COURSE 4

SOME QUANTUM-TO-CLASSICAL ASYMPTOTICS

MICHAEL BERRY

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

M.-J. Giannoni, A. Voros and J. Zinn-Justin, eds.
Les Houches, Session LII, 1989
Chaos et Physique Quantique /Chaos and Quantum Physics


BERRY'S FACE
i
e
PS
89

Contents

1. Theories as limits of other theories 255
2. The semiclassical limit 256
   2.1. Nonchaotic semiclassical nonanalyticities 256
   2.2. Long-time, small-$\hbar$ 259
   2.3. Algorithmic complexity 260
   2.4. The challenge of quantum chaology 262
3. Spectra 264
   3.1. Expectations 264
   3.2. Formalism 265
   3.3. Spectral asymptotics: principles 267
   3.4. Spectral asymptotics: formulae 270
4. Long orbits (asymptotics of asymptotics) 273
   4.1. Gaussian random waves? 273
   4.2. Phase space democracy 274
   4.3. Energy-level statistics 278
   4.4. Application(?) to the Riemann zeros 285
5. Divergence and resurgence 288
   5.1. Making the periodic-orbit sum converge 288
   5.2. Resurgence 1: exponentially small waves 292
   5.3. Resurgence 2: Riemann–Siegel 295
   5.4. Resurgence 3: bootstrapping the closed orbits 297
6. Last words 299

References 299

1. Theories as limits of other theories

In these lectures I do not wish to attempt a systematic, comprehensive and even-handed review of chaos in quantum physics, as I did in 1981 [1] (extending my earlier, almost entirely classical review [2]). The reason is certainly not a lack of progress – on the contrary, there have been many advances in almost every aspect of the subject, well summarised by Eckhardt [3]. Instead, I intend to emphasize some issues that lie at the foundations, and bring out a major difficulty that continues to obstruct our understanding.

There is controversy and confusion at the heart of the subject and even about its name. This is reflected in the title of the present school. 'Chaos and quantum physics' invites the abbreviation 'quantum chaos', but this carries the implication that chaos can be present in quantum evolution, which is unfortunate because (with one curious exception [124]) all evidence [4] indicates that there never is any. But if there is no quantum chaos, what is the real subject of the conferences [5–7], books [8,9], and many papers reporting discoveries about it? And if there is no quantum chaos, there certainly is classical chaos, so what becomes of the correspondence principle? I think that underlying these uncertainties are much deeper issues about what it means for one theory to be the limit of another, and I will begin by discussing these.

We usually envisage a situation where a more general theory $\mathcal{G}$ and a more special theory $\mathcal{S}$ are connected by a dimensionless parameter δ:

$$\mathcal{G} \rightarrow \mathcal{S} \quad \text{as} \quad \delta \rightarrow 0\,. \tag{1.1}$$

In the simplest case, quantities in $\mathcal{G}$ are analytic at $\delta = 0$. A paradigm is the mechanics of a particle with speed v, where $\mathcal{G}$ = special relativity, $\mathcal{S}$ = Newtonian mechanics, and $\delta = v^2/c^2$. The reduction of Einstein to Newton involves the elementary asymptotics

$$\sqrt{\left(1-\frac{v^2}{c^2}\right)} = \sqrt{(1-\delta)} = 1 - \frac{1}{2}\delta - \cdots\,. \tag{1.2}$$

In this case we are justified in writing the more general theory as a Taylor expansion in δ about the more special one:

$$\text{“}\mathcal{G} = \mathcal{S} + \delta\mathcal{S}_1 + \delta^2\mathcal{S}_2 + \cdots.\text{”} \tag{1.3}$$

This state of affairs is, however, neither the commonest nor the most interesting. Usually, $\mathcal{G}$ is nonanalytic at $\delta = 0$, so that the limit $\mathcal{G} \to \mathcal{S}$ is singular and the reduction of $\mathcal{G}$ to $\mathcal{S}$ is not straightforward. An example is the inviscid limit of fluid mechanics, that is the reduction of the Navier–Stokes equation to the Euler equation, with δ being 1/Reynolds' number. The highly singular nature of the limit – still far from being understood – is revealed by *turbulence* [10]. Instead of the dissipation vanishing smoothly at zero viscosity, as the naive formula of eq. (1.3) might suggest, it concentrates onto irregular regions that become fractal as $\delta \to 0$ [11].

Sometimes the $\delta \to 0$ limit is analytic except where some other parameter (X, say) takes a critical value X^*. An example is the thermodynamic limit, where $\mathcal{G}$ = statistical mechanics, $\mathcal{S}$ = thermodynamics, $\delta = 1/N$ for a system of N particles, X = temperature T and X^* = the critical temperature T_c. Away from T_c, statistical mechanics goes over smoothly into thermodynamics as described in textbooks [12]. At T_c, however, there is no continuum limit but rather a new state of matter, the critical state [13], which is neither continuous nor discrete, and for which thermodynamic quantities (e.g., compressibility) are singular.

These nonanalyticies, obstructions to naive reduction as embodied in eq. (1.3), should not be regarded as a nuisance. On the contrary, they are pointers to new physics, important features of the world like turbulence and critical behaviour, inhabiting the asymptotic borderland between theories.

2. The semiclassical limit

2.1. Nonchaotic semiclassical nonanalyticities

For us, the general theory is $\mathcal{G}$ = quantum mechanics, the special theory is $\mathcal{S}$ = classical mechanics, and the parameter is δ = Planck's constant $\hbar$. Of course $\hbar$ is not dimensionless; it has the dimensions of action. Therefore, '$\hbar$' is really shorthand for 'a dimensionless combination of physical quantities with $\hbar$ in the numerator'. In the course of these lectures I will draw attention to many such semiclassical parameters.

