoscillations' using quantum-mechanical arguments[2]. Here I will exhibit a large class of them, and use asymptotics and numerics to study their strange properties in detail.

Consider functions $f(x)$ whose spectrum of frequencies k is band-limited, say by $|k|\leq 1$, so that on a conventional view f should oscillate no faster than $\cos(x)$. But we wish f to be superoscillatory, that is to vary as $\cos(Kx)$, where K can be arbitrarily large, for an arbitrarily long interval in x. A representation that achieves this is

$$f(x,A,\delta)=\frac{1}{\delta\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{d}u\,\exp\{\mathrm{i}xk(u)\}\exp\left\{-\frac{1}{2\delta^2}(u-\mathrm{i}A)^2\right\} \qquad (1)$$

where the wavenumber function $k(u)$ is even, with $k(0)=1$ and $|k|\leq 1$ for real u, A is real and positive, and δ is small. Examples are

$$k_1(u)=\frac{1}{1+\frac{1}{2}u^2},\quad k_2(u)=\operatorname{sech}u,\quad k_3(u)=\exp\left\{-\tfrac{1}{2}u^2\right\},\quad k_4(u)=\cos u \qquad (2)$$

Aharonov's reasoning (he suggested Eq.(1) with k_4) was that when δ is small the second exponential would act like a 'complex delta-function' and so project out the value of the first exponential at u=iA. Thus f should vary as

$$f \approx \exp\{iKx\} \quad \text{where } K = k(iA) \tag{3}$$

Under the conditions above Eq.(2), k increases from u=0 along the imaginary axis, so that K>1, (and for the given examples can be arbitrarily large), and so corresponds to superoscillations. What follows is a study of the small-δ asymptotics of the integral representing f. As well as justifying Aharonov's argument, this will dissolve the paradox posed by superoscillations, by showing that when x>O($1/\delta^2$) they get replaced by the expected cos(x), and f gets exponentially large.

2. Asymptotics

The aim is to get an asymptotic approximation for small δ to the integral defining f, Eq.(1), which is valid uniformly in x. To achieve this, it is convenient to define

$$\xi \equiv x\delta^2 \tag{4}$$

so that Eq.(1) can be written

$$f\left(\xi/\delta^2, A, \delta\right) = \frac{1}{\delta\sqrt{2\pi}} \int_{-\infty}^{\infty} du\, \exp\left\{-\frac{1}{\delta^2}\Phi(u,\xi,A)\right\} \quad \text{where } \Phi \equiv \tfrac{1}{2}(u - iA)^2 - i\xi\, k(u) \tag{5}$$

For small δ, f can now be approximated by the saddle-point method, that is by deforming the path of integration through saddles u_s of the exponent and replacing Φ by its quadratic approximation near u_s. f is dominated by the saddle with smallest ReΦ. Saddles, whose location depends on ξ (and also A) are defined by

$$\frac{d\Phi}{du} = 0, \quad \text{i.e. } u_s = i\left[\xi\, k'(u_s) + A\right] \tag{6}$$

Application of the saddle-point method now gives the main result:

$$f \approx \frac{\exp\left\{ix\,k(u_s) - \frac{1}{2\delta^2}(u_s - iA)^2\right\}}{\sqrt{1 - ix\delta^2 k''(u_s)}} \tag{7}$$

To interpret this formula, it is necessary to understand the behaviour of the dominant saddle as ξ varies.

When ξ<<1, that is x<<δ^{-2}, Eq.(6) gives u_s≈iA, and (7) reduces to Eq.(3); this is the regime of superoscillations. When ξ>>1, that is x>>δ^{-2}, the saddles are the zeros of $k'(u)$; assuming for simplicity that k has a single maximum at u=0 (as in the first three functions in Eq.(2)), this is the only real saddle, and (7) reduces to

$$f \approx \frac{1}{\delta\sqrt{x|k''(0)|}} \exp\left\{ix - \tfrac{1}{4}\pi\right\} \exp\left\{\frac{A^2}{2\delta^2}\right\} \tag{8}$$

This is the behaviour to be expected conventionally, that is on the basis of the frequency content of f; in the infinite range of validity of Eq.(8), f is O(exp{$A^2/2\delta^2$}) and so exponentially amplified relative to the superoscillation regime.

As x increases, the saddle moves from iA to 0 along a curved track, illustrated in figure 1. This is the dominant saddle u_s; its track resembles figure 1 for all $k(u)$ of this type that I have studied. There are other solutions of Eq.(6), whose arrangement and motion are complicated and depend on the details of $k(u)$, but they are not dominant and so do not compromise the validity of Eq.(7) as the leading-order approximation to the integral defining f, Eq.(1).

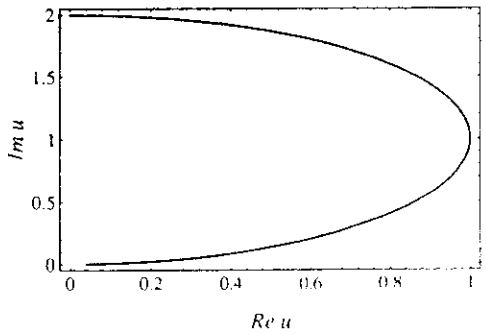


Figure 1. Track of leading saddle u_s as ξ increases from 0 to ∞, for the wavenumber function $k_5(u)$ in Eq.(10), for A=2 (the track is similar for any $k(u)$ with a single maximum)

In understanding the oscillations, it is helpful to study the local wavenumber, defined as

$$q(\xi) \equiv -\mathrm{Im}\frac{\partial\Phi\left\{u_s(\xi),\xi,A\right\}}{\partial\xi} = \mathrm{Re}\,k\left(u_s(\xi)\right) \tag{9}$$

As illustrated in figure 2, $q(\xi)$ decreases smoothly from k(iA) (which is real) to 1 as ξ increases. Note that the decrease is rapid (this is true for all $k(u)$ that I have studied). This has the important implication that to observe superoscillations it is necessary to keep ξ

much smaller than unity, and if we want to allow x to be large, in order to observe *many* superoscillations, δ must be correspondingly smaller, Eq.(4), and the exponential amplification in the regime of conventional oscillation, Eq.(8), will be correspondingly larger.

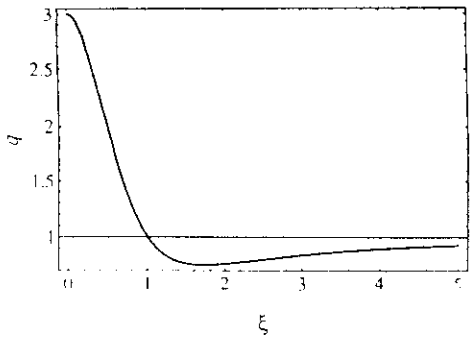


Figure 2. Local wavenumber $q(\xi)$, Eq.(9), for the $k_S(u)$ in Eq.(10), for $A=2$

None of the wavenumber functions in Eq.(2) gives an f whose integral representation can be evaluated exactly in terms of special functions. However, if we choose the wavenumber function

$$k_S(u) = 1 - \tfrac{1}{2}u^2 \tag{10}$$

we can ensure that it is band-limited ($|k|<1$) by restricting the range of integration in Eq.(1) to $|u|<2$. The resulting truncated integral is

$$f(x,A,\delta) = \frac{1}{\delta\sqrt{2\pi}}\int_{-2}^{2} du \exp\left\{ix\left(1-\tfrac{1}{2}u^2\right)\right\}\exp\left\{-\frac{1}{2\delta^2}(u-iA)^2\right\} \tag{11}$$

which be expressed in terms of error functions:

$$\begin{aligned} f(x,A,\delta) = {} & \frac{1}{2\sqrt{1+ix\delta^2}}\exp\left\{\frac{ix\left(2+A^2+2ix\delta^2\right)}{2\left(1+ix\delta^2\right)}\right\}\times \\ & \times\left[\operatorname{erf}\left\{\frac{2+iA+2ix\delta^2}{\delta\sqrt{2+2ix\delta^2}}\right\}+\operatorname{erf}\left\{\frac{2-iA+2ix\delta^2}{\delta\sqrt{2+2ix\delta^2}}\right\}\right] \end{aligned} \tag{12}$$

It is instructive to examine this in detail. The superoscillation wavenumber, Eq.(3), is

$$K = k_S(iA) = 1 + \tfrac{1}{2}A^2 \tag{13}$$

There is a single saddle, at (figure 1)

$$u_S(\xi) = \frac{iA}{1+i\xi} \tag{14}$$

and the local wavenumber is (figure 2)

$$q(\xi) = 1 + \frac{A^2\left(1-\xi^2\right)}{2\left(1+\xi^2\right)^2} \tag{15}$$

For this case, the saddle-point approximation, Eq.(7) gives

$$f(x,A,\delta) = \frac{1}{\sqrt{1+ix\delta^2}}\exp\left\{ix\left[1+\frac{A^2}{2\left(1+x^2\delta^4\right)}\right]\right\}\exp\left\{\frac{A^2\delta^2x^2}{2\left(1+x^2\delta^4\right)}\right\} \tag{16}$$

However, the asymptotics of (11) includes contributions from the end-points $u=\pm2$ as well as the saddle u_S. This can be seen by realising that the steepest path between -2 and +2 runs from -2 to infinity in the negative half-plane, through u_S to infinity in the positive half-plane, and back to +2. The end-point contributions oscillate conventionally, with the wavenumber -1, so we must be sure that they do not mask the superoscillations that exist for small ξ. The condition for this is that the absolute value of the Gaussian in (11) must not exceed unity at the end-points. Thus

$$\exp\left\{\frac{A^2-4}{2\delta^2}\right\} \le 1, \qquad \text{i.e.} \qquad A \le 2 \tag{17}$$

(we include the equality because the end-point contribution is smaller than that from the saddle by a factor δ). Eq.(13) now implies that the maximum rate of superoscillation obtainable with this model is $K=3$. (It is worth remarking that $x=0$, $A=2$ lies on the anti-Stokes line for the error functions in Eq.(12), that is, where the exponential contribution from the saddle exchanges dominance with those from the end-points.)

The representation Eq.(1) does not have the form of a Fourier transform, namely (for a band-limited function)

$$f(x,A,\delta) = \int_{-1}^{1} dq \exp\{ixq\}\tilde{f}(q) \tag{18}$$

It is however easy to cast it into this form. The transform $\tilde{f}(q)$ depends on the inverse function of $k(u)$; this is multivalued, and the path of integration can be deformed into a loop around a cut extending along the real axis negatively from the branch point at q=1 (the ends of the loop are pinned to the cut, at q=-1 for k_5 and at the essential singularity q=0 for k_1, k_2, and k_3). Again there is a dominant saddle, which for small ξ lies at q=K, and the loop can be expanded to pass through this. All previous results can be reproduced in this way.

3. Numerics

The aim here is twofold: to compare the saddle-point approximation Eq.(7) with the exact integral (1), and to exhibit the superoscillations. I carried out computations of f for the wavenumber functions k_1, k_2, and k_3 (Eq.(2)), but will display results only for Re f (Im f is similar) for k_5 (Eq.(10)), with the truncated integral of Eq.(11), for which the results are very similar. The computations will be exhibited for the fastest superoscillations, namely K=3, that is A=2 (Eq.(17)), choosing δ=0.2.

Figure 3 shows the results. The superoscillations for small x, with period $2\pi/3$, are shown on figure 3a, and figure 3b shows a range of x where there are conventional oscillations, with period more than 3 times greater (actually about 8.4 - cf. figure 2, where $\xi \sim 1.6$ corresponds to $x \sim 40$). In both cases, the approximation (in this case Eq.(16)) agrees well with the exact expression, Eq.(12). For example, the fractional error is 0.18 for x=2, and 2.8×10^{-18} for x=42. Note the enormous ratio of the sizes of f for large and small x; from Eq.(16), this can be estimated as $\exp(36)\sim10^{16}$ (the asymptotic ratio of Eq.(8) is not attained in figure 3b). The transition between the superoscillation and conventional regimes is clearly shown in figure 3c.

In these computations, the value A=2 is the largest for which the saddle dominates the end-points. The competition between contributions shows up most clearly at x=0, for which (12) gives

$$f(0,A,\delta) = \mathrm{Re}\,\mathrm{erf}\left\{\frac{1}{\delta}\left(\sqrt{2} + \mathrm{i}\frac{A}{\sqrt{2}}\right)\right\} \tag{19}$$

For A<2, f is well approximated by the saddle contribution of unity, for A>2, the end-points dominate and f increases exponentially, Eq.(17), masking the superoscillations for small x. This is illustrated in figure 4. Even at the critical value A=2, that is, on the anti-Stokes line for the function (19), the exact value f=0.945 is close to the saddle-point value f~1.

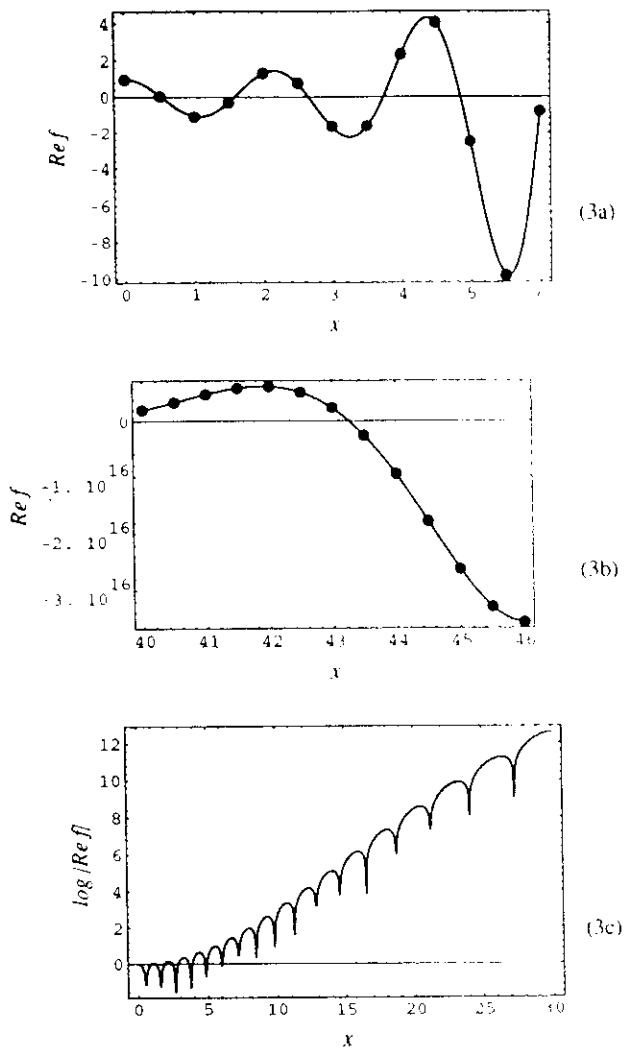


Figure 3. Computations of $f(x,2,0.2)$ for the truncated integral, Eq.(11), showing (a), superoscillations, and (b) conventional oscillations. Circles: exact expression, Eq.(12); full lines: saddle-point approximation, Eq.(16). In (c) the logarithms are base 10

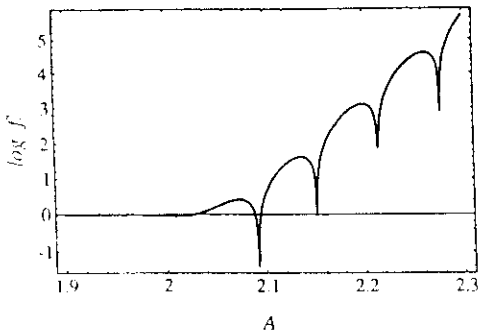


Figure 4. Computations of $\log|f(0, A, 0.2)|$, Eq.(19), for the truncated integral Eq.(11); logarithms are base 10. Note the exponential growth after crossing the anti-Stokes line at A=2.

4. Beethoven at 1Hz

Professor I. Daubechies has informed me that superoscillations are known in signal processing, in the context of oversampling. This is sampling a function faster than the Nyquist rate, i.e. at points $x=n\pi$ where the function is band-limited by $|k|\le 1$. If a function is oversampled in a finite range, extrapolation outside this range is exponentially unstable[2]. She quotes B. Logan as saying that it is possible in principle to design a bandlimited signal with a bandwidth of 1Hz that would reproduce Beethoven's ninth symphony exactly. With the superoscillatory functions described in this paper it is possible to give an explicit recipe for constructing this signal, as I now explain.

We require superoscillations for the duration T (~4000s) of the symphony. Therefore the desired signal $B(t)$ can be represented as periodic outside this interval, namely

$$B(t) \quad \sum_{-N}^{N} B_n \exp\left\{\mathrm{i}\frac{2\pi nt}{T}\right\} \tag{20}$$

Here N is the order of the Fourier component corresponding to the highest frequency $\nu_{\max} = N/T \sim 20\text{kHz}$ it is desired to reproduce.

To approximate this with a signal band-limited by frequency ν_0 (=1Hz) we make the replacement

$$\exp\left\{\mathrm{i}\frac{2\pi nt}{T}\right\} \to \Phi_n(t) \tag{21}$$

where (cf.Eq.(1)) Φ_n is the superoscillatory function

$$\Phi_n(t) \equiv \frac{1}{\delta_n\sqrt{2\pi}} \int_{-\infty}^{\infty} du \exp\{\mathrm{i}2\pi t\nu(u)\}\exp\left\{-\frac{1}{2\delta_n^2}(u-\mathrm{i}A_n)^2\right\} \tag{22}$$

Here the frequency function $\nu(u)$ never exceeds (for real u) its band-limited value $\nu(0)\equiv\nu_0$, and A_n and δ_n will now be determined by the requirement that Φ_n superoscillates with frequency n/T for time T.

The superoscillation frequency of $\Phi_n(t)$ is $\nu(\mathrm{i}A_n)$ (cf. Eq.(3)). Thus from Eq.(21) A_n must satisfy

$$\nu(\mathrm{i}A_n) = \frac{n}{T} \tag{23}$$

We fix δ_n by requiring that the superoscillations are maintained for time T, in the sense that the replacement of Eq.(21) remains a good approximation. For this we require the next correction to the superoscillatory exponential that $\Phi_n(t)$ represents. Expanding the saddle-point approximation to Eq.(22) (analogous to Eq.(7)) for small t, we find

$$\Phi_n(t) \approx \exp\left\{\mathrm{i}\frac{2\pi nt}{T}\right\}\exp\left\{2\pi^2\delta_n^2\left[-\nu''(\mathrm{i}A_n)\right]t^2\right\} \tag{24}$$

The second factor is an increasing exponential, because $\nu''(\mathrm{i}A_n)$ is imaginary, and must remain close to unity for $0<t<T$. Thus

$$\delta_n << \left[2\pi\left|\nu''(\mathrm{i}A_n)\right|T\right]^{-1} \tag{25}$$

Choosing A_n and δ_n as in Eqs.(23) and (25) guarantees that the signal $B_n(t)$, with its frequencies up to $\nu_{\max}$, will be imitated for time T. When $t>T$ the imitation will grow rapidly in strength, and eventually, that is when it is oscillating at the frequency ν_0 corresponding to its Fourier content, it will acquire an amplification factor corresponding to its largest Fourier component $n=N$. An argument analogous to that leading to Eq.(8) gives this factor as

$$F = \exp\left\{\frac{A_N^2}{2\delta_N^2}\right\} >> \exp\left\{A_N^2\pi^2T^2\left|\nu_N''(\mathrm{i}A_N)\right|^2\right\} \tag{26}$$

with A_N determined by Eq.(23) with the right-hand side set equal to $\nu_{\max}$.

Let us calculate this amplification for the model frequency function

$$\nu(u) = \nu_0\exp\left\{-u^2\right\} \tag{27}$$

(cf. $k_3(u)$ in Eq.(2)). We find

$$A_N^2 = \log\left\{\frac{\nu_{\max}}{\nu_0}\right\} \tag{28}$$

and hence, from Eq.(26),

$$F >> \exp\left\{4\pi^2 \log^2\left(\frac{\nu_{\max}}{\nu_0}\right)\nu_{\max}^2 T^2\right\} \tag{29}$$

For Beethoven's ninth symphony this gives

$$F >> \exp\left\{10^{19}\right\} \tag{30}$$

This amplification will not be achieved until a time t_F, which can be estimated by the argument preceding Eq.(8) as

$$t_F \sim \left[\nu_0 \delta_N^2\right]^{-1} \sim \frac{\nu_{\max}^2 T^2}{\nu_0} \sim 10^8 \text{years} \tag{31}$$

Other choices for $w(u)$ give similar expressions and numerical estimates.

The estimate of Eq.(30) indicates that to reproduce music as superoscillations requires a signal with so much energy as to be hopelessly impractable, but more modest bandwidth compression might be feasible.

5. Concluding remarks

Aharonov's discovery, elaborated here, could have applications in several branches of physics. One possibility is the use of superoscillations for bandwidth compression as discussed in §4. Another example, also in signal processing, concerns the observation of oscillations faster than those expected on the basis of applied or inferred filters. These would conventionally be interpreted as high frequencies leaking through imperfect filters, but the arguments presented here show that the phenomenon could have a quite different origin, namely superoscillations compatible with perfect filtering.

Perhaps more interesting are the possible applications of superoscillatory functions of two variables, representing images. One envisages new forms of microscopy, in which structures much smaller than the wavelength λ would be resolved by representing them as superoscillations. (This is different from conventional superresolution, which is based on the fact that Fourier components larger than $2\pi/\lambda$ can be present in the field near the surface of an object, but decay exponentially away from the object because the wavenumber in the perpendicular direction is imaginary. With superoscillations, the larger Fourier components are not present.)

Superoscillations can probably exist in random functions $f(x)$: arbitrarily long intervals, in which f is exponentially small relative to elsewhere, could superoscillate. Consider how this might be achieved. If f is Gauss-distributed, its statistics are completely described by its autocorrelation function, which by the Wiener-Khinchin theorem is the Fourier transform of the power spectrum $S(q)$ of f. Even if f is band-limited, it ought to be possible to choose $S(q)$ with analytic structure (saddles with Re q >1, etc.) such that the autocorrelation superoscillates as it falls from its initial value. This idea is worth pursuing.

On the purely mathematical side, it is clear that superoscillations carry a price: the function is exponentially smaller than in the regime of conventional oscillations, with the exponent increasing with the size of the interval of superoscillations. We have seen examples of this, but there ought to be a general theorem (perhaps based on a version of the uncertainty principle).

6. Acknowledgments

I thank Professors Jeeva Anandan and John Safko for arranging, and generously supporting my participation in, Yakir Aharonov's birthday meeting, and inviting me to write this paper, and Professor Ingrid Daubechies for suggestions leading to the calculation of §4.

7. Reference

1. H.J. Landau, IEEE Trans. Inf. Theory, **IT-32** (1986) 464
2. Y. Aharonov, J. Anandan, S. Popescu and L. Vaidman, *Phys.Rev.Lett.* **64** (1990) 2965.

The geometric phase for chaotic systems†

By J. M. Robbins and M. V. Berry

H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, U.K.

The geometric phase acquired by the eigenstates of cycled quantum systems is given by the flux of a two form through a surface in the system's parameter space. We obtain the classical limit of this two-form in a form applicable to systems whose classical dynamics is chaotic. For integrable systems the expression is equivalent to the Hannay two form. We discuss various properties of the classical two-form, derive semiclassical corrections to it (associated with classical periodic orbits), and consider implications for the semiclassical density of degeneracies.

1. Introduction

Since its discovery (Berry 1984), considerable attention has been devoted to the geometric phase γ_n acquired by the eigenstates $|n\rangle$ of quantum systems $\hat{H}(R)$ whose parameters R are taken through a cycle C. According to one of a number of equivalent expressions

$$\gamma_n = -\frac{1}{\hbar}\int_S \Gamma_n \tag{1.1}$$

where S is a surface in parameter space bounded by C, and

$$\Gamma_n = -i\hbar\langle \mathrm{d}n| \wedge |\mathrm{d}n\rangle \tag{1.2}$$

is the two form whose flux through S is the geometric phase. There exist several extensive reviews of the geometric phase (see, for example, Shapere & Wilczek 1989). (Note: multiplying by $-i$ in (2) is equivalent to taking the imaginary part, and the factor of $\hbar$ introduced will render Γ_n $\hbar$ independent in the classical limit.)

Given Γ_n, a quantum mechanical quantity of geometric origin, it is natural to ask what it corresponds to classically. For integrable systems this question was answered by Hannay (1985), who discovered a classical anholonomy for cycled integrable systems, analogous to the geometric phase. Hannay's two-form was subsequently shown to correspond to Γ_n in the classical limit (Berry 1985).

More generally, the correspondence principle, couched in geometrical language, asserts that in the classical limit, the spectral invariants of quantum systems correspond to the invariant manifolds of classical systems. For integrable systems the invariant manifolds are tori, and this correspondence is embodied in the torus wave functions and quantization conditions (Berry 1983) which form the basis for the semiclassical analysis of Berry (1985). The invariant manifolds of chaotic systems are the energy shells and the isolated periodic orbits contained therein. Semiclassical quantization conditions in terms of these are fundamentally more difficult than for the integrable case; the quest for such conditions lies at the heart of *quantum chaology* (see, for example, Gutzwiller 1990; Berry 1991; Keating 1991).

† This paper was accepted as a rapid communication

Proc. R. Soc. Lond. A (1992) **436**, 631–661
Printed in Great Britain

As Hannay's two-form is associated with invariant tori, one would expect the classical two-form for chaotic systems to be associated with the energy shell and periodic orbits. A theory along these lines, along with some of the difficulties encountered, was previously discussed informally in Berry (1990). Here we present a complete account of the classical limit of the two-form in a form applicable to chaotic systems.

Before proceeding, let us mention an interesting though quite distinct extension of the Hannay two-form due to Montgomery (1988) and Golin *et al.* (1989). These authors consider classical hamiltonian systems with parameter-dependent continuous symmetries (though not necessarily integrable), and establish the existence of a unique (hamiltonian) connection, i.e. a prescription for lifting curves from parameter space to phase space (with certain additional properties). Crudely speaking, the connection describes the geometrical component of motion in the 'ignorable' coordinates (i.e. those conjugate to the momenta which generate the symmetries); in the integrable case these momenta are the actions, and the connection determines the geometrical component of the angle evolution, namely the Hannay angles.

Here we are considering a different problem; our concern is the intrinsic anholonomy, defined quantally but so far not classically, associated with ergodic hamiltonians with no symmetries at all. One might attempt to apply the preceding formalism to such systems, by regarding the dynamics itself as the symmetry, but the associated connection is ill defined, as the expressions for it diverge exponentially.

The paper is organized as follows. We introduce a time-dependent formalism for the quantum two-form (§2), from which its classical limit, our principal result, follows directly (§3). The cases of anticanonical symmetries and additional constants of the motion (in particular, integrable systems) are considered (§4), along with some specific examples (§5). We then develop an alternative formalism for both the quantum and classical two-forms (§6), which is used to establish formally that the two form is closed (§7). Finally, we calculate the periodic orbit contributions to the two form (§8) and its derivative, the density of degeneracies (§9). In the interest of maintaining continuity, the derivations of some results have been placed in Appendixes. Throughout we use the notation of differential forms. Arnold (1978) provides a good general reference.

2. Time-dependent quantum formalism

(a) *Derivation*

For chaotic systems, the classical limit of (1.2) is not directly accessible, because (and in contrast to integrable systems), semiclassical eigenstates are not known. As is customary in quantum chaology, we proceed by expressing the spectral property of interest (in this case, Γ_n) within a time-dependent formalism. Taking the classical limit is then a straightforward matter.

We start with the equation (Berry 1984)

$$\Gamma_n = -i\hbar \sum_{m\neq n} \frac{\langle n|\mathrm{d}\hat{H}|m\rangle \wedge \langle m|\mathrm{d}\hat{H}|n\rangle}{(E_n - E_m)^2}. \tag{2.1}$$

Here, as elsewhere, d is the exterior derivative with respect to parameters R. Throughout the paper the R dependences are usually left implicit, though occasionally in the interest of clarity they are indicated explicitly.

The energy denominator in (2.1) may be expressed as a time integral:

$$(E_n - E_m)^{-2} = -\frac{1}{\hbar^2}\lim_{\epsilon \to 0}\int_0^\infty \mathrm{d}t\, \mathrm{e}^{-\epsilon t}\cos\omega_{nm}t, \tag{2.2}$$

where $\omega_{nm} = (E_n - E_m)/\hbar$. (Usually the convergence factor $\lim_{\epsilon \to 0}\exp(-\epsilon t)$ is left implicit.) The oscillations $\cos\omega_{nm}t$ may then be incorporated into time-dependent matrix elements, as in

$$\cos\omega_{nm}t\,\langle n|\mathrm{d}\hat{H}|m\rangle \wedge \langle m|\mathrm{d}\hat{H}|n\rangle = \tfrac{1}{2}(\langle n|(\mathrm{d}\hat{H})_t|m\rangle \wedge \langle m|\mathrm{d}\hat{H}|n\rangle + \langle n|\mathrm{d}\hat{H}|m\rangle \wedge \langle m|(\mathrm{d}\hat{H})_t|n\rangle). \tag{2.3}$$

Here $(\mathrm{d}\hat{H})_t = U^\dagger(t)\,(\mathrm{d}\hat{H})\,U(t)$ is the time-evolved operator, in which $U(t)$ is the evolution operator at fixed R, namely

$$U(t) = \mathrm{e}^{-\mathrm{i}\hat{H}t/\hbar}. \tag{2.4}$$

With the substitution of (2.3) and (2.2) into (2.1), the restriction $m \neq n$ on the sum is no longer necessary, and $\sum_m |m\rangle\langle m|$ gives the identity. Therefore

$$\Gamma_n = \frac{\mathrm{i}}{2\hbar}\int_0^\infty \mathrm{d}t\,\langle n|(\mathrm{d}\hat{H})_t \wedge \mathrm{d}\hat{H} + \mathrm{d}\hat{H} \wedge (\mathrm{d}\hat{H})_t|n\rangle. \tag{2.5}$$

The sum of operators appearing in (2.5) is actually a commutator, i.e.

$$(\mathrm{d}\hat{H})_t \wedge \mathrm{d}\hat{H} + \mathrm{d}\hat{H} \wedge (\mathrm{d}\hat{H})_t = [(\mathrm{d}\hat{H})_t \wedge \mathrm{d}\hat{H}]. \tag{2.6}$$

At first this might appear surprising, since the commutator of two scalar operators is antisymmetric in its arguments. However, the commutator of operator one-forms is symmetric, the two antisymmetries cancelling as it were. To clarify this point, let us consider (as we often will in what follows) the 'reference' area element $\square_R$ in parameter space drawn in figure 1*a*, spanned by infinitesimal displacements r_1 and r_2 from R. The flux through $\square_R$ of $\mathrm{d}\hat{A} \wedge \mathrm{d}\hat{B} + \mathrm{d}\hat{B} \wedge \mathrm{d}\hat{A}$, as computed from the usual rules for two forms, is

$$(\hat{A}_1\hat{B}_2 - \hat{A}_2\hat{B}_1) + (\hat{B}_1\hat{A}_2 - \hat{B}_2\hat{A}_1) \tag{2.7}$$

(here $\hat{A}_i \overset{\mathrm{def}}{=} \mathrm{d}\hat{A}\cdot r_i$ and similarly for $\hat{B}$), whereas the flux of $[\hat{A} \wedge \hat{B}]$ through $\square_R$ is $[\hat{A}_1, \hat{B}_2] - [\hat{A}_2, \hat{B}_1]$, or

$$(\hat{A}_1\hat{B}_2 - \hat{B}_2\hat{A}_1) - (\hat{A}_2\hat{B}_1 - \hat{B}_1\hat{A}_2). \tag{2.8}$$

Clearly (2.7) and (2.8) are the same.

Substituting (2.6) into (2.5) we obtain

$$\Gamma_n = \frac{\mathrm{i}}{2\hbar}\int_0^\infty \mathrm{d}t\,\langle n|[(\mathrm{d}\hat{H})_t \wedge \mathrm{d}\hat{H}]|n\rangle, \tag{2.9}$$

our principal formula for the quantum two-form. We may write it in a form more symmetrical with respect to time. Since expectation values of eigenstates are time invariant, $\langle n|(\mathrm{d}\hat{H})_t \wedge \mathrm{d}\hat{H}|n\rangle = \langle n|\mathrm{d}\hat{H} \wedge (\mathrm{d}\hat{H})_{-t}|n\rangle$. Therefore

$$\Gamma_n = \frac{\mathrm{i}}{4\hbar}\int_0^\infty \mathrm{d}t\,\langle n|[(\mathrm{d}\hat{H})_t + (\mathrm{d}\hat{H})_{-t}) \wedge \mathrm{d}\hat{H}]|n\rangle. \tag{2.10}$$

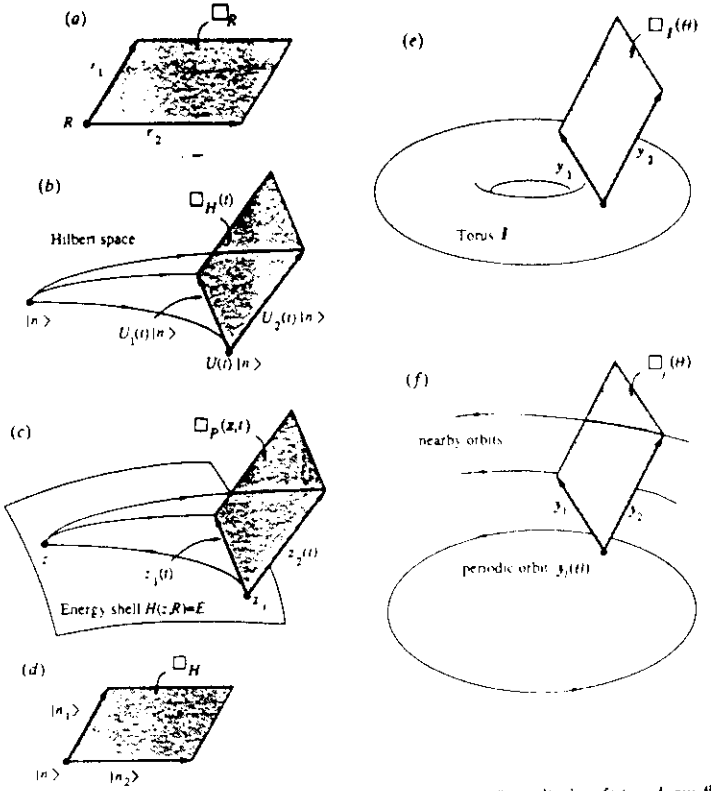


Figure 1. (*a*) area element $\square_R$ spanned by displacements r_1 and r_2 from R. (*b*) (*f*) two-form fluxes through $\square_R$ as follows. (*b*) The time-dependent formula (6.2) relates $\Gamma_n \cdot \square_R$ to the time average of the symplectic area of $\square_H(t)$ in Hilbert space. (*c*) The time-dependent formula (6.8) relates $\Gamma^\epsilon(E)\cdot\square_R$ to the time and microcanonical averages of the symplectic area of $\square_p(z,t)$ in phase space. (*d*) The time-independent formula (1.2) relates $\Gamma_n \cdot \square_R$ to the symplectic area of $\square_H$ in Hilbert space. (*e*) The time-independent formula (G.1) relates $\Gamma_J \cdot \square_R$ to the symplectic area of $\square_J(\theta)$ in phase space. (*f*) The periodic orbit two-form (8.10).

(*b*) *Antiunitary symmetries*

An antiunitary symmetry $\hat{K}$, such as time reversal, takes inner products to their complex conjugates and commutes with the hamiltonian:

$$\langle \hat{K}\cdot\psi\,|\,\hat{K}\cdot\phi\rangle = \langle\psi\,|\,\phi\rangle^*, \tag{2.11a}$$

$$\hat{H}\hat{K} = \hat{K}\hat{H}. \tag{2.11b}$$

If (2.11*b*) holds for all parameters, the two-form satisfies

$$\Gamma_{\bar{n}} = -\Gamma_n, \tag{2.12}$$

where $|\bar{n}\rangle = \hat{K}|n\rangle$. In particular, if $|n\rangle$ is invariant under $\hat{K}$, then Γ_n vanishes.

The symmetry property (2.12) is easily obtained from (1.2) and can be also made manifest in the new formula (2.10). We note that

$$\hat{K}(\mathrm{d}\hat{H})_t = (\mathrm{d}\hat{H})_{-t}\hat{K}, \tag{2.13}$$

because (i) time evolved operators $\hat{A}_t$ transform according to $\hat{K}\hat{A}_t = \hat{\bar{A}}_{-t}\hat{K}$, where $\hat{\bar{A}} = \hat{K}\hat{A}\hat{K}^{-1}$, and (ii) from (2.11*b*) $\mathrm{d}\hat{\bar{H}} = \mathrm{d}\hat{H}$. Turning to the expectation value in (2.10), computed for $|\bar{n}\rangle$ rather than $|n\rangle$,

$$\begin{aligned}\langle\bar{n}|[(\mathrm{d}\hat{H})_t + (\mathrm{d}\hat{H})_{-t}\wedge\mathrm{d}H]|\bar{n}\rangle &= \langle\hat{K}n|[(\mathrm{d}\hat{H})_t + (\mathrm{d}\hat{H})_{-t}\wedge\mathrm{d}\hat{H}]\hat{K}n\rangle \\ &= \langle\hat{K}n,\hat{K}|(\mathrm{d}\hat{H})_t + (\mathrm{d}\hat{\bar{H}})_{-t}\wedge\mathrm{d}\hat{H}]n\rangle = \langle n|[(\mathrm{d}\hat{H})_t + (\mathrm{d}\hat{H})_{-t}\wedge\mathrm{d}\hat{H}]|n\rangle^*, \end{aligned}\tag{2.14}$$

where in the third equality we have used (2.13), and in the last equality the antiunitary property (2.11*a*). Since the expectation value of the commutator in the last term is pure imaginary, it follows that

$$\begin{aligned}\langle\bar{n}|[(\mathrm{d}\hat{H})_t + (\mathrm{d}\hat{H})_{-t}\wedge\mathrm{d}\hat{H}]|\bar{n}\rangle &= \langle n|[(\mathrm{d}\hat{H})_t + (\mathrm{d}\hat{H})_{-t}\wedge\mathrm{d}\hat{H}]n\rangle^* \\ &= -\langle n|[(\mathrm{d}\hat{H})_t + (\mathrm{d}\hat{H})_{-t}\wedge\mathrm{d}\hat{H}]n\rangle, \end{aligned}\tag{2.15}$$

and (2.12) follows immediately.

3. Classical limit

(*a*) *Wigner–Weyl formalism*

The classical limit of (2.9) is carried out within the Wigner–Weyl formalism, the defining relation of which is the following correspondence between operators $\hat{A}$ and phase space functions $A(z)$.

$$A(z) = \int \mathrm{d}^N s\,\langle q + \tfrac{1}{2}s|\hat{A}|q - \tfrac{1}{2}s\rangle\, \mathrm{e}^{-\mathrm{i}p\cdot s/\hbar}, \tag{3.1}$$

where $z = (q\ p)$. $A(z)$ is called the Weyl symbol, or simply the symbol of $\hat{A}$; sometimes we write $(\hat{A})_W(z)$ for $A(z)$. There is nothing intrinsically semiclassical or approximate about the Wigner–Weyl correspondence: it is simply another formalism for exact quantum mechanics. However, it lends itself to semiclassical approximations in which the canonical structure of classical mechanics is manifest. For discussions and reviews, see Groenewold (1946), Moyal (1949), Baker (1958).

Γ_n is expressed in (2.9) as the time integral of an expectation value. In the Wigner–Weyl formalism expectation values of operators correspond to phase space averages of symbols. In particular,

$$\langle n|[\mathrm{d}\hat{H}]_t\wedge\mathrm{d}\hat{H}]|n\rangle = \int \mathrm{d}^{2N}z\, W_n(z)\,([(\mathrm{d}\hat{H})_t\wedge\mathrm{d}\hat{H}])_W(z). \tag{3.2}$$

$W_n(z)$ is the symbol of the projection $|n\rangle\langle n|$, and is also called the Wigner function.

For chaotic systems, the simplest and crudest semiclassical approximation is to replace $W_n(z)$ by the microcanonical density, namely

$$W_n(z) \approx \delta(E - H(z))/\Omega'(E) \tag{3.3}$$

Here $H(z)$, an abbreviation for $H(z,R)$, is the classical hamiltonian, i.e. the symbol of $\hat{H}$. The normalization factor

$$\Omega'(E) = \int \mathrm{d}^{2N}z\,\delta(E - H(z))$$

is the volume of the energy shell, whose energy E is quantized in this lowest approximation according to the rule

$$\Omega(E_n) = nh^N, \tag{3.4}$$

which associates a quantum state to each phase space cell of volume h^N.

$$\Omega(E) = \int \mathrm{d}^{2N}z\,\Theta(E - H(z))$$

is the phase space volume with energy less than E. Refinements to (3.3), involving classical periodic orbits, are discussed in §8.

Next we consider $([(\mathrm{d}\hat{H})_t\wedge\mathrm{d}\hat{H}])_W$. To lowest order in $\hbar$, the symbol of the commutator of two operators is given by $\mathrm{i}\hbar$ times the Poisson bracket of their symbols. Thus

$$([(\mathrm{d}\hat{H})_t\wedge\mathrm{d}\hat{H}])_W \to \mathrm{i}\hbar\{(\mathrm{d}H)_t\wedge\mathrm{d}H\}. \tag{3.5}$$

Like its commutator analogue, the Poisson bracket of one-forms is symmetric rather than antisymmetric in its arguments. (Explicitly, the flux of $\{\mathrm{d}A\wedge\mathrm{d}B\}$ through $\square_R$ is given by $\{A_1,B_2\} + \{A_2,B_1\}$, where $A_i \overset{\mathrm{def}}{=} \mathrm{d}A\cdot r_i$, $B_i \overset{\mathrm{def}}{=} \mathrm{d}B\cdot r_i$.) The symbol of a time-evolved operator is given, to lowest order in $\hbar$, by its classically time-evolved symbol.

If we define the time-evolved function $A_t(z) \overset{\mathrm{def}}{=} A(z_t)$, where z_t is the trajectory from z at time t, then

$$((\mathrm{d}\hat{H})_t)_W \to (\mathrm{d}H)_t. \tag{3.6}$$

(An explication of notation might be helpful at this point. $(\mathrm{d}H)_t$, evaluated at z and R, is just $H(z_t,R+\mathrm{d}R) - H(z_t,R)$ to first order in $\mathrm{d}R$.) Then from (3.5) and (3.6),

$$([(\mathrm{d}\hat{H})_t\wedge\mathrm{d}\hat{H}])_W \to \mathrm{i}\hbar\{(\mathrm{d}H)_t\wedge\mathrm{d}H\} \tag{3.7}$$

Let us mention that both (3.5) and (3.6) give the lowest order terms in formal power series expansion in $\hbar$. The next terms are of order $\hbar^2$ higher than the leading one (see Voros 1976), so that the next term in (3.7) is of order $\hbar^3$.

Substituting (3.3) and (3.7) into (3.2) and (2.9) we obtain the classical limit of the two form

$$\Gamma_n \to \Gamma^{\mathrm{cl}}(E) = -\frac{1}{2}\int_0^t \mathrm{d}t\,\langle(\mathrm{d}H)_t\wedge\mathrm{d}H\rangle_E \tag{3.8}$$

(In general $\langle f\rangle_E$ denotes the microcanonical average

$$\int \mathrm{d}^{2N}z\,\delta(E - H)f/\Omega(E)$$

When there is no risk of confusion we simply write $\langle f\rangle$, leaving the energy dependence implicit.) For given n, E is quantized according to the approximation (3.4).

(*b*) *Convergent formula*

For chaotic systems, it is not clear that the expression in (3.8) is convergent. The reason is that the Poisson bracket $\{(\mathrm{d}H)_t\wedge\mathrm{d}H\}(z)$ grows exponentially in time. To see this, note that in general

$$\{A_t,B\}(z) = \boldsymbol{A}'(z_t)\cdot S(z,t)J\cdot\boldsymbol{B}'(z) \tag{3.9}$$

where $\boldsymbol{A}'(\boldsymbol{z})$ and $\boldsymbol{B}(\boldsymbol{z})$ are phase space gradients, $\boldsymbol{J}$ is the Poisson tensor

$$\boldsymbol{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \tag{3.10}$$

and

$$S_{ij}(\boldsymbol{z}, t) = \partial Z_i(\boldsymbol{z}, t)/\partial z_j \tag{3.11}$$

is the linearized flow, where $\boldsymbol{Z}(\boldsymbol{z}, t) \overset{\text{def}}{=} \boldsymbol{z}_t$ is the flow. The exponential growth of $\boldsymbol{S}(\boldsymbol{z}, t)$ (which would imply the same of (3.9)) is the very definition of chaos.

To make sense of the classical limit of the two-form we require a manifestly convergent formula. This can be obtained by means of an identity which eliminates the $\boldsymbol{z}$ derivatives in (3.8).

$$\langle (\mathrm{d}A)_t \wedge \mathrm{d}B \rangle_E = (1/\Omega')(\Omega' \langle (\mathrm{d}\dot{A})_t \wedge \mathrm{d}B \rangle_E)' \tag{3.12}$$

The derivation is given in Appendix A. The prime ($'$) denotes the derivative with respect to energy, and the dot ($\dot{}$) the derivative with respect to time. While (3.9) implies that $\langle A, B \rangle(\boldsymbol{z})$ is exponentially divergent in t, (3.12) implies that it is oscillatory in $\boldsymbol{z}$ and that the exponential oscillations cancel in the main when averaged over the energy shell. Substituting (3.12) into (3.8),

$$\Gamma'(E) = \frac{1}{2\Omega}\left(\Omega \int_0^\infty \mathrm{d}t \langle (\mathrm{d}\dot{H})_t \wedge \mathrm{d}H \rangle_E\right) \tag{3.13}$$

($(\mathrm{d}\dot{H})_t$ denotes the derivative of $(\mathrm{d}H)_t$ with respect to t.) We integrate by parts,

$$\int_0^\infty \mathrm{d}t (\mathrm{d}\dot{H})_t = \lim_{\tau \to \infty} \int_0^\tau \mathrm{d}t\, \mathrm{e}^{-\varepsilon t} (\mathrm{d}H)_t = \int_0^\infty \mathrm{d}t (\mathrm{d}H)_t \tag{3.14}$$

(the reinstated convergence factor of (2.2) justifies the neglect of the boundary terms) and obtain

$$\Gamma'(E) = \frac{1}{2\Omega}\left(\Omega \int_0^\infty \mathrm{d}t \langle (\mathrm{d}H)_t \wedge \mathrm{d}H \rangle_E\right) \tag{3.15}$$

This is our principal formula for the classical two-form. Like the quantum formula (2.9) it can be made more symmetrical with respect to time. Since microcanonical averages are time invariant, we get that $\langle (\mathrm{d}H)_t \wedge \mathrm{d}H \rangle = \langle \mathrm{d}H \wedge (\mathrm{d}H)_{-t} \rangle = -\langle (\mathrm{d}H)_{-t} \wedge \mathrm{d}H \rangle$. Therefore

$$\Gamma'(E) = \frac{1}{4\Omega}\left(\Omega \int_0^\infty \mathrm{d}t \langle (\mathrm{d}H)_t - (\mathrm{d}H)_{-t} \wedge \mathrm{d}H \rangle_E\right) \tag{3.16}$$

From (3.15) the convergence of $\Gamma'(E)$ depends on the behaviour of the correlation function $\langle (\mathrm{d}H)_t \wedge \mathrm{d}H \rangle$. If the dynamics is *mixing* (Arnold & Avez 1989), then

$$\lim_{t \to \infty} \langle (\mathrm{d}H)_t \wedge \mathrm{d}H \rangle = \langle \mathrm{d}H \rangle \wedge \langle \mathrm{d}H \rangle = 0 \tag{3.17}$$

In fact, we shall assume the rate of mixing (i.e. the rate at which $\langle A_t B \rangle - \langle A \rangle \langle B \rangle$ is sufficiently rapid so that

$$\int_0^\infty \mathrm{d}t \langle (\mathrm{d}H)_t \wedge \mathrm{d}H \rangle$$

converges. This is certainly true for hyperbolic (or Axiom A) systems, for which the mixing rate is exponential (see Ruelle 1986), but clearly holds for power law mixing rates $t^{-\nu}$ with $\nu > 1$. (As discussed in §4*c*, (3.15) also holds for integrable systems, which are not mixing at all, essentially because the correlation functions are quasi-periodic.)

Because the dynamics is ergodic,

$$\langle (\mathrm{d}H)_t \wedge \mathrm{d}H \rangle_E = \lim_{T \to \infty} \frac{1}{T} \int_0^T \mathrm{d}\tau\, \mathrm{d}H(\boldsymbol{z}_{t+\tau}) \wedge \mathrm{d}H(\boldsymbol{z}_\tau) \tag{3.18}$$

for almost all initial conditions $\boldsymbol{z}$. Therefore, the phase space integrals of (3.15) may be replaced by time integrals along a single trajectory, and the energy derivative obtained by varying the energy of the initial condition. Use of (3.18) would considerably facilitate numerical computations, as $\Gamma'(E)$ could be computed from just a pair of trajectories with slightly separated energies E and $E + \varepsilon$.

Finally, let us mention an equivalent form of (3.15) of some theoretical interest. Considering the quantum two-form for the moment, we note that because it depends only on the eigenstates of $\hat{H}$ (and not on the energy levels), it is unchanged if $\hat{H}$ is replaced by a (possibly parameter-dependent) function of itself. The same is true of the classical two-form; it is easy to verify that Γ' remains unchanged if $H(\boldsymbol{z}, R)$ is replaced by

$$G(\boldsymbol{z}, R) \overset{\text{def}}{=} g(H(\boldsymbol{z}, R), R), \tag{3.19}$$

where $g = g(E, R)$ is a function of energy and parameters. (The effect on the dynamics of this substitution is simply to rescale the time.) A natural representative for the family of 'hamiltonians' defined by (3.19) is the volume function

$$\Omega_E(\boldsymbol{z}, R) \overset{\text{def}}{=} \Omega(H(\boldsymbol{z}, R), R)$$

(whose quantum analogue is the counting operator $\hat{\Omega} = \sum_n n |n\rangle\langle n|$), in terms of which

$$\Gamma'(\omega) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\omega} \int_0^\omega \mathrm{d}\sigma \langle (\mathrm{d}\Omega_E)_\sigma \wedge \mathrm{d}\Omega_E \rangle_\omega \tag{3.20}$$

Here the expectation value $\langle \cdot \rangle_\omega$ is taken over that energy shell which contains phase volume ω, and $(\mathrm{d}\Omega_E)_\sigma$ denotes $\mathrm{d}\Omega_E$ evolved under the dynamics of Ω_E for a fictitious time σ. For one dimensional systems, (3.20) is closely related to equation (9) of Hannay (1985).

4. Symmetries

(*a*) *Anticanonical symmetries*

The classical analogue of an antiunitary symmetry $\hat{K}$ (cf. (2.11)) is an anticanonical symmetry K (Robnik & Berry 1986), a phase space transformation which reverses the sign of Poisson brackets and commutes with the hamiltonian:

$$\{A \circ K, B \circ K\} = -\{A, B\} \circ K, \tag{4.1a}$$

$$H \circ K = H \tag{4.1b}$$

(Here $\circ$ denotes composition, so that $(H \circ K)(\boldsymbol{z}) = H(K(\boldsymbol{z}))$.) Time reversal, in the form $K(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}, -\boldsymbol{p})$, is the prototypical example.

Suppose a system possesses an anticanonical symmetry K. In general, the energy shell $\{z|H(z)=E\}$ will be composed of distinct connected components mapped into each other by K. Let Γ_E and $\bar{\Gamma}_E$ be two such components, related by $\bar{\Gamma}_E = K(\Gamma_E)$ (it could be that $\Gamma_E = \bar{\Gamma}_E$, i.e. that Γ_E is invariant under K). Let $\Gamma^{\mathrm{c}}(E)$ and $\bar{\Gamma}^{\mathrm{c}}(E)$ be the associated two-forms obtained by restricting the microcanonical average in (3.16) to Γ_E and $\bar{\Gamma}_E$ respectively. Then, in analogy with (2.12), the classical two-form obeys

$$\bar{\Gamma}^{\mathrm{c}}(E) = -\Gamma^{\mathrm{c}}(E). \tag{4.2}$$

In particular, if $\Gamma_E = \bar{\Gamma}_E$, then $\Gamma^{\mathrm{c}}(E)$ vanishes.

The proof of (4.2) is quite similar to that of (2.12). Calculating $\bar{\Gamma}^{\mathrm{c}}(E)$ from (3.16) and integrating over $K(z)$ rather than z (a volume-preserving change of variable),

$$\begin{aligned}\bar{\Gamma}^{\mathrm{c}}(E) &= \frac{1}{4\Omega}\int_0^\tau \mathrm{d}t \int \mathrm{d}^{2N} z\, \delta_{\bar{\Gamma}}(E-H)\,((\mathrm{d}H)_t - (\mathrm{d}H)_{-t}) \wedge \mathrm{d}H \\ &= \frac{1}{4\Omega}\int_0^\tau \mathrm{d}t \int \mathrm{d}^{2N} z\, \delta_{\bar{\Gamma}}(E-H\circ K)\,((\mathrm{d}H)_t \circ K - (\mathrm{d}H)_{-t}\circ K) \wedge (\mathrm{d}H \circ K)\end{aligned} \tag{4.3}$$

Here $\delta_{\bar{\Gamma}}(E-H)$ denotes the restriction of $\delta(E-H)$ to $\bar{\Gamma}_E$, and $\Omega(E) = \bar{\Omega}(E)$ is the volume of Γ_E or $\bar{\Gamma}_E$. Since anticanonical symmetries reverse the sense of time (easily shown), $\boldsymbol{Z}(t)\circ K = K\circ \boldsymbol{Z}(-t)$, or more explicitly $\boldsymbol{Z}(K(z),t) = K(\boldsymbol{Z}(z,-t))$. Therefore $(\mathrm{d}H)_t \circ K \overset{\mathrm{def}}{=} \mathrm{d}H\circ \boldsymbol{Z}(t)\circ K = \mathrm{d}H\circ K\circ \boldsymbol{Z}(-t)$. But $\mathrm{d}H\circ K = \mathrm{d}H$ (K is a symmetry), so that

$$(\mathrm{d}H)_t \circ K = \mathrm{d}H\circ \boldsymbol{Z}(-t) = (\mathrm{d}H)_{-t} \quad (\mathrm{d}H)_{-t}\circ K = (\mathrm{d}H)_t \tag{4.4}$$

Substituting the preceding into (4.3),

$$\bar{\Gamma}^{\mathrm{c}}(E) = \frac{1}{4\Omega}\int_0^\tau \mathrm{d}t \int \mathrm{d}^{2N} z\, \delta_{\bar{\Gamma}}(E-H\circ K)\,((\mathrm{d}H)_t - (\mathrm{d}H)_{-t}) \wedge \mathrm{d}H. \tag{4.5}$$

Since $\delta_{\bar{\Gamma}}(E-H\circ K) = \delta_{\Gamma}(E-H)$, the right hand side is just $-\Gamma^{\mathrm{c}}(E)$, and the symmetry property (4.2) follows directly.

(b) *Additional constants of the motion*

Before considering integrable systems we first consider the more general case in which there are k commuting constants of the motion, with $1 \leqslant k \leqslant N$. Ergodic systems correspond to $k = 1$, integrable systems to $k = N$. Let $\boldsymbol{F} = (F_1, \ldots, F_k)$ denote the constants of motion.

Assuming the dynamics is ergodic on the invariant manifold $\{z|\boldsymbol{F}(z) = f\}$, there is a straightforward generalization of the classical two-form. Equation (3.8) is still valid if the microcanonical density is reinterpreted to be $\delta^k(f-\boldsymbol{F})/D(f)$, where

$$D(f) = \int \mathrm{d}^{2N} z\, \delta^k(f-\boldsymbol{F})$$

is the volume of the invariant manifold. Thus

$$\Gamma^{\mathrm{c}}(f) = -\frac{1}{2}\int_0^\tau \mathrm{d}t \langle (\mathrm{d}H)_t \wedge \mathrm{d}H\rangle_f \tag{4.6}$$

From a straightforward generalization of (3.12) (derived in Appendix A),

$$\langle (\mathrm{d}H)_t \wedge \mathrm{d}H\rangle_f = (1/D)\nabla_f \cdot (D\langle (\mathrm{d}H)_t \cdot \boldsymbol{F}\} \wedge \mathrm{d}H\rangle_f). \tag{4.7}$$

But $\{(\mathrm{d}H)_t, \boldsymbol{F}\} = \{\mathrm{d}H, \boldsymbol{F}_{-t}\}_t$, since Poisson brackets are preserved by the dynamics, and $\{\mathrm{d}H, \boldsymbol{F}_{-t}\}_t = \{\mathrm{d}H, \boldsymbol{F}\}_t$, since $\boldsymbol{F}$ is a constant of the motion. However, $\{\mathrm{d}H, \boldsymbol{F}\} = \mathrm{d}\{H, \boldsymbol{F}\} - \{H, \mathrm{d}\boldsymbol{F}\}$, and $\mathrm{d}\{H, \boldsymbol{F}\} = 0$, again because $\boldsymbol{F}$ is a constant of the motion. Therefore

$$\{(\mathrm{d}H)_t, \boldsymbol{F}\} = -\{H, \mathrm{d}\boldsymbol{F}\}_t = (\mathrm{d}\dot{\boldsymbol{F}})_t \tag{4.8}$$

Substituting (4.8) and (4.7) into (4.6) and integrating by parts over t, we obtain

$$\Gamma^{\mathrm{c}}(f) = \frac{1}{2D}\nabla_f \left(D\int_0^\tau \mathrm{d}t \langle (\mathrm{d}F)_t \wedge \mathrm{d}H\rangle_f\right), \tag{4.9}$$

the required generalization of (3.15).

(c) *Integrable systems*

Take the constants of motion $\boldsymbol{F}$ to be the actions $\boldsymbol{I}$. Then $D(I) = (2\pi)^N$ (obtained from integration over the angles $\boldsymbol{\theta}$), and

$$\Gamma^{\mathrm{c}}(I) = \frac{1}{2}\int_0^\tau \mathrm{d}t \nabla_I \cdot \langle (\mathrm{d}I)_t \wedge \mathrm{d}H\rangle_I \tag{4.10}$$

We obtain a more explicit formula by expanding $\mathrm{d}H$ in a Fourier series,

$$\mathrm{d}H(\boldsymbol{\theta}, \boldsymbol{I}) = \sum_{\boldsymbol{m}} h_{\boldsymbol{m}}(\boldsymbol{I}) \exp(\mathrm{i}\boldsymbol{m}\cdot\boldsymbol{\theta}). \tag{4.11}$$

in which the coefficients $h_{\boldsymbol{m}}$ are one-forms. (Note that $\mathrm{d}H$ is the derivative of $H(z, R)$ with z, rather than $(\boldsymbol{\theta}, \boldsymbol{I})$, held fixed.) The Fourier coefficients $i_{\boldsymbol{m}}$ of $\mathrm{d}I$ may be expressed in terms of the $h_{\boldsymbol{m}}$s. Expanding the relation $\{\mathrm{d}I, H\} = \{\mathrm{d}H, I\}$ (the derivative of $\{I, H\} = 0$) in a Fourier series, we obtain $\mathrm{i}(\boldsymbol{m}\cdot\boldsymbol{\omega})i_{\boldsymbol{m}} = \mathrm{i}\boldsymbol{m}h_{\boldsymbol{m}}$, where $\boldsymbol{\omega} = \nabla_I H$ are the frequencies. Therefore

$$i_{\boldsymbol{m}} = h_{\boldsymbol{m}}\boldsymbol{m}/(\boldsymbol{m}\cdot\boldsymbol{\omega}) \tag{4.12}$$

The dynamics is simply $\boldsymbol{\theta}_t = \boldsymbol{\theta} + \boldsymbol{\omega}t$. Substituting the Fourier series for $(\mathrm{d}I)_t$ and $\mathrm{d}H$ into (4.9), we readily carry out the torus average $\langle\cdots\rangle_I$ and time integral (the latter after reinstating the convergence factor of (2.2)) and obtain

$$\Gamma^{\mathrm{c}}(I) = \tfrac{1}{2}\mathrm{i}\sum_{\boldsymbol{m}\cdot\boldsymbol{\omega}\neq 0} (\boldsymbol{m}\cdot\nabla_I)\frac{h_{\boldsymbol{m}}\wedge h_{-\boldsymbol{m}}}{(\boldsymbol{m}\cdot\boldsymbol{\omega})^2} \tag{4.13}$$

As $h_{-\boldsymbol{m}} = h_{\boldsymbol{m}}^*$, $\Gamma^{\mathrm{c}}(I)$ is real. In Appendix B we show that (4.13) is equivalent to the Hannay two-form.

5. Examples

(a) *Uniform magnetic field*

In appropriate units the hamiltonian of a three-dimensional charged particle in a uniform magnetic field $\boldsymbol{B}$ is given by

$$H = \tfrac{1}{2}(\boldsymbol{p}-\boldsymbol{A})^2 + V(\boldsymbol{r}), \quad \boldsymbol{A} = \tfrac{1}{2}\boldsymbol{B}\times\boldsymbol{r}. \tag{5.1}$$

We take the parameters of H to be the components of the magnetic field and use vector notation for parameter space, writing $\nabla_{\boldsymbol{B}}$ instead of d. Straightforward calculation gives

$$\nabla_{\boldsymbol{B}} H = -\tfrac{1}{2}\boldsymbol{M} \tag{5.2}$$

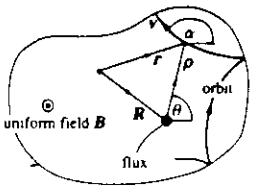


Figure 2. Aharonov–Bohm billiard. The particle at r moves with velocity $v = (v, \alpha)$ in uniform background field B, and a flux line is located at R. $\rho = r - R = (\rho, \theta)$

where $l = r \times v$ is the mechanical angular momentum and $v = p - A$ is the velocity $\dot{r}$. The two-form $V^c(E, B)$ is a vector field in B-space and is given by

$$V^c(E, B) = \frac{1}{4\Omega'}\left(\Omega' \int_0^\infty \mathrm{d}t \langle (l_t - l_{-t}) \times l \rangle_{E,B}\right)'. \tag{5.3}$$

At $B = 0$ the hamiltonian is invariant under the time reversal transformation $(r, v) \to (r, -v)$. Under this transformation $l \to -l$ and $l_t \to -l_{-t}$. Since microcanonical averages are invariant under time reversal,

$$\langle (l_t - l_{-t}) \times l \rangle = -\langle (l_t - l_{-t}) \times l \rangle, \tag{5.4}$$

which in turn implies that $V^c(E, B)$ vanishes when $B = 0$. Note that this conclusion is *not* a consequence of (4.2). Equation (4.2) is derived for an anticanonical symmetry which holds for all parameters, whereas the hamiltonian (5.1) is time-reversal invariant only for $B = 0$.

$V^c(E, B)$ is not invariant under parameter-dependent gauge transformations of the vector potential A, in spite of its expression in terms of the mechanical angular momentum. Under the gauge transformation $A(r) \to A(r) + \nabla_r \psi(r, B)$,

$$\nabla_B H \to \nabla_B H - \nabla_B (v \cdot \nabla_r \psi). \tag{5.5}$$

and $V^c(E, B)$ transforms accordingly. The analogous behaviour of the quantum two-form is discussed in Mondragon & Berry (1988). (There it is noted that, although the geometrical and dynamical phases are not separately gauge invariant, their sum is.)

(b) *Aharonov–Bohm billiard in uniform magnetic field.*

A particle is confined to a two-dimensional billiard threaded by an infinitely thin unit solenoid (e.g. the flux line of a magnetic monopole) in a constant background magnetic field. In suitable units the hamiltonian is

$$H = \tfrac{1}{2}(p - A_s - A_b)^2, \quad A_s = (\hat{z} \times \rho)/\rho^2, \quad A_b = \tfrac{1}{2}B\hat{z} \times r, \tag{5.6}$$

where (see figure 2) $r = (x, y)$ are the particle coordinates, $R = (X, Y)$ are the solenoid coordinates, and $\rho = r - R$. A_s and A_b are the vector potentials of the solenoid and the background field, respectively. One can verify that $B_s = 2\pi\delta^2(\rho)\hat{z}$ and $B_b = B\hat{z}$.

We take R, the coordinates of the solenoid, as the parameters of the system. (We could, if we wished include B, the background field strength, and introduce another parameter for the solenoid strength.) We use vector notation for parameter space, writing ∇_R instead of d. Omitting straightforward calculations, we get that

$$\nabla_R H = \hat{z} \times (v_\perp - v_\parallel)/\rho^2 \tag{5.7a}$$

$$= (-v/\rho^2)(\sin(\alpha - 2\theta), \cos(\alpha - 2\theta)), \tag{5.7b}$$

where $v = p - A_s - A_b$ is the velocity $\dot{r}$. Equation (5.5a) is expressed in terms of the components of v parallel and perpendicular to $\hat{\rho}$, namely $v_\parallel = (\hat{\rho} \cdot v)\hat{\rho}$ and $v_\perp = v - v_\parallel$. Equation (5.5$b$) is expressed in terms of the polar coordinates of v and ρ, namely $\rho = \rho(\cos\theta, \sin\theta)$ and $v = v(\cos\alpha, \sin\alpha)$.

For this system Ω' is a constant equal to $2\pi A$, where A is the area of the billiard, and $v = |v|$ is a constant of the motion. The two-form is a scalar field in R-space given by

$$V^c(E, R) = \left(\int_0^\infty \mathrm{d}t \left\langle \frac{(v_\perp)_t - (v_\parallel)_t}{\rho_t^2} \times \frac{v_\perp - v_\parallel}{\rho^2} \right\rangle\right)' \tag{5.8a}$$

$$= \left(-v^2 \int_0^\infty \mathrm{d}t \left\langle \frac{\sin(2(\theta_t - \theta) - (\alpha_t - \alpha))}{\rho_t^2 \rho^2} \right\rangle\right)'. \tag{5.8b}$$

Let us point out some features of (5.8). For trajectories which either start (resp. end) at the solenoid, ρ (resp. ρ_t) vanishes, and the integrand is singular. However, the contribution of these singularities to $V^c(E, R)$ is finite. (See Appendix C.) Next, in the absence of a background field (i.e. $B = 0$), the dynamics is time-reversal invariant, and with an argument similar to the one in §5a, one can show that $V^c(E, R)$ vanishes identically. Thus, even though the solenoid breaks the time-reversal invariance of the hamiltonian, the background field is needed to produce a non-zero two form. Finally, one can show that $V^c(E, R)$ vanishes if the solenoid lies outside the billiard. (This is intuitively clear but is not obvious from (5.8). It does follow immediately from the alternative formula (6.8) derived in the next section.) Thus R space is effectively the billiard itself.

The Aharonov–Bohm billiard in a uniform background field is perhaps the simplest example of a chaotic system for which the classical two form is non-trivial. It is two-dimensional, the minimum required for chaos. The dynamics may be computed without having to solve differential equations (the trajectories are circular arcs specularly reflected at the billiard boundary.) Finally, the dynamics is independent of the parameters R (as B_s vanishes everywhere but at a point, only a zero measure set of trajectories is affected by it.) In light of the discussion following (3.18), a numerical calculation would require only a pair of trajectories, with slightly separated energies, to determine $V^c(E, R)$ for all R. A variant of this example (not quite as simple) is a billiard in crossed uniform electric and magnetic fields. The magnetic field is normal to the billiard, the electric field tangent to it, and the two field strengths and the direction of the electric field are natural parameters.

6. Alternative form

We obtain a useful alternative expression for the classical two form, (6.8) below. Instead of deriving it directly from (3.15), we begin with an alternative expression for the quantum two-form (6.2) below) which is of independent interest.

(a) *Quantum formula*

The alternative quantum formula follows from an identity,

$$-\tfrac{1}{2}i\hbar\langle \mathrm{d}U(t) \cdot n| \wedge |\mathrm{d}U(t) \cdot n\rangle = V_n + \tfrac{1}{2}i\hbar \sum_{j \neq n} \langle \mathrm{d}n|j\rangle \wedge \langle j|\mathrm{d}n\rangle \cos\omega_{nj}t, \tag{6.1}$$

derived in Appendix D. $\mathrm{d}U(t)$ is the derivative of the propagator. Upon averaging over time the oscillatory terms vanish, and we obtain

$$V_n = -\tfrac{1}{2}i\hbar\,\overline{\langle \mathrm{d}U(t) \cdot n| \wedge |\mathrm{d}U(t) \cdot n\rangle}. \tag{6.2}$$

(Here and hereafter, $\int$ denotes the time average

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathrm{d}t f(t).)$$

With a similar calculation one can also show that

$$\Gamma_n = -i\hbar\overline{\langle \mathrm{d}U^{\dagger}(t)\cdot n|\wedge|\mathrm{d}U^{\dagger}(t)\cdot n\rangle}. \tag{6.2'}$$

In (6.2′) the time averages over the bra and ket are performed separately, or incoherently, whereas in (6.2) they are performed simultaneously, or coherently; the factor of $\frac{1}{2}$ accounts for this difference. One can develop the formalism starting from either the coherent or the incoherent expression: for brevity we present the coherent version only.