In the limit $\hbar \to 0$, nonanalyticities are abundant. I will give a series of preliminary examples that do not involve chaos. One does not have to

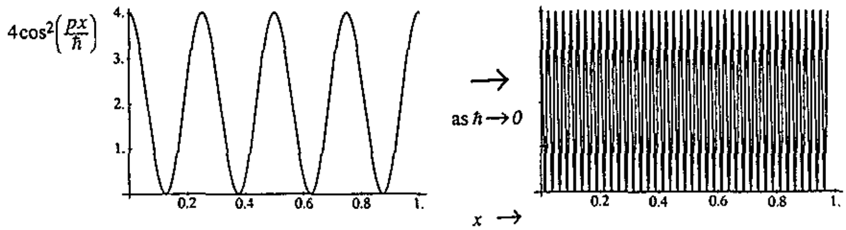

Figure 2.1.

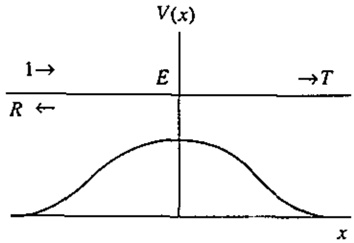

Figure 2.2.

search far to find nonanalyticity. In *two-beam interference*, for example (as in the double-slit experiment), the wave is a superposition of beams Ψ_+, with momentum p, and Ψ_-, with momentum $-p$:

$$\Psi_+ = \exp(\mathrm{i}px/\hbar), \quad \Psi_- = \exp(-\mathrm{i}px/\hbar). \tag{2.1}$$

The intensity, observable as the probability density of the wave, is

$$|\Psi_+ + \Psi_-|^2 = 4\cos^2(px/\hbar). \tag{2.2}$$

The limit $\hbar \to 0$ is pathological: the intensity oscillates infinitely fast and takes all values between 0 and 4 (fig. 2.1). This is because there is an essential singularity, of exponential type, at $\hbar = 0$. Only after averaging over a finite interval Δx, representing the practical impossibility of making measurements with perfect precision, do we regain the classical result [14],

$$1 + 1 = 2, \tag{2.3}$$

expected on the basis of the correspondence principle. (It is said that the averaging in this argument – though not the connection with the semiclassical limit – was Lord Rayleigh's answer to the question: why are two candles twice as bright as one?)

Although the two-beam example has a singular classical limit, it is special in that its semiclassical limiting form given by eq. (2.2) is exact rather

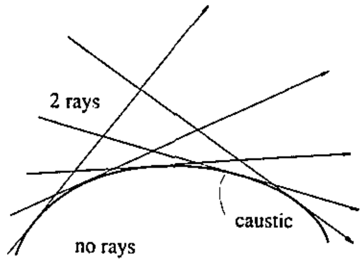

Figure 2.3.

than asymptotic. In this respect a more typical example, also involving an exponential essential singularity, is *reflection above a barrier* (fig. 2.2). As is well-known [15] for energy E and analytic potential $V(x)$ the semiclassical limit of the reflection coefficient is

$$|R|^2 \approx \exp(-1/\delta). \tag{2.4}$$

Here, δ - the dimensionless semiclassical parameter - is

$$\delta = \frac{4\hbar}{\mathrm{Im} \int_x^{z^*} \mathrm{d}z \sqrt{2m[E - V(x)]}}. \tag{2.5}$$

where x is any point on the real axis and z^* is the complex turning point (zero of $E - V$) closest to the real axis. This example illustrates the difference between the classical limit, which is zero, and the semiclassical limit, which is eq. (2.4).

Now we will see how semiclassical nonanalyticities can be enriched by changing a parameter (as thermodynamics is enriched by taking T through T_c). This will be important when we get to chaos. The best known example, whose optical manifestation was explained in 1838 by Airy in his theory of the rainbow [16], is *waves near caustics* [17,18] of the classical paths. A caustic (fig. 2.3) is an envelope of paths: on one side ('lit'), there are two rays through each point, and on the other ('shadow') there are none. If X is a parameter that passes through zero on the caustic, the semiclassical approximation to the wave is

$$\Psi(X) \approx \frac{1}{\hbar^{1/6}} \mathrm{Ai}\left(-\frac{KX}{\hbar^{2/3}}\right), \tag{2.6}$$

where Ai is the Airy function [19] and K is a constant. This is a much more complicated limit than our two previous examples. Indeed, it encapsulates both of them, because the large-$|X|$ asymptotics of Ai shows that on the lit side (X large and positive) Ψ has an oscillatory exponential semiclassical singularity, as in eq. (2.2). and in the shadow (X large and negative) Ψ has a real exponential semiclassical singularity. as in eq. (2.4). For waves near caustics, a rich hierarchy of higher singularities. classified as mathematical catastrophes and obeying scaling laws involving a series of exponents [20–22] can be generated by varying more parameters.

2.2. Long-time, small-$\hbar$

In these lectures we will be interested in a particular aspect of the semiclassical limit, namely the ways in which quantum quantities reflect the presence of classical chaos. Chaos is exponential sensitivity to initial conditions. in a bounded region (both conditions are necessary). It leads to stretching and folding in classical phase space and hence to ergodicity and mixing, and to the algorithmic complexity [4] of orbits (to be further discussed in the next section). Now, these classical phenomena emerge over *infinite time t*. so any discussion of the semiclassical limit must take account of the simultaneous limit $t \to \infty$ of a parameter independent of $\hbar$. Here $t \to \infty$ is analogous to the earlier X crossing a caustic or temperature passing T_c.

It is natural to ask whether the two limits $t \to \infty$ and $\hbar \to 0$, which are individually so complicated. commute. The answer is that they do not. Long-time quantum evolution is fundamentally different from long-time classical evolution.

Combining the semiclassical and long-time limits can lead to intricate behaviour even in the absence of chaos. An instructive example [23,24] is a *spin in a quadrupole field*. The classical spin is a vector $\boldsymbol{S}$ with energy $S_z^2/2I$, where I is a moment of inertia. The motion is trivially integrable: the length S and z-component S_z of $\boldsymbol{S}$ are both conserved, so $\boldsymbol{S}$ precesses about the z-axis with period $2\pi I/S_z$. The shortest period is $2\pi I/S$. If we allow classical superpositions (Liouville densities in phase space), classical motion need not be periodic.