Equation (6.2) has an interesting geometrical interpretation. Suppose we want the flux of Γ_n through $\square_h$ (see figure 1*a*). We apply $U(t,R)$, $U(t,R+r_1)$ and $U(t,R+r_2)$ to $|t,R\rangle$. The states obtained describe an area element $\square_H(t)$ in Hilbert space (see figure 1*b*), spanned by displacements $U_1(t)|n\rangle$ and $U_2(t)|n\rangle$ from $U(t)|n\rangle$ (here $U_i(t) \overset{\text{def}}{=} \mathrm{d}U(t)\cdot r_i$). By virtue of the canonical structure of quantum mechanics, $\square_H(t)$ has a (naturally defined) symplectic area $-2\hbar\,\mathrm{Im}\langle U_1^{\dagger}(t)\cdot n|U_2(t)\cdot n\rangle$ (see Appendix E). According to (6.2) the time average of the symplectic area of $\square_H(t)$ is just (minus twice) the required flux. (Note there is a factor of two which arises from the definition of the wedge product.)

The equivalence of (6.2) and (2.9) can be established by expressing $\mathrm{d}U(t)$ in terms of $(\mathrm{d}\hat{H})_t$. These are related by the formula for the derivative of an exponential (see Bellman 1960),

$$\mathrm{d}U(t) = -\frac{i}{\hbar}U(t)\int_0^t \mathrm{d}\tau(\mathrm{d}\hat{H})_\tau. \tag{6.3}$$

Substituting into (6.2),

$$\Gamma_n = -\frac{1}{4\hbar}\overline{\int_0^t \mathrm{d}\tau\int_0^t \mathrm{d}\tau'\langle n|[(\mathrm{d}\hat{H})_\tau\wedge(\mathrm{d}\hat{H})_{\tau'}]|n\rangle}, \tag{6.4}$$

in which, because the τ and τ' integrals appear symmetrically, we have replaced $(\mathrm{d}\hat{H})_\tau\wedge(\mathrm{d}\hat{H})_{\tau'}$ by its symmetrized part, $\frac{1}{2}[(\mathrm{d}\hat{H})_\tau\wedge(\mathrm{d}\hat{H})_{\tau'}]$ (cf. (2.6)). Since expectation values of eigenstates are time invariant,

$$\langle n|[(\mathrm{d}\hat{H})_\tau\wedge(\mathrm{d}\hat{H})_{\tau'}]|n\rangle = \langle n|[(\mathrm{d}\hat{H})_{\tau-\tau'}\wedge\mathrm{d}\hat{H}]|n\rangle.$$

Making this substitution in (6.4) enables the τ' integral to be performed, and (2.9) follows from a few more manipulations.

(*b*) *Classical formula*

The corresponding classical alternative is obtained directly from (6.4). Taking its classical limit as in §3, we get

$$\Gamma_n \to \Gamma^{\mathrm{c}}(E) = \frac{1}{4}\overline{\int_0^t \mathrm{d}\tau\int_0^t \mathrm{d}\tau'\langle|(\mathrm{d}H)_\tau\wedge(\mathrm{d}H)_{\tau'}|\rangle_E}. \tag{6.5}$$

In Appendix F we show that

$$\int_0^t \mathrm{d}\tau\int_0^t \mathrm{d}\tau'|(\mathrm{d}H)_\tau\wedge(\mathrm{d}H)_{\tau'}| = -[\mathrm{d}\boldsymbol{Z}(t)\wedge\mathrm{d}\boldsymbol{Z}(t)]. \tag{6.6}$$

where $\mathrm{d}\boldsymbol{Z}(t)$ is the derivative of the flow with respect to parameters, in which the dependence on initial conditions $\boldsymbol{z}$ has been left implicit. The square brackets denote the symplectic inner product (Arnold 1978), defined as follows: Given vectors $\boldsymbol{u} = (\boldsymbol{q},\boldsymbol{p})$ and $\boldsymbol{v} = (\boldsymbol{q}',\boldsymbol{p}')$ in phase space,

$$[\boldsymbol{u},\boldsymbol{v}] \overset{\text{def}}{=} \boldsymbol{u}\cdot\boldsymbol{J}^{-1}\cdot\boldsymbol{v} = \boldsymbol{p}\cdot\boldsymbol{q}' - \boldsymbol{q}\cdot\boldsymbol{p}', \tag{6.7}$$

where $\boldsymbol{J}$ is given by (3.10). (The symplectic inner product $[\boldsymbol{u},\boldsymbol{v}]$ should not be confused with the commutator $[\hat{A},\hat{B}]$; in the latter the quantum operators are distinguished by carats.) Substituting (6.6) into (6.5) we obtain

$$\Gamma^{\mathrm{c}}(E) = -\tfrac{1}{4}\overline{\langle[\mathrm{d}\boldsymbol{Z}(t)\wedge\mathrm{d}\boldsymbol{Z}(t)]\rangle_E} = -\tfrac{1}{2}\overline{\langle[\mathrm{d}\boldsymbol{P}(t)\wedge\mathrm{d}\boldsymbol{Q}(t)]\rangle_t}. \tag{6.8}$$

where in the last expression we have taken $\mathrm{d}\boldsymbol{Z}(t) = (\mathrm{d}\boldsymbol{Q}(t),\mathrm{d}\boldsymbol{P}(t))$. This is the alternative formula for the classical two-form.

Equation (6.8) is a precise version of a formula (eq. (4.18)) derived in Berry (1990) (there the time dependence was integrated over, but the interpretation of the differentials $\mathrm{d}\boldsymbol{Q}$ and $\mathrm{d}\boldsymbol{P}$ was left ambiguous). Equation (6.8) might appear to be the simplest expression for the classical two-form, but its simplicity is deceptive (cf. the discussion in Berry (1990)). As with (3.8), the fact that (6.8) converges is not obvious, because if the dynamics is chaotic, both $\mathrm{d}\boldsymbol{Z}(\boldsymbol{z},t)$ and $[\mathrm{d}\boldsymbol{Z}(\boldsymbol{z},t)\wedge\mathrm{d}\boldsymbol{Z}(\boldsymbol{z},t)]$ grow exponentially in time for fixed $\boldsymbol{z}$. (We shall not give the somewhat involved general argument here — none of our results depend on it — but the particular case of periodic orbits is treated in appendix K.) However, since the microcanonical average $\overline{\langle|\mathrm{d}\boldsymbol{Z}(t)\wedge\mathrm{d}\boldsymbol{Z}(t)|\rangle}$ does not in fact diverge, it follows that $[\mathrm{d}\boldsymbol{Z}(\boldsymbol{z},t)\wedge\mathrm{d}\boldsymbol{Z}(\boldsymbol{z},t)]$ oscillates with $\boldsymbol{z}$, and that the exponentially large oscillations must cancel in the main when averaged over the energy shell.

Like (6.2), (6.8) has a geometrical interpretation. Suppose we want the flux of $\Gamma^{\mathrm{c}}(E)$ through $\square_h$. To a point $\boldsymbol{z}$ on the energy shell we apply the hamiltonians $H(R)$, $H(R+r_1)$ $H(R+r_2)$ for a time t. The resulting trajectories describe an area element $\square_P(\boldsymbol{z},t)$ in phase space drawn in figure 1*c*, spanned by displacements $\boldsymbol{Z}_1(\boldsymbol{z},t)$ and $\boldsymbol{Z}_2(\boldsymbol{z},t)$ from $\boldsymbol{z}_t$, where $\boldsymbol{Z}_i(\boldsymbol{z},t) \overset{\text{def}}{=} \mathrm{d}\boldsymbol{Z}(\boldsymbol{z},t)\cdot r_i$. The symplectic area of $\square_P(\boldsymbol{z},t)$ is $[\boldsymbol{Z}_1(\boldsymbol{z},t),\boldsymbol{Z}_2(\boldsymbol{z},t)]$, and according to (6.8), its microcanonical and time average is just (minus twice) the required flux. In light of figure 1*b* and *c*, the correspondence between quantum and classical two-forms, (6.2) and (6.8), is immediate.

(*c*) *An equivalent form*

Differentiating Hamilton's equations $\dot{\boldsymbol{Z}}(t) = \boldsymbol{J}\cdot(\boldsymbol{H}')_t$ with respect to parameters, we find that $\mathrm{d}\boldsymbol{Z}(t)$ satisfies the linear inhomogeneous equation

$$\mathrm{d}\dot{\boldsymbol{Z}}(t) = \boldsymbol{J}(\boldsymbol{H}'')_t\cdot\mathrm{d}\boldsymbol{Z}(t) + \boldsymbol{J}\cdot(\mathrm{d}\boldsymbol{H}')_t, \tag{6.9}$$

with initial conditions $\mathrm{d}\boldsymbol{Z}(\boldsymbol{z},0) = 0$. (Here $(\boldsymbol{H}'')_{ij} = \partial^2 H/\partial z_i\partial z_j$.) It is often useful to express $\mathrm{d}\boldsymbol{Z}(t)$ in terms of other solutions $\mathrm{d}\boldsymbol{Y}(t)$ of (6.9), to be specified later, which will of course satisfy different initial conditions. (Here we make a slight abuse of notation, as we will not assume that $\mathrm{d}\boldsymbol{Y}(t)$ is an exact differential.) In general, any two solutions of (6.9) differ by a solution of the homogeneous equation $\mathrm{d}\dot{\boldsymbol{X}}(t) = \boldsymbol{J}(\boldsymbol{H}'')_t\cdot\mathrm{d}\boldsymbol{X}(t)$, and solutions of the homogeneous equation are of the form $\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{S}(t)\cdot\mathrm{d}\boldsymbol{X}(0)$. (Here $\boldsymbol{S}(t)$ is an abbreviation for $\boldsymbol{S}(\boldsymbol{z},t)$, the linearized flow of (3.11).) Therefore

$$\mathrm{d}\boldsymbol{Z}(t) = \mathrm{d}\boldsymbol{Y}(t) - \boldsymbol{S}(t)\cdot\mathrm{d}\boldsymbol{Y}(0). \tag{6.10}$$

Substituting (6.10) into (6.8) we obtain

$$\Gamma^c(E) = -\tfrac{1}{4}\overline{\langle[\mathrm{d}Y(t),\wedge\mathrm{d}Y(t)]\rangle}+\tfrac{1}{4}\overline{\langle[\mathrm{d}Y(t),\wedge S(t)\cdot\mathrm{d}Y(0)]\rangle}-\tfrac{1}{4}\overline{\langle[\mathrm{d}Y(0),\wedge\mathrm{d}Y(0)]\rangle}, \tag{6.11}$$

an equivalent expression for the classical two-form. In the last term we have used

$$[S(t)\cdot\mathrm{d}Y(0),\wedge S(t)\cdot\mathrm{d}Y(0)] = [\mathrm{d}Y(0),\wedge\mathrm{d}Y(0)], \tag{6.12}$$

the canonical invariance of the symplectic inner product. Equation (6.11) is particularly useful in deriving the two-form for integrable systems (Appendix G) and periodic orbits (§§8 and 9).

7. Is the classical two-form closed?

The question as to whether $\Gamma^c(E)$ is closed is of central importance, but it is not easily answered. Here we present a formal derivation of closedness.

$$\mathrm{d}\Gamma^c(E) = 0. \tag{7.1}$$

First, let us point out that it is not correct to argue, on the basis of the correspondence principle, that because the quantum two-form is closed, so must be its classical limit. The reason is that Γ_n is not closed, as $\mathrm{d}\Gamma_n$ is singular at eigenvalue degeneracies. (Indeed this property was one of the motivations underlying the discovery of the geometric phase (see Shapere & Wilczek 1989, p. 26).) Thus $\mathrm{d}\Gamma_c(E)\neq 0$ would have implications for the distribution of degeneracies in the classical limit, as will be explained in §9. On the other hand, $\mathrm{d}\Gamma_c(E)=0$ would imply that (at least locally) $\Gamma^c(E)$ is the derivative of a one-form, whose integral around a closed loop in parameter space one might expect to describe a classical anholonomy for adiabatically cycled chaotic systems, analogous to the Hannay angles for integrable systems.

From the formulas derived so far it is not even clear that the three-form $\mathrm{d}\Gamma^c(E)$ converges. For example, the derivative of (3.15) introduces the two-form $\mathrm{d}((\mathrm{d}H)_t)$. While $\mathrm{d}(\mathrm{d}H)$ vanishes, $\mathrm{d}((\mathrm{d}H)_t)$ does not, due to the parameter dependence of the dynamics. In fact, for fixed $\mathbf{z}$, $\mathrm{d}((\mathrm{d}H)_t)(\mathbf{z})$ grows exponentially in time, since

$$\mathrm{d}((\mathrm{d}H)_t)(\mathbf{z}) = \mathrm{d}H^t(\mathbf{z}_t)\wedge\cdot\,\mathrm{d}\mathbf{Z}(\mathbf{z},t) \tag{7.2}$$

and $\mathrm{d}\mathbf{Z}(\mathbf{z},t)$ grows exponentially.

The alternative form (6.8) turns out to be the most convenient for calculating $\mathrm{d}\Gamma^c(E)$. To proceed, we first note that upon differentiating an ensemble average such as $\langle\phi\rangle_E$ (here ϕ is any differential form), account must be taken of both the explicit parameter dependence of ϕ and the implicit parameter dependence of the ensemble. Also, the derivative of the ensemble is to be taken at fixed volume rather than fixed energy (cf. of the Weyl rule (3.4)). As shown in Appendix H,

$$\mathrm{d}\langle\phi\rangle_E = \langle\mathrm{d}\phi\rangle_E+(1/\Omega')(\Omega'\langle(\mathrm{d}E-\mathrm{d}H)\wedge\phi\rangle_E)',\quad\text{where}\quad \mathrm{d}E=\langle\mathrm{d}H\rangle_E. \tag{7.3}$$

Thus differentiation of (6.8) gives

$$\mathrm{d}\Gamma^c(E) = -(1/4\Omega')(\Omega'\overline{\langle(\mathrm{d}E-\mathrm{d}H)\wedge[\mathrm{d}\mathbf{Z}(t),\wedge\mathbf{Z}(t)]\rangle_E})' \tag{7.4}$$

(we have used the closedness of $[\mathrm{d}\mathbf{Z}(t),\wedge\mathrm{d}\mathbf{Z}(t)]$). In what follows we show that

$$\tfrac{1}{4}\overline{\langle\mathrm{d}H\wedge[\mathrm{d}\mathbf{Z}(t),\wedge\mathrm{d}\mathbf{Z}(t)]\rangle} = \tfrac{1}{4}\mathrm{d}E\wedge\overline{\langle[\mathrm{d}\mathbf{Z}(t),\wedge\mathrm{d}\mathbf{Z}(t)]\rangle_E}, \tag{7.5}$$

which together with (7.4) implies that $\Gamma^c(E)$ is closed. To streamline the presentation we have left some details to Appendix I.

In Appendix I we show that

$$\tfrac{1}{4}\overline{\langle\mathrm{d}H\wedge[\mathrm{d}\mathbf{Z}(t),\wedge\mathrm{d}\mathbf{Z}(t)]\rangle} = \lim_{s\to 0}G(s,s), \tag{7.6}$$

where

$$\left.\begin{aligned} G(s,s') &= -\int_0^\infty \mathrm{d}\tau\,\frac{\mathrm{e}^{-s\tau}}{s}s'\int_0^\infty \mathrm{d}\tau'\,\mathrm{e}^{-s'\tau'}F(\tau,\tau'),\\ F(\tau,\tau') &= \tfrac{1}{4}\langle(\mathrm{d}H)_{-\tau}\wedge[(\mathrm{d}H)_{\tau'}\wedge\mathrm{d}H]\rangle \end{aligned}\right\} \tag{7.7}$$

The principal steps leading to (7.6) involve replacing $[\mathrm{d}\mathbf{Z}(t),\wedge\mathrm{d}\mathbf{Z}(t)]$ by its expression in (6.6) and writing its time average as the residue of its Laplace transform at the origin (as in Appendix J). Instead of taking both arguments of $G(s,s')$ to zero simultaneously, let us take the limit $s'\to 0$ first. From Appendix J

$$\lim_{s'\to 0}s'\int_0^\infty \mathrm{d}\tau'\,\mathrm{e}^{-s'\tau'}F(\tau,\tau') = \lim_{T\to\infty}\frac{1}{T}\int_0^T \mathrm{d}\tau'F(\tau,\tau'), \tag{7.8}$$

because the right-hand side is the τ'-average of $F(\tau,\tau')$ and the left-hand side is the residue of its τ'-Laplace transform at the origin. But from (7.7) the τ'-average of $F(\tau,\tau')$ is $\frac{1}{2}\mathrm{d}E\wedge\langle[(\mathrm{d}H)_\tau\wedge\mathrm{d}H]\rangle$, as implied by the *weak mixing property* $\overline{\langle A,B\rangle} = \langle A\rangle\,\langle B\rangle$ (Arnold & Avez 1989) (weak mixing is implied by mixing, which we have already assumed the dynamics to be). Therefore

$$\lim_{s\to 0}G(s,s') = -\mathrm{d}E\wedge\frac{1}{2}\int_0^\infty \mathrm{d}\tau\,\frac{\mathrm{e}^{-s'\tau}}{s'}\langle[(\mathrm{d}H)_\tau\wedge\mathrm{d}H]\rangle \tag{7.9}$$

The remaining limit $s'\to 0$ is straightforward. As shown in Appendix I,

$$\lim_{s'\to 0}\frac{1}{2}\int_0^\infty \mathrm{d}\tau\,\frac{\mathrm{e}^{-s'\tau}}{s'}\langle[(\mathrm{d}H)_\tau\wedge\mathrm{d}H]\rangle_E = \Gamma^c(E) \tag{7.10}$$

Thus $\lim_{s'\to 0}(\lim_{s\to 0}G(s,s')) = -\mathrm{d}E\wedge\Gamma^c(E)$, which in turn is equal to

$$\tfrac{1}{4}\mathrm{d}E\wedge\overline{\langle[\mathrm{d}\mathbf{Z}(t),\wedge\mathrm{d}\mathbf{Z}(t)]\rangle_E}.$$

Together with (7.6) and (7.4), this implies the closedness of the classical two-form.

The preceding derivation is purely formal in that we have not established the convergence of the expressions involved nor justified the interchange of limits. These difficulties might yield to a more technically rigorous treatment but might also conceal some interesting behaviour. One possibility is that $\Gamma^c(E)$ is in a sense exact but not closed; it might be the derivative of a one-form (in which case its integral over a closed surface would vanish), but it might not be differentiable itself.

8. The spectral two-form

Just as the Weyl density of states $\bar{\mathrm{d}}(E)=\Omega'/h^N$ describes smooth variations in the exact density of states $\mathrm{d}(E)=\sum_n\delta(E-E_n)$, so too $\Gamma^c(E)$ describes smooth variations in Γ_n on a classical energy scale. Similarly, just as quantum fluctuations in the density of states are described by classical periodic orbits, so too are fluctuations in the two-form. The starting point for these considerations is not the two-form itself but rather the *spectral two-form*

$$D(E) = \sum_n\delta(E-E_n)\,\Gamma_n. \tag{8.1}$$

We can write the previously derived quantum formula (6.2) as

$$\Gamma_n = -\tfrac{1}{2}i\hbar\,\overline{\mathrm{Tr}\,[P_n\,\mathrm{d}U^{\dagger}(t)\wedge\mathrm{d}U(t)]}. \tag{6.2'}$$

where $P_n = |n\rangle\langle n|$: this is a time-dependent version of the manifestly gauge-invariant formula $\Gamma_n = -i\hbar\,\mathrm{Tr}\,[P_n\,\mathrm{d}P_n\wedge\mathrm{d}P_n]$ (see Avron *et al.* 1987, 1989) in which the two-form is expressed in terms of spectral projections. Thus we may express the spectral two-form in a form more suitable for semiclassical approximation.

$$\Omega(E) = -\tfrac{1}{2}i\hbar\,\overline{\mathrm{Tr}\,[\delta(E-\hat{H})\,\mathrm{d}U^{\dagger}(t)\wedge\mathrm{d}U(t)]}. \tag{8.2}$$

Here $\delta(E-\hat{H}) = \sum_n \delta(E-E_n)P_n$ is the *spectral operator*.

The classical limit of (8.2) is taken exactly as in §§3 and 6, with the result

$$\Omega(E) \sim \Gamma^{c}(E) = -\frac{1}{4}\overline{\int \mathrm{d}^{2N}z\,W(z,E)\,[\mathrm{d}\boldsymbol{Z}(z,t)_0\wedge\mathrm{d}\boldsymbol{Z}(z,t)]} \tag{8.3}$$

In place of the Wigner function $W_n(z)$ (taken to be the microcanonical density in §§3 and 6) there appears the *spectral Wigner function* (Berry 1989)

$$W(z,E) = \langle\delta(E-\hat{H})\rangle_n(z), \tag{8.4}$$

the symbol of the spectral operator, whose semiclassical approximation is given by

$$W(z,E) \sim \frac{1}{h^N}\delta(E-H(z))\left(1+2h^{N-1}\sum_j A_j(E)\,\delta_j(z)\right). \tag{8.5}$$

The first term in (8.5) is just the microcanonical density weighted by the Weyl density of states. The additional terms, whose amplitudes are of order h^{N-1} less than the leading one, are the periodic orbit contributions. $\delta_j(z)$ is a normalized δ function on the jth periodic orbit, and

$$A_j(E) = \frac{T_j}{|\det(M_j - I)|^{\frac{1}{2}}}\cos(S_j/\hbar - \tfrac{1}{2}\mu_j\pi) \tag{8.6}$$

are the oscillatory amplitudes of the Gutzwiller trace formula (Gutzwiller 1971, 1990). S_j is the action, T_j the period of a single repetition, M_j the linearized Poincaré map and μ_j the Maslov index of the jth orbit. (In fact, (8.5) is a limiting form of a more refined expression, in which the delta functions are replaced by smooth functions localized on the energy shell and the periodic orbits.)

Substituting (8.5) into (8.3), we obtain the classical limit of the spectral two form,

$$\Gamma^{c}(E) = \bar{\Gamma}^{c}(E) + \sum_j \Gamma_j^{c}(E). \tag{8.7}$$

The smooth contribution

$$\bar{\Gamma}^{c}(E) = (\bar{\Omega}/h^N)\,\Gamma^{c}(E) \tag{8.8}$$

is simply the classical two form weighted by the Weyl density of states. Our interest here is in the periodic orbit contributions $(2/h)\,A_j(E)\,\Gamma_j^{c}(E)$, where

$$\Gamma_j^{c}(E) = -\tfrac{1}{4}\langle[\mathrm{d}\boldsymbol{Z}(t)_0\wedge\mathrm{d}\boldsymbol{Z}(t)]\rangle_{jE}. \tag{8.9}$$

(In general $\langle f\rangle_{jE}$ denotes the average of f round the jth orbit at energy E.)

There is a natural two form associated with periodic orbits, analogous to the Hannay two form for one-dimensional systems. Periodic orbits belong to continuous

families $\boldsymbol{y}_j(\theta,S,R)$, labelled by action S (in preference to the energy) and parameters R. θ, the coordinate along the orbits, is the scaled time ('angle'), in terms of which $\boldsymbol{y}_j(\theta,S,R)$ is 2π-periodic. Then the periodic orbit two-form is given by

$$\Gamma_j^{c}(S) = -\tfrac{1}{2}\langle[\mathrm{d}\boldsymbol{y}_j\wedge\mathrm{d}\boldsymbol{y}_j]\rangle_{jS}. \tag{8.10}$$

(Here $\langle\cdot\rangle_{jS}$ denotes the average around the jth orbit at action S.) As shown in Appendix K, $\Gamma_j^{c}(S)$ is well-defined. Like (6.2) and (6.8), it has a geometrical interpretation. Suppose we want the flux of $\Gamma_j^{c}(S)$ through $\square_R$. We draw vectors from $\boldsymbol{y}_j(\theta,S,R)$ to points on neighbouring orbits at the same action and scaled time, but with parameters $R+r_1$ and $R+r_2$. These vectors span an area element $\square_j(\theta)$ in phase space drawn in figure 1*f*. The symplectic area of $\square_j(\theta)$ averaged round the orbit is (minus) the required flux.

In Appendix K we show that the two-forms (8.9) and (8.10) are the same, i.e.

$$\Gamma_j^{c}(S_j(E)) = \Gamma_j^{c}(E) \tag{8.11}$$

Thus the periodic orbit contribution to the spectral two form is

$$\Omega_j^{c}(E) = (2/h)\,A_j(E)\,\Gamma_j^{c}(S_j) \tag{8.12}$$

For unstable periodic orbits the derivation of (8.11) is not straightforward. As shown in Appendix K, $|\mathrm{d}\boldsymbol{Z}(z,t)_0\wedge\mathrm{d}\boldsymbol{Z}(z,t)|_{jE}$ diverges exponentially in time, and while the divergent behaviour disappears when z is averaged over the energy shell, it does *not* when z is averaged only over a periodic orbit. Thus $\langle|\mathrm{d}\boldsymbol{Z}(t)_0\wedge\mathrm{d}\boldsymbol{Z}(t)|\rangle_{jE}$ grows exponentially with t, and its time average must be defined by analytic continuation (as in Appendix J). The origin of the divergence is the singular nature of the periodic orbit delta function, itself an artefact of the semiclassical approximation (8.5). We would like a derivation of (8.12) free of all divergences (possibly based on the Airy function smoothing of Berry (1989)), but have not yet found one.

9. Semiclassical density of degeneracies

In this section, we consider systems without time reversal symmetry and for the sake of explicitness take parameter space to be three dimensional.

The distribution of energy level degeneracies in parameter space is of considerable interest. While degeneracies are exceptional – according to a well known theorem of Von Neumann & Wigner (1929), for systems without time reversal symmetry at least three parameters must be varied to find one – they provide a mechanism for dissipation in adiabatic processes. As a hamiltonian is varied in time, its path through parameter space passes near degeneracies; these near approaches violate the conditions of the quantum adiabatic theorem and generate transitions of Landau–Zener type between states. This subject has received and continues to receive much attention, as described in Hill & Wheeler (1952) and Wilkinson (1990).

As discussed in the original work on the subject, the geometric phase is intimately connected to degeneracies (Berry 1984). It turns out that $\mathrm{d}\Gamma_n(\boldsymbol{R})$ (a scalar density in three-dimensional $\boldsymbol{R}$-space) has δ-function singularities at degeneracies (generically these occur at isolated points) and is zero elsewhere. Explicitly, letting $\boldsymbol{R}_{n,a}$ denote the degeneracies between the states $|n\rangle$ and $|n+1\rangle$,

$$\mathrm{d}\Gamma_n(\boldsymbol{R}) = 2\pi\sum_a(\sigma_{n,a}\delta^3(\boldsymbol{R}-\boldsymbol{R}_{n,a}) - \sigma_{n-1,a}\delta^3(\boldsymbol{R}-\boldsymbol{R}_{n-1,a}))\,\mathrm{d}^3R. \tag{9.1}$$

where $\rho = \mathrm{d}R_1 \wedge \mathrm{d}R_2 \wedge \mathrm{d}R_3$ is the coordinate volume form and the σs denote ± 1, as we discuss presently. Note that degeneracies with both the state above and below $|n\rangle$ contribute to $\mathrm{d}\Gamma_n$, and that the degeneracies between, say, $|n\rangle$ and $|n+1\rangle$ contribute to $\mathrm{d}\Gamma_n$ and $\mathrm{d}\Gamma_{n+1}$ with opposite signs.

The σs are defined as in Simon (1983). Assuming the eigenstates are continuous functions of parameters, let $|+\rangle \overset{\text{def}}{=} |n+1(\boldsymbol{R}_{n,a})\rangle$, $|-\rangle \overset{\text{def}}{=} |n(\boldsymbol{R}_{n,a})\rangle$ denote the pair of degenerate states at $\boldsymbol{R} = \boldsymbol{R}_{n,a}$. We construct a two-dimensional hermitian matrix $H(\boldsymbol{R})$ with matrix elements $H_+(\boldsymbol{R}) \overset{\text{def}}{=} \langle +|\hat{H}(\boldsymbol{R})|-\rangle$, $H_-(\boldsymbol{R}) \overset{\text{def}}{=} \langle -|\hat{H}(\boldsymbol{R})|+\rangle$, etc. The expansion of H in terms of the Pauli matrices, $H = AI + \boldsymbol{B}\cdot\hat{\boldsymbol{\sigma}}$, determines a vector field $\boldsymbol{B}(\boldsymbol{R})$ on parameter space. Then $\sigma_{n,a}$ is given by $-\operatorname{sgn}\det(\partial B_i/\partial R_j)|_{\boldsymbol{R}_{n,a}}$; that is, $\sigma_{n,a}$ is negative if at $\boldsymbol{R}_{n,a}$ the mapping from $\boldsymbol{R}$ to $\boldsymbol{B}$ is orientation preserving, and is positive if it is not.

The quantity we will consider is not $\mathrm{d}\Gamma_n$ itself but rather the sum

$$M_n = \sum_{m=1}^{n} \mathrm{d}\Gamma_m. \tag{9.2}$$

From (9.1) we get that

$$M_n = 2\pi \sum_a \sigma_{n,a}\, \delta^3(\boldsymbol{R} - \boldsymbol{R}_{n,a})\, \rho, \tag{9.3}$$

as the alternating contributions from $m < n$ cancel each other. Thus M_n gives the algebraic or signed density of degeneracies between $|n\rangle$ and $|n+1\rangle$. M_n should be distinguished from the absolute density of degeneracies,

$$|M|_n \overset{\text{def}}{=} 2\pi \sum_a \delta^3(\boldsymbol{R} - \boldsymbol{R}_{n,a})\, \rho.$$

An interesting question (we will not pursue it here) is which of the two densities, M_n or $|M|_n$, determines the rate of Landau–Zener transitions; does each play a distinctive role in the description? Let us just mention that $|M|_n$ can also be expressed in terms of the two form, explicitly

$$|M|_n(\boldsymbol{R}) = \lim_{\varepsilon \to 0} \frac{1}{2\pi} \left(\int_{R'\ R\ <\ \varepsilon} \mathrm{d}^3\boldsymbol{R}'\, M_n(\boldsymbol{R}') \right) M_n(\boldsymbol{R}). \tag{9.4}$$

In terms of the spectral two form (8.1),

$$M_n = \lim_{\varepsilon \to 0^+} \mathrm{d} \int_{-\infty}^{E_n(\boldsymbol{R}) - \varepsilon} \mathrm{d}E\, D(E), \tag{9.5}$$

an expression whose classical limit $M^{\mathrm{c}}(E)$ is readily obtained from (8.7), (8.8) and (8.12). $M^{\mathrm{c}}(E)$, like $D^{\mathrm{c}}(E)$, contains smooth and oscillatory terms, but since $\Gamma^{\mathrm{c}}(E)$ is closed according to (7.1), it follows directly from (9.2) that the smooth contribution vanishes.

The periodic orbit contributions $M_j^{\mathrm{c}}(E)$ are given by

$$M_j^{\mathrm{c}}(E) = \frac{\partial}{\hbar}\mathrm{d} \int_{-\infty}^{E} \mathrm{d}\epsilon\, A_j(\epsilon)\, \Gamma_j^{\mathrm{c}}(S_j(\epsilon)), \tag{9.6}$$

where A_j is given by (8.6) and Γ_j^{c} by (8.10). Because the factor $\cos(S_j/\hbar - \frac{1}{2}\mu_j\pi)$ oscillates rapidly, to lowest order in $\hbar$ we may neglect the energy dependence of the other factors in the integrand. (In so doing we ignore singularities in $\det|M_j - I|^{-\frac{1}{2}}$ at bifurcations.) Then

$$M_j^{\mathrm{c}}(E) = \frac{2}{\hbar}\mathrm{d}\left(\hbar \frac{\sin(S_j/\hbar - \frac{1}{2}\mu_j\pi)}{\det|M_j - I|^{\frac{1}{2}}}\, \Gamma_j^{\mathrm{c}}(S_j) \right). \tag{9.7}$$

To lowest order in $\hbar$, d acts only on the oscillatory factor and $\mathrm{d}\sin(S_j/\hbar - \frac{1}{2}\mu_j\pi) = \cos(S_j/\hbar - \frac{1}{2}\mu_j\pi)\,\mathrm{d}S_j/\hbar$. Also

$$\mathrm{d}S_j = T_j(\mathrm{d}E - \langle \mathrm{d}H \rangle_j). \tag{9.8}$$

as shown in Appendix K. Combining these results, we obtain

$$M_j^{\mathrm{c}}(E) = (2/h)\, A_j(E)\, (\mathrm{d}E - \langle \mathrm{d}H \rangle_j) \wedge \Gamma_j^{\mathrm{c}}(S_j) \tag{9.9}$$

(Note that for long periodic orbits $\mathrm{d}E - \langle \mathrm{d}H \rangle_j$ approaches zero, as $\langle \mathrm{d}H \rangle_j \to \langle \mathrm{d}H \rangle_k = \mathrm{d}E$.) Thus the density of degeneracies, while neutral on a classical scale, is resolved semiclassically into oscillations described by classical periodic orbits.

10. Discussion

Our principal result (3.15) is an explicit and explicitly finite expression for the classical limit of the geometric phase two-form which is valid for chaotic systems. In the derivation we have assumed the dynamics is ergodic and mixing at a sufficiently rapid rate. We have given a formal derivation of the closedness of the classical two-form, obtained semiclassical corrections to it associated with periodic orbits, and derived a semiclassical expression for the algebraic density of degeneracies in parameter space. We have also discussed the case of additional constants of the motion and specific examples including the Aharonov–Bohm billiard in a uniform magnetic field.

From this investigation there emerge a number of questions to be pursued. One would like to test these formulas numerically, particularly the periodic orbit contributions. The Aharonov-Bohm billiard is one candidate system, although for the version we are considering the quantum calculations might not be simple. Maps (classical and quantum) present alternative and possibly simpler test cases; the necessary modifications to the formalism presented here should be straightforward. It would also be interesting to see if the periodic orbit two form plays some role in purely classical mechanics, for instance in the study of bifurcations.

The most important question is whether the classical two form itself has any intrinsic significance in classical mechanics. Does it describe an anholonomy in adiabatically cycled chaotic systems, as the Hannay two form does for integrable systems? If so, it must be derivable purely within classical mechanics (as the Hannay two-form is). In this connection there remains the related question of the closedness of the two-form (another question amenable to numerical investigation). The formal argument of §7 should be right in some sense, but precisely how requires further study, perhaps facilitated by consideration of the purely classical problem.

It is a pleasure to acknowledge important contributions by M. Wilkinson in the early stages of this work. We also thank J. H. Hannay for helpful discussions. J. M. R. was supported by a NATO postdoctoral fellowship awarded in 1989, and by the SERC.

Appendix A. Derivations of (3.12) and (4.7)

By definition

$$\langle (\mathrm{d}A)_t \wedge \mathrm{d}B \rangle_E = \frac{1}{\Omega'} \int \mathrm{d}^{2N} z \, (\mathrm{d}A)_t \wedge \mathrm{d}B \, \delta(E - H). \tag{A 1}$$

Using the Leibniz rule for Poisson brackets, $\{F, G\}K = \{F, GK\} - \{F, K\}G$, the integrand $(\mathrm{d}A)_t \wedge \mathrm{d}B\,\delta(E-H)$ may be written as

$$(\mathrm{d}A)_t \wedge \delta(E-H)\,\mathrm{d}B - (\mathrm{d}A)_t\,\delta(E-H) \wedge \mathrm{d}B \tag{A 2}$$

The first term vanishes when integrated over phase space because (i) it is a Poisson bracket and therefore a pure divergence (explicitly, $\{F, G\} = \nabla_z \cdot (F J \nabla G)$, and (ii) its surface integral at infinity vanishes because $\delta(E - H(z))$ does. As for the second term

$$(\mathrm{d}A)_t\,\delta(E-H) = -(\mathrm{d}A)_t H_t \delta'(E-H) = -(\mathrm{d}A)_t \delta'(E-H), \tag{A 3}$$

where the prime denotes the derivative with respect to energy and the dot the derivative with respect to time. From (A 1), (A 3) we obtain

$$\langle (\mathrm{d}A)_t \wedge \mathrm{d}B \rangle_E = \frac{1}{\Omega'} \int \mathrm{d}^{2N} z \, \delta(E - H)(\mathrm{d}A)_t \wedge \mathrm{d}B = \frac{1}{\Omega'} (\Omega' \langle (\mathrm{d}A)_t \wedge \mathrm{d}B \rangle_t), \tag{A 4}$$

the required result (3.12).

A similar argument establishes (4.7). By definition

$$\langle (\mathrm{d}A)_t \wedge \mathrm{d}B \rangle_f = \frac{1}{D} \int \mathrm{d}^{2N} z \, \delta^N(f - F)(\mathrm{d}A)_t \wedge \mathrm{d}B. \tag{A 5}$$

The integrand $\delta^N(f - F)(\mathrm{d}A)_t \wedge \mathrm{d}B$ may be expressed as

$$(\mathrm{d}A)_t \wedge \delta^N(f - F)\,\mathrm{d}B - (\mathrm{d}A)_t\,\delta^N(f-F) \wedge \mathrm{d}B \tag{A 6}$$

The first term is a pure divergence which does not contribute to the phase space integral, and the second is

$$\nabla_f \delta^N(f - F) \cdot (\mathrm{d}A)_t F \wedge \mathrm{d}B \tag{A 7}$$

Therefore $\langle (\mathrm{d}A)_t \wedge \mathrm{d}B \rangle_f = (1/D)\nabla_f \cdot (D \langle (\mathrm{d}A)_t F \rangle \wedge \mathrm{d}B \rangle_f$. (A 8)

Appendix B. The integrable case: first derivation

We establish the equivalence of the classical two-form $\Gamma^{\mathrm{cl}}(I)$ and the Hannay two-form

$$\Gamma^{\mathrm{cl}}(I) = -\langle \mathrm{d}p \wedge \mathrm{d}q \rangle_I \tag{B 1}$$

for integrable systems. In (B 1), q and p are functions of θ, I and R, and the average is taken over an invariant torus; for convenience a minus sign has been introduced into the usual definition. For the sake of simplicity, we restrict ourselves to one degree of freedom; the generalization to higher dimensions is straightforward.

Hamilton's equations and (4.11) give

$$\left.\begin{aligned} \mathrm{d}p &= -\mathrm{d}\partial_q H = -\partial_q \mathrm{d}H = -\partial_q \sum_m h_m \exp(\mathrm{i}m\theta), \\ \mathrm{d}q &= \partial_p \sum_m h_m^* \exp(-\mathrm{i}m\theta) \end{aligned}\right\} \tag{B 2}$$

where we have used the fact that phase space and parameter differentials commute. Inserting the time dependence $\theta_t = \theta + \omega t$ and integrating gives

$$\left.\begin{aligned} \mathrm{d}p &= \mathrm{i}\partial_q \sum_{m \neq 0} h_m \exp(\mathrm{i}m\theta)/m\omega + \partial_q F(I), \\ \mathrm{d}q &= \mathrm{i}\partial_p \sum_{m \neq 0} h_m^* \exp(-\mathrm{i}m\theta)/m\omega + \partial_p F(I), \end{aligned}\right\} \tag{B 3}$$

where F is a non-oscillatory 1-form describing the mean displacement of the torus labelled I.

These expressions for $\mathrm{d}q$ and $\mathrm{d}p$ must be substituted into (B 1) and integrated over θ. The terms involving F cancel because

$$\partial_q F \wedge \partial_p F = \tfrac{1}{2}[F, \wedge F] = \partial_q F \wedge \partial_p F = 0 \tag{B 4}$$

The other terms give

$$\begin{aligned} \Gamma^{\mathrm{cl}}(I) = &-\sum_m \left[\partial_I \left(\frac{h_m}{\omega} \right) \wedge \frac{1}{m^2} \partial_I \left(\frac{h_m^*}{\omega} \right) \partial_q I \partial_p I + \frac{1}{\omega^2} h_m \wedge h_m^* \partial_q \theta \partial_p \theta \right] \\ &- \mathrm{i} \sum_m \frac{1}{m\omega} \left[\left(\frac{h_m}{\omega} \right) \wedge \partial_I \left(\frac{h_m^*}{\omega} \right) \partial_q \theta \partial_p I - \partial_I \left(\frac{h_m}{\omega} \right) \wedge \left(\frac{h_m^*}{\omega} \right) \partial_p \theta \partial_q I \right] \end{aligned} \tag{B 5}$$

Now $h_m \wedge h_m^*$ is odd in m, causing the terms in the first line to cancel. The remaining terms can be replaced by their average over m and $-m$, which gives

$$\Gamma^{\mathrm{cl}}(I) = -\frac{1}{\omega} \sum_m \partial_I (h_m \wedge h_m^*/m\omega) \partial_q \theta \partial_p I - \partial_p \theta \partial_q I)/m \tag{B 6}$$

which is the same as (4.13), because (θ, I) are canonical variables.

Appendix C. Convergence of billiard two-form

First we write (5.8*b*) in a more explicit form. The microcanonical average $\langle \cdot \rangle$ is given by

$$\frac{1}{\Omega'} \int_0^{2\pi} \mathrm{d}\alpha \int_{\mathcal{A}} \mathrm{d}^2 r$$

(the r integral is taken over the billiard), and it is readily shown that $\Gamma^{\mathrm{cl}}(E, R) = -2(E\Gamma)/\Omega'$, where

$$\Gamma = \int_0^\infty \mathrm{d}t \int_0^{2\pi} \mathrm{d}\alpha \int_{\mathcal{A}} \mathrm{d}^2 r \frac{\sin(2(\theta_t - \theta) - (\alpha_t - \alpha))}{\rho^2 t^2}. \tag{C 1}$$

Equation (C 1) is an integral in four dimensional (α, t, r) space, and the integrand exhibits four types of singularities: (*a*) $\rho = 0$; (*b*) $\rho_t = 0$; (*c*) $\rho = \rho_t = 0$, $t \neq 0$; (*d*) $\rho = 0, t = 0$. We consider these in turn.

(*a*) $\rho = 0$. The singularity is a two-dimensional surface in (α, t, r) space and corresponds to trajectories which begin at the solenoid. Regarding α and t as fixed, we consider the contribution $I_a(\alpha, t)$ of a two-dimensional pencil of trajectories beginning near $\rho = 0$ in direction α. Changing integration variables from r to (ρ, θ),

$$I_a(\alpha, t) = \int_0^\infty \mathrm{d}\rho \int_0^{2\pi} \mathrm{d}\theta \frac{\sin(2(\theta_t - \theta) - (\alpha_t - \alpha))}{t^2 \rho}. \tag{C 2}$$

As our concern is the singularity at $\rho = 0$, the upper limit of the ρ integral is left

indefinite. Let ρ', θ' and α' denote the final coordinates of the central trajectory from $\rho = 0$. Since final conditions depend smoothly on initial conditions, $\rho_t = \rho' + O(\rho)$, and analogous relations hold for θ_t and α_t. Expanding the integrand about $\rho = 0$,

$$\frac{\sin(2(\theta_t-\theta)-(\alpha_t-\alpha))}{\rho_t^2\rho} = \frac{\sin(\delta-2\theta)}{\rho'^2\rho} + \frac{O(\rho)}{\rho}, \tag{C 3}$$

where $\delta = 2\theta' - \alpha' + \alpha$. Only the first term is singular, and it vanishes upon integration over θ. Therefore (C 2) converges conditionally. The divergence itself is only logarithmic.

(*b*) $\rho_t = 0$. Again the singularity is two dimensional, and it corresponds to trajectories which end at the solenoid. The analysis proceeds exactly as in (*a*) and is therefore omitted.

(*c*) $\rho = \rho_t = 0$, $t \neq 0$. These occur at isolated points in $(\alpha, t, \boldsymbol{r})$ space, and correspond to trajectories which begin and end at the solenoid. In the neighbourhood of such a point it is convenient to change variables from α and t to $\rho' \overset{\text{def}}{=} \rho_t$ and $\theta' \overset{\text{def}}{=} \theta_t$. I_c, the contribution of the neighbourhood of the singularity, is given by

$$I_c = \int_0 \mathrm{d}\rho \int_0^{2\pi} \mathrm{d}\theta \int_0 \mathrm{d}\rho' \int_0^{2\pi} \mathrm{d}\theta' \frac{1}{J} \frac{\sin(2(\theta'-\theta)-(\alpha_t-\alpha))}{\rho'\rho}, \tag{C 4}$$

where J is the jacobian $|\partial \boldsymbol{r}_t/\partial(\alpha, t)|$. Expanding about $\rho = 0$ and $\rho' = 0$, we get that $\alpha_t - \alpha$ is of the form $(\text{const.} + f(\theta)\rho + g(\theta')\rho' + O(\rho^{1-m}\rho'^{1-n}))$, where $m, n \geqslant 0$. Assuming that J does not vanish at $\rho = \rho' = 0$, the singular terms in the integrand vanish upon integration over θ and θ', and the divergences in ρ and ρ' are logarithmic. (The case where J does vanish corresponds to a coincidence of closed orbits and caustics, and occurs only on a one-dimensional set in parameter space, and therefore not for generic $\boldsymbol{R}$. This set includes self-conjugate points along periodic orbits.)

(*d*) $\rho = 0$, $t = 0$. The singularity is one dimensional (it is parametrized by the initial direction α) and is the strongest of the four. As in (*c*) we change variables from α and t to $\rho' \overset{\text{def}}{=} \rho_t$ and $\theta' \overset{\text{def}}{=} \theta_t$. To first order in $L \overset{\text{def}}{=} |\boldsymbol{\rho}' - \boldsymbol{\rho}|$, i.e. short times,

$$t = \frac{L}{v}, \quad \alpha = \arctan\left(\frac{\rho'\sin\theta' - \rho\sin\theta}{\rho'\cos\theta' - \rho\cos\theta}\right), \quad J \overset{\text{def}}{=} \left|\frac{\partial(\alpha, t)}{\partial(\rho', \theta')}\right| = \frac{\rho'}{vL}. \tag{C 5}$$

Also, $\alpha_t - \alpha = \omega t = \omega L/v$, where ω is the Larmor frequency ($= B$ in our units.) Thus to first order in L,

$$I_d = \int_0 \mathrm{d}\rho \int_0^{2\pi} \mathrm{d}\theta \int_0 \mathrm{d}\rho' \int_0^{2\pi} \mathrm{d}\theta \left[\frac{\sin 2(\theta'-\theta)}{v\rho\rho'L} - \frac{\omega\cos 2(\theta'-\theta)}{v^2\rho\rho'}\right]. \tag{C 6}$$

Both terms vanish on integrating over $\theta' - \theta$, the first because the θ and θ' dependence in L is through $\cos(\theta' - \theta)$. Thus I_d is conditionally convergent, although the leading-order divergence is stronger than logarithmic.

Appendix D. Derivation of (6.1)

Differentiating the spectral resolution of the propagator

$$U(t) = \sum_j P_j \exp(-\mathrm{i}\omega_j t) \tag{D 1}$$

with respect to parameters (here $P_j = |j\rangle\langle j|$ and $\omega_j = E_j/\hbar$), we get that

$$\langle \mathrm{d}U(t)\cdot n| \wedge |\mathrm{d}U(t)\cdot n\rangle = \sum_{jk} \langle n|(\mathrm{d}P_j + \mathrm{i}tP_j\,\mathrm{d}\omega_j) \wedge (\mathrm{d}P_k - \mathrm{i}tP_k\,\mathrm{d}\omega_k)|n\rangle\, \mathrm{e}^{\mathrm{i}\omega_{jk}t}. \tag{D 2}$$

where $\omega_{jk} = \omega_j - \omega_k$. The terms quadratic in t are of the form $\delta_{j,n}\delta_{k,n}(\mathrm{d}\omega_j \wedge \mathrm{d}\omega_k)$ and vanish by antisymmetry. The terms linear in t are of the form $\delta_{k,n}\langle n|\mathrm{d}P_j|n\rangle \wedge \mathrm{d}\omega_k$. These vanish because $\langle n|\mathrm{d}P_j|n\rangle = 0$. We are left with

$$\langle \mathrm{d}U(t)\cdot n| \wedge |\mathrm{d}U(t)\cdot n\rangle = \sum_{jk} \langle n|\mathrm{d}P_j \wedge \mathrm{d}P_k|n\rangle\, \mathrm{e}^{\mathrm{i}\omega_{jk}t}. \tag{D 3}$$

Next we substitute $|\mathrm{d}j\rangle\langle j| + |j\rangle\langle \mathrm{d}j|$ for $\mathrm{d}P_j$ and similarly for $\mathrm{d}P_k$ to obtain

$$\begin{aligned}\langle \mathrm{d}U(t)\cdot n| \wedge |\mathrm{d}U(t)\cdot n\rangle = {} & \langle \mathrm{d}n| \wedge |\mathrm{d}n\rangle + \sum_j \langle n|\mathrm{d}j\rangle \wedge \langle \mathrm{d}j|n\rangle \\ & + \sum_j (\langle n|\mathrm{d}j\rangle \wedge \langle j|\mathrm{d}n\rangle\, \mathrm{e}^{\mathrm{i}\omega_{jn}t} + \langle \mathrm{d}n|j\rangle \wedge \langle \mathrm{d}j|n\rangle\, \mathrm{e}^{\mathrm{i}\omega_{nj}t}).\end{aligned} \tag{D 4}$$

Differentiating $\langle n|j\rangle = \delta_{nj}$, we get that $\langle n|\mathrm{d}j\rangle = -\langle \mathrm{d}n|j\rangle$. This implies that (i) the second sum may be written as $\sum_j \langle \mathrm{d}n|j\rangle \wedge \langle j|\mathrm{d}n\rangle$ and the sum performed to give $\langle \mathrm{d}n| \wedge |\mathrm{d}n\rangle$, and (ii) the prefactors of the two exponentials in the third sum are the same, and the $j = n$ term vanishes. Thus we obtain

$$\langle \mathrm{d}U(t)\cdot n| \wedge |\mathrm{d}U(t)\cdot n\rangle = 2\langle \mathrm{d}n| \wedge |\mathrm{d}n\rangle - 2\sum_{j\neq n} \langle \mathrm{d}n|j\rangle \wedge \langle j|\mathrm{d}n\rangle \cos\omega_{jn}t. \tag{D 5}$$

which when multiplied by $-\frac{1}{2}\mathrm{i}\hbar$ gives (6.1).

Appendix E. Symplectic form on Hilbert space

It is well known that the equations of quantum mechanics can be cast in hamiltonian form (see, for example, Abraham & Marsden 1978). Our purpose here is to do so in a manner motivated by the correspondence principle: we make the convention that the hamiltonian functional $\mathscr{H}(\psi)$ (which plays the role of H in Hamilton's equations) should be given by the energy expectation value $\langle\psi|\hat{H}|\psi\rangle$ (usually it is taken to be half of this). As we now show, this convention implies the following definition of the symplectic form:

$$[\phi, \chi] \overset{\text{def}}{=} -2\hbar\,\mathrm{Im}\langle\phi|\chi\rangle. \tag{E 1}$$

Our main interest is to explain the origin of the numerical factor $-2\hbar$, so we do not take pains to introduce a precise notation. Formally Hamilton's equations are $\dot{\psi} = J\cdot\mathscr{H}'(\psi)$. Therefore $J^{-1}\cdot\dot{\psi} = \mathscr{H}'(\psi)$, so that for arbitrary ϕ,

$$[\phi, \dot{\psi}] \overset{\text{def}}{=} \langle\phi|J^{-1}\cdot\dot{\psi}\rangle = \mathscr{H}'(\psi)\cdot\phi.$$

But $$\mathscr{H}'(\psi)\cdot\phi \overset{\text{def}}{=} (\mathrm{d}/\mathrm{d}\epsilon)_0 \mathscr{H}(\psi + \epsilon\phi) = 2\,\mathrm{Re}\langle\phi|\hat{H}\cdot\psi\rangle,$$

and from Schrödinger's equation, $\hat{H}\cdot\psi = \mathrm{i}\hbar\dot{\psi}$. Therefore

$$[\phi, \dot{\psi}] = 2\,\mathrm{Re}\langle\phi|\mathrm{i}\hbar\dot{\psi}\rangle = -2\hbar\,\mathrm{Im}\langle\phi|\dot{\psi}\rangle.$$

As $\dot{\psi}$ is arbitrary, (E 1) follows.

Appendix F. Derivation of (6.6)

Consider the quantity $|S^{-1}(\tau)\cdot d\boldsymbol{Z}(\tau)\wedge S^{-1}(\tau')\cdot d\boldsymbol{Z}(\tau')|$ and its derivative with respect to τ and τ'. (Here $\boldsymbol{Z}(\tau)$ and $S(\tau)$ are abbreviations for the flow $\boldsymbol{Z}(\boldsymbol{z},\tau)$ and the linearized flow $S(\boldsymbol{z},\tau)$ of (3.11); the $\boldsymbol{z}$ dependence is left implicit.) From (6.9) and from $\dot{S}^{-1}(\tau)=-S^{-1}(\tau)J(H'')$, we get that

$$\frac{\partial}{\partial\tau}(S^{-1}(\tau)\cdot d\boldsymbol{Z}(\tau))=S^{-1}(\tau)J\cdot(d\boldsymbol{H}')_t=J\cdot\nabla_{\boldsymbol{z}}((dH)_t). \tag{F 1}$$

as the terms in H'' cancel. In the last equality we have used $S^{-1}=-JS^{\mathrm{T}}J$. (We remark in passing that (F 1) is the classical analogue of the t-derivative of (6.3).) Therefore

$$\frac{\partial^2}{\partial\tau\,\partial\tau'}|S^{-1}(\tau)\cdot d\boldsymbol{Z}(\tau)\wedge S^{-1}(\tau')\cdot d\boldsymbol{Z}(\tau')|=|J\cdot\nabla_{\boldsymbol{z}}((dH)_t)\wedge J\cdot\nabla_{\boldsymbol{z}}((dH)_t)|$$
$$=-\nabla_{\boldsymbol{z}}((dH)_t)\cdot J\cdot\nabla_{\boldsymbol{z}}((dH)_t) \tag{F 2}$$

by using (6.7) and $J=J^{\mathrm{T}}$. But the last expression is just $-|(dH)_t\wedge(dH)_t|$. Therefore

$$|(dH)_t\wedge(dH)_t|=-\frac{\partial^2}{\partial\tau\,\partial\tau'}|S^{-1}(\tau)\cdot d\boldsymbol{Z}(\tau)\wedge S^{-1}(\tau')\cdot d\boldsymbol{Z}(\tau')| \tag{F 3}$$

Integrating τ and τ' from 0 to t, and noting that $d\boldsymbol{Z}(0)=0$ and

$$|S^{-1}(t)\cdot d\boldsymbol{Z}(t)\wedge S^{-1}(t)\cdot d\boldsymbol{Z}(t)|=|d\boldsymbol{Z}(t)\wedge d\boldsymbol{Z}(t)| \tag{F 4}$$

(the invariance of the symplectic inner product under canonical transformations) we obtain (6.6).

Appendix G. The integrable case: second derivation

We give an alternative derivation of the equivalence of the classical two-form $\Gamma^{\mathrm{C}}(I)$ and the Hannay two-form $\Gamma^{\mathrm{H}}(I)$ based on the formalism of §6, in terms of which (B 1) may be rewritten as

$$\Gamma^{\mathrm{H}}(I)=-\tfrac{1}{2}\langle|d\boldsymbol{y},d\boldsymbol{y}|\rangle_I. \tag{G 1}$$

Here $\boldsymbol{y}(\boldsymbol{\theta},I,R)=(q(\boldsymbol{\theta},I,R),p(\boldsymbol{\theta},I,R))$. As in Appendix B, we restrict ourselves to one degree of freedom, but the generalization to higher dimensions is straightforward.

Let $\boldsymbol{v}=\partial\boldsymbol{y}/\partial\theta$. Then $\frac{1}{2}\langle|\boldsymbol{y},\boldsymbol{v}|\rangle_I=\langle p\partial_\theta q\rangle_I=I$. Differentiating with respect to R, $\langle|d\boldsymbol{y},\boldsymbol{v}|\rangle_I+\langle|\boldsymbol{y},d\boldsymbol{v}|\rangle_I=0$. But $\langle|\boldsymbol{y},d\boldsymbol{v}|\rangle_I=\langle|d\boldsymbol{y},\boldsymbol{v}|\rangle_I$ (this follows from integration by parts over θ and interchanging the arguments of the symplectic inner product), so that

$$\langle|d\boldsymbol{y},\boldsymbol{v}|\rangle_I=0. \tag{G 2}$$

Let $\boldsymbol{Y}(\theta,t)\overset{\mathrm{def}}{=}\boldsymbol{y}(\theta+\omega t)$ (the I and R dependence is left implicit). Equivalently $\boldsymbol{Y}(\theta,t)\overset{\mathrm{def}}{=}\boldsymbol{y}_t(\theta)$ (since in general $f_t(\theta)=f(\theta+\omega t)$), or more simply $\boldsymbol{Y}(t)\overset{\mathrm{def}}{=}\boldsymbol{y}_t$. Since $\boldsymbol{Y}(t)$ satisfies Hamilton's equations, its derivative with respect to parameters,

$$d\boldsymbol{Y}(t)=d(\boldsymbol{y}_t)=(d\boldsymbol{y})_t+d\omega\,t\boldsymbol{v}_t. \tag{G 3}$$

is a solution of (6.9). Substituting (G 3) into (G 11) we get that

$$\Gamma^{\mathrm{C}}=-\tfrac{1}{4}\overline{\langle|(d\boldsymbol{y})_t\wedge(d\boldsymbol{y})_t|\rangle_I}+\tfrac{1}{2}\overline{\langle|d\boldsymbol{Y}(0)\wedge S(\boldsymbol{y}_t)\cdot d\boldsymbol{y}|\rangle_I}-\tfrac{1}{4}\overline{\langle|d\boldsymbol{y}_0\wedge d\boldsymbol{y}|\rangle_I}. \tag{G 4}$$

as terms of the form $\langle|d\boldsymbol{y},\boldsymbol{v}|\rangle_I$ vanish in light of (G 2). The first and third terms in (G 4) are both equal to $\frac{1}{2}\Gamma^{\mathrm{H}}(I)$. It remains to show that the second term vanishes.

Letting $\boldsymbol{w}\overset{\mathrm{def}}{=}\partial\boldsymbol{y}/\partial I$, and resolving $d\boldsymbol{y}$ into its $\boldsymbol{v}$ and $\boldsymbol{w}$ components,

$$d\boldsymbol{y}=\alpha^{\mathrm{T}}\boldsymbol{v}+\alpha^{n}\boldsymbol{w}. \tag{G 5}$$

It is straightforward to verify that $S(\boldsymbol{y},t)\cdot\boldsymbol{v}=\boldsymbol{v}_t$ (or more explicitly, $S(\boldsymbol{y}(\theta),t)\cdot\boldsymbol{v}(\theta)=\boldsymbol{v}(\theta+\omega t)$) and $S(\boldsymbol{y},t)\cdot\boldsymbol{w}=\boldsymbol{w}_t+(\partial\omega/\partial I)t\boldsymbol{v}_t$. Therefore

$$S(\boldsymbol{y},t)\cdot d\boldsymbol{y}=(\alpha^{\mathrm{T}}+(\partial\omega/\partial I)t\alpha^{n})\boldsymbol{v}_t+\alpha^{n}\boldsymbol{w}_t \tag{G 6}$$

Also, since $\omega\boldsymbol{v}=J\cdot\boldsymbol{H}'(\boldsymbol{y})$, it follows that

$$|\boldsymbol{w},\boldsymbol{v}|=(1/\omega)\,\boldsymbol{w}\cdot\boldsymbol{H}'=(1/\omega)\,\partial H/\partial I=1 \tag{G 7}$$

By using (G 3) and (G 5)–(G 7), the second term in (G 4) may be written as half the time average of

$$\langle(\alpha^{n})_t\wedge\alpha^{\mathrm{T}}\rangle_I+(\partial\omega/\partial I)t\langle(\alpha^{n})_t\wedge\alpha^{n}\rangle_I-\langle(\alpha^{\mathrm{T}})_t\wedge\alpha^{n}\rangle_I-\mathrm{d}\omega\wedge\langle\alpha^{n}\rangle_I t \tag{G 8}$$

From (G 2), (G 5) and (G 7) it follows that

$$\langle\alpha^{n}\rangle_I=\overline{(\alpha^{n})_t}=0 \tag{G 9}$$

One can show that (G 9) implies the vanishing of the time average of (G 8), which in turn implies that $\Gamma^{\mathrm{C}}(I)=\Gamma^{\mathrm{H}}(I)$. (Note that $\overline{(\alpha^{n})_t}$ vanishes in the generalized sense of Appendix J.) A similar though more involved argument appears in Appendix K for the periodic orbit two-form.

Let us point out that (G 1) has a geometrical interpretation similar to those described in §6. Suppose we want the flux of $\Gamma^{\mathrm{H}}(I)$ through $\square_R$. From a point $\boldsymbol{y}$ on an invariant torus we draw vectors $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ to points on neighbouring tori with the same actions and angles but with parameter $R+\tau_1$ and $R+\tau_2$. $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ span an area element $\square_I(\boldsymbol{\theta})$ in phase space, as drawn in figure 1*c*. According to (G 1), the required flux is (minus) the symplectic area of $\square_I(\boldsymbol{\theta})$ averaged over the torus. There is a similar construction for the fundamental formula (1.2) for the quantum two-form Γ_n. From $|n(R)\rangle$ we draw vectors $|n_1\rangle$ and $|n_2\rangle$ in Hilbert space to $|n(R+\tau_1)\rangle$ and $|n(R+\tau_2)\rangle$. $|n_1\rangle$ and $|n_2\rangle$ span an area element $\square_H$ in Hilbert space, as in figure 1*d*. According to (1.2), the flux of Γ_n through $\square_R$ is just (minus) the symplectic area of $\square_H$ (as defined in Appendix E). In this light, the correspondence of the quantum and classical two forms, (1.2) and (G 1), is immediate.

Appendix H. Derivative at constant volume

The microcanonical average, regarded as a function of phase volume ω rather than energy E, is given by

$$\langle\phi\rangle_\omega=\int\mathrm{d}^{2N}\boldsymbol{z}\,\delta(\omega-\Omega_p)\phi. \tag{H 1}$$

where $\Omega_p(\boldsymbol{z})\overset{\mathrm{def}}{=}\Omega(H(\boldsymbol{z}))$ is the volume of phase space with energy less than $H(\boldsymbol{z})$. (More explicitly, $\Omega_p(\boldsymbol{z},R)\overset{\mathrm{def}}{=}\Omega(H(\boldsymbol{z},R),R)$.) Equation (H 1) is correctly normalized since

$$\delta(\Omega(E)-\Omega_p(\boldsymbol{z}))=\delta(E-H(\boldsymbol{z}))/\Omega'(E).$$

Differentiating (H 1) with respect to parameters at fixed ω, we get

$$\mathrm{d}\langle\phi\rangle_\omega = \langle\mathrm{d}\phi\rangle_\omega - \int \mathrm{d}^{2N}z\,\delta'(\omega-\Omega_J)\,\mathrm{d}\Omega_J\wedge\phi = \langle\mathrm{d}\phi\rangle_\omega - \frac{\mathrm{d}}{\mathrm{d}\omega}\langle\mathrm{d}\Omega_J\wedge\phi\rangle_\omega. \quad \text{(H 2)}$$

But $\mathrm{d}\Omega_J = \mathrm{d}\Omega + \Omega'\,\mathrm{d}H$, and

$$\mathrm{d}\Omega = -\int \mathrm{d}^{2N}z\,\delta(E-H)\,\mathrm{d}H = -\Omega'\langle\mathrm{d}H\rangle.$$

Therefore

$$\mathrm{d}\Omega_J = -\Omega'(\mathrm{d}E - \mathrm{d}H) \quad \text{(H 3)}$$

where $\mathrm{d}E \overset{\mathrm{def}}{=} \langle\mathrm{d}H\rangle$. We substitute (H 3) into (H 2), and express the result in terms of E rather than ω via $\omega = \Omega(E)$. Noting that $\mathrm{d}/\mathrm{d}\omega = (\Omega')^{-1}\,\mathrm{d}/\mathrm{d}E$,

$$\mathrm{d}\langle\phi\rangle = \langle\mathrm{d}\phi\rangle + (1/\Omega')\{\Omega'\langle(\mathrm{d}E-\mathrm{d}H)\wedge\phi\rangle\}'. \quad \text{(H 4)}$$

Appendix I. Derivations of (7.6) and (7.10)

First we derive (7.6). Starting with the expression $\frac{1}{2}\langle\overline{\mathrm{d}H\wedge[\mathrm{d}\boldsymbol{Z}(t)\cdot\wedge\mathrm{d}\boldsymbol{Z}(t)]}\rangle$ on the left side of (7.6), we express the time average as the residue of the Laplace transform at the origin (as in Appendix J), and replace $[\mathrm{d}\boldsymbol{Z}(t)\cdot\wedge\mathrm{d}\boldsymbol{Z}(t)]$ by its expression in (6.6). The result is

$$-\tfrac{1}{2}\lim_{s\to 0} s\int_0^\infty \mathrm{d}t\,\mathrm{e}^{-st}\int_0^t \mathrm{d}\tau\int_0^t \mathrm{d}\tau'\,\langle\mathrm{d}H\wedge\{(\mathrm{d}H)_\tau\wedge(\mathrm{d}H)_{\tau'}\}\rangle. \quad \text{(I 1)}$$

In writing (I 1) we have used the symmetry of the τ and τ' integrals to restrict the domain of integration to $\tau>\tau'$; this restriction is compensated by an additional factor of 2. Reversing the order of the integrations allows the t integral to be performed, with the result

$$\tfrac{1}{2}\lim_{s\to 0} s\int_0^\infty \mathrm{d}\tau'\int_{\tau'}^\infty \mathrm{d}\tau\,\mathrm{e}^{-s\tau}\langle\mathrm{d}H\wedge\{(\mathrm{d}H)_\tau\wedge(\mathrm{d}H)_{\tau'}\}\rangle \quad \text{(I 2)}$$

Since microcanonical averages are time invariant, $\langle\mathrm{d}H\wedge\{(\mathrm{d}H)_\tau\wedge(\mathrm{d}H)_{\tau'}\}\rangle$ may be replaced by $\langle\mathrm{d}H_{\tau-\tau'}\wedge\{(\mathrm{d}H)_0\wedge(\mathrm{d}H)_{\tau-\tau'}\}\rangle$. After changing variables from τ to $\tau-\tau'$, (I 2) becomes

$$-\lim_{s\to 0}\int_0^\infty \mathrm{d}\tau'\,\mathrm{e}^{-s\tau'}\int_0^\infty \mathrm{d}\tau\,\mathrm{e}^{-s\tau}\langle(\mathrm{d}H)_{-\tau'}\wedge\{(\mathrm{d}H)_0\wedge\mathrm{d}H\}\rangle. \quad \text{(I 3)}$$

This expression may be written as $\lim_{s\to 0}G(s,s)$, where

$$\left.\begin{aligned} G(s,s') &= -\int_0^\infty \mathrm{d}\tau\,\frac{\mathrm{e}^{-s\tau}}{s}\,s'\int_0^\infty \mathrm{d}\tau'\,\mathrm{e}^{-s'\tau'}F(\tau,\tau'),\\ F(\tau,\tau') &= \langle(\mathrm{d}H)_{-\tau}\wedge\{(\mathrm{d}H)_0\wedge\mathrm{d}H\}\rangle. \end{aligned}\right\} \quad \text{(I 4)}$$

as asserted in (7.6) and (7.7). In passing from (I 3) to (I 4) we have in effect multiplied and divided by s.

Next we derive (7.10). Consider

$$\lim_{s\to 0}\frac{1}{2}\int_0^\infty \mathrm{d}\tau\,\frac{\mathrm{e}^{-s\tau}}{s}\langle\{(\mathrm{d}H)_\tau\wedge\mathrm{d}H\}\rangle \quad \text{(I 5)}$$

From (3.12), $\langle\{(\mathrm{d}H)_\tau\wedge\mathrm{d}H\}\rangle = (\Omega'\langle(\mathrm{d}H)_\tau\wedge\mathrm{d}H\rangle)'/\Omega'$. Making this substitution in (I 5) and integrating by parts over τ, we obtain

$$\lim_{s\to 0}\frac{1}{2\Omega'}\int_0^\infty \mathrm{d}\tau\,\mathrm{e}^{-s\tau}(\Omega'\langle(\mathrm{d}H)_\tau\wedge\mathrm{d}H\rangle)' = \frac{1}{2\Omega'}\int_0^\infty \mathrm{d}\tau(\Omega'\langle(\mathrm{d}H)_\tau\wedge\mathrm{d}H\rangle)'. \quad \text{(I 6)}$$

But from (3.15), this last expression is just $\Gamma'(E)$, as asserted in (7.10).