For the quantum spin. the conserved quantities are both quantized:

$$S = \hbar\sqrt{j(j+1)} \approx \hbar j, \qquad S_z = \hbar n \quad (-j \leqslant n \leqslant j). \tag{2.7}$$

Quantum evolution over time t is governed by a unitary operator, whose simplest property is its trace, easily shown to be

$$T_U \equiv \operatorname{Tr}\hat{U} = \sum_{n=-j}^{j} \exp\left(-\mathrm{i}\frac{tn^2\hbar}{2I}\right) \approx \sum_{n=-j}^{j} \exp\left(-\mathrm{i}\frac{tSn^2}{2Ij}\right). \tag{2.8}$$

The quantum motion is always recurrent, with period $4\pi Ij/S$ – greater by $2j$ than the shortest classical period.

The semiclassical limit is $j \to \infty$, corresponding to $\hbar \to 0$ with S and t fixed ($1/j$ is the dimensionless semiclassical parameter δ). The long-time limit is $t \to \infty$ with j and S fixed. These limits are discordant, as can be seen from the very different patterns obtained by plotting T_U in its complex plane. Having given details elsewhere [24], I confine myself to a brief description here. In the long-time limit, T_U, being periodic, repeatedly traces a loop as t increases; the loops depend on j, of course, and are more complicated for larger j. In the semiclassical limit, T_U is not periodic, but can be transformed into a sum whose number of terms is independent of j (and proportional to t) and which, therefore, does not get more complicated as $j \to \infty$.

Faced with such a discordance, the natural way to attempt a resolution is to take both limits together, that is

$$j \to \infty\,,\ t \to \infty\,, \tau \equiv \frac{tS}{2\pi Ij} \text{ fixed}\,. \tag{2.9}$$

In the combined limit, T_U traces patterns which depend on the arithmetic of τ. If τ is irrational, the patterns have an infinite hierarchy of scales whose unraveling involves a renormalization [23] of the Gaussian sum of eq. (2.8), in which τ changes chaotically. Thus even in this very simple problem the origin of the plane with coordinates $1/j$, $1/t$, i.e., the semiclassical long-time limit, is a singularity much more powerful than the exponential and Airy singularities of our earlier examples.

2.3. Algorithmic complexity

Now we turn to the situation that will concern us for the rest of these lectures, in which the classical long-time limit is chaotic. It appears that in quantum mechanics, long-time evolution is not chaotic. This is an important discovery, first made by Casati, Chirikov, Ford and Izraelev with the kicked rotator [25]. Recently, Ford [4], and his colleagues [59, 60] have formulated the clash between classical and quantum long-time evolution in a striking and original way, with the aid of ideas from algorithmic complexity theory.

The calculation of the time-development of a system involves running a program. The input is a digit string, in the form of an algorithm for solving Newton's or Schrödinger's equation plus initial conditions, and the output is a digit string describing the orbit for a time t. Algorithmic complexity can be defined as the ratio

$$C = \frac{\text{input string length (in bits)}}{\text{output string length (in bits)}}. \tag{2.10}$$

This quantity can obviously be defined for any t, but the distinction between chaotic and regular motion, i.e., between random and nonrandom digit strings, emerges sharply only in the limit $t \to \infty$. If there is chaos, C tends to a finite constant: if not, C tends to zero.

To see why this is so, note first that the output string length increases as t (provided the same question about the orbit is being asked at each instant). The input string consists of three parts. First is the statement of the scheme for solving the equations of motion; the length of this is independent of t. Next is the specification of the time t for which the program is to run: this part of the string has length $\log t$ (number of digits in t). Finally, and crucially, there is the specification of the initial conditions to the accuracy required to produce the desired output. For a chaotic system, whose errors increase exponentially, this part is itself proportional to t and so dominates the input string, so that C tends to a constant. In that case the output string is random because it can only be generated by an input as long as itself; each new digit comes as a surprise – the string is 'algorithmically incompressible'. For a nonchaotic system, the initial condition part of the input string increases slower than t (for example as $\log t$ when the errors increase linearly rather than exponentially, or – as in the example to follow – not at all) and the limiting C is zero – the output string is algorithmically compressible.

A trivial example will serve to illustrate the importance of the limit $t \to \infty$. Consider a system with only two possible states $x = 0$ and $x = 1$, evolving in discrete time $t = 1, 2, 3, \ldots$ according to the simplest possible evolution law, in which the initial state never changes. The output string is then simply $0, 0, 0, 0, \ldots$ or $1, 1, 1, 1, \ldots$, which are obviously nonrandom. One program that generates these orbits is

$$\text{print (0 or 1) (digit string representing } t\text{) times} \tag{2.11}$$

This is longer than the output (t binary digits) until $t = 6$: thereafter, it is smaller, with $C \sim (\log t)/t$ which vanishes only in the limit $t \to \infty$. The futility of attempting to distinguish randomness from chaos in finite sequences, i.e., without asymptotics, was well dramatised by Van Kampen's question (in a lecture): is 1 a random number?

For two quantum systems, one classically not chaotic [59] and one classically chaotic [60] Ford and his colleagues show that the evolution has zero algorithmic complexity. In the latter case, they show also that the complexity of the sequence of (quasi)energy eigenstates is zero. Both systems have special features, but it begins to look as though the same conclusions hold for all quantum systems. The reason – still not quite clear – might be that the evolution of systems that can display classical chaos is determined quantally by a discrete spectrum of eigenstates (or, in the case of driven systems like the kicked rotator, of localized quasi-energy states).

This discordance between complexity in classical and quantum mechanics is not surprising, in view of the unavoidable long-time limit in the definition of algorithmic randomness and the known clash between $t \to \infty$ and $\hbar \to 0$ (section 2.2). One should not regard the discordance as a failure of the correspondence principle, for two reasons. First, that principle asserts that quantum expectation values tend to their classical limits as $\hbar$ alone tends to zero, other quantities being held fixed, whereas here we are taking $t \to \infty$ first. Second, it is not clear that complexity is an observable (is there a 'program length operator'?). I would rather think of the discordance in the spirit of Bohr, as an example of complementarity whose extreme expression can be formulated as an antinomy:

Classical evolution is deterministic and can be random;
Quantum evolution is indeterministic and nonrandom. (2.12)

2.4. *The challenge of quantum chaology*

If, as I expect, somebody eventually proves that persistent stretching and folding – i.e., chaos – never occurs in quantum mechanics, one possible reaction would be simply to agree that there is no quantum chaos, and stop studying the subject. But this would be a retreat from the challenge of unraveling the quantum phenomena concealed in the small $\hbar$, large t regime – in my view a misjudgement equivalent to ignoring other singular reductive phenomena such as turbulence and critical behaviour.