Appendix J. Time average as residue of Laplace transform

Assuming that $\bar f$ exists, we show that

$$\lim_{s\to 0} sF(s) = \bar f. \quad \text{(J 1)}$$

where F is the Laplace transform of f. The result is true for constant functions (easily verified). Then taking $\tilde f = f-\bar f$ to be the oscillatory part of f, it suffices to show that

$$\lim_{s\to 0} s\tilde F(s) = 0, \quad \text{(J 2)}$$

where $\tilde F$ is the Laplace transform of $\tilde f$. Integrating the left-hand side of (J 2) by parts, we get

$$\lim_{s\to 0} s\tilde F(s) = \lim_{s\to 0} s^2\int_0^\infty \mathrm{d}t\,\tilde g(t)\,\mathrm{e}^{-st}, \quad \text{(J 3)}$$

where $\tilde g(t) = \int_0^t \mathrm{d}\tau\,\tilde f(\tau)$. Since the time average of $\tilde f$ vanishes, $|\tilde g(t)|/t\to 0$ as $t\to\infty$. Thus for any $\epsilon>0$ we can take T sufficiently large so that $|\tilde g(t)|<\epsilon t$ for $t>T$. Then dividing the integral in (J 3) between $[0,T]$ and $[T,\infty]$,

$$\lim_{s\to 0}|s\tilde F(s)| < \lim_{s\to 0} s^2\int_0^T \mathrm{d}t\,|\tilde g(t)|\,\mathrm{e}^{-st} + \lim_{s\to 0} s^2\int_T^\infty \mathrm{d}t\,\epsilon t\,\mathrm{e}^{-st} \quad \text{(J 4)}$$

Taking the $s\to 0$ limit on the right-hand side, we get $\lim_{s\to 0}|s\tilde F(s)|<\epsilon$. Since ϵ is arbitrary, (J 2) follows. If F has a meromorphic extension to a neighbourhood of the origin, (J 1) is equivalent to

$$\bar f = \mathrm{Res}_0 F(s). \quad \text{(J 5)}$$

where $\mathrm{Res}_0 F(s)$ denotes the residue of F at the origin. If $\bar f$ does not exist, we may regard (J 5) as its definition. In this way we can say that t^n $(n>0)$ and $\mathrm{e}^{\omega t}$ have time averages equal to zero.

Appendix K. The periodic orbit two-form

$\boldsymbol{y}_j(\theta,S,R)$ denotes a family of periodic orbits parametrized by an angle θ (proportional to the time), action S and parameters R. For convenience we drop the subscript j from $\boldsymbol{y}_j$. Usually the R dependence is left implicit, and sometimes the S and θ dependence is left implicit as well. T_j is the period and $\omega_j = 2\pi/T_j$ is the frequency of the orbit; $\boldsymbol{v} \overset{\mathrm{def}}{=} \partial\boldsymbol{y}/\partial\theta$ is proportional to the velocity.

The periodic orbit two-form is given by

$$\Gamma_j^\circ(S) = -\langle[\mathrm{d}\boldsymbol{y}\cdot\wedge\mathrm{d}\boldsymbol{y}]\rangle_{jS}. \quad \text{(K 1)}$$

$\langle\cdot\rangle_{jS}$ denotes the orbit average $(2\pi)^{-1}\int_0^{2\pi}\mathrm{d}\theta$. While we will not use this result, let us

point out that $\Gamma_j'(S)$ is closed, simply because $\mathrm{d}\mathbf{y}(\theta, S)$ is closed. Note that in §8, the periodic orbit two-form is a function of energy rather than action, as S is set equal to $S_j(E)$. In general $\mathrm{d}(\Gamma_j'(S_j(E))) \neq 0$.

The θ origin along each periodic orbit is arbitrary and may be shifted by the transformation

$$\mathbf{y}(\theta, S, R) \to \mathbf{y}(\theta + F(S, R), S, R). \tag{K 2}$$

However $\Gamma_j'(S)$ remains invariant under this transformation, as we now show. Under (K 2) $\mathrm{d}\mathbf{y}(\theta) \to \mathrm{d}\mathbf{y}(\theta + F) + \mathrm{d}F\,\dot{\mathbf{y}}(\theta + F)$ and the two-form (K 1) acquires an additional term $-2\langle[\mathrm{d}\mathbf{y}\ \mathbf{r}]\rangle_{jS} \wedge \mathrm{d}F$. We have that $\frac{1}{2}\langle[\mathbf{y}, \mathrm{d}\mathbf{r}]\rangle_{jS} = \langle \boldsymbol{p} \cdot \partial_\alpha \boldsymbol{q}\rangle_{jS} = S/2\pi$. Differentiating $\mathrm{d}\langle[\mathbf{y}\ \mathbf{r}]\rangle_{jS} = \langle[\mathrm{d}\mathbf{y}, \mathbf{r}]\rangle_{jS} + \langle[\mathbf{y}, \mathrm{d}\mathbf{r}]\rangle_{jS} = 0$. But $\langle[\mathbf{y}, \mathrm{d}\mathbf{r}]\rangle_{jS} = \langle[\mathrm{d}\mathbf{y}, \mathbf{r}]\rangle_{jS}$ (this follows from integration by parts over θ and reversing the arguments of the symplectic inner product). Therefore

$$\langle[\mathrm{d}\mathbf{y}\ \mathbf{r}]\rangle_{jS} = 0, \tag{K 3}$$

which in turn implies the invariance of the two form.

Next we derive (9.8). The energy of an orbit $E_j(S, R) = H(\mathbf{y}(\theta, S, R), R)$ is independent of θ. Therefore its variation with R and S

$$\delta E = \langle \mathrm{d}H \rangle_{jS}\,\delta R + (1/2\pi)\langle \boldsymbol{H}(\mathbf{y})\cdot \mathrm{d}\mathbf{y}\rangle_{jS}\,\delta R + (1/2\pi)\langle \boldsymbol{H}(\mathbf{y})\cdot \mathbf{w}\rangle_{jS}\delta S, \tag{K 4}$$

where $\mathbf{w} \overset{\mathrm{def}}{=} 2\pi\,\partial\mathbf{y}/\partial S$ is also θ independent. Averaging (K 4) around the orbit and using the fact that $\mathrm{d}\boldsymbol{H}(\mathbf{y}) = \omega_j\,\mathbf{r}$, we obtain

$$\delta E = \langle \mathrm{d}H \rangle_{jS}\,\delta R + \omega_j\langle[\mathrm{d}\mathbf{y}\ \mathbf{r}]\rangle_{jS}\cdot\delta R + (\omega_j/2\pi)\langle[\mathbf{w}, \mathbf{r}]\rangle_{jS}\,\delta S. \tag{K 5}$$

From (K 3), $\langle[\mathrm{d}\mathbf{y}\ \mathbf{r}]\rangle_{jS}$ vanishes. Also $[\mathbf{w}\ \mathbf{r}] = (2\pi/\omega_j)\boldsymbol{H}(\mathbf{y})\cdot\partial\mathbf{y}/\partial S = T_j\,\partial E_j/\partial S = 1$. Therefore (K 5) becomes $\delta E = \langle \mathrm{d}H \rangle_{jS}\,\delta R + \delta S/T_j$, or if S is regarded as a function of E and R

$$\mathrm{d}S = T_j(\mathrm{d}E - \langle \mathrm{d}H \rangle_{jS}) \tag{9.8}$$

the required result (9.8).

The last result to be derived is (8.11). The starting point is the alternative formula (6.11) for $\Gamma_j'(E)$, which remains valid if the microcanonical average is replaced by an orbit average. Let $\boldsymbol{Y}(\theta, t) \overset{\mathrm{def}}{=} \mathbf{y}(\theta + \omega_j t)$ (the S and R dependence is left implicit). Equivalently, $\boldsymbol{Y}(\theta, t) \overset{\mathrm{def}}{=} \mathbf{y}_j(\theta)$ (since in general $f_j(\theta) = f(\theta + \omega_j t)$) or more simply $\boldsymbol{Y}(t) \overset{\mathrm{def}}{=} \mathbf{y}_t$. Since $\boldsymbol{Y}(t)$ satisfies Hamilton's equations, its derivative with respect to parameters

$$\mathrm{d}\boldsymbol{Y}(t) = \mathrm{d}(\mathbf{y}_t) = (\mathrm{d}\mathbf{y})_t + \mathrm{d}\omega_j t\,\mathbf{r}_t \tag{K 6}$$

is a solution of (6.9). From (K 6), (6.11) and (8.9) we obtain

$$\Gamma_j'(E) = -\tfrac{1}{4}\langle[\mathrm{d}\mathbf{y}_t \wedge \mathrm{d}\mathbf{y}_t]\rangle_{jE} + \tfrac{1}{2}\langle[\mathrm{d}\boldsymbol{Y}(t) \wedge \overline{\boldsymbol{S}(\mathbf{y}, t)\cdot\mathrm{d}\mathbf{y}}]\rangle_{jE} - \tfrac{1}{4}\langle[\overline{\mathrm{d}\mathbf{y}} \wedge \overline{\mathrm{d}\mathbf{y}}]\rangle_{jE}, \tag{K 7}$$

where the average $\langle\ \rangle_{jE}$ is taken over the orbit with energy E, and terms of the form $\langle[\mathrm{d}\mathbf{y}, \mathbf{r}]\rangle_{jE}$ vanish in light of (K 3). Both the first and third terms of (K 7) are equal to $\frac{1}{2}\Gamma_j'(S_j(E))$. It remains to show that the second term vanishes.

Expressing its time average as the residue of the Laplace transform at the origin (as in Appendix J), we may write the second term of (K 7) as $\mathrm{Res}_0 F(s)$, where $F(s)$ is the Laplace transform of

$$f(t) = \langle[\mathrm{d}\boldsymbol{Y}(t), \boldsymbol{S}(\mathbf{y}, t)\cdot\mathrm{d}\boldsymbol{Y}(0)]\rangle_{jE}. \tag{K 8}$$

For the explicit evaluation of $f(t)$ we introduce a Floquet basis along the orbit, in terms of which the action of the linearized flow $\boldsymbol{S}(\mathbf{y}, t)$ is simply expressed. The Floquet basis consists of $\mathbf{r} = \partial\mathbf{y}/\partial\theta$ tangent to the orbit, $\mathbf{w} = 2\pi\,\partial\mathbf{y}/\partial S$ transverse to the energy shell on which the orbit lies, and $\boldsymbol{\xi}^{\lambda\pm}$, $\boldsymbol{\xi}^{\lambda}$, $k = 1, \ldots, N-1$, which span the orbit's stable and unstable manifolds, respectively. $\mathbf{r}$, $\mathbf{w}$ and $\boldsymbol{\xi}^{\lambda\pm}$ are functions of θ, S, R. Then one can show that

$$\begin{aligned} \boldsymbol{S}(\mathbf{y}, t)\cdot\mathbf{r} &= \mathbf{r}_t, \quad \boldsymbol{S}(\mathbf{y}, t)\cdot\mathbf{w} = \mathbf{w}_t + (\omega_j/T_j)t\mathbf{r}_t \\ \boldsymbol{S}(\mathbf{y}, t)\cdot\boldsymbol{\xi}^{\lambda\pm} &= \mathrm{e}^{\pm\lambda_k t}\boldsymbol{\xi}^{\lambda\pm}_t, \end{aligned} \tag{K 9}$$

(The first equation is written more explicitly as $\boldsymbol{S}(\mathbf{y}(\theta), t)\cdot\mathbf{r}(\theta) = \mathbf{r}(\theta + \omega_j t)$, and similarly for the others.) Both λ_k and $\boldsymbol{\xi}^{\lambda\pm}$ may be complex, but we assume that

$$\lambda_k \neq n\omega_j \tag{K 10}$$

(i.e. that the stability exponents are either elliptic or hyperbolic).

The Floquet basis can be chosen to be symplectic, so that

$$[\boldsymbol{\xi}^{\lambda+}, \boldsymbol{\xi}^{\lambda'+}] = 0, \quad [\boldsymbol{\xi}^{\lambda}\ \boldsymbol{\xi}^{\lambda'}] = 0 \quad [\mathbf{r}\ \boldsymbol{\xi}^{\lambda\pm}] = 0 \quad [\mathbf{w}, \boldsymbol{\xi}^{\lambda\pm}] = 0 \tag{K 11a}$$

$$[\boldsymbol{\xi}^{\lambda+}\ \boldsymbol{\xi}^{\lambda'-}] = \delta_{\lambda\lambda'} \quad [\mathbf{w}\ \mathbf{r}] = 1 \tag{K 11b}$$

Among these relations, the homogeneous equations (K 11*a*) are a direct consequence of (K 9). For example,

$$[\boldsymbol{\xi}^{\lambda+}\ \boldsymbol{\xi}^{\lambda'+}] = [\boldsymbol{S}(\mathbf{y}, T_j)\ \boldsymbol{\xi}^{\lambda+}, \boldsymbol{S}(\mathbf{y}, T_j)\ \boldsymbol{\xi}^{\lambda'+}] = \exp[-(\lambda_\lambda + \lambda_{\lambda'})T_j][\boldsymbol{\xi}^{\lambda+}\ \boldsymbol{\xi}^{\lambda'+}];$$

(the first equality follows from the invariance of the symplectic inner product and the second from (K 9)). Therefore

$$[\exp[-(\lambda_\lambda + \lambda_{\lambda'})T_j] - 1][\boldsymbol{\xi}^{\lambda+}\ \boldsymbol{\xi}^{\lambda'+}] = 0.$$

Since $\exp[-(\lambda_\lambda + \lambda_{\lambda'})T] \neq 1$ (cf. (K 10)), this implies that $[\boldsymbol{\xi}^{\lambda+}, \boldsymbol{\xi}^{\lambda'+}] = 0$. Of the inhomogeneous equations (K 11*b*) $[\boldsymbol{\xi}^{\lambda+}, \boldsymbol{\xi}^{\lambda-}] = 1$ is simply a normalization convention, whereas $[\mathbf{w}\ \mathbf{r}] = 1$ was shown in the discussion following (K 5).

Expanding $\mathrm{d}\mathbf{y}$ in the Floquet basis,

$$\left.\begin{aligned} \mathrm{d}\mathbf{y} &= \alpha^{\mathrm{r}}\mathbf{r} + \alpha^{\mathrm{w}}\mathbf{w} + \sum_{k=1}^{N-1}(\alpha^{\lambda+}\boldsymbol{\xi}^{\lambda+} + \alpha^{\lambda-}\boldsymbol{\xi}^{\lambda-}) \\ \alpha^{\mathrm{r}} &= [\mathbf{w}, \mathrm{d}\mathbf{y}], \quad \alpha^{\mathrm{w}} = -[\mathbf{r}\ \mathrm{d}\mathbf{y}], \quad \alpha^{\lambda\pm} = \mp[\boldsymbol{\xi}^{\lambda\pm}, \mathrm{d}\mathbf{y}]. \end{aligned}\right\} \tag{K 12}$$

the expressions for the coefficients α (which are functions of θ, S, R) follow from (K 11). Substituting (K 9), (K 11) and (K 12) into (K 8), one can show that

$$\begin{aligned} f(t) = \phi^{\mathrm{r}\mathrm{r}}(t) + \phi^{\mathrm{w}\mathrm{r}}(-t) + (\omega_j/T_j)\phi^{\mathrm{w}\mathrm{w}}(t)t + \alpha^{\mathrm{r}}{}_t \wedge \mathrm{d}\omega_j t \\ + \sum_{k=1}^{N-1}(\phi^{\lambda}(t)\mathrm{e}^{-\lambda_k t} + \phi^{\lambda}(-t)\mathrm{e}^{-\lambda_k t}). \end{aligned} \tag{K 13}$$

where

$$\phi^{\mathrm{r}\mathrm{r}}(t) = \langle(\alpha^{\mathrm{r}})_t \wedge \alpha^{\mathrm{r}}\rangle_{jE}, \quad \phi^{\mathrm{w}\mathrm{r}}(t) = \langle(\alpha^{\mathrm{w}})_t \wedge \alpha^{\mathrm{r}}\rangle_{jE}, \quad \phi^{\lambda}(t) = \langle(\alpha^{\lambda+})_t \wedge \alpha^{\lambda-}\rangle_{jE}. \tag{K 14}$$

From (K 3) and (K 12) it follows that

$$\langle\alpha^{\mathrm{w}}\rangle_{jE} = \overline{(\alpha^{\mathrm{w}})}_t = 0 \tag{K 15}$$

Equation (K 15) in turn implies that the Laplace transform of the first three terms in (K 13) have no poles at the origin. As for the remaining terms, since $\phi^k(t)$ is T_j-periodic, the Laplace transform of $\phi^k(\pm t)\exp(\pm\lambda_k t)$ can have poles only at $\pm\lambda_k + in\omega_j$; from (K 10) none of these lie at the origin. Thus $\mathrm{Res}_0 F(s) = 0$, and the second term in (K 7) vanishes, as claimed. This implies in turn than $\Gamma_j^{\mathrm{tr}}(E) = \Gamma_j^{\mathrm{tr}}(S_j(E))$, the required result (8.11).

References

Abraham R & Marsden J E 1978 *Foundations of mechanics* p 461 Reading: Benjamin/Cummings

Arnold, V I 1978 *Mathematical methods of classical mechanics* New York: Springer

Arnold, V I & Avez A 1989 *Ergodic problems of classical mechanics* Redwood City: Addison-Wesley

Avron, J E, Seiler R & Jaffe L G 1987 *Commun math Phys* **110** 33–49

Avron, J E Raveh A & Zur B 1988 *Rev mod Phys* **60** 873–915

Baker G A Jr 1958 *Phys Rev* **109** 2198–2206

Bellman, R 1960 *Introduction to matrix analysis* p 171 New York: McGraw-Hill

Berry, M V 1983 In *Les Houches Lecture Series 36* (ed G Iooss R H G Helleman & R Stora) pp 171–271 Amsterdam: North-Holland

Berry, M V 1984 *Proc R Soc Lond* A **392** 45–57

Berry, M V 1985 *J Phys* A **18** 15–27

Berry, M V 1989 *Proc R Soc Lond* A **423** 219–231

Berry, M V 1990 In *Anomalies, phases, defects* (ed M Bregola G Marmo & G Morandi) pp 125–181 Naples: Bibliopolis

Berry, M V 1991 In *Les Houches Lecture Series 52* (ed M J Giannoni & A Voros) pp 251–303 Amsterdam: North-Holland

Gohm, S Knauf A & Marmo S 1989 *Commun math Phys* **123** 95–122

Groenewold, H J 1946 *Physica* **12** 405–460

Gutzwiller, M C 1971 *J math Phys* **12** 343–358

Gutzwiller, M C 1990 *Chaos in quantum mechanics* Berlin: Springer

Hannay, J H 1985 *J Phys* A **18** 221–230

Hill, D L & Wheeler J A 1952 *Phys Rev* **89** 1102–1145

Keating, J P 1991 In *Lecture notes for International School of Physics Enrico Fermi course CXIX* Amsterdam: North-Holland

Mondragon, R J & Berry, M V 1989 *Proc R Soc Lond* A **424** 263–278

Montgomery, R 1988 *Commun math Phys* **120** 269–294

Moyal, J E 1949 *Proc Camb phil Soc* **45** 99–124

Robnik, M & Berry, M V 1986 *J Phys* A **19** 669–682

Ruelle D 1986 *Phys Rev Lett* **56** 405–407

Shapere, A & Wilczek F (eds) 1989 *Geometric phases in physics* Singapore: World Scientific

Simon, B 1983 *Phys Rev Lett* **51** 2167–2170

Von Neumann, J & Wigner, E 1929 *Phys Z* **30** 467–470

Voros, A 1976 *Ann Inst Henri Poincaré* A **24** 31–90

Wilkinson, M 1990 *Phys Rev* A **41** 4645–4652

Received 11 November 1991; accepted 30 December 1991

For the explicit evaluation of $f(t)$ we introduce a Floquet basis along the orbit, in terms of which the action of the linearized flow $S(y,t)$ is simply expressed. The Floquet basis consists of $v = \partial y/\partial\theta$ tangent to the orbit, $w = 2\pi\,\partial y/\partial S$ transverse to the energy shell on which the orbit lies, and ξ^{k+}, ξ^{k-}, $k = 1, \ldots, N-1$, which span the orbit's stable and unstable manifolds respectively. v, w and $\xi^{k\pm}$ are functions of θ, S, R. Then one can show that

$$\left.\begin{aligned} S(y,t)\cdot v &= v_t, \quad S(y,t)\cdot w = w_t + (\partial\omega_j/\partial S)\,t v_t, \\ S(y,t)\cdot\xi^{k\pm} &= e^{\pm\lambda_k t}(\xi^{k\pm})_t. \end{aligned}\right\} \tag{K 9}$$

(The first equation is written more explicitly as $S(y(\theta)\,t)\cdot v(\theta) = v(\theta+\omega_j t)$ and similarly for the others.) Both λ_k and $\xi^{k\pm}$ may be complex, but we assume that

$$\lambda_k \neq in\omega_j \tag{K 10}$$

i.e. that the stability exponents are either elliptic or hyperbolic.

The Floquet basis can be chosen to be symplectic, so that

$$[\xi^{k+}, \xi^{l-}] = 0, \quad [\xi^{k+}, \xi^{l+}] = 0, \quad [v, \xi^{k\pm}] = 0 \quad [w, \xi^{k\pm}] = 0, \tag{K 11a}$$

$$[\xi^{k+}, \xi^{l-}] = \delta_{kl}, \quad [w, v] = 1 \tag{K 11b}$$

Among these relations, the homogeneous equations (K 11*a*) are a direct consequence of (K 9). For example,

$$[\xi^{k+}, \xi^{l-}] = [S(y,T_j)\;\xi^{k+}, S(y,T_j)\;\xi^{l-}] = \exp(-(\lambda_k+\lambda_l)T_j)[\xi^{k+}, \xi^{l-}].$$

(the first equality follows from the invariance of the symplectic inner product and the second from (K 9)). Therefore

$$(\exp(-(\lambda_k+\lambda_l)T_j) - 1)[\xi^{k+}, \xi^{l-}] = 0$$

Since $\exp(-(\lambda_k+\lambda_l)T_j) \neq 1$ (cf. (K 10)) this implies that $[\xi^{k+}, \xi^{l-}] = 0$. Of the inhomogeneous equations (K 11*b*), $[\xi^{k+}, \xi^{k-}] = 1$ is simply a normalization convention, whereas $[w, v] = 1$ was shown in the discussion following (K 5).

Expanding dy in the Floquet basis,

$$\left.\begin{aligned} dy &= \alpha^v v + \alpha^w w + \sum_{k=1}^{N-1}(\alpha^{k+}\xi^{k+} + \alpha^{k-}\xi^{k-}) \\ \alpha^v &= [w, dy], \quad \alpha^w = -[v, dy], \quad \alpha^{k\pm} = \mp[\xi^{k\pm}, dy]. \end{aligned}\right\} \tag{K 12}$$

the expressions for the coefficients α (which are functions of θ, S, R) follow from (K 11). Substituting (K 9), (K 11) and (K 12) into (K 8) one can show that

$$f(t) = \phi^{vv}(t) + \phi^{vv}(-t) + (\partial\omega_j/\partial S)\,\phi^{vw}(t)\,t + \langle\alpha^v\rangle_t \wedge d\omega_j t + \sum_{k=1}^{N-1}(\phi^k(t)e^{\lambda_k t} + \phi^k(-t)e^{-\lambda_k t}), \tag{K 13}$$

where

$$\phi^{vv}(t) = \langle(\alpha^v)_t \wedge \alpha^v\rangle_{fE}, \quad \phi^{vw}(t) = \langle(\alpha^v)_t \wedge \alpha^w\rangle_{fE}, \quad \phi^k(t) = \langle(\alpha^{k+})_t \wedge \alpha^{k-}\rangle_{fE}. \tag{K 14}$$

From (K 3) and (K 12) it follows that

$$\langle\alpha^v\rangle_{fE} = \overline{(\alpha^v)_t} = 0. \tag{K 15}$$

Equation (K 15) in turn implies that the Laplace transform of the first three terms in (K 13) have no poles at the origin. As for the remaining terms, since $\phi^k(t)$ is T_j-periodic, the Laplace transform of $\phi^k(\pm t)\exp(\pm\lambda_k t)$ can have poles only at $\pm\lambda_k + in\omega_j$, from (K 10) none of these lie at the origin. Thus $\mathrm{Res}_0 F(s) = 0$, and the second term in (K 7) vanishes, as claimed. This implies in turn that $V_j''(E) = V_j''(S_j(E))$, the required result (8 11)

References

Abraham R & Marsden J E 1978 *Foundations of mechanics* p 461 Reading Benjamin Cummings

Arnold V I 1978 *Mathematical methods of classical mechanics* New York Springer

Arnold V I & Avez A 1989 *Ergodic problems of classical mechanics* Redwood City Addison Wesley

Avron J E Seiler R & Jaffe L G 1987 *Commun math Phys* **110** 33 49

Avron J E Raveh A & Zur B 1988 *Rev mod Phys* **60** 873 913

Baker G A Jr 1958 *Phys Rev* **109** 2198 2206

Bellman R 1960 *Introduction to matrix analysis* p 171 New York McGraw Hill

Berry M V 1983 In *Les Houches Lecture Series* 36 ed G Iooss R H G Helleman & R Stora pp 171 271 Amsterdam North Holland

Berry M V 1984 *Proc R Soc Lond* A **392** 45 57

Berry M V 1985 *J Phys* A **18** 15 27

Berry M V 1989 *Proc R Soc Lond* A **423** 219 231

Berry M V 1990 In *Anomalies phases defects* ed M Bregola G Marmo & G Morandi pp 125 181 Naples Bibliopolis

Berry M V 1991 In *Les Houches Lecture Series* 52 ed M J Giannoni & A Voros pp 251 303 Amsterdam North Holland

Golin S Knauf A & Marmi S 1989 *Commun math Phys* **123** 95 122

Greenwood H J 1946 *Physica* **12** 405 460

Gutzwiller M C 1971 *J math Phys* **12** 343 358

Gutzwiller M C 1990 *Chaos in quantum mechanics* Berlin Springer

Hannay J H 1985 *J Phys* A **18** 221 230

Hill D L & Wheeler J A 1953 *Phys Rev* **89** 1102 1145

Keating J P 1991 In *Lecture notes for International School of Physics Enrico Fermi course CXIX* Amsterdam North Holland

Mondragon R J & Berry M V 1989 *Proc R Soc Lond* A **424** 263 278

Montgomery R 1988 *Commun math Phys* **120** 269 294

Moyal J E 1949 *Proc Camb phil Soc* **45** 99 124

Robnik M & Berry M V 1986 *J Phys* A **19** 669 682

Ruelle D 1986 *Phys Rev Lett* **56** 405 407

Shapere A & Wilczek F (eds) 1989 *Geometric phases in physics* Singapore World Scientific

Simon B 1983 *Phys Rev Lett* **51** 2167 2170

Von Neumann J & Wigner E 1929 *Phys Z* **30** 467 470

Voros A 1976 *Ann Inst Henri Poincaré* A **24** 31 90

Wilkinson M 1990 *Phys Rev* A **41** 4645 4652

Received 11 November 1991 accepted 30 December 1991

Universal transition prefactors derived by superadiabatic renormalisation

M V Berry and R Lim

H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, U.K.

Abstract

For time-dependent two-state quantum systems, the transition probability is exponentially small in the adiabatic parameter ε, with the exponent determined by a transition point t_c in the complex time plane. Here we study the ε-independent prefactors associated with different sorts of transition point (which need not correspond to complex degeneracies of the adiabatic energy). Unlike previous approaches the method we use does not make use of special functions. It consists of applying first-order perturbation theory to the Schrödinger equation obtained by transforming to a series of 'superadiabatic' bases clinging ever more closely to the evolving state. If the original matrix elements share a leading singularity $(t-t_c)^r$, and their fractional deviation from this is $(t-t_c)^s$, the prefactor is

$$4\sin^2\left\{\frac{\pi s}{2(2r+s+2)}\right\}$$

This is universal in the sense of being invariant under time reparameterisation and quantum changes of frame.

Short title: Transition prefactors

Submitted to: *J. Phys. A.*, April 1993

PACS: 03.65

1. Introduction

In the simplest model for quantum transitions, a system with two states evolves under a time-dependent Hamiltonian operator $H(t)$. This has many physical applications (see e.g. Garraway et al. 1993). It is well known (see e.g. Davis and Pechukas 1976) that for Hamiltonians which are analytic on the real time axis the probability for a transition after infinite time from one of the instantaneous eigenstates to the other is exponentially small in the adiabatic parameter ε describing the speed with which $H(t)$ varies. The exponent involves

$$w_c = 2\int_0^{t_c} \mathrm{d}t\, E(t) \qquad (1)$$

where $E(t)$ is the instantaneous energy, defined as the eigenvalue of $H(t)$ which is positive for real t (we assume $E(t)$ has no real zeros), and t_c is a point in the complex plane where the adiabatic transition can be considered to originate. Therefore these transitions describe real physics in the complex plane. In the most familiar case, the transition point t_c is a simple zero of $E^2(t)$, and (when $H(t)$ is real symmetric) the ε-independent prefactor multiplying the exponential is unity.

Here we are concerned with the different prefactors that can occur when the transition point is not a simple zero of $E^2(t)$. Demkov et al (1978) calculated a class of such prefactors, in which t_c is a higher-order zero, and Joye (1993) has provided a rigorous treatment that also covers more general sorts of transition point. Earlier, Pokrovskii and Khalatnikov (1961) found the prefactors for the analogous problem of above-barrier reflection in the semiclassical approximation. All these authors use a comparison-equation technique, in which $H(t)$ is approximated near t_c and the resulting approximate Schrödinger equation solved exactly in terms of hypergeometric functions.

We have two reasons for presenting another calculation of these prefactors: first, to emphasise their wide universality class, and second, because of the independent interest of the method we use. This is first-order perturbation theory, applied not to states in the usual adiabatic basis (which is known to give the wrong prefactor) but to a sequence of 'superadiabatic' bases that cling ever more closely to the evolving state;

the corresponding sequence of prefactors renormalises onto the correct value. No knowledge of special functions is required in this method, which can therefore be regarded as elementary. It was introduced by Berry (1990a) and applied to obtain the prefactor of unity for simple transition points, in a paper whose main purpose was to study the history of the transition, that is the growth of the probability amplitude from zero to its exponentially small final value (see also Berry 1990b and Lim and Berry 1991).

Confusion should be avoided between the prefactors we study here and the recently-discovered 'geometric amplitudes' (Berry 1990c, Joye et al 1991, Zwanziger et al 1991). Geometric amplitudes are also independent of ε, but arise from spinor rotations associated with the complex Hermitian nature of $H(t)$, rather than from more complicated transition points.

2. Preliminaries

We seek approximate solutions of the Schrödinger equation

$$i\varepsilon\dot{\Psi}(t) = H(t)\Psi(t) \tag{2}$$

in the adiabatic limit of small ε. Here the dot denotes differentiation with respect to time, and

$$\Psi(t) \equiv \begin{pmatrix} \Psi_1(t) \\ \Psi_2(t) \end{pmatrix},$$
$$H(t) \equiv \begin{pmatrix} Z(t) & X(t) \\ X(t) & -Z(t) \end{pmatrix} \equiv E(t) \begin{pmatrix} \cos\theta(t) & \sin\theta(t) \\ \sin\theta(t) & -\cos\theta(t) \end{pmatrix} \tag{3}$$

We assume that $\theta(t)$ is asymptotically constant as $t \to \pm\infty$. The adiabatic states (proportional to instantaneous eigenstates of $H(t)$) are

$$\psi_{\pm}(t) = \exp\left\{\mp\frac{i}{\varepsilon}\int_0^t dt' E(t')\right\} u_{\pm}(t), \text{ where}$$
$$u_+(t) \equiv \begin{pmatrix} \cos\frac{1}{2}\theta(t) \\ \sin\frac{1}{2}\theta(t) \end{pmatrix}, \quad u_-(t) \equiv \begin{pmatrix} -\sin\frac{1}{2}\theta(t) \\ \cos\frac{1}{2}\theta(t) \end{pmatrix} \tag{4}$$

Using these as a basis, we write the exact solutions of (2) as

$$\Psi(t) = c_+(t)\psi_+(t) + c_-(t)\psi_-(t) \tag{5}$$

As initial state we choose

$$c_+(-\infty) = 1, \quad c_-(-\infty) = 0 \tag{6}$$

so that the desired transition probability is

$$P = |c_-(+\infty)|^2 \tag{7}$$

It is easy to see that the conventional view, in which adiabatic transitions originate in complex zeros of $E(t)$, misses the essence of the problem. For all zeros can be eliminated by the transformation

$$t \to w, \quad \text{where} \quad w(t) \equiv 2\int_0^t dt' E(t') \tag{8}$$

because this converts (2) into

$$i\varepsilon \begin{pmatrix} \Psi_1' \\ \Psi_2' \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix} \tag{9}$$

(where primes denote differentiation with respect to w), in which the instantaneous energy is constant. We will obtain a formula for P which is invariant under all time reparameterisations, not just the choice (8). Two particular Hamiltonians for which E is constant, and whose prefactors we will study in §5, are H_A and H_B, defined by

$$X_A(t) = \frac{1}{\sqrt{1+t^2}}, \quad Z_A(t) = \frac{t}{\sqrt{1+t^2}} \quad \text{and}$$
$$X_B(t) = \operatorname{sech}\left(\tfrac{1}{2}\pi t\right), \quad Z_B(t) = \tanh\left(\tfrac{1}{2}\pi t\right) \tag{10}$$

H_A was also studied by Suominen et al (1991); H_B was introduced by Demkov and Kunike (1969) (see also Suominen and Garraway 1992).

The true origin of adiabatic transitions is revealed by the further transformation to the adiabatic basis (5). The evolution law for the amplitudes $c_\pm$ (now regarded as functions of w) is

$$c'_{\pm}(w) = \pm\tfrac{1}{2}\theta'(w)\exp\left\{\pm i\frac{w}{\varepsilon}\right\}c_{\mp}(w) \qquad (11)$$

What causes adiabatic transitions are therefore the (complex) singularities of $\theta'(w)$, and these will play a central role in what follows. In §4 we shall find that for a very wide class of functions $X(t)$ and $Z(t)$ the singularities at $w=w_c$ (cf. (1)) are simple poles, that is

$$\theta'(w) \to -\frac{i\gamma}{w - w_c} \qquad (12)$$

where γ is a real constant with the universality property that it depends only on the type of singularity and not on any associated coefficients. This universality and its significance were first appreciated by Davis and Pechukas (1976) for the particular case of a simple zero of $E^2(t)$, where they found $\gamma=1/3$.

From (7) and (11), the transition probability is

$$P = \left| -\tfrac{1}{2}\int_{-\infty}^{\infty} dw\, c_+(w)\theta'(w)\exp\left\{-i\frac{w}{\varepsilon}\right\}\right|^2 \qquad (13)$$

In a first approximation to P, we use lowest-order perturbation theory, in which it is assumed that $c_+(t)$ preserves its initial value of unity. Then (13) depends on the singularities of $\theta'(w)$ in the lower half-plane. Assuming that there is one closest to the real axis (that is, ignoring the 'Stückelberg oscillations' arising from the interference of singularities with equal Im w), we use (12) to obtain

$$P \approx P_0 = \pi^2\gamma^2\exp\left\{-2\frac{|\mathrm{Im}\, w_c|}{\varepsilon}\right\} \qquad (14)$$

As is well known, in this lowest-order perturbation result the prefactor $\pi^2\gamma^2$ is wrong. One way to get the right result is to iterate the equations (11) to obtain the full perturbation series for c_-. This was done by Davis and Pechukas (1976) for a simple zero, and by Berry (1982) for above-barrier reflection from a turning point of arbitrary order. For small ε all terms involve the same exponential but different (ε-independent)

multipliers, whose sum is the correct adiabatic prefactor. In what follows, we employ a different procedure.

3. Superadiabatic renormalisation

As explained by Berry (1990a - to which we refer for many details of the argument of this section), first-order perturbation theory fails because the quantity being calculated is exponentially small and therefore beyond all orders in the small parameter ε. However, first-order perturbation theory can be used if the representation (5), in terms of the adiabatic basis (4) is replaced by the n'th order superadiabatic representation

$$\Psi(w) = c_{n+}(w)\psi_{n+}(w) + c_{n-}(t)\psi_{n-}(w) \qquad (15)$$

for sufficiently large n. Here $\psi_{n\pm}$ are the series solutions of (2) in powers of ε, truncated at ε^n, namely

$$\psi_{n\pm}(w) = \exp\left\{\mp\frac{iw}{2\varepsilon}\right\}\sum_{m=0}^{n}\varepsilon^m u_{m,\pm}(w) \qquad (16)$$

where the vectors $u_{m,\pm}(w)$ will shortly be determined. The zero-order states in this sequence, namely $\psi_{0\pm}$, involve the adiabatic eigenvectors $u_\pm$ in (4).s The infinite series corresponding to (16) diverges, and as is well known this is associated with the existence of the transitions we seek to describe.

In §§2 and 3 of Berry (1990a) it is shown that to lowest order in ε the first-order perturbation solution of the Schrödinger equation satisfied by $c_{n\pm}(w)$ is

$$c_{n-}(w) \approx -i\varepsilon^n\int_{-\infty}^{w} dw'\exp\left\{-i\frac{w'}{\varepsilon}\right\}u_-^{\mathrm{T}}(w')\cdot u_{n+1,+}(w') \qquad (17)$$

where T denotes the transpose. Therefore we require $u_{n+1,+}(w)$, which we find by expanding in the complete set $u_\pm(w)$, that is

$$u_{n,+}(w) = a_n(w)u_+(w) + b_n(w)u_-(w) \qquad (18)$$

and substituting (16) into (2). This gives

$$b_n = 2\frac{a_n'}{\theta'} \tag{19}$$

and the recurrence relation

$$a_n' = -\mathrm{i}\left[\tfrac{1}{4}\theta'^2 a_{n-1} + a_{n-1}'' - \frac{\theta''}{\theta'}a_{n-1}'\right] \qquad \left(a_0(w)=1,\quad a_{n>0}(-\infty)=0\right) \tag{20}$$

Thus the first-order transition probability in the nth superadiabatic basis is

$$P_n \approx \left|2\mathrm{i}\varepsilon^n \int_{-\infty}^{\infty} \mathrm{d}w \frac{a_{n+1}'(w)}{\theta'(w)} \exp\left\{-\mathrm{i}\frac{w}{\varepsilon}\right\}\right|^2 \tag{21}$$

Just as with the adiabatic basis (cf. (13) with $c_+=1$), P depends on singularities w_c of the integrand in the lower half-plane. Therefore it is necessary to solve (20) for the coefficients a_n near w_c, where (§4) θ' has the form (12). The exact solution of

$$a_n' = \mathrm{i}\left[\frac{\gamma^2}{4(w-w_c)^2}a_{n-1} - a_{n-1}'' - \frac{1}{w-w_c}a_{n-1}'\right] \qquad \left(a_0(w)=1,\quad a_{n>0}(-\infty)=0\right) \tag{22}$$

is

$$a_n(w) \approx \frac{\mathrm{i}^n\left(n-1-\frac{1}{2}\gamma\right)!\left(n-1+\frac{1}{2}\gamma\right)!}{(w-w_c)^n\left(-1-\frac{1}{2}\gamma\right)!\left(-1+\frac{1}{2}\gamma\right)!n!} \qquad (w\approx w_c) \tag{23}$$

With this we obtain, from (21),

$$P_n \approx A_n^2 \exp\left\{-2\frac{|\mathrm{Im}\,w_c|}{\varepsilon}\right\} \quad \text{where}$$
$$A_n = 2\pi\frac{\left(n-\frac{1}{2}\gamma\right)!\left(n+\frac{1}{2}\gamma\right)!}{n!^2\left(-\frac{1}{2}\gamma\right)!\left(-1+\frac{1}{2}\gamma\right)!} = 2\sin\left\{\tfrac{1}{2}\pi\gamma\right\}\frac{\left(n-\frac{1}{2}\gamma\right)!\left(n+\frac{1}{2}\gamma\right)!}{n!^2} \tag{24}$$

In the adiabatic basis $n=0$, the prefactor is $A_0=\pi\gamma$, and (24) reproduces the incorrect result (14). As n increases, the prefactor renormalises onto $2\sin\{\pi\gamma/2\}$, giving the transition probability

$$P \approx P_\infty = 4\sin^2\left\{\tfrac{1}{2}\pi\gamma\right\}\exp\left\{-2\frac{|\mathrm{Im}\,w_c|}{\varepsilon}\right\} \tag{25}$$

This result could also have been obtained by resumming the divergent tail of the infinite series corresponding to (16) (Berry 1990b).

4. Complex singularities

The universal form (12) occurs when $X(t)$ and $Z(t)$ have identical leading singularities at $t=t_c$ which give cancelling contributions to the energy $E(t)$. A sufficiently general form is

$$X(t) = f(t-t_c)\left(1+A(t-t_c)^s+\ldots\right)$$
$$Z(t) = \pm\mathrm{i}f(t-t_c)\left(1+B(t-t_c)^s+\ldots\right), \quad s>0 \tag{26}$$

Thus

$$E^2(t) = 2\left[f(t-t_c)\right]^2(t-t_c)^s(A-B+\ldots) \tag{27}$$

and, from (8),

$$w-w_c = 2\sqrt{2(A-B)}\int_0^{t-t_c}\mathrm{d}\tau f(\tau)\tau^{s/2}+\ldots \tag{28}$$

We assume that the integral converges.

Before proceeding, we make several remarks about this formulation. First, a simple zero in E^2 arises with f constant (i.e. not singular at all) and $s=1$; this case is generic in not requiring any conspiracy of singularities in X and Z. Second, some of the singular cases can be made generic by considering a family of Hamiltonians depending on several parameters, and varying these. Third, the cancellation of

leading singularities need not imply that E has a zero; from (27), E can be constant at t_c, or diverge there.

Now we must calculate the central quantity $\theta'(w)$ near the singularity w_c. From (3) and (8),

$$\theta' = \frac{\cos^2\theta}{2E}\frac{\mathrm{d}}{\mathrm{d}t}\tan\theta = -\frac{Z^2}{2E^{3/2}}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{X}{Z}\right) \tag{30}$$

Subsituting (26), we find

$$\begin{aligned}\theta' &= \frac{\pm i s}{4\sqrt{2(A-B)}\,(t-t_c)^{s/2+1} f(t-t_c)}(1+\ldots) \\ &= \frac{\pm i s}{2(w-w_c)}\frac{\int_0^{t-t_c} \mathrm{d}\tau\, f(\tau)\tau^{s/2}}{(t-t_c)^{s/2+1} f(t-t_c)}(1+\ldots)\end{aligned} \tag{31}$$

This indeed has the form (12), with the constant given by

$$\gamma = \mp\tfrac{1}{2} s \lim_{\tau\to 0}\frac{\int_0^{\tau}\mathrm{d}\tau f(\tau)\tau^{s/2}}{f(\tau)\tau^{s/2+1}} \tag{32}$$

whenever the limit exists.

A class that includes all interesting Hamiltonians we know is

$$f(\tau) = C\tau^r(1+\ldots) \tag{33}$$

for which

$$\gamma = \mp\frac{s}{2r+s+2} \tag{34}$$

Thus the adiabatic transition probability (25) is

$$P_\infty = 4\sin^2\left\{\frac{\pi s}{2(2r+s+2)}\right\}\exp\left\{-2\frac{|\mathrm{Im}\, w_c|}{\varepsilon}\right\} \tag{35}$$

Transition probabilities must be invariant under arbitrary transformations to new time variables and orthogonal transformations to new quantum reference frames, that is under

$$\left.\begin{aligned} t &\to t'(t) && \text{(a)} \\ \begin{Bmatrix} X(t) \\ Z(t) \end{Bmatrix} &\to \begin{Bmatrix} X'(t) \\ Z'(t) \end{Bmatrix} \equiv \begin{Bmatrix} X(t)\cos\phi + Z(t)\sin\phi \\ -X(t)\sin\phi + Z(t)\cos\phi \end{Bmatrix} && \text{(b)} \\ &\text{i.e. } \theta(t) \to \theta'(t) \equiv \theta(t) + \phi \end{aligned}\right\} \tag{36}$$

In appendix A we show that (34) indeed enjoys these invariance properties.

It should be emphasised that although (26), in which $X(t)$ and $Z(t)$ have the same leading singularities at $t=t_c$, generates the universal simple-pole formula (12), it represents a class of special situations. It is not difficult to find cases where the singularity of θ' is not a simple pole. For example, if $X(t)=B(t-t_c)^\mu$, $Z(t)=1+A(t-t_c)$, then $\theta'\sim(w-w_c)^{\mu-1}$ if $\mu>0$, and $(w-w_c)^{(-1-2\mu)/(\mu+1)}$ if $\mu<0$, with coefficients depending on B. In such nonuniversal cases, the prefactor in P depends on ε , and there is probably no simple general theory.

5. Examples

In the familiar Landau-Zener case,

$$X_{LZ} = 1, \quad Z_{LZ} = t, \quad \text{i.e.} \quad E^2 = 1 + t^2 \tag{37}$$

The form (26) is obtained in the lower half-plane by expanding about the simple zero of E^2 at t_c = -i, and it is easy to see that f= constant and (because Z deviates linearly from its value at t_c) s=1, so r=0 in (33) and (34) gives γ=1/3. Thus the prefactor is unity and the adiabatic transition probability is

$$P_{\infty LZ} = \exp\left\{-\frac{\pi}{\varepsilon}\right\} \tag{38}$$

Next, consider the Hamiltonian H_A, defined in (10). Again t_c = -i, and (8) gives w_c=-2i. Now f=1/$\sqrt{}$(1+t^2), so the singularity in X and Z is a square root branch point, with r= -1/2 in (33), and (because Z deviates

linearly from its value at t_c) s=1. Thus γ=1/2 and the adiabatic transition probability is

$$P_{\infty A} = 2\exp\left\{-\frac{4}{\varepsilon}\right\} \tag{39}$$

The invariance under time reparameterisation is nicely illustrated by the transformation

$$t \to \sinh t \tag{40}$$

(suggested by Dr Alain Joye). This preserves the structure of (2) and gives

$$X_A = 1, \quad Z_A = \sinh t, \quad \text{i.e. } E^2 = \cosh^2 t \tag{41}$$

Now the form (26) is obtained by expanding about the double zero of E^2 at t_c = -iπ/2, giving r=0 in (33) and s=2. These are different values from those generated by the formulae in (10), but of course refer to the same Hamiltonian H_A and so generate the same $\gamma$and the same transition probability.

Our last example is the Hamiltonian H_B, defined in (10). Again t_c = -i, and (8) gives w_c=-2i. Now f=sech$t$, so the singularity in X and Z is a simple pole, with r= -1 in (33), and (because Z deviates quadratically from its value at t_c) s=2. Thus γ=1 and the adiabatic transition probability is

$$P_{\infty B} = 4\exp\left\{-\frac{4}{\varepsilon}\right\} \tag{42}$$

Each of these three Hamiltonians has a different status in adiabatic theory. H_{LZ} is exactly solvable, and the solution (Zener 1932) shows that the adiabatic formula (38) is in fact exact. H_A seems to have no exact solution, and we suppose that the formula (39) is the first term in an adiabatic expansion in powers of ε. This is supported by numerical exploration. (Computational solution of (2) for small ε is not trivial; the method we employed is outlined in the appendix B.) H_B does have an exact solution (described by Suominen and Garraway 1992), namely

$$P_B = \frac{1}{\sinh^2\left(\frac{2}{\varepsilon}\right)} = \frac{4\exp\left\{-\frac{4}{\varepsilon}\right\}}{\left(1-\exp\left\{-\frac{4}{\varepsilon}\right\}\right)^2} = P_{\infty B}\left(1+2\exp\left\{-\frac{4}{\varepsilon}\right\}+\ldots\right) \tag{43}$$

Therefore the error is exponentially small, and results not from an adiabatic expansion associated with the singularity at w_c=-2i but from contributions associated with other singularities (Suominen 1992).

Figure 1 shows the exact prefactors for the three Hamiltonians as functions of ε, indicating clearly the considerably lower accuracy of the adiabatic approximation for H_A.

Acknowledgment

We are grateful for Dr Alain Joye for sending us his different and independent derivation of the prefactors.

Appendix A

First we show that the formula (34) for γ, and hence the adiabatic transition probability (35), is invariant under (36a), i.e. time reparameterisation. From (2) and (3), the new Hamiltonian involves

$$\begin{Bmatrix} X'(t') \\ Z'(t') \end{Bmatrix} = \frac{dt}{dt'}\begin{Bmatrix} X(t(t')) \\ Z(t(t')) \end{Bmatrix} \tag{A1}$$

Now suppose

$$t - t_c \propto (t' - t_c')^{\mu} \tag{A2}$$

The quantities in (26) and (33) rescale as

$$s \to s' = \mu s, \quad r \to r' = r\mu + \mu - 1 \tag{A3}$$

and this leaves (34) invariant.

Now we demonstrate the invariance under (36b), i.e. quantum orthogonal transformation. This changes (26) to

$$\begin{aligned} X'(t) &= \exp\{i\phi\} f(t-t_c) \\ &\qquad \left[1+(t-t_c)\exp\{-i\phi\}(A\cos\phi + iB\sin\phi)\right] \\ Z'(t) &= i\exp\{i\phi\} f(t-t_c) \\ &\qquad \left[1+(t-t_c)\exp\{-i\phi\}(B\cos\phi + iA\sin\phi)\right] \end{aligned} \tag{A4}$$

Clearly the form of (26) is preserved, with

$$\begin{aligned} f(t-t_c) \rightarrow f'(t-t_c) &= \exp\{i\phi\} f(t-t_c) \\ A \rightarrow A' &= A\cos\phi + iB\sin\phi \\ B \rightarrow B' &= B\cos\phi + iA\sin\phi \end{aligned} \tag{A5}$$

Appendix B.

The numerical solution of (2) for small ε, given a Hamiltonian specified by functions $X(t)$, $Z(t)$ is complicated by the fact that the desired final transition amplitude is an exponentially small quantity emerging from rapid oscillations that are much larger. This numerical instability can be reduced by solving the Schrödinger equation not in the original basis, or the adiabatic basis defined by (4) and (5), but in one of the superadiabatic bases. As explained by Berry (1990a) and Lim and Berry (1991), the optimal basis would be the one whose order is the nearest integer to $|w_0|/\varepsilon$ (because in the nth basis the oscillations of the transition amplitude $c_n(t)$ are of order $\varepsilon^{n+1}/n!$)

However, it is not necessary to use the optimal basis; in the computations with H_A illustrated in figure 1, only the first-order (n=1) superadiabatic basis was employed, and this was dramatically superior to the ordinary adiabatic basis (n=0).

For numerical purposes the most convenient sequence of superadiabatic bases is not that defined by the perturbation expansion (16) (although this is useful for theoretical purposes), but that generated by adiabatic iteration Berry (1987). In this procedure, the system is

transformed to a basis specified by the eigenstates of the instantaneous Hamiltonian, and the process is repeated. If we define

$$x_0(t) \equiv \frac{2X(t)}{\varepsilon}, \quad z_0(t) \equiv \frac{2Z(t)}{\varepsilon} \tag{B1}$$

then with a particular choice of phases the successive Hamiltonians are determined by

$$\begin{aligned} x_{n+1}(t) &\equiv \frac{x_n(t)\dot{z}_n(t) - z_n(t)\dot{x}_n(t)}{x_n^2(t) + z_n^2(t)} \\ z_{n+1}(t) &\equiv \sqrt{x_n^2(t) + z_n^2(t)} \end{aligned} \tag{B2}$$

These functions can easily be found by repeated differentiation.

It is also convenient (see also Suominen 1992) to solve not the Schrödinger equation but the equivalent real equation for the spin vector (expectation value of the vector of Pauli spin matrices) on the Bloch sphere. For the nth superadiabatic basis this is

$$\begin{aligned} \dot{S}_n(t) &= r_n(t) \wedge S_n(t), \text{ where} \\ S_n(t) &\equiv \{S_{n1}(t), S_{n2}(t), S_{n3}(t)\}, \\ r_n(t) &\equiv \{x_n(t), 0, z_n(t)\} \text{ and } S_n(0) = \{0,0,1\} \end{aligned} \tag{B3}$$

and we used the Runge-Kutta method for its numerical solution. In this formulation, the transition probability is

$$P = \tfrac{1}{2}\left[1 - S_{n3}(+\infty)\right] \tag{B4}$$

References

Berry M V 1982 *J.Phys.A.* **15** 3693-3704

Berry M V 1987 *Proc.Roy.Soc.* **A414** 31-46

Berry M V 1990a *Proc.Roy.Soc.Lond* **A429** 61-72

Berry M V 1990b *Proc.Roy.Soc.Lond* **A427** 265-280

Berry M V 1990c *Proc.Roy.Soc.Lond.* **A430** 405-411

Davis J P and Pechukas P 1976 *J. Chem. Phys.* **64** 3129-3137

Demkov Yu N and Kunike M 1969 *Vestn Leningr. Univ. Ser.Fiz.Khim.* **16** 39

Demkov Yu N, Ostrovskii V N and Solov'ev E A 1978 *Phys. Rev.* **A18** 2089-2096

Garraway B M, Stenholm S and Suominen K-A 1993 *Physics World* **6** (no.4) 46-51

Joye A 1993 Preprint CPT-93/P.2891 (CNRS Marseille)

Joye A, Kunz H and Pfister C-E 1991 *Ann. Phys.* **208** 299-332

Lim R and Berry M V 1991 *J. Phys. A.* **24** 3255-3264

Pokrovskii V L and Khalatnikov I M 1961 *Sov. Phys. JETP* **13** 1207-1210

Suominen K-A 1992 Thesis (University of Helsinki Theoretical Physics Report HU-TFT-IR-92-1)

Suominen K-A, Garraway B M and Stenholm S 1991 *Optics Commns* **82** 260-266

Suominen K-A and Garraway B M 1992 *Phys. Rev.* **A45** 374-386

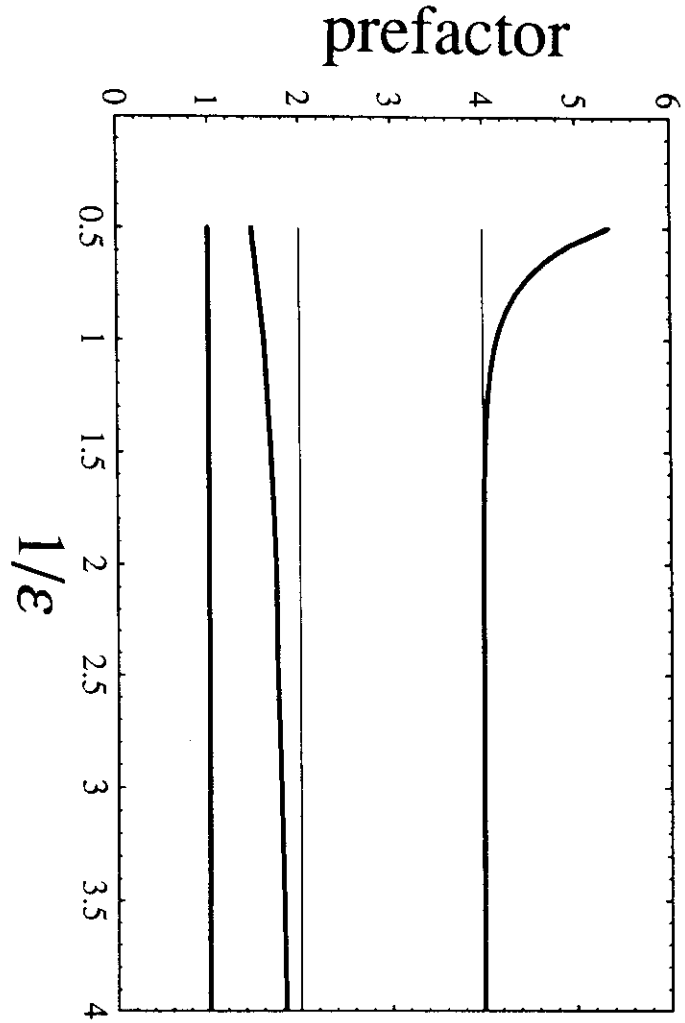
Zener C 1932, *Proc. Roy. Soc. Lond.* **A317**, 696-702

Zwanziger J W, Rucker S P and Chingas G C 1991 *Phys. Rev.* **A43** 3233-3240

Figure Caption

Figure 1. Prefactors (thick lines) multiplying the leading exponential in the transition probability, as functions of the adiabatic parameter ε, for the Hamiltonians H_{LZ} (equation 36), H_A and H_B (equation 10). The LZ prefactor is exactly unity, and the A and B prefactors are asymptotic to 2 and 4 respectively.

figure 1



QUANTUM ADIABATIC ANHOLONOMY

Michael Berry

H.H.Wills Physics Laboratory, Tyndall Avenue,
Bristol BS8 1TL, UK

Lectures given at the Ferrara School of Theoretical Physics on 'Anomalies, defects, phases.....', June 1989

~~To be~~ published by Bibliopolis (Naples)

published 1990

Contents

(Illustrations can be found at the end of lectures 1 and 4)

Lecture 1

My aim is to give a self-contained account of the geometric phase, which is helping to sweep clean a corner of quantum mechanics that was for a long time dusty and obscure. The treatment will be thoroughly elementary - there will be no fibre bundles and Chern numbers - but nevertheless in the spirit of Einstein's injunction: physics should be made as simple as possible, but not simpler.

I do not intend to give comprehensive coverage of what has become a large subject, and so will omit entirely several major topics that have been treated elsewhere, such as nonadiabatic corrections (ref [1]) and the reaction of quantum systems on their environment [2]. Nor will I give a complete list of references, because with the assistance of Richard Lim I am compiling a comprehensive phase bibliography that will be freely available, and because a reprint collection [3] has just been published.

Details of many of the arguments will be left as exercises for the student. In these notes such exercises will be denoted by [E].

The geometric phase is based on two ideas: *adiabaticity* and *anholonomy*. Adiabaticity here refers to quantum physics on the border between statics and dynamics. Statics is concerned with *things*, that is persisting entities. In quantumspeak these are eigenstates of the Hamiltonian describing the system's environment. Dynamics is concerned with *happenings*, in this case those induced by changes in the environment. On the border are things in environments that change slowly; such changes are the province of adiabatic theory.

Anholonomy is a geometric concept: the failure of some quantities to come back to their original values when others, which drive them, are forced to return. The failure derives from nonintegrability of the driving law. In this lecture I will concentrate on an example of

anholonomy, introducing it as pure geometry in a way that generalizes easily to quantum mechanics.

Let a vector e, lying in the surface of a sphere, be transported round a circuit C (fig.1.1). "Transported" means that the unit radius vector r, to which e is attached, is forced round a loop $r(T)$ with $r(T)=r(0)$) and "in the surface" means $e\cdot r=0$. The law of transport is *parallel transport*: e never twists about r. After the circuit, it is found that e has rotated, by an angle θ(C) that we wish to calculate. θ(C) embodies anholonomy: e has not returned, even though r, which drives it, has.

This parallel transport anholonomy is easy to demonstrate. Hold a pointer at arm's length above your head and pointing forwards. Move your arm down till it is horizontal, then rotate it sideways through a right angle, and finally bring it back up again, taking care never to twist the pointer. You will find that the pointer now points sideways, that is, it has rotated, in spite of never having been turned! I have done this little trick many times, and yet it still seems magical.

To calculate θ(C) we must give mathematical expression to the law of parallel transport. Let the orthonormal frame r,e_1,e_2 (with e any fixed combination of e_1 and e_2) rotate with angular velocity Ω, i.e.

$$\dot{e} = \Omega \wedge e \tag{1.1}$$

(the overdot denotes differentiation). Ω has the general form

$$\Omega = ar + b\dot{r} + c\, r \wedge \dot{r} \tag{1.2}$$

Parallel transport means that Ω has no component along r, so $a=0$. To determine b and c we impose the requirement that e remain perpendicular to r, i.e. $(e\cdot r)^{\cdot}=0$. This gives [E] $b=0$, $c=1$, so

$$\Omega = r \wedge \dot{r} \tag{1.3}$$

The law of parallel transport is therefore

$$\dot{e} = (r \wedge \dot{r}) \wedge e = - e \cdot \dot{r}\, r \tag{1.4}$$

We express the law in a form suitable for later generalization to quantum mechanics. Define a complex unit vector on the sphere by

$$\phi \equiv (e_1 + i e_2)/\sqrt{2} \tag{1.5}$$

Thus $\phi^* \cdot \phi = 1$. From the fact that e_1, e_2 separately transport according to (1.4), it follows easily {E} that

$$\phi^* \cdot \dot{\phi} = 0 \tag{1.6}$$

Now we can calculate the anholonomy $\theta(C)$. Chart the passage of e_1, e_2 relative to a *local basis* of unit vectors $u(r)$, $v(r)$ defined at each point r; (fig.1.2), and so singlevalued round C by construction. For example u and v could point along lines of latitude and longitude. Let the corresponding complex unit vector be

$$n(r) = (u + i v)/\sqrt{2} \tag{1.7}$$

The relation between the transported basis $\phi(t)$ and the local basis $n(r(t))$ is {E} that they differ by a phase factor:

$$\phi(t) = \exp\{- i \theta(t)\}\ n(r(t)) \tag{1.8}$$

where θ is the angle by which u, v must be rotated to coincide with e_1, e_2.

The desired $\theta(C)$ is the increment of θ round C, that is the total rotation of e_1, e_2 relative to u, v.

We find θ using (1.6):

$$0 = \phi^* \cdot \dot{\phi} = \exp\{i \theta\}(- i \dot{\theta}\, n^* \cdot n + n^* \cdot \dot{n}) \exp\{- i \theta\} \tag{1.9}$$

Now $n^* \cdot n = 1$ and so $n^* \cdot \dot{n}$ is imaginary, so that

$$\dot{\theta} = \mathrm{Im}\ n^* \cdot \dot{n} \tag{1.10}$$

Thus

$$\begin{aligned} \theta(C) &= \mathrm{Im} \int_0^T n^* \cdot \dot{n}\, \mathrm{d}\, t = \mathrm{Im} \oint_C n^* \cdot \mathrm{d}\, n \\ &= - \oint_C v \cdot \mathrm{d}\, u \end{aligned} \tag{1.11}$$

where the last equality is a simple {E}.

To obtain the explicit form of $\theta(C)$ for any circuit $r(t)$ we use Stokes' theorem:

$$\theta(C) = \iint_{\partial S = C} V \tag{1.12}$$

where V is the *2-form*

$$\begin{aligned} V &= \mathrm{Im}\, \mathrm{d}\, n^* \wedge \cdot\, \mathrm{d}\, n = \mathrm{Im}\, \nabla n^* \wedge \cdot\, \nabla n \cdot \mathrm{d}\, S \\ &= - \nabla v \cdot \wedge \nabla u \cdot \mathrm{d}\, S \end{aligned} \tag{1.13}$$

Here dS is an element of area on the sphere, and the integration is over any surface whose boundary is C. The scalar products $\cdot$ act between n^* and n, and u and v, and the vector products $\wedge$ act between the ∇s. Of course the anholonomy $\theta(C)$ is independent of the basis $u(r)$, $v(r)$. A

convenient choice is the unit vectors corresponding to spherical polar angles, that is

$$u \equiv \frac{r \wedge e_z}{|r \wedge e_z|} \quad , \quad v \equiv \frac{r \wedge u}{r} \tag{1.14}$$

where e_z is a fixed vector and we include the length r of r because it is convenient to have formulae valid in the whole three-dimensional r space rather than just on the unit sphere. From (1.12), a straightforward calculation [E] gives

$$V = r \cdot dS / r^3 \tag{1.15}$$

We thus obtain the final result, from (1.12):

$$\begin{aligned} \theta(C) &= \iint dS \cdot r / r^3 \\ &= \text{flux through C of unit monopole at } r = 0 \\ &= \text{solid angle } \Omega(C) \text{ subtended by C at } r = 0 \end{aligned} \tag{1.16}$$

Anholonomy is the fact that θ does not vanish. Its origin is the nonintegrability of the parallel transport law (1.4).

An amusing observation (by V J Smith) is that (1.16) equates a plane angle, measured in radians, to a solid angle, measured in steradians=(radians)^2. This shows how unnatural is the decision of the International Committee of Weights and Measures [4] that radians and steradians shall be supplementary units in the SI system, with different dimensions, rather than being dimensionless as any physicist would think.

With these preparations we can now consider anholonomy in the transport of quantum states. Let $X \equiv \{X_1, X_2, \ldots\}$ be parameters influencing the quantum state $|\phi\rangle$ of a system. For given X, $|\phi\rangle$ is determined up to a phase. Now let X be varied round a cycle C:

$(X=X(t):X(T)=X(0))$ (fig.1.3). We seek the phase $\gamma(C)$ accumulated when $|\phi\rangle$ is parallel-transported round C. The definition of transport is obvious from the analogy with vectors: X is analogous to position r, and $|\phi\rangle$, a complex unit vector in many-dimensional Hilbert space, is analogous to the two-dimensional complex unit vector ϕ. Parallel transport of $|\phi\rangle$ is now defined by generalizing (1.6) to

$$\langle \phi | \dot{\phi} \rangle = 0 \tag{1.17}$$

To find $\gamma(C)$we develop the analogy further. Let $|n(X)\rangle$ be a local basis state, defined to be single-valued on and inside C, coinciding with $|\phi\rangle$ up to a phase.Thus $|n\rangle$ generalizes n, and instead of (1.8) we now have

$$|\phi(t)\rangle = \exp\{i\,\gamma(t)\}\,|n(X(t))\rangle \tag{1.18}$$

(1.17) gives

$$\dot{\gamma} = i\langle n|\dot{n}\rangle = -\operatorname{Im}\langle n|d\,n\rangle/dt \tag{1.19}$$

where we have used the fact that $\langle n|\dot{n}\rangle$is imaginary (which follows from $\langle n|n\rangle=1$. Thus

$$\gamma(C) = -\oint \operatorname{Im}\langle n|d\,n\rangle \tag{1.20}$$

(cf. 1.11), and application of Stokes' theorem (cf.1.12) now gives

$$\gamma(C) = -\iint_{\partial S = C} V \tag{1.21}$$

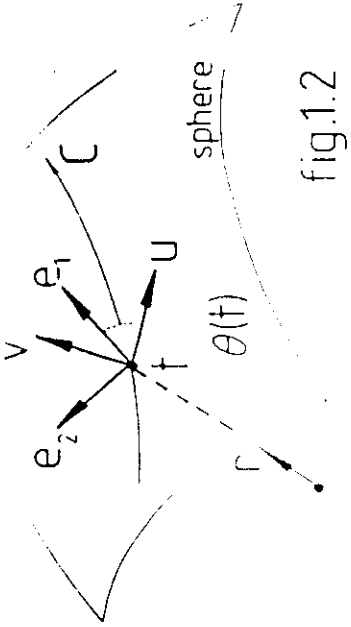


fig.1.2

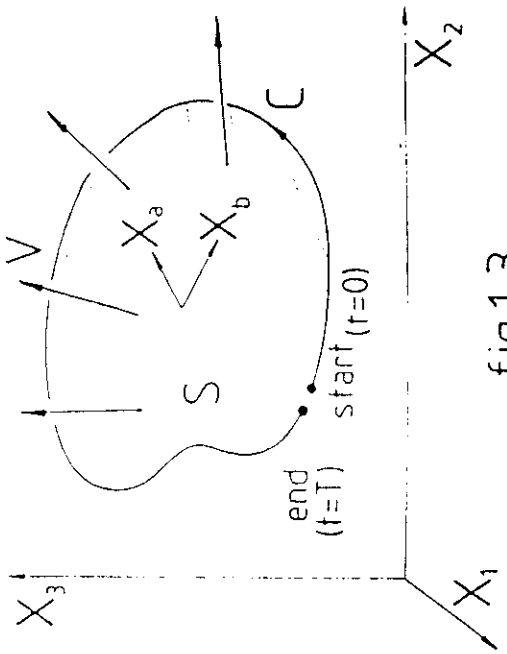


fig.1.3

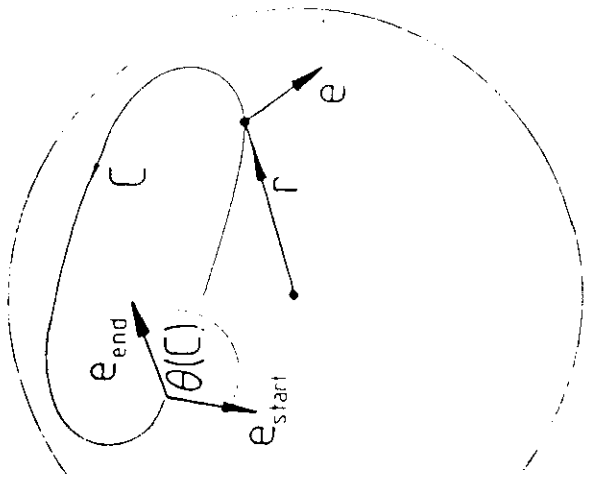


fig.1.1

Lecture 2

In 'reality', quantum states are transported not by any convenient mathematical rule but by a Hamiltonian operator H acting via the time-dependent Schrödinger equation. One way [6] to implement the parallel-transport rule (1.17) is to incorporate the changing parameters X into H and make the changes occur *slowly*. Thus we have the slowly-cycled Hamiltonian

$$H = H(X(t)) \qquad (0 \le t \le T, T \to \infty, \quad X(0) = X(T)) \tag{2.1}$$

and we can invoke the quantum *adiabatic theorem*.