What are these quantum phenomena? In the kicked rotator itself, they are the fact that the quantum energy does not persistently diffuse as it does classically, and the underlying intricate quasi-energy level distribution and localization of eigenfunctions. I will not discuss these, but rather concentrate on another class of phenomena.

I will seek to understand the *spectrum* of autonomous bound systems, by which I will mean its energy levels and wavefunctions. The spectrum concerns the stationary states, which are the states that never change. Therefore, the study of spectra is par excellence a $t \to \infty$ subject, and the study of semiclassical spectra unavoidably involves the clashing limits $\hbar \to 0$, $t \to \infty$. Because of this clash, spectra provide some of the hardest semiclassical problems. (An example of a relatively easy problem is the description of nonstationary wavepackets at finite t, whose semiclassical limit is an application of Ehrenfest's theorem.) Two of the new quantum phenomena in semiclassical spectra are the applicability of random-matrix ensembles as models for some aspects of the statistics of energy levels [26,27,44], and Heller's wavefunction scars [28–31].

My point of view, then, is that the clash of limits should be regarded as an opportunity for quantum mechanics, rather than a contradiction or even [4] a shortcoming of that theory. I invented the term 'quantum chaology' [27,32] for the study of these new quantum phenomena, but it has not been widely taken up – perhaps one reason is that its use by others appears to have been censored by *Physical Review Letters*.

Here is an example of a result in quantum chaology, obtained by Keating [33–35] for the quantized version of Arnold's cat map [36] on the phase-space 2-torus. In this problem, Planck's constant is quantized to $h = 1/N$ where N is an integer [37], so that the semiclassical limit is $N \to \infty$. The classical map corresponds to a kicked oscillator, so the spectrum of the quantum map is that of a unitary operator, with eigenvalues on the unit circle. It can be shown [37] that the eigenvalues are distributed on $n(N)$ sites equally spaced around the circle. $n(N)$ is an erratic arithmetic function whose semiclassical limiting statistics can be found by the methods of asymptotic number theory. One important quantity is the mean site occupancy (inverse degeneracy) n/N, for which Keating finds the semiclassical average

$$\left\langle \log\left(\frac{n(N)}{N}\right)\right\rangle \approx -\frac{2\sqrt{\pi}}{e^2}(\log\log N)(\log\log\log N)\,. \tag{2.13}$$

Therefore, the quasi-energies very slowly become infinitely degenerate, on the average, in the semiclassical limit. This disproves an old conjecture with Hannay [37] that the mean degeneracy would tend to a constant. Keating's result is quantum chaology because it describes a quantum (i.e., nonclassical) phenomenon, in the semiclassical limit, reflecting the chaos in the classical system (for related non-chaotic maps, the formula does not hold).

The above system has an unhappy feature: the quantum cat, although it corresponds to an honestly chaotic classical map, is riddled with symmetries that make its semiclassical limit very special (for example, its quasi-energy distribution is unrelated to random matrices). From now on we will try to get results that hold for systems whose genericity is not spoiled by symmetries.

3. Spectra

3.1. Expectations

We consider a quantum system whose classical counterpart is bounded and has N freedoms ($N \geqslant 2$), so that its orbits inhabit a finite region in the $2N$-dimensional phase space

$$\boldsymbol{x} \equiv (\boldsymbol{q}, \boldsymbol{p}) \equiv (q_1, \ldots, q_N, p_1 \ldots, p_N) . \tag{3.1}$$

The orbits $\boldsymbol{x}(t)$ are generated by the Hamiltonian function $H(\boldsymbol{x})$. The aim is to understand the semiclassical limit of the discrete eigenvalues E_n and eigenfunctions $|n\rangle$ of the Hamiltonian operator $\hat{H}$ obtained by replacing $\boldsymbol{x}$ by operators $\hat{\boldsymbol{x}}$ whose noncommutation of course introduces $\hbar$. In this limit, quantum objects are built on a skeleton of classical ones.

Our expectations for the semiclassical E_n and $|n\rangle$ are guided by the following intuition. The spectrum of the Hamiltonian consists of stationary states and so the corresponding classical objects should be stationary too. For a state with energy E the classical stationary objects are the *invariant sets* with energy E in phase space. For a chaotic system, these are the *whole energy surface* $H(\boldsymbol{x}) = E$, with dimension $N-1$, and the individual *periodic orbits* on it, with dimension 1. (For a completely or partially integrable system there are also invariant N-tori, which play an important and well-understood part in the quantization of such systems [1,38,39], but which we will not consider here because the main focus of our interest is chaos.)

Previously [1, 40] (see also [120, 41]) I thought that the energy surface plays the most important part in determining the spectrum, and the periodic orbits are relatively insignificant. The argument was that quantum states, being in some sense averages, ought to correspond to generic invariant sets – those which are explored by almost all the classical orbits – and for chaotic systems these are precisely the energy surfaces. By contrast, the periodic orbits are of measure zero and so should be less important. With this hypothesis it is possible to predict the average ('Weyl') spectral density [27], and some interesting properties of the states (average local spatial probability density, behaviour near classical boundaries, and spatial autocorrelation of the wavefunction [40,41]). And the picture of a spectrum dominated by the energy surface has some theoretical support in theorems by Shnirelman [42] and Colin de Verdière [43] which state that the semiclassical limit of the expectation values of an operator is almost always the microcanonical average of the classical function corresponding to the operator. Moreover, the effects of periodic orbits can be calculated, and as we shall see there is a formal sense in which they are small.