This states that the time-dependent Schrödinger equation is satisfied by eigenstates of the frozen Hamiltonian H at each instant, multiplied by the usual oscillatory time factor. Thus the adiabatic states (labelled n) are

$$|\Psi_n(t)\rangle = \exp\{-i\int_0^t dt\, E_n(X(t))/\hbar\}|\phi_n(t)\rangle \tag{2.2}$$

where $|\phi_n\rangle$ (assumed nondegenerate) satisfies the eigenequation at t:

$$H(X)|\phi_n\rangle = E_n(X)|\phi_n\rangle \tag{2.3}$$

The phase $-\int dt E_n(t)/\hbar$ of the oscillatory factor is the *dynamical phase*; it generalizes the familiar "$-\omega t$" of any oscillatory process, and is present even if X is held fixed. (Of course the simple form (2.2), in which $|\Psi(t)\rangle$ clings to individual eigenstates $|\phi_n\rangle$, is an exceptional case, valid only in the adiabatic limit $T\to\infty$. Otherwise the changing H couples different $|\phi_n\rangle$ via transitions, whose study is the usual business of time-dependent quantum mechanics.)

where (cf.1.13)

$$V \equiv \mathrm{Im}\langle \mathrm{d}\, n| \wedge |\mathrm{d}\, n\rangle \tag{1.22}$$

and $\wedge$ now denotes the wedge product.

This is the main result. It gives the phase anholonomy $\gamma(C)$ as the flux through C of the *2-form* V. Do not be afraid of 2-forms; if you are mathematically innocent (like me), think of them as objects which deliver a number when integrated over an area. A more explicit representation of the 2-form can be written by defining coordinates X_a, X_b on the spanning surface S and expanding $|n\rangle$ in an X-dependent superposition of *fixed* orthonormal basis states $|\chi_m\rangle$, i.e.

$$|n(X)\rangle = \sum_m a_m(X)\,|\chi_m\rangle \tag{1.23}$$

(in position representation, for example, m would label position and $a_m(X)$ would be the X-dependent (complex) wavefunction). Then

$$\begin{aligned} V &= \mathrm{Im} \sum_m \mathrm{d}\, a_m^* \wedge \mathrm{d}\, a_m \\ &= \mathrm{d}X_a \mathrm{d}X_b\, \mathrm{Im} \sum_m \left(\frac{\partial a_m^*}{\partial X_a} \frac{\partial a_m}{\partial X_b} - \frac{\partial a_m^*}{\partial X_b} \frac{\partial a_m}{\partial X_a} \right) \end{aligned} \tag{1.24}$$

So far our considerations have been rather abstract and mathematical. To turn them into physics we have to specify a realistic mechanism for parallel-transporting the state $|\phi\rangle$. This will be done in the next lecture. For now I simply mention a case where the abstractions can be interpreted literally, namely the Aharonov-Bohm effect [5.6]. Here $|\phi\rangle$ is the state of an electron, the parameters X are its position in ordinary 3-space, the 2-form V is proportional to the field of a

single line of ordinary magnetic flux, and the anholonomy γ is the phase change of the electron round a circuit C of the flux line.

Now comes the central point. Equation (2.2) does not constitute a complete specification of the adiabatic state $|\Psi(t)\rangle$, because the eigenequation (2.3) provides no means to connect its solutions $|\phi\rangle$ at different parameters X: we need a transport law. This is provided by the time-dependent Schrodinger equation, projected on the subspace $|\phi_n\rangle$:

$$\begin{aligned} 0 &= \langle \Psi_n(t)|(H - i\hbar\partial_t)|\Psi_n(t)\rangle \\ &= \langle \Psi_n(t)|(H_n - i\hbar\partial_t)\exp\{-i\int_0^t dt' E_n/\hbar\}|\phi_n(t)\rangle \\ &= \langle \phi_n(t)|\dot{\phi}_n(t)\rangle \end{aligned} \tag{2.4}$$

This is exactly the parallel-transport law (1.17), which we found to be nonintegrable. At the end of the cycle, $|\Psi\rangle$ as given by (2.2) therefore acquires a phase from the non-return of $|\phi_n(t)\rangle$ as well as the dynamical phase: thus

$$|\Psi_n(T)\rangle = \exp\{-i\int_0^T dt\, E_n(X(t/T))/\hbar\}\exp\{i\gamma_n(C)\}|\Psi_n(0)\rangle \tag{2.5}$$

Here $\gamma_n(C)$ is the *geometric phase*, given by (1.21) and (1.22) with $|n(X)\rangle$ now being any solution of (2.2) that is single-valued in X space on C and on the chosen spanning surface S.

The reason for calling $\gamma_n(C)$ geometric is that it depends only on the geometry of C in X space (and on which state n is being transported), and not on the rates with which different parts of C are traversed (assuming of course that the transport is slow). In particular, γ is independent of the adiabatic parameter T, unlike the dynamical phase which $\{E\}$ increases linearly with T.

It is amusing to see how the geometric phase is contained in the path-integral representation of the evolving state. This was done with J.Hannay and M.Wilkinson in 1983 but never published (but see [45]). Introducing the time-ordered product for the evolution over the cycle, and dividing the time interval into many small steps of duration $\delta \equiv T/N$ where $N \to \infty$, we have

$$\begin{aligned} |\Psi(T)\rangle &= T\exp\{-i\int_0^T H(t)/\hbar\}|\Psi(0)\rangle \\ &= \prod_{k=1}^{N} \exp\{-i\,\delta\, H(t_k)/\hbar\}\,|\Psi(0)\rangle \end{aligned} \tag{2.6}$$

At each t_k we introduce the complete set of instantaneous eigenstates $|m(X_k)\rangle$ of $H(t_k)$ (single-valued solutions of (2.3)). Thus

$$|\Psi(T)\rangle = \prod_{k=1}^{N}\sum_{m}\exp\{-i\,\delta\, E_m(t_k)/\hbar\}|m(X_k)\rangle\langle m(X_k)|\times \times|\Psi(0)\rangle \tag{2.7}$$

This is exact, but now we invoke the adiabatic approximation to eliminate terms $m \neq n$ in all the intermediate sums, where $|\Psi(0)\rangle = |n(X(0))\rangle$ is the initial state. Then all the factors involving the E_m combine to give the dynamical phase, leaving

$$\begin{aligned} |\Psi(T)\rangle &= |\Psi(0)\rangle\exp\{-i\int_0^T dt\, E_n(t)/\hbar\}\times \\ &\quad \times\langle n(X_N)|n(X_{N-1})\rangle\ldots\langle n(X_1)|n(X_0)\rangle \end{aligned} \tag{2.8}$$

Terms in the product have the form

$$\begin{aligned} \langle n(X_{t+1})|n(X_t)\rangle &= \langle n(X(t+\delta))|n(X(t))\rangle \\ &= (\langle n| + \delta\langle\dot{n}|)|n\rangle \\ &= 1 + \delta\langle\dot{n}|n\rangle \\ &= \exp\{\delta\langle n|\dot{n}\rangle\} = \exp\{-\delta\langle\dot{n}|n\rangle\} \end{aligned} \tag{2.9}$$

By (1.20), the accumulation of these factors gives precisely $\gamma_n(C)$.

By (1.22), the geometric phase for the n'th state is the flux through C of a 2-form that we now call $V_n(X)$:

$$V_n(X) = \mathrm{Im} < \mathrm{d}n| \wedge |\mathrm{d}n > \tag{2.10}$$

This is the mathematical object at the heart of the whole subject. It sits in parameter space, waiting to be brought to life as a phase when H is cycled.

Now we describe some interesting properties of V_n. The first concerns its singularities. To see where these are (in X space), we introduce the complete set of eigenstates $|m(X)>$ of $H(X)$.

$$V_n = \mathrm{Im} \sum_{m \neq n} < \mathrm{d}n|m > \wedge < m|\mathrm{d}n > \tag{2.11}$$

Note the exclusion of the state $m=n$ [E]. From the eigenequation for H it is possible [E] to derive

$$< m|\mathrm{d}n > = - \frac{< m|\mathrm{d}H|n >}{E_n - E_m} \qquad (m \neq n) \tag{2.12}$$

so that

$$V_n = \mathrm{Im} \sum_{m \neq n} \frac{< n|\mathrm{d}H|m > \wedge < m|\mathrm{d}H|n >}{(E_n - E_m)^2} \tag{2.13}$$

This shows that the singularities of V_n occur where the spectrum of $H(X)$ has *degeneracies* involving the transported state $|n>$. Later we shall determine the precise nature of the singularity.

The other properties of V_n concern *gauge invariance*. In its simplest form, this is the fact that V_n is independent of the choice of single-

valued eigenstates $|n(X)>$. Different choices are related by a single-valued X-dependent phase factor, and we have [E]

$$\begin{aligned} &\mathrm{Im} < \mathrm{d}n' \wedge \mathrm{d}n' > = \mathrm{Im} < \mathrm{d}n \wedge \mathrm{d}n > \\ &\text{if } |n'(X)> \equiv \exp\{\mathrm{i}\chi(X)\}|n(X)> \end{aligned} \tag{2.14}$$

By contrast, the 1-form $<n|\mathrm{d}n>$ (cf 1.20) does not possess this gauge invariance.

Another quantity is invariant under the gauge transformation $|n> \rightarrow |n'>$ To find out what it is, introduce coordinates X_i in parameter space and write

$$V = V_{ij}\, \mathrm{d}X_i \wedge \mathrm{d}X_j \tag{2.15}$$

where V_{ij} is the antisymmetric second-rank tensor [E]

$$V_{ij} = 2\mathrm{Im} < \partial_i n | \partial_j n > \tag{2.16}$$

What about the real (symmetric) part of the tensor? It is easy to show [E] that this is not gauge-invariant, but that the following quantity is:

$$g_{ij} = \mathrm{Re} < \partial_i n | (1 - |n> < n|) | \partial_j n > \tag{2.17}$$

The interpretation [2] of g_{ij} is as a *metric tensor* in X space, measuring distance ds between states $|n>$ at neighbouring points X and $X+\mathrm{d}X$ in the most natural gauge-invariant way:

$$\mathrm{d}s^2 = g_{ij}\, \mathrm{d}X_i \mathrm{d}X_j = 1 - |< n(X)|n(X + \mathrm{d}X)>|^2 \tag{2.18}$$

(Both tensors, V_{ij} and g_{ij} also contribute in important ways [2] to the reaction of the quantum system on its environment, in the form of the dy-

namics of the X_j when regarded self-consistently not as parameters but as quantum variables.)

In elementary physics the most familiar gauge invariance is that of *magnetic field*

$$B(r) = \nabla \wedge A(r) \tag{2.19}$$

under the transformation of the vector potential A to A' where

$$A' = A + \nabla_r \Lambda(r) \tag{2.20}$$

and Λ is a single-valued scalar function of r. This type of gauge transformation has to be considered when studying the geometric phase for a charged particle whose slowly-cycled environment (parameterized by X) includes a magnetic field.

We expect $\gamma_n(C)$ to depend on A because A occurs in the Hamiltonian. But all physics must be invariant under (2.20), even when this transformation is parameter-dependent, i.e. $\Lambda=\Lambda(r,X)$. The effect of Λ is to contribute a phase factor to the wavefunctions $\langle r|n(X)\rangle$ in position representation. This is so similar to the type of transformation in (2.14) - a phase factor multiplying the Hilbert-space vector - that it came as a surprise to find that not only the 2-form V_n but also $\gamma_n(C)$ itself *change* under (2.20), the transformation law for V_n being [E]

$$V'_n(X) = V_n + d_X \langle n|d\Lambda|n\rangle \tag{2.21}$$

Physics is saved from inconsistency, however, by the fact that a parameter-dependent A also generates an *electric* field (through the slow change in X) unless it is allowed to transform the scalar potential as well. In the resulting complete gauge transformation, the *dynamical* phase is changed in a way that exactly compensates the effect of (2.21). Details are given in [7].

Lecture 3

At first the geometric phase appears unobservable because its detection based on (2.5) seems to require superposing the system at $t=T$ on its former self (at $t=0$), which is impossible. There are however at least two ways in which $\gamma_n(C)$ can be - and has been - measured.

The way that was originally suggested [6] was by interferometry. A system in the state $|n\rangle$ (e.g. a coherent beam of particles) is split at $t=0$ into two subsystems. One is slowly cycled and the other not. Both subsystems will acquire dynamical phases, say $\gamma_{1\mathrm{dyn}}$ and $\gamma_{2\mathrm{dyn}}$, but the cycled one will, in addition, gain a geometric phase. If the subsystems are subsequently recombined, the intensity of their superposition is

$$I \propto \left|\exp\{i(\gamma_{1\,\mathrm{dyn}} + \gamma_n(C)\} + \exp\{i\,\gamma_{2\,\mathrm{dyn}}\}\right|^2$$
$$= 4\cos^2\{\tfrac{1}{2}|\gamma_{1\mathrm{dyn}} - \gamma_{2\,\mathrm{dyn}} + \gamma_n(C)|\} \qquad (3.1)$$

Therefore $\gamma_n(C)$ can be detected as a shift of interference fringes - as in the Aharonov-Bohm experiments [5].

One can say that the interferometric experiments involve the same state and two Hamiltonians (one for each subsystem). The second class of experiments, on the other hand, involves (at least) two states and the same Hamiltonian. Let the initial state be a superposition of two eigenstates $|m\rangle$ and $|n\rangle$ of $H(0)$:

$$|\Psi(0)\rangle = a_m|m\rangle + a_n|n\rangle \qquad (3.2)$$

This is a non-stationary state, which after the cycle has become, in an obvious notation,

$$|\Psi(T)\rangle = a_m|m\rangle\exp\{i|\gamma_{m\mathrm{dyn}} + \gamma_m(C)|\} +$$
$$+ a_n|n\rangle\exp\{i|\gamma_{n\mathrm{dyn}} + \gamma_n(C)|\} \qquad (3.3)$$

Now measure the expectation value of some operator A that does not commute with H; we find [E]

$$\langle\Psi(T)|A|\Psi(T)\rangle = |a_n|^2\langle n|A|n\rangle + |a_m|^2\langle m|A|m\rangle +$$
$$+ 2\,\mathrm{Re}\,a_n^*a_m\langle n|A|m\rangle\exp\{i|\gamma_{n\mathrm{dyn}} - \gamma_{m\mathrm{dyn}} + \gamma_n(C) - \gamma_m(C)|\} \qquad (3.4)$$

The interference term reveals the difference of the phase shifts experience by the two constituent states, and of course this includes the difference of their geometric phases.

Many of the experiments that have been carried out [3] involve the *turning of spinning particles*. I will now work out the underlying theory, which is useful in several other applications as well. Consider a particle with spin I (integer or half-integer), described by the vector of three $(2I+1)\times(2I+1)$dimensional angular-momentum matrices σ satisfying the familiar commutation rule

$$\sigma \wedge \sigma = i\,\sigma \qquad (3.5)$$

Let H at each instant be rotationally symmetric about some direction described by a vector in the parameter space $R=(X,Y,Z)$, i.e.

$$H(t) = F\{R(t)\cdot\sigma\} \qquad (3.6)$$

An example is a particle with magnetic moment μ in a magnetic field μR, for which F is linear and $R = \mu B$.

To find the phase 2-form $V_n(R)$ (here regarded as a vector in R space), we use the sum-over-states (2.13), replacing d by ∇. The states $|n>$ are $(2l+1)$-component spinors labelled by the component of angular momentum along R (n runs from $-l$ to $+l$) and the eigenvalues are $F(nR)$ where $R=|R|$. In (2.13) the dependence on F cancels and we obtain {E}.

$$V_n(R) = \frac{1}{R^2} \operatorname{Im} \sum_{m \neq n=-l}^{l} \frac{<nl\,\sigma\wedge|m><m|\,\sigma\wedge|n>}{(n-m)^2} \qquad (3.7)$$

Now, the matrix elements are zero unless $m=n$ or $m=n\pm1$ {E} and the contribution $m=n$ is excluded by the $\wedge$. Therefore we can take $(n-m)^2=1$ out of the sum, eliminate the sum-over states and use (3.5) to get

$$V_n(R) = \frac{1}{R^2} \operatorname{Im} <nl\,\sigma\wedge\,\sigma\wedge|n> = \frac{1}{R^2} <nl\,\sigma\wedge|n>$$
$$= n\,R/R^3 \qquad (3.8)$$

where the last equality follows from the fact that $|n>$ is an eigenstate of the component of σ along R, in which the perpendicular components have zero expectation value.

The 2-form is therefore the field of a *monopole* of strength n, situated at $R=0$. The geometric phase is minus the flux through C of this monopole, that is

$$\gamma_n(C) = -n\,\Omega(C) \qquad (3.9)$$

The simplest case is spin 1/2, for which σ are the three Pauli matrices. This describes (for example) *neutrons*, whose geometric phase was measured in an ingenious experiment by Bitter and Dubbers [8]. They sent a beam of neutrons along the x direction in a helical magnetic field:

$$B = B\{\cos\theta, \sin\theta\cos(2\pi x/L), \sin\theta\sin(2\pi x/L)\}$$
$$(0 \le x \le L) \qquad (3.10)$$

The moving neutrons see a B that changes with time, sweeping out a cone with solid angle {E}

$$\Omega(C) = 2\pi(1-\cos\theta) \qquad (3.11)$$

Their experiment was of the second type described above, with the initial state being polarized along z. Thus

$$|\Psi(0)> = \begin{pmatrix}1\\0\end{pmatrix} = \cos(\theta/2)|+> + \sin(\theta/2)|->$$
(3.12)

where

$$|+> = \begin{pmatrix}\cos(\theta/2)\\ \sin(\theta/2)\end{pmatrix} \quad \text{and} \quad |-> = \begin{pmatrix}\sin(\theta/2)\\ -\cos(\theta/2)\end{pmatrix} \qquad (3.13)$$

are the eigenstates of $H=\mu\sigma\cdot B$ along the initial direction of B {E}. At the end of the cycle,

$$|\Psi(T)> = \cos(\theta/2)\exp\{-i\alpha\}|+> + \sin(\theta/2)\exp\{i\alpha\}|->$$
$$= \begin{pmatrix}\cos\alpha - i\sin\alpha\cos\theta\\ -i\sin\theta\cos\alpha\end{pmatrix} \qquad (3.14)$$

where

$$\alpha \equiv \mu BT/2\hbar = \Omega/2 \qquad (3.15)$$

(the first term is the dynamical phase). Bitter and Dubbers measured σ_z by again passing the beam through a polarizer. Its expectation value is {E}

$$< \sigma_z > = < \Psi(T) | \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} | \Psi(T) >$$
$$= \tfrac{1}{2}(\cos^2\theta + \cos 2\alpha \sin^2\theta) \qquad (3.16)$$

They detected Ω (and verified the formula 3.11) through its effect on α (3.15), by measuring the phase of the oscillations of the $\cos 2\alpha$ term as B was varied.

The spin-1/2 case has the wider importance that it describes any 2-state quantum system. The restriction to rotational symmetry, and the generality embodied in the function F in (3.6), are here unnecessary because the most general 2-state Hamiltonian is (up to a trivial multiple of the identity)

$$H = \sigma \cdot R = \tfrac{1}{2}\begin{pmatrix} Z & X + iY \\ X - iY & -Z \end{pmatrix} \qquad (3.17)$$

An interesting recent application is to the *electron microscopy of crystal dislocations* [9] in the 2-beam approximation: a dislocation causes distortion and disruption of micrograph fringes, which can be interpreted entirely as an effect of the geometric phase.

Another application is to *degeneracies*. From the sum-over-states (2.13) it was clear that degeneracies of the transported state $|n>$ are singularities of $V_n(X)$. We can discover the nature of the singularities by realising that close to the parameters X^* where they occur the dominant contribution m to (2.13) comes from the state degenerating with n (we assume the typical situation where there is only one such state). Therefore we have, locally, a 2-state problem, for which a linear change of parameters brings the non-trivial part of H to the form (3.17). Application of (3.8) for n=1/2 shows at once that the singularity of V_n is a monopole with strength ± 1/2. (The sign depends on whether $|n>$ degenerates with the state above or below, and whether the transformation to local parameters R is proper or improper. Originally [6] I left out the latter condition; it was stated correctly by Simon [10].) Mondragon and I [7] have explored the details of the monopole singularities in several numerical examples.

An important special case of degeneracy occurs when H is real (for example when the dynamics at each instant has time-reversal symmetry). Then in the local model (3.17), Y=0, and circuits C lie in the XZ plane with their spanning surfaces like hemispheres, which have solid angle Ω=2π (if C encloses the degeneracy). The geometric phase $\gamma = n\Omega = \pm(2\pi)/2 = \pm\pi$ therefore contributes a sign change (which is of course the only phase change that a real eigenfunction can have). Elsewhere [2] I have described some early history associated with this sign change, in the differential geometry of surfaces and in molecular physics.

Now I want to discuss some experiments involving *photons*, whose interpretation has been controversial. These particles have spin 1, so σ are 3x3 matrices; the eigenvalues of $\sigma \cdot R$ are $+R$, 0, $-R$. Photons have no magnetic moment and so cannot be turned with a magnetic field. But they have the property of *helicity*: along their propagation direction e_k they may have states with $\sigma \cdot e_k = \pm 1$ but not zero. Therefore the photon spinor can be turned by turning its propagation vector k. Chiao and Wu [11] and Tomita and Chiao [12] had the clever idea of achieving this with the light in a coiled optical fibre. k is the forward tangent direction of the fibre, and can be cycled with a coil whose ends are parallel. The geometric phases for the two helicities would then be (cf.3.9)

$$\gamma_{\pm 1}(C) = \pm \Omega(C) \qquad (3.18)$$

where Ω is the solid angle swept out by e_k on its unit sphere.

Momentarily choosing the z axis along $\boldsymbol{k}$, we can write the helicity states as

$$|+1\rangle = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad |-1\rangle = \begin{pmatrix}0\\0\\1\end{pmatrix} \tag{3.19}$$

Classically these correspond to *beams of circularly polarized light*, whose (generally complex) unit polarization vector e, in the electric field

$$E = |E|\,\mathrm{Re}\, e\, \exp\{i(kz - \omega t)\} \tag{3.20}$$

is related to the spinor $|\Psi\rangle$ by

$$|\Psi\rangle = \begin{pmatrix} e_x - i e_y \\ 0 \\ e_x + i e_y \end{pmatrix}/\sqrt{2} \tag{3.21}$$

The states $|+\rangle$ and $|-\rangle$ correspond to right and left polarization, with $e_y = +ie_x$ and $-ie_x$.

Tomita and Chiao fed their fibre with light *linearly* polarized in direction α, i.e. $e_x = \cos\alpha$, $e_y = \sin\alpha$. This is a superposition of the two helicity states:

$$|\Psi(0)\rangle = \begin{pmatrix} \exp\{-i\alpha\} \\ 0 \\ \exp\{+i\alpha\} \end{pmatrix}/\sqrt{2} = \exp\{-i\alpha\}|+\rangle + \exp\{i\alpha\}|-\rangle \tag{3.22}$$

They were careful to coil their fibre without twisting it. Because of this, there was no stress-induced circular birefringence and the two helicities propagated at the same speed. In other words, the two components in (3.22) acquire *identical* dynamical phases (equal to kL, where L is the length of the fibre). But their geometric phases are equal and opposite, leading to emergent light with state

$$|\Psi(L)\rangle = \exp\{i\,kL\}\{\exp\{-i(\alpha + \Omega)\}|+\rangle + \exp\{i(\alpha + \Omega)\}|-\rangle\} \tag{3.23}$$

This is again linearly polarized, but along $\alpha+\Omega$ rather than α.

So the effect of the geometric phase is to rotate the direction of linear polarization by Ω. In other words, the coiling induces 'geometric optical activity' or 'geometric circular birefringence'. Experiment [12] verifies the effect very accurately. Recalling now the geometry of lecture 1, we can rephrase the description in yet another way: quantum anholonomy of photon eigenstates is equivalent to *parallel transport* of linear polarization along the fibre.

This raises the question: is the effect quantum or classical? Several authors have [13,14] argued that it is classical. Chiao and Wu

> "would rather think of these effects as topological features of classical Maxwell theory which originate at the quantum level, but survive the correspondence principle limit ($\hbar \to 0$) into the classical level".

Further questions now arise: *where* in Maxwell's theory is the anholonomy? Why is it so tricky to understand the effect classically, yet so straightforward quantum-mechanically? I have answered the first question [15] by deriving the parallel transport of linear polarization from Maxwell's equations for a fibre (the result cannot be justified by appeal to the known parallel transport along curved *rays* because these

experiments involve monomode fibres, for which geometrical optics is not a valid approximation).

The second question is a pseudo-problem, and so in fact is the whole discussion of whether the fibre effect is classical or quantum, because for the optical processes considered here the quantum and classical descriptions are the same! This is the view of Feynman [16]

> "The photon equation is just the same as Maxwell's equations...."

To appreciate the intuition underlying this assertion, consider Maxwell's equations in a fibre:

$$\dot{D} = \nabla \wedge H;\ \dot{B} = -\nabla \wedge E;$$
$$B = \mu_0 H;\ D = \varepsilon(r) E \qquad (3.24)$$

Here the dielectric permittivity function $\varepsilon(r)$ describes the glass refractive index which confines the light inside the fibre. These equations have the form

$$\partial_t(\text{fields}) = (\text{matrix linear in } \nabla) \times (\text{fields}) \qquad (3.25)$$

Multiplying by $i\hbar$ gives

$$i\hbar\partial_t(\text{fields}) = (\text{matrix linear in } p = -i\hbar\nabla) \times (\text{fields}) \qquad (3.26)$$

and makes Maxwell's classical equations look like Schrödinger's quantum equation.

To make this interpretation legitimate, we have to ensure that the operator on the right of (3.26) is Hermitian. Several authors [17-20] have carried out this programme, but their results are useless here

because they are restricted to propagation in free space, for which the permittivity is constant so there can be no guiding of the light. In discussion with A.Pines I have however found the following exact (and essentially unique) Schrödinger implementation of the fibre equations (3.24) (which also allows the magnetic permeability μ_0 to be replaced by a function $\mu(r)$.

Define the six-component spinor

$$|\Psi(r,t)\rangle \equiv \begin{pmatrix} M_+ \\ M_- \end{pmatrix} \qquad (3.27)$$

where

$$M_\pm = \varepsilon^{1/2} E \pm i\mu^{1/2} H \qquad (3.28)$$

In terms of the refractive index

$$n(r) = \left[\frac{\varepsilon(r)\,\mu(r)}{\varepsilon_0\,\mu_0}\right]^{1/2} \qquad (3.29)$$

we define the modified momentum operator

$$\Pi \equiv n^{-1/2}(r)\, p\, n^{-1/2}(r) \qquad (3.30)$$

and the 'inhomogeneity vector'

$$\xi \equiv \frac{1}{4n(r)}\nabla \log\left(\frac{\varepsilon(r)}{\mu(r)}\right) \qquad (3.31)$$

It can now be shown [E] that $|\Psi\rangle$ obeys a time-dependent Schrödinger equation with Hamiltonian

$$H = c\begin{pmatrix} \Pi\cdot\sigma & i\hbar\xi\cdot\sigma \\ -i\hbar\xi\cdot\sigma & \Pi\cdot\sigma \end{pmatrix} \qquad (3.32)$$

where σ is the following vector of spin-1 matrices:

$$\sigma = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.33}$$

H is manifestly Hermitian.

To apply this 'Schrodinger lookalike' formalism to the fibre experiments, we replace $|\Psi(r,t)\rangle$ by a wavepacket travelling with the speed of light in the fibre. This allows the replacement

$$cH \rightarrow c\, \hbar k/n = \hbar\omega\, c_k(t) \tag{3.34}$$

where ω is the frequency of the light and $c_k(t)$ the fibre direction at the place reached by the packet at the time t. Then we reject the off-diagonal terms in (3.32), on the grounds that the inhomogeneity vector ξ in (3.31) is perpendicular to that helicity component σ which in the adiabatic (guiding) approximation is along the fibre. Now M_+ and M_- are uncoupled in $|\Psi\rangle$, giving the two separate Schrödinger equations

$$i\hbar\partial_t M_\pm(t) = \pm\, \hbar\omega\, c_k(t) \cdot \sigma\, M_\pm(t) \tag{3.35}$$

This describes 'photons' (light particles in the sense Newton meant) with energy $E=\hbar\omega$. The eigenstates with $M_\pm$ correspond (E) to light with the two hands of circular polarization. Each of (3.35) also has an eigenstate with $E=-\hbar\omega$. These 'antiphotons' are not new particles but redundant descriptions of the backward-travelling particles with positive energy and k reversed (the redundancy can be traced to the fact that the physical fields are real - cf. 3.20).

With the two equations (3.35) we have distilled from Maxwell's equations a description of light in a fibre as a stream of spin-1 particles driven by a time-dependent Hamiltonian of the form (3.6). The anholonomy (3.18) follows at once, justifying the argument of Chaio and Wu.

Optical anholonomy of a different sort was discovered long ago by Pancharatnam [21] in strikingly original work whose significance went unappreciated until recently [22,23]. Here I have space only for a brief description. Instead of cyclically changing the direction k of light in a fixed state of polarization, Pancharatnam was concerned with cyclic changes in the polarization of light travelling in a fixed direction.

Such light can be represented as a 2-component spinor, whose state of polarization is an eigenstate of (3.17) determined by the direction R/R, which is a point on the Poincaré sphere. A polarization cycle is then a loop C on this sphere. Pancharatnam showed that there is an associated geometric phase $\Omega(C)/2$, where Ω is the solid angle subtended by C at the centre of the sphere, and thereby anticipated our result (3.9) for $n=1/2$.

I have given elsewhere [23] the 'reconciliation' of the Pancharatnam $\Omega/2$ (on the Poincaré sphere) and the Chiao-Wu-Tomita Ω (on the sphere of directions e_k). It is of course possible to combine polarization and direction cycles; Bhandari [24] describes how to calculate the resulting geometric phases.

Lecture 4

Suppose the cycled Hamiltonian has a classical limit, corresponding to a system with N freedoms. Then instead of being an operator, H is a function

$$H = H(x; X(t)) \tag{4.1}$$

where

$$x \equiv (q,p) = (q_1 \cdots q_N, p_1 \cdots p_N) \tag{4.2}$$

is position in the $2N$-dimensional phase space. It is natural to expect the quantum phase anholonomy $\gamma_n(C)$, and the underlying 2-form $V_n(X)$, to be mirrored by anholonomy in the classical system. Mystery still shrouds the nature of that anholonomy in the general case, but Hannay [25, see also 26] discovered what it is in the important special case of systems whose motion at each fixed X is *multiply periodic*.

Hannay reasoned as follows. A quantum eigenstate for fixed X is like an oscillator:

$$|\psi> = |n> \exp[-i\omega t], \text{ where } \omega = E_n/\hbar \tag{4.3}$$

We have seen that it exhibits anholonomy when X is cycled. Now there are of course oscillators in classical mechanics too, so we should likewise expect them to possess anholonomy. Instead of $|\psi>$ we will have the oscillator coordinate, and instead of $|n>$ we will have the oscillator amplitude. The phase will now represent an *angle* θ. This may be literally an angle in space - as with a wheel - or, more commonly, an abstract angle variable [27] chosen to make the motion uniform in phase space - as with a swinging pendulum. In an adiabatically cycled system, the anholonomy should show up as a shift in the total change of θ:

$$\theta(t) - \theta(0) = \int_0^T dt\, \omega(X(t)) + \Delta\theta(C) \tag{4.4}$$

The first term is the dynamical angle change, which is the obvious generalization of ωT for changing ω. The second term is the geometric angle shift, now called Hannay's angle. Of course for multiply-periodic motion there is more than one angle, the maximum (fully integrable motion) being N; we shall denote the j'th angle by $\Delta\theta_j$.

By geometric arguments, Hannay found a formula [25] for $\Delta\theta_j$ which has spawned a considerable amount of new classical phase-space geometry [28-30]. I reformulated Hannay's derivation, and showed [26] how, in the semiclassical limit, $\Delta\theta_j$ is related to the quantum phase. Here I will not repeat those arguments,but will instead outline a new, general approach to the semiclassical limit of the 2-form, developed last year after conversations with M.Wilkinson. As well as reproducing known results this gives a hint of what might happen in the *nonintegrable* case, where the motion is chaotic rather than periodic, and no angle variables exist.