Nevertheless, the argument outlined above is misleading and in a sense wrong. It is misleading because although the individual periodic orbits are formally insignificant because their strength is of higher order in $\hbar$, these orbits influence the spectrum in striking ways (e.g., Heller's scars [28,29]) because their contributions are essentially singular functions of $\hbar$, unlike the contribution from the energy surface which is smooth. It is wrong because the periodic orbits are infinitely numerous (even for fixed E) and they contribute collectively as well as individually (one collective effect is the applicability of formulae from random-matrix theory). Many aspects of the collective contribution are still mysterious, as I will discuss in the last two lectures.

3.2. Formalism

In the unified approach to the spectrum which I am going to outline (based on [31]), a useful object is the spectral Wigner function $W(x;E,\varepsilon)$, defined as

$$W(x;E,\varepsilon) \equiv h^N \operatorname{Tr} \hat{\Delta}(E,\varepsilon)\delta(x-\hat{x})\,. \tag{3.2}$$

Here the operator $\hat{\Delta}(E,\varepsilon)$ is defined as

$$\hat{\Delta}(E,\varepsilon) = \delta_\varepsilon(E-\hat{H}) \equiv \frac{\varepsilon}{\pi}\frac{1}{\left[\left(E-\hat{H}\right)^2+\varepsilon^2\right]}\,, \tag{3.3}$$

the operator δ-function is defined by its Fourier transform, and ε is an energy smoothing. Two important values of ε will be the inner scale $\varepsilon_{\min}$, defined as the mean spacing between energy levels, i.e.,

$$\varepsilon_{\min} \equiv \frac{\hbar^N}{\int \mathrm{d}x\, \delta\left[E-H\left(x\right)\right]} \tag{3.4a}$$

and the outer scale $\varepsilon_{\max}$, defined in terms of the shortest closed orbit period by

$$\varepsilon_{\max} \equiv \frac{h}{T_{\min}}\,. \tag{3.4b}$$

W is useful because it describes both the E_n and $|n\rangle$ in ways that are amenable to semiclassical approximation. The levels E_n come from the smoothed *spectral density*

$$d(E,\varepsilon) = \sum_n \delta_\varepsilon (E - E_n) = \frac{1}{h^N} \int \mathrm{d}\boldsymbol{x}\, W(\boldsymbol{x}; E, \varepsilon)\,. \tag{3.5}$$

The function W describes the states $|n\rangle$ through their individual Wigner functions $W_n(\boldsymbol{x})$, namely [44–48],

$$\begin{aligned} W_n(\boldsymbol{x}) &= \mathrm{Tr}\,|n\rangle\langle n|\delta(\boldsymbol{x} - \hat{\boldsymbol{x}}) \\ &= \frac{1}{h^N} \int \mathrm{d}\boldsymbol{q}' \exp\left(-\frac{\mathrm{i}}{\hbar}\boldsymbol{q}'\cdot\boldsymbol{p}\right)\langle \boldsymbol{q} + \tfrac{1}{2}\boldsymbol{q}'|n\rangle\langle n|\boldsymbol{q} - \tfrac{1}{2}\boldsymbol{q}'\rangle. \end{aligned} \tag{3.6}$$

The relation is that W is a weighted average of the Wigner functions W_n of states whose energies E_n lie within ε of E:

$$W(\boldsymbol{x}; E, \varepsilon) = h^N \sum_n W_n(\boldsymbol{x})\delta_\varepsilon (E - E_n)\,. \tag{3.7}$$

As ε decreases through $\varepsilon_{\min}$, the average reveals individual states.

Several familiar quantities can be regained from W. The spatial probability density is the projection of W 'down' $\boldsymbol{p}$. And the Husimi function [49,50] can be obtained by Gaussian smoothing of W. Unlike W, this is positive-definite, but [31] the smoothing obliterates a great deal of the quantum detail we seek to understand.

If we would try to obtain a semiclassical approximation for W directly, we would run straight into the clash of limits $\hbar \to 0$, $t \to \infty$. The reason is that W involves, by construction, the infinitely long-lived, stationary, states. We can eliminate this difficulty – or at least delay it until we have some explicit formulae that will show more clearly what is involved – by representing W as an integral over time, i.e., as

$$W(\boldsymbol{x}, E, \varepsilon) = \frac{2}{\hbar}\,\mathrm{Re} \int_0^\infty \mathrm{d}t \exp\left[-\frac{\mathrm{i}}{\hbar}t(E - \mathrm{i}\varepsilon)\right] K_\mathrm{W}(\boldsymbol{x}, t)\,. \tag{3.8}$$

Here K_W is the Wigner propagator, that is the Weyl representation of the time-evolution operator

$$\begin{aligned} K_\mathrm{W}(\boldsymbol{x}, t) &\equiv h^N\,\mathrm{Tr}\exp\left(-\mathrm{i}\frac{\hat{H}t}{\hbar}\right)\delta(\boldsymbol{x} - \hat{\boldsymbol{x}}) \\ &= \int \mathrm{d}\boldsymbol{q}' \exp\left(-\frac{\mathrm{i}}{\hbar}\boldsymbol{q}'\cdot\boldsymbol{p}\right) K\left(\boldsymbol{q} - \frac{1}{2}\boldsymbol{q}', \boldsymbol{q} + \frac{1}{2}\boldsymbol{q}', t\right), \end{aligned} \tag{3.9}$$

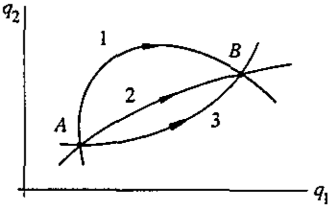

Figure 3.1.

involving the familiar coordinate propagator

$$K(q_A, q_B, t) = \left\langle q_B \left| \exp\left(-\mathrm{i}\frac{\hat{H}t}{\hbar}\right) \right| q_A \right\rangle . \tag{3.10}$$

3.3. Spectral asymptotics: principles

The asymptotics of the Wigner propagator (3.9) can be obtained [31] from those of the coordinate propagator (3.10). For fixed t there is no clash of limits, and, as is well-known [51, 15], the semiclassical limit of K depends only on the classical paths from position q_A to position q_B in time t (fig. 3.1). The phase of each contribution is the classical action divided by $\hbar$, namely

$$\frac{1}{\hbar}\left[\int_A^B p \cdot \mathrm{d}q - H(x^*)\,t\right] . \tag{3.11}$$

where x^* is any point on the path. Obviously these are not the paths contributing to K_W, because this function is specified by a single phase-space point rather than two positions. To find which paths are important now, we determine K_W by evaluating the integral in eq. (3.9) using the method of stationary phase, as is appropriate semiclassically.

The phase to be made stationary is (3.11) minus $q' \cdot p/\hbar$. An elementary calculation [31] gives the condition

$$\tfrac{1}{2}(x_A + x_B) = x . \tag{3.12}$$

Thus we have the *midpoint rule* (fig. 3.2): the paths that contribute to K_W are those linking x_A to x_B in time t which have x as midpoint. Elementary geometry gives the phase of the contribution as

$$\frac{1}{\hbar}\left[\mathcal{A}(x,t) - H(x^*)t\right] , \tag{3.13}$$

where $\mathcal{A}$ is the phase-space (symplectic) area [52] enclosed between the classical path and the chord between A and B. Similar midpoint and chord rules occur in other Wigner functions [53, 54, 8].

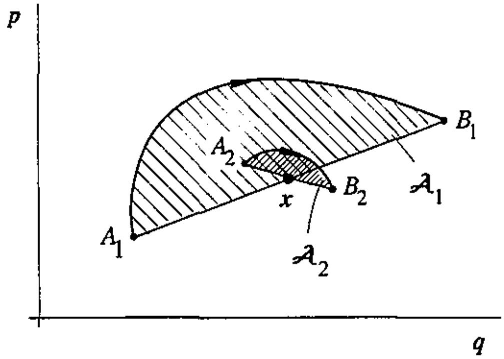

Figure 3.2.

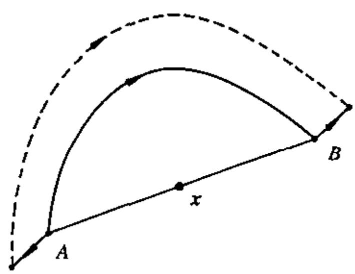

Figure 3.3.

A complete formula for K_W must include an amplitude for each path. We can expect this to depend on the $2N \times 2N$ stability matrix

$$\mathbf{m} \equiv \frac{\mathrm{d}\boldsymbol{x}_B}{\mathrm{d}\boldsymbol{x}_A} \tag{3.14}$$

and be singular whenever $\mathbf{m}$ has an eigenvalue -1, because then the mid-point rule is also satisfied for some first-order variations away from $\boldsymbol{x}_A$, $\boldsymbol{x}_B$ (fig. 3.3): these singularities are the 'Wigner caustics'. Moreover, in the short-time limit, when only one path contributes, K_W must reduce to the classical function corresponding to the time-evolution operator, namely [see eq. (3.9)],

$$K_W(\boldsymbol{x}, t) \approx \exp\left(-\frac{\mathrm{i}H(\boldsymbol{x})t}{\hbar}\right) \quad \text{as} \quad t \to 0\,. \tag{3.15}$$

These requirements are satisfied by the following formula, obtained [31] by a more careful evaluation of the integral in eq. (3.9):

$$K_W(\boldsymbol{x}, t) \approx 2^N \sum_j \frac{\exp\{(\mathrm{i}/\hbar)\,[A_j(\boldsymbol{x}, t) - H(\boldsymbol{x}^*(\boldsymbol{x}, t))t] + \mathrm{i}\gamma_j\}}{\{\det\,[\mathbf{m}_j(\boldsymbol{x}, t) + \mathbf{I}]\}^{1/2}}\,. \tag{3.16}$$

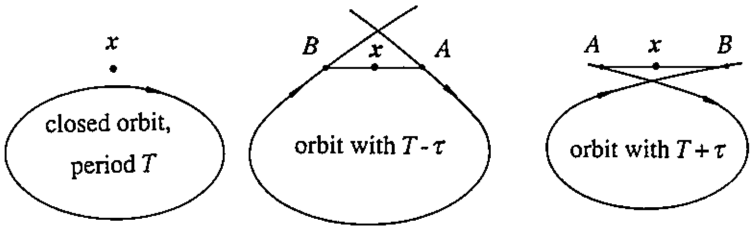

Figure 3.4.

Here j labels the different contributing paths, and γ_j is a phase counting the Wigner caustics between 0 and t ; henceforth the same symbol γ_j will denote all such focal phases.

To get the semiclassical spectral Wigner function W from eq. (3.8). we have to take the Fourier transform of eq. (3.16). There are two sorts of contribution. First is the neighbourhood of the origin $t = 0$. For this, we substitute the limiting form (3.15), and obtain

$$W(\boldsymbol{x}; E.\varepsilon) \sim \delta_\varepsilon[E - H(\boldsymbol{x})] , \tag{3.17}$$

giving the contribution from very short paths linking neighbouring points A and B (fig. 3.2) close to the energy surface E. This is the simple microcanonical classical limit of W, in which the states near E are spread uniformly over the energy surface. There is no dependence on $\hbar$. The corresponding spectral density, obtained from eq. (3.5). is the average

$$d(E; \varepsilon) \sim \bar{d}(E) \equiv \frac{1}{h^N} \int \mathrm{d}\boldsymbol{x}\, \delta[E - H(\boldsymbol{x})] . \tag{3.18}$$

This is the 'Weyl rule' [55] corresponding to the old idea that each quantum state occupies a phase-space volume h^N. Setting $\varepsilon = 0$ and integrating. we obtain a useful approximate quantization formula: the number $\mathcal{N}(E)$ of states with energy less that E (i.e., the spectral staircase) equals $n + 1/2$ for energy E_n. i.e..

$$\mathcal{N}(E_n) = n + \frac{1}{2} \sim \frac{1}{h^N} \int \mathrm{d}\boldsymbol{x}\, \Theta\, [E_n - H(\boldsymbol{x})] \tag{3.19}$$

where Θ is the unit step function.