We start with the following formula (intermediate between 1.21 and 1.22):

$$V_n(X) = \mathrm{Im}\, d\wedge <n|d|n> \tag{4.5}$$

to find $d|n>$ we differentiate the eigenequation for $H(X)$:

$$(dE_n - dH)|n> + (E_n - H)|d n> = 0 \tag{4.6}$$

where

$$dH \equiv H(X + dX) - H(X) \tag{4.7}$$

Thus

$$|\mathrm{d}\,n> = \lim_{\varepsilon\to 0} \frac{1}{H - E_n - \mathrm{i}\,\varepsilon}(\mathrm{d}\,E_n - \mathrm{d}\,H)|\,n> \tag{4.8}$$

where ε is necessary to provide a temporary resolution of the essential ambiguity in $<n|\mathrm{d}n>$. Substituting into (4.5), and introducing an integral representation for the first operator, we obtain {E}

$$\mathrm{V}_n = \lim_{\varepsilon\to 0} \frac{1}{h} \mathrm{Re}\int_0^\infty \mathrm{d}\,t \exp\{-\varepsilon t/h\}\,\mathrm{d}\wedge < n|[(\mathrm{d}\,H)_t - \mathrm{d}\,E_n]|\,n> \tag{4.9}$$

in which $(\mathrm{d}H)_t$ denotes the Heisenberg-evolved operator, namely

$$(\mathrm{d}\,H)_t \equiv \exp\{\mathrm{i}\,Ht/h\}\,\mathrm{d}\,H\exp\{-\mathrm{i}\,Ht/h\} \tag{4.10}$$

The purpose of these dubious formal manipulations was to get V in terms of an expectation value. Now we can use the correspondence principle: the classical limit of the expectation $<n|A|n>$ of any observable A is the average of the corresponding classical phase-space function $A(x)$over the manifold corresponding [31-33] to the state $|n>$. Because $|n>$ is a stationary state, the manifold must be *invariant* under the dynamics. For an integrable system, it is a *phase-space torus* with given quantized values of the actions $I=(I_1,\ldots,I_N)$. For an ergodic system, the manifold may be - at least in some averaged sense - the whole *energy surface* with the energy $E_n=H(x)$ of the state. It might also happen that the manifold is a single closed orbit. Henceforth we denote such classical averages, which replace $<n|A|n>$, by

$$< A(X) > = \int \mathrm{d}\,\alpha\, A(x(\alpha)) \tag{4.11}$$

where $\alpha=(\alpha_1,\alpha_2\ldots)$ is a set of coordinates on the manifold with $\mathrm{d}\alpha$ an invariant measure; the choice of α will be discussed later.

We also need the classical counterpart of the operator (4.10). this must incorporate the X-dependence of the classical manifolds. Let $x(t,\alpha;X)$ (fig.4.1) be the phase point at time t on the orbit which starts at $t=0$ from the point x with coordinates α on the manifold at parameters X. Then corresponding to (4.10) we have (cf 4.7)

$$(\mathrm{d}\,H)_t = H(x(t,\alpha;X);X + \mathrm{d}\,X) - H(x(t,\alpha;X);X) \tag{4.12}$$

The quantity $\mathrm{d}E_n$ in (4.9) corresponds to the energy difference $\mathrm{d}E$ between manifolds at X and $X+\mathrm{d}X$. This is the same for any pair of points, one on each manifold, and we choose points with the same α and t. Thus

$$\mathrm{d}\,E = H(x(t,\alpha;X + \mathrm{d}\,X);X + \mathrm{d}\,X) - H(x(t,\alpha;X);X) \tag{4.13}$$

Subtracting, we obtain

$$(\mathrm{d}\,H)_t - \mathrm{d}\,E = -\frac{\partial H}{\partial x}\mathrm{d}\,x = \dot{p}_t\cdot\mathrm{d}\,q_t - \dot{q}_t\cdot\mathrm{d}\,p_t \tag{4.14}$$

where

$$\mathrm{d}\,x_t = x(t,\alpha;X + \mathrm{d}\,X) - x(t,\alpha;X) \tag{4.15}$$

Now, the operator d in (4.9) commutes with the average (4.11) so we can allow it to act on (4.14)

$$\begin{aligned} \mathrm{d} \wedge \{(\mathrm{d} H)_t \cdot \mathrm{d} E\} &= \mathrm{d} p_t \wedge \cdot \mathrm{d} q_t - \mathrm{d} q_t \wedge \cdot \mathrm{d} p_t \\ &= \mathrm{d} p_t \wedge \cdot \mathrm{d} q_t + \mathrm{d} q_t \wedge \cdot \mathrm{d} p_t \\ &= \frac{\mathrm{d}}{\mathrm{d} t} \mathrm{d} p_t \wedge \cdot \mathrm{d} q_t \end{aligned} \tag{4.16}$$

The classical limit of the 2-form now becomes

$$\begin{aligned} V_n \underset{h \to 0}{\to} & \lim_{\varepsilon \to 0} \frac{1}{h} \mathrm{Re} < \int_0^\infty \mathrm{d} t \exp\{-\varepsilon t/h\} \frac{\mathrm{d}}{\mathrm{d} t} \mathrm{d} p_t \wedge \cdot \mathrm{d} q_t > \\ & \equiv - W(X)/h \end{aligned} \tag{4.17}$$

where the *classical 2-form* $W(X)$ is easily found to be

$$W(X) = < \mathrm{d} p \wedge \cdot \mathrm{d} q > \tag{4.18}$$

The wedge $\wedge$ acts between the d's in X space, and the $\cdot$ acts between the vectors p and q. There is no longer any time-dependence: $\mathrm{d}p$ and $\mathrm{d}q$ refer to displacements linking points labelled α at X and $X+\mathrm{d}X$, that is $x(0,\alpha; X+\mathrm{d}X)$ and $x(0,\alpha;X)$ (fig. 4.1).

The result (4.18) has two important invariance properties. First {E}, W is invariant under *canonical transformations* of the phase-space variables x, provided the transformation does not involve the parameters X. Second {E, not easy} W is invariant under arbitrary X-dependent shifts of the manifold coordinates α; that is, under the change to coordinates

$$\alpha'_j = \alpha_j + F_j(X) \tag{4.19}$$

This latter invariance is the classical analogue of the quantum gauge invariance (2.14).

Because of the simple appearance of (4.18), and these two invariances, it seems that the result we have found is surely the correct classical limit of the phase 2-form. But appearances can deceive, and in fact (4.18) is a subtle and slippery formula whose meaning is proving hard to extract.

Consider first the manifold corresponding to $|n>$. We do not know what this is in the general case. The correspondence principle, combined with the quantum adiabatic theorem, strongly suggests that the manifold is labelled by the (quantized) value of some classical adiabatic invariant, which is conserved as X varies slowly. However, no such invariant is known for a general system, whose motion is neither integrable nor ergodic. For ergodic (e.g. completely chaotic) systems there is an adiabatic invariant, namely [34] the phase space volume within the energy surface $H(x)=E$. When quantized, this invariant yields the Weyl rule [31] for the energy levels, which gives quite a good semiclassical description of the average behaviour of the spectrum. But the difficulty in an ergodic system is that there seems no sensible choice of coordinates α on the energy surface; this is necessary in order to be able to associate phase points for different X and so give meaning to $\mathrm{d}p$ and $\mathrm{d}q$.

No problems arise for integrable systems - the case considered by Hannay, where motion is multiply periodic. Corresponding to $|n>$ is a phase-space torus, labelled by the N action variables I, which are not only invariant under the motion for fixed X but adiabatically conserved when X changes slowly. And the natural coordinates α, labelling each torus with an invariant measure, are the N angular variables $\theta = (\theta_1 \ldots \theta_N)$ conjugate to I. Therefore the classical 2-form $W(X)$ is given unambiguously for this case.

Using semiclassical analysis that I will not repeat here, it is possible to show [26] that the j'th Hannay angle (cf.4.4) corresponding to a circuit C is the following flux through C:

$$\Delta\theta_j(C) = -\frac{\partial}{\partial I_j}\iint W(X) \tag{4.20}$$

We can express this very simply in terms of the quantum phase by noting (4.17) and the fact that for integrable systems states are labelled by N quantum numbers $n=(n_1...n_N)$, one corresponding to each action I_j, which is quantized in units of $\hbar$. The result is

$$\Delta\theta_j(C) = -\frac{\partial}{\partial n_j}\gamma_n(C) \tag{4.21}$$

After all this abstraction, it is healthy to do a concrete calculation. I will illustrate the inner workings of (4.18) by evaluating W (and $\Delta\theta$) for a *classical spin*. This application was mentioned briefly in Hannay's paper [25] and worked out in detail later [35]. Consider an angular momentum vector S whose dynamics is determined by an energy function $E(S)$ through the equation of motion

$$\dot{S} = \nabla_S E(S) \wedge S \tag{4.22}$$

This conserves the length $S=|S|$, so S moves on the surface of a sphere.

To make contact with our earlier analysis of quantum spins, we choose E as a function of $S\cdot R$, where R is a given vector whose components (X,Y,Z) are the parameters, soon to be cycled. For fixed R the component of S along R is conserved {E}. Therefore the motion is *precession* about R, and S moves uniformly round a circle

(fig.4.2) on its sphere. This motion can be described by the evolution of an angle θ.

Now let R be slowly cycled. At the end of the cycle C, S is back on its original circle, at a position shifted by a Hannay angle $\Delta\theta$. These assertions follow from the fact that (4.22) describes a Hamiltonian system with one freedom, whose phase space is the S sphere. To see this, choose a fixed direction z in S space and canonical variables

$$p = S_z; \quad q = \tan^{-1}\{S_y/S_x\}$$
$$= \text{azimuth angle of polar coordinates with axis } z \tag{4.23}$$

Thus $dq\,dp$ is the area element on the S sphere. As Hamiltonian choose the energy in q,p variables, i.e.

$$H(q,p) = E\left(\sqrt{S^2-p^2}\cos q, \sqrt{S^2-p^2}\sin q, p\right) \tag{4.24}$$

Then Hamilton's equations reproduce the dynamics (4.22) {E}. Moreover in our case where E is a function of $S\cdot R(t)$, the adiabatically conserved action variable is {E}

$$I = S(t)\cdot r(t) \tag{4.25}$$

where r is the unit vector R/R. I confines S to a 1-torus on its sphere; this is just the circle in fig.4.2, and conjugate to I is the angle θ.

It is worth remarking that the anholonomy $\Delta\theta$ involves a hierarchy of three levels of rotation: the *spin* (axis S), the *precession* (of S round the axis R), and the *turn* (of R around C)

Now we calculate W(R) from (4.18). We need

$$\mathrm{d}p \wedge \mathrm{d}q = \mathrm{d}S_z \wedge \mathrm{d}(\tan^{-1}\{S_y/S_x\}) = \frac{\mathrm{d}S_z \wedge (S_x\,\mathrm{d}S_y - S_y\,\mathrm{d}S_x)}{S_x^{\,2} + S_y^{\,2}} \tag{4.26}$$

Introduce a unit triad r,u,v on the R sphere. With these local axes (fig.4.3)

$$S = I\,r + \sqrt{S^2 - I^2}\cos\theta\,u + \sqrt{S^2 - I^2}\sin\theta\,v \tag{4.27}$$

The aim now is to express $<\mathrm{d}p \wedge \mathrm{d}q>$ in terms of $\mathrm{d}u$ and $\mathrm{d}v$ (and ultimately in terms of R through $r = u \wedge v$).

It helps to choose instantaneous axes x,y,z along u,v,r, i.e.

$$u = (1,0,0); \quad v = (0,1,0); \quad r = (0,0,1) \tag{4.28}$$

Then

$$\begin{aligned} \mathrm{d}u &= (0, \mathrm{d}u_y, \mathrm{d}u_z) \\ \mathrm{d}v &= (\mathrm{d}v_x, 0, \mathrm{d}v_z) \\ \mathrm{d}r &= (-\mathrm{d}u_z, -\mathrm{d}v_z, 0) \end{aligned} \tag{4.29}$$

From (4.27) with θ fixed, we find

$$\begin{aligned} \mathrm{d}S_x &= -I\,\mathrm{d}u_z + \sqrt{S^2 - I^2}\sin\theta\,\mathrm{d}v_x \\ \mathrm{d}S_y &= -I\,\mathrm{d}v_z + \sqrt{S^2 - I^2}\cos\theta\,\mathrm{d}u_y \\ \mathrm{d}S_z &= \sqrt{S^2 - I^2}(\cos\theta\,\mathrm{d}u_z + \sin\theta\,\mathrm{d}v_z) \end{aligned} \tag{4.30}$$

Substitution into (4.26) and averaging over θ leads to [E]

$$<\mathrm{d}p \wedge \mathrm{d}q> = -\frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\,\mathrm{d}p \wedge \mathrm{d}q = -I\,\mathrm{d}u_z \wedge \mathrm{d}v_z \tag{4.31}$$

Reinstating general axes gives the result:

$$W = <\mathrm{d}p \wedge \mathrm{d}q> = -I\,\mathrm{d}u \wedge \mathrm{d}v \tag{4.32}$$

Apart from the factor $-I$, this is the same as the 2-form (1.13) that occurred in our earlier study of parallel transport. Therefore we can use (1.15) to give the monopole formula, which (with d replaced by ∇_R) is

$$W = -I\,R/R^3 \tag{4.33}$$

For quantized actions $I = n\hbar$ this immediately confirms the correctness of the relation (4.17), because it reproduces the quantum spin 2-form (3.8) (here 4.17 is exact, rather than being a semiclassical approximation).

From (4.20), Hannay's angle is the flux through C of a unit monopole, namely

$$\Delta\theta = \Omega(\mathrm{C}) \tag{4.34}$$

This dynamical angle anholonomy is exactly the same as the geometric anholonomy in the parallel transport of a vector (lecture 1). We encountered a similar identity in the optical fibre experiment (lecture 3) in the duality between photon spin phase anholonomy and the parallel transport of linear polarization. A purely mechanical illustration of the duality is the *Foucault pendulum*.

Imagine first that the pendulum is rotating conically, with angular velocity ω ('circular polarization'), rather than swinging to and fro ('linear polarization') as it usually does. Our general classical spin analysis can be applied, with the local upward vertical $-g$ playing the role of the parameter R. As the earth turns, the vertical turns with it.

After a day ($t=T$) R has cycled, and the angle θ of the conical pendulum has increased by the solid angle

$$\Omega = 2\pi\,(1 - \sin(\text{latitude})) \tag{4.35}$$

as well as the dynamical ωT. This anholonomy is the same for both senses $\pm\,\omega$ of conical rotation.

Now let the pendulum swing linearly, and regard this as the superposition of two opposite conical rotations. If the bob swings in the xy (horizontal) plane, and is initially x polarized, we have

$$\begin{aligned}
&\text{at the start: } x + i\,y = \exp(i\,\omega t) + \exp(-i\,\omega t),\\
&\qquad\qquad \text{i.e. } (x\ ,y\ ) = 2\cos\,\omega t\ (1,0)\\
&\text{at the end: } x + i\,y = \exp(i\,\Omega)2\cos\,\omega t\,,\\
&\qquad\qquad \text{i.e. } (x\ ,y\ ) = 2\cos\,\omega t\ (\cos\,\Omega,\sin\,\Omega)
\end{aligned} \tag{4.36}$$

The effect of cycling is therefore to rotate the direction of swing by Ω. In other words, the direction is *parallel-transported*. From the rotating earth this appears as a slow rotation, at a rate

$$\omega_{\text{Foucault}} = (2\pi - \Omega)/T = \omega_{\text{earth}}\sin(\text{latitude}) \tag{4.37}$$

(in Bristol this is 11.7° an hour). A clever mechanical analogue of the Foucault pendulum, whose anholonomy can be seen without waiting a day, was developed by Kugler and Shtrikman [36].

It seems that we have come full circle in these lectures. We started with parallel transport on a sphere, and now once again we encounter parallel transport on a sphere. But as with our other cycles, the end is subtly different from the beginning. In lecture 1, parallel transport was introduced as a purely mathematical construction. Now we find that what was mathematically natural is enforced physically by the laws of Nature (in this case Newton's).

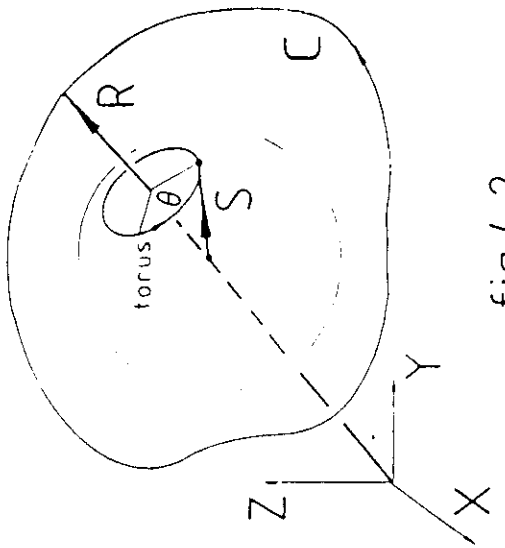


fig.4.2

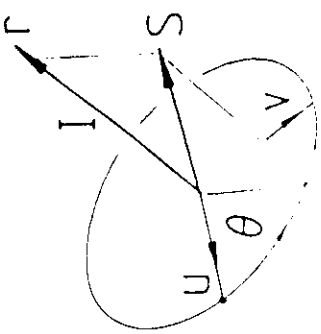


fig.4.3

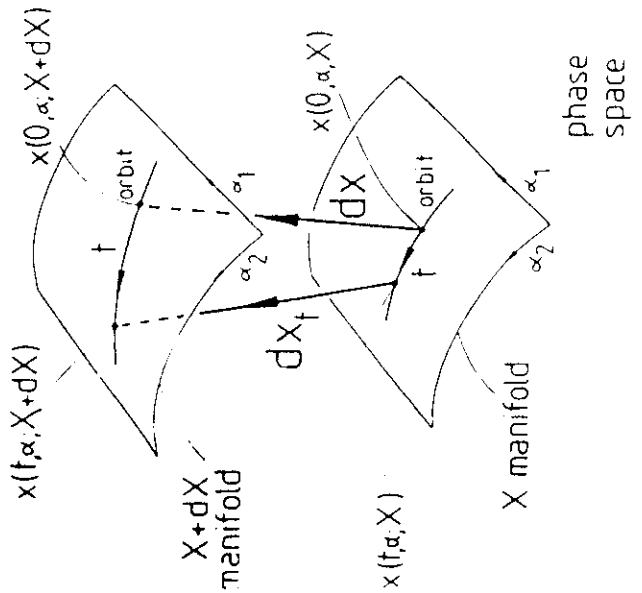


fig.4.1

Lecture 5

Here I will describe three generalizations of quantum adiabatic anholonomy. The first, from Wilczek and Zee [37], allows the transported states to be degenerate. Consider a group of N (orthonormal) states $|1(x)>$, $|2(x)>$ ..... $|N(X)>$ which are degenerate for all X on C, with energy $E(X)$. Such a situation usually arises when $H(X)$ has some symmetry. Because of the degeneracy, adiabatically evolving states $|\Psi(t)>$ will not cling to individual members $|n>$ of the group. All that can be said is that the $|\Psi>$ will remain a superposition of members of the group, (i.e. there will be no transitions to states outside the group) so that the adiabatic ansatz generalizing (2.2), is

$$|\Psi(t)> = \exp\left\{-\mathrm{i}\int_0^t \mathrm{d}t' E(X(t))/\hbar\right\} \sum_1^N a_n(t)\,|n(X(t))> \qquad (5.1)$$

To find the evolution of the coefficients $a_n(t)$, it is necessary to use the Schrödinger equation. The result is that the final superposition is the result of a unitary matrix acting on the initial superposition:

$$a_n(T) = \sum_{m=1}^{N} U_{nm}(C)\, a_m(0) \qquad (5.2)$$

where [E]

$$U_{nm}(C) = \mathrm{P}\exp\left\{\mathrm{i}\int_C A_{nm}\right\} \qquad (5.3)$$

in which P denotes path-ordering and $A_{nm}(=A^*_{mn})$ is the Hermitian matrix 1-form

$$A_{nm} = \mathrm{i} < n|\mathrm{d}|m > \qquad (5.4)$$

This generalises our previous case, where $N=1$ and U is a unit complex number whose phase is $\gamma_n(C)$. Even when $N>1$ it can

happen, exceptionally, that A_{nm} is diagonal; then the constituent states remain uncoupled and so acquire separate geometric phases $\gamma_n(C)$ just as though they were non-degenerate. This occurs in the optical fibre experiment (lecture 3), because the change in one helicity state, induced by change in the fibre direction, has no overlap with the other [E].

Usually, though, $A_{nm}(X)$ is not diagonal, and the A_{nm} at different X do not commute, a 'nonAbelian' property that makes it impossible to write U as the flux of anything simple. But it causes the final coefficients a_n to have different amplitudes as well as phases. In other words, the *populations* of degenerate levels can change, without any transitions involving the absorption or emission of energy. Segert [38] has proposed an interesting spectroscopic experiment, involving cycling the direction of parallel electric and magnetic fields whose magnitudes are tuned so as to make an atomic level degenerate (by cancellation of Stark and Zeeman shifts).

The second generalization, from Aharonov and Anandan [39], provides a setting in which the geometric phase can appear in evolutions that are not adiabatic. Let $H(T)$ be chosen to make the state $|\psi(t)>$ return *exactly*, apart from a phase, i.e.

$$|< \psi(0)|\,\psi(T)>| = 1 \qquad (5.5)$$

Such 'cyclic evolution' can be made to occur in many ways [39], even with an H that does not change at all.

It will usually be the case that $|\psi(t)>$ is not an eigenstate of $H(t)$. Nevertheless, it is possible to define the dynamical phase as the integral of the instantaneous expectation value of H. This can be factored out by defining (cf 2.2)

$$|\psi(t)> = \exp\left\{-\mathrm{i}\int_0^t \mathrm{d}t' < \psi(t')|\,H(t')\,|\psi(t')> /\hbar\right\} |\phi(t)> \qquad (5.6)$$

As in our earlier examples, the anholonomy of $|\phi>$ is determined by the Schrödinger equation, which again gives exactly [E] the parallel-transport law (cf 2.4)

$$< \phi | \dot{\phi} > = 0 \tag{5.7}$$

The anholonomy - that is, the geometric phase - is conveniently calculated in terms of any base state $|\psi(t)>$ that coincides with $|\psi(t)>$ up to phase but is singlevalued round C. Thus $|\psi>$ plays the same role as the singlevalued eigenstate $|n>$ in the adiabatic theory, and leads to the same result as before (cf. 1.20) namely

$$|\psi(T)> = \exp\left\{ - i \int_0^T dt' < \psi(t') | H(t') | \psi(t') > / \hbar + i\gamma(C) \right\} |\psi(0)> \tag{5.8}$$

where

$$\gamma(C) = - \mathrm{Im} \oint < \psi | d | \psi > \tag{5.9}$$

Note that in this formulation there is no parameter space. The circuit C is in the *Hilbert space* of states without phase. Sometimes this is called ray space, or density-matrix space, or projective Hilbert space. It differs from the full Hilbert space of all states $|\psi>$ by regarding as identical any states differing by complex scalar multipliers.

Aharonov and Anandan's theory is both richer and poorer than the adiabatic theory. It is richer in the sense that the adiabatic theory is a special case, where the base states $|\psi>$ are the n'th eigenstates of a family of Hamiltonians labelled by parameters X. Thus parameter space is a submanifold of projective Hilbert space. However, in applications (e.g. Born-Oppenheimer theory) parameters occur naturally, and can have a richer geometry (cf. the 2-form and its singularities at

degeneracies, and the geodesics generated by the metric 2.17) than the big Hilbert space - just as geometry on a curved surface can be richer than the geometry of the Euclidian 3-space in which it lives. Moreover, in the adiabatic framework, where the *Hamiltonian* is cycled exactly and states follow as best they can, there is a hierarchy of *corrections* to the geometric phase, of higher order in the adiabaticity parameter, which reveal [1] additional rich anholonomy (of Hamiltonians obtained by successive transformations to moving frames).

The third generalization, from Garrison and Wright [40], removes the restriction to unitary evolution. Consider a vector $|\Psi(t)>$ driven by a first order differential equation with a general time-dependent operator. For convenience we can still write this in 'Schrödinger' form, but $H(X)$ is now not a Hamiltonian but an arbitrary and usually non-Hermitian operator, with eigenvalues $E_n(X)$, possibly complex, corresponding to which are left eigenvectors $<\pi(X)|$ as well as right eigenvectors $|n(X)>$, chosen singlevalued on and within the circuit C in X space.

The calculation of adiabatic anholonomy is almost the same as before. We make the adiabatic ansatz (cf 2.2 and 1.18).

$$|\Psi(t)> = \exp\left\{ - i \int_0^t dt\, E_n(X(t))/\hbar + \gamma(t) \right\} |n(X(t))> \tag{5.10}$$

with the expectation that now $\gamma(t)$, and its value at $t=T$ which is the anholonomy, will be complex. A simple argument [E] gives

$$\gamma_n(C) = i \oint \frac{< \pi | d | n >}{< \pi | n >} \tag{5.11}$$

To conclude, here is an interesting application of the non-Hermitian theory. In the semiclassical asymptotics of the time-*independent* Schrödinger equation, there occur phase shifts of the

WKB-Maslov type [31], appearing as multiples of $\pi/2$ in quantization conditions and reflection amplitudes. People have often wondered whether these phases can be interpreted as anholonomy. Long ago, Voros gave one such interpretation in his thesis [41], and Littlejohn [42] has recently published a similar argument. Here I give a different interpretation, achieved after a conversation with A.Shapere.

A quantum particle with mass m and energy W, moving along a line in a potential $V(x)$, satisfies

$$d_z^2 u(z) + \frac{P^2(z)}{\hbar^2} u(z) = 0 \tag{5.12}$$

involving the classical momentum

$$P^2(z) = 2m(W - V(z)) \tag{5.13}$$

We write z rather than x because we want to continue the wavefunction u into the complex plane.

Define the two-component 'spinor' state vector

$$|\Psi(z)\rangle \equiv \begin{pmatrix} u(z) \\ \hbar u'(z) \end{pmatrix} \tag{5.14}$$

Then (5.12) is equivalent to the 'Schrödinger' evolution

$$i\hbar |\Psi'(z)\rangle = \begin{pmatrix} 0 & i \\ -iP^2(z) & 0 \end{pmatrix} |\Psi(z)\rangle \tag{5.15}$$

The 'Hamiltonian' in this equation is not Hermitian, even on the real axis $z=x$, although its eigenvalues

$$E_\pm(z) = \pm P(z) \tag{5.16}$$

are real in classically allowed regions of the real axis. The eigenstates $|\pm(z)\rangle$, and their duals $\langle\mp(z)|$ (corresponding to forward (-) and backward (+) travelling WKB waves in allowed regions), are [E]

$$|\pm(z)\rangle = \begin{pmatrix} 1 \\ \mp iP(z) \end{pmatrix} \qquad \langle\mp(z)| = (\mp iP(z), 1) \tag{5.17}$$

Degeneracies ($E_+ = E_-$) correspond to *classical turning points*, which are real or complex zeros of $W-V(z)$. In the usual case of simple zeros, these are *branch points* of $P(z)$ and hence of the spectrum $E_\pm(z)$ and the eigenvectors.

It is easy [E] to show that

$$i\frac{\langle\mp|\pm\rangle'}{\langle\mp|\pm\rangle} = i\frac{P'}{2P} = \frac{i}{4}(\log P(z))' \tag{5.18}$$

Thus the anholonomy (5.11) associated with a complex C can be written as the contour integral

$$\gamma(C) = \frac{i}{4}\oint_C dz\,(\log P)'$$
$$= -\frac{\pi}{2} \times (\text{number of zeros of } P^2(z) \text{ inside } C) \tag{5.19}$$

(If the turning points are not simple, they must be counted with their multiplicity.) Note that γ is real, that is the anholonomy takes the form of a phase shift, in spite of H being nonHermitian.

The two-state formalism based on (5.15) bears a superficial resemblance to the spin-1/2 problem considered in Lecture 3, but its nonunitarity is responsible for two important differences. First, the geometric phase is the same for the two states |+> and |->, in contrast to spin where the $n=\pm 1/2$ have opposite phases (cf 3.9). Second, the phase associated with a planar circuit of a degeneracy is $\pi/2$ rather than π. This is because degeneracies are branch points for nonHermitian

operators, and diabolical points ([43], especially the final remarks) for Hermitian ones.

An immediate application is to *oscillators* (e.g. harmonic, which have two real turning points bounding a classically allowed region. For a circuit of this region enclosing both branch points, the states return, with phase shifts (dynamical plus geometric)

$$\frac{1}{2} \cdot \frac{1}{\hbar} \oint P \, \mathrm{d}z = \pi \tag{5.20}$$

Single-valuedness of u (or $|\Psi\rangle$) requires that this phase be $2n\pi$. Thus we reproduce the well-known quantization condition

$$\oint P \, \mathrm{d}z = (n + 1/2)\, h \tag{5.21}$$

with the half-integer appearing as a consequence of the nonHermitian geometric phase. In this formulation the '1/2' is the combined effect of the branch points - as though the oscillator were a composite of two particles with spin 1/4, i.e. half-Fermions [46].

Another application is to the amplitude for *reflection above a barrier*. As is well known [44], this process is classically forbidden, and the reflection diminishes exponentially as $\hbar \to 0$. This is because there are no real turning points and we have to take C around the nearest complex one, at z^*, say. After the circuit of this branch point, P has changed sign (cf 5.16) and 5.17) and so $|+\rangle$ changes into $|-\rangle$ (and vice versa); physically this means the transformation of an incoming wave into a reflected wave. We immediately identify the reflection amplitude

$$r = -i \exp\left\{ 2i \int_{z_0}^{z^*} P \, \mathrm{d}z / \hbar \right\} \tag{5.22}$$

where z_0 is the point on the real axis from which the phase of the incident and reflected waves is reckoned. This is exponentially small because the integral in the exponential - the dynamical 'phase' - is in fact

not a phase because it has a positive imaginary part. The phase factor -i is a consequence of nonHermitian anholonomy.

Of course these one-dimensional semiclassical turning-point problems have been solved long ago by other means [44]. Our main result (5.19) is just a fancy way of dealing with the multivaluedness of the amplitude factor $P^{-1/2}$ in the WKB solutions of (5.12). Nevertheless, it is pleasant to discover how easily and naturally these phases appear, without any sign ambiguities, when interpreted as anholonomy.

References

[1] Berry, M.V. (1987), *Proc.Roy.Soc. A* 414, 31-46

[2] Berry, M.V. (1989), *'The quantum phase, five years after'*, in [3]

[3] Shapere, A and Wilczek,F, (1989), (eds) *'Geometric Phases in Physics'* (World Scientific: Singapore)

[4] Kaye, G.W.C and Laby, T.H, (1973), *Tables of Physical and Chemical Constants* (London: Longman; 14th ed.)

[5] Olariu, S, Popescu,I and Iovitsu I.I, (1985),*Revs.Mod.Phys.* 57, 339-436.

[6] Berry,M.V, (1984) *Proc.Roy.Soc.Lond., A* 392, 45-57.

[7] Mondragon,R.J and Berry,M.V, (1989),*Proc.Roy.Soc.Lond.* In press

[8] Bitter, T and Dubbers D, (1987),*Phys.Rev.Lett.* 59, 251-254

[9] Bird,D and Preston, A.R, (1988), *Phys.Rev.Lett.* 61, 2863-2866

[10] Simon, B, (1983), *Phys.Rev.Lett.* 51, 2167-2170

[11] Chiao R. Y and Wu,Y.S, (1986), *Phys.Rev.Lett.*, 57, 933-936

[12] Tomita, A and Chiao R.Y, (1986), *Phys.Rev.Lett.*, 57, 937-940

[13] Haldane, F.D.M, (1986), *Opt. Lett.* 11, 730-732

[14] Segert, J, (1987), *Phys.Rev.A.* 36, 10-15

[15] Berry, M.V, (1987), *Nature*, 326, 277-278

[16] Feynman,R, Leighton,R.B and Sands,M, (1965), *The Feynman lectures on physics'*, (Addison-Wesley) Volume 3, p21-6

[17] Laporte,O and Uhlenbeck,G.E, (1931), *Phys.Rev.* 37, 1380-1397

[18] Oppenheimer, J.R, (1931), *Phys.Rev.* 38, 725-746

[19] Archibald,W.J, (1955), *Can.J.Phys.* 33, 565-572

[20] Good,R.H Jr, (1957), *Phys.Rev.* 105, 1914-1919

[21] Pancharatnam,S, (1956), *Proc.Ind.Acad.Sci. A* 44, 247-262, also in *'Collected works of S.Pancharatnam'* (Oxford University Press 1975)

[22] Ramaseshan, S and Nityananda, R, (1986), *Curr.Science* 55, 1225-1226

[23] Berry,M.V, (1987), *J.Modern Optics*, 34, 1401-1407

[24] Bhandari, R, (1989), *Phys.Lett.A* 135, 240-244

[25] Hannay, J.H, (1985), *J.Phys.A* 18, 221-230

[26] Berry, M, (1985), *J.Phys.A* 18, 15-27

[27] Arnol'd, V.I, (1978), *Mathematical Methods of Classical Mechanics* (Springer: New York)

[28] Gozzi,E and Thacker,W.D, (1987), *Phys.Rev.D* 35, 2388-2397, 2398-2406

[29] Montgomery, R, (1988), *Commun.Math.Phys.* 120, 269-294

[30] Weinstein, A, (1988), Berkeley Mathematics preprint

[31] Berry, M.V, (1983), *Les Houches Lecture Series Session XXXVI*, eds.G.Iooss, R.H.G.Helleman and R.Stora, North Holland, Amsterdam, 171-271

[32] Heller,E.J, (1986), in *Quantum Chaos and Statistical Nuclear Physics* (ed: T-H Seligman and H.Nishioka) (Springer lecture notes in physics No. 263) pp 162-181

[33] Berry,M.V, (1989), *Proc.Roy.Soc.* A423, 219-231

[34] Lochak,P and Meunier,C, (1988), *'Multiphase averaging for classical systems'*, (Springer-Verlag, applied mathematical sciences series, No. 72)

[35] Berry, M.V, (1986), in *Fundamental aspects of quantum theory*, eds.V.Gorini and A.Frigerio, Plenum, NATO ASI series vol. 144, 267-278

[36] Kugler, M and Shtrikman,S, (1988), *Phys.Rev.D* 37, 934-937

[37] Segert, J, (1987), *Ann.Phys.* 179, 294-312

[38] Wilczek,F and Zee A, (1984), *Phys.Rev.Lett.* 52, 2111-2114

[39] Aharonov,Y and Anandan, (1987), *Phys.Rev.Lett.* 58, 1593-1596

[40] Garrison, J.C and Wright E.M, (1988), *Phys.Lett. A* **128**, 177-181

[41] Voros,A, (1977), *Développements semi-classiques*'(Thesis No. 1843 Orsay, France)

[42] Littlejohn, R.G, (1988), *Phys.Rev.Lett.* **61**, 2159-2166

[43] Berry, M.V and Mount,K.E, (1972), *Reps.Prog.Phys.*, **35**, 315-97

[44] Berry, M.V. and Wilkinson, M, (1984), *Proc.Roy.Soc.Lond.,A* **392**, 15-43

[45] Kuratsuji, H. and Iida, S. (1985), *Prog. Theor. Phys.*, **74**, 439-445

[46] Chapter 5 of reference [3]