The second sort of contribution will be the focus of all our subsequent interest. It comes from those finite t for which the phase in eq. (3.8) is stationary. It is not hard to show that these t correspond to classical paths of finite length, with energy E, linking points A and B with midpoint $\boldsymbol{x}$ (see fig. 3.2). There are such contributions to W from all $\boldsymbol{x}$, even far from the energy surface. However, points $\boldsymbol{x}$ which are close to the energy surface give particularly strong contributions because they correspond to the degenerate stationary points of the t integral in (3.8), and points which are in addition close to periodic orbits give distinctive patterns of oscillation. Figure 3.4 illustrates how the degeneracy happens: the closed orbit has period T, and the two nonperiodic orbits through a nearby point $\boldsymbol{x}$ take times $T - \tau$ and $T + \tau$, where $\tau \to 0$ as $\boldsymbol{x}$ moves onto the orbit. These closed-orbit contributions to W are the *Wigner scars*, the phase space manifestation of the configuration-space scars discovered by Heller [28,29] in the stadium billiard. Thus we arrive at the following semiclassical representation:

$$W(\boldsymbol{x}; E, \varepsilon) \approx \delta_\varepsilon \left[E - H(\boldsymbol{x})\right] + \sum_j W^j_{\text{scar}}(\boldsymbol{x}; E, \varepsilon), \qquad (3.20)$$

in which j labels all closed orbits (including repetitions) with energy E.

3.4. Spectral asymptotics: formulae

In [31] can be found the derivation of the formula for the contribution W_{scar} from points $\boldsymbol{x}$ near a closed orbit. Here I present the formula, and then explain its ingredients:

$$\begin{aligned} W_{\text{scar}}(\boldsymbol{x}; E, \varepsilon) = {} & \exp\left(-\frac{\varepsilon T}{\hbar}\right) \\ & \times \frac{2^N}{\sqrt{\det(\mathbf{M} + \mathbf{I})}} \cos\left[\frac{1}{\hbar}\left(S + \tilde{X}\mathbf{J}\frac{\mathbf{M}-\mathbf{I}}{\mathbf{M}+\mathbf{I}}X\right) + \gamma\right] \\ & \times \frac{2}{(\hbar^2|\dot{\boldsymbol{x}} \times \ddot{\boldsymbol{x}}|)^{1/3}} \mathrm{Ai}\left\{\frac{2\left[H(x) - E\right]}{\left(\hbar^2 \dot{\boldsymbol{x}} \times \ddot{\boldsymbol{x}}\right)^{1/3}}\right\}. \end{aligned} \qquad (3.21)$$

It is convenient to discuss the three factors separately. We shall consider only the typical case, where the periodic orbits are isolated on each energy surface but exist for a range of energy.

In the first factor, T is the period of the closed orbit. If the smoothing exceeds $\varepsilon_{\max}$ [eq. (3.4)] all scars are damped out, and the spectral Wigner function and spectral density are given by the simple classical formulae (3.17) and (3.18). As ε is reduced, longer orbits contribute, until when $\varepsilon = 0$ all orbits must be considered.

The second and third factors describe the variation of W as $\boldsymbol{x}$ deviates from the closed orbit, and as $\boldsymbol{x}$ moves off the energy surface E (fig. 3.5).

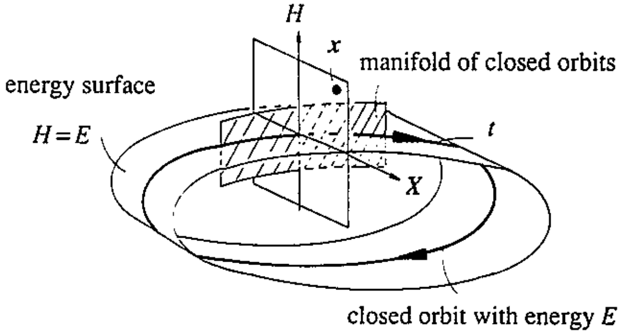

Figure 3.5.

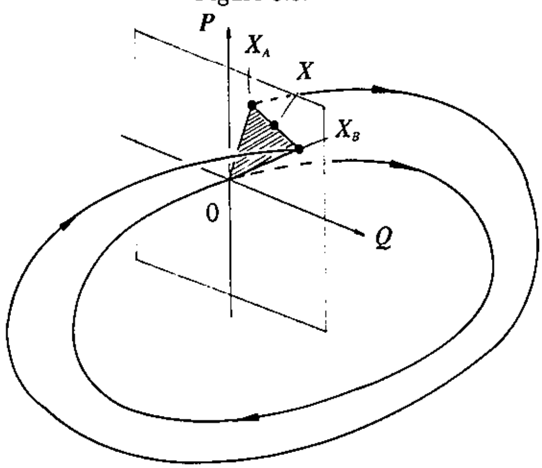

Figure 3.6.

It is convenient to employ phase-space variables

$$x = \{H, t, X \equiv (Q, P) = (Q_1, \ldots, Q_{N-1}, P_1, \ldots, P_{N-1})\}\ , \tag{3.22}$$

where H is the Hamiltonian, X are $2N-2$ Poincaré section coordinates, and t is time along orbits, measured from the section.

The second factor involves the action

$$S = \oint p \cdot \mathrm{d}q\ , \tag{3.23}$$

of the periodic orbit, and the $(2N-2) \times (2N-2)$ stability matrix M of the linearized Poincaré map for successive intersections X_A, X_B with the section (fig. 3.6), defined by

$$X_B = \mathsf{M} X_A; \tag{3.24}$$

$\mathbf{J}$ is the unit symplectic matrix [52]. The quadratic form in the cosine is the symplectic area $\mathcal{A}$ (shaded in fig. 3.6). Thus, the second factor in eq. (3.21) describes quadratic fringes as $\boldsymbol{x}$ moves off the closed orbit, whose width is of order $\hbar^{1/2}$ near the orbit, and of order $\hbar$ far from the orbit. The pattern of fringes varies along the orbit, because $\mathbf{M}$ depends on time t along the orbit ($0 \leqslant t \leqslant T/k$, where k is the number of repetitions of the primitive orbit), but its amplitude does not (because $\det(\mathbf{M}+\mathbf{I})$ does not depend on t).

The third factor involves the phase-space velocity and acceleration on the orbit at $\boldsymbol{x}$, in a combination whose geometric significance has been explained by Balazs [56]. It describes Airy fringes, whose width is of order $\hbar^{2/3}$, as $\boldsymbol{x}$ moves off the energy surface (see also [8]).

Note that the quantities T, S and $\mathbf{M}$ in eq. (3.21) all depend on the energy E.

Each of the three factors in eq. (3.21) is essentially singular at $\hbar = 0$, in ways we have met before: an exponential singularity, in the first factor (cf. 2.4): an oscillatory singularity (cf. 2.2) and an Airy singularity (cf. 2.6). There are several ways to take the classical limit of eq. (3.21) to clarify its content. The most elementary, that is $\hbar \to 0$ with fixed smoothing ε, simply obliterates all the scars. This effect can be avoided by making ε vanish along with $\hbar$, for example by choosing ε between $\varepsilon_{\min}$ and $\varepsilon_{\max}$ [eq. (3.4)].

Alternatively, we can consider both the Airy and cosine fringes as too fine to resolve (cf. 2.3), in which case we can use the limit

$$\frac{1}{\varepsilon}\mathrm{Ai}\left(\frac{x}{\varepsilon}\right) \to \delta(x) \quad \text{as} \quad \varepsilon \to 0\,, \tag{3.25}$$

to obtain

$$W_{\text{scar}}(\boldsymbol{x}; E, \varepsilon) = \frac{2h^{N-1}}{\sqrt{\det(\mathbf{M}-\mathbf{I})}} \exp\left(-\frac{\varepsilon T}{\hbar}\right) \times \cos\left(\frac{S}{\hbar} + \gamma\right) \delta[E - H(\boldsymbol{x})]\delta(\boldsymbol{X})\,. \tag{3.26}$$

With eq. (3.20) this gives the picture of W as a distribution painted uniformly on the energy surface, with the scars drawn as sharp lines of strength $\hbar^{N-1}$ along the periodic orbits. This is the sense in which the scars are weak.

Another possibility is to regard the cosine fringes ($\sim \hbar^{1/2}$) as resolvable, but not the smaller Airy fringes ($\sim \hbar^{2/3}$). This leads to

$$W_{\rm scar}(\boldsymbol{x};E,\varepsilon) = \frac{2^N}{\sqrt{\det(\mathsf{M}+\mathsf{I})}} \exp\left(-\frac{\varepsilon T}{\hbar}\right) \times \cos\left[\frac{1}{\hbar}\left(S + \tilde{X}\mathsf{J}\frac{\mathsf{M}-\mathsf{I}}{\mathsf{M}+\mathsf{I}}X\right) + \gamma\right] \delta[E - H(\boldsymbol{x})]. \quad (3.27)$$

Now we see that relative to the ergodic background the scar height is independent of $\hbar$ and dependent only on the stability of the orbit (through M), a conclusion previously reached by Heller [28,29].

By projecting the scar formula (3.21) 'down' $\boldsymbol{p}$, it is possible to obtain the formula of Bogomolny [30] for the scars on the probability density in configuration space. By further projecting 'across' $\boldsymbol{q}$, we obtain the contribution of the periodic orbit to the spectral density given in eq. (3.5) [actually the calculation is easier in the variables (3.22)]:

$$d_{\rm scar}(E) = \frac{T}{k\pi\hbar\sqrt{\det(\mathsf{M}-\mathsf{I})}} \exp\left(-\frac{\varepsilon t}{\hbar}\right) \cos\left(\frac{S}{\hbar} + \gamma\right). \quad (3.28)$$

Here k is the number of repetitions of the orbit. This is the celebrated formula of Gutzwiller [57, see also 58], about which I shall say a great deal later.

4. Long orbits (asymptotics of asymptotics)

4.1. Gaussian random waves?

From eqs. (3.20) and (3.21) we see that the spectrum of eigenvalues and eigenfunctions is determined by a 'background' contribution from the whole energy surface, and oscillatory decorations from the periodic orbits. The short periodic orbits combine to give the striking patterns observed by Heller [28,29]. The long periodic orbits can be regarded as individually weak, but they are infinitely numerous, so their collective effect cannot be ignored.

Consider first the spectral Wigner function on phase space. As we shall discuss in more detail in the next section, the long orbits ($T \to \infty$) of a chaotic system explore the energy surface in a way that is asymptotically uniform. But it is not the case that the scars get more concentrated as T increases; in the second factor of eq. (3.21) M stretches along its eigendirections and cancels from the cosine, giving limiting fringes with always the same phase-space width: $\hbar^{1/2}$ near the orbit and $\hbar$ far from the orbit [31]. Therefore, the scars of the long orbits overlap.

At present we have no idea how to calculate the effect of these overlapping scars and so can only speculate. Because of the asymptotic uniformity of the long orbits, it might appear that the collectivity of scars would cover the energy surface uniformly. But we already have such a contribution, in the form of the first, ergodic, term of eq. (3.20): a second, similar, contribution would be an embarrassment. A more reasonable possibility, in view of the fact that the collectivity is the superposition of many oscillations, is that the separate phases [$S + \gamma$ in eq. (3.21)] can be regarded as random, so that the resultant is a *Gaussian random function* decorating the smooth background. This would project to give Gaussian random fluctuations on the scale of the de Broglie wavelength ($\sim \hbar$) in the spatial probability density, in accordance with my old conjecture [53], for which support, while not overwhelming [61] is not nonexistent either [62].

4.2. Phase space democracy

The contribution of eq. (3.28) to the spectral density from a periodic orbit oscillates with energy, because the action S is a function of E. Now $\mathrm{d}S/\mathrm{d}E =$ orbit period T, so the wavelength of the energy oscillation is

$$\varepsilon(T) = \frac{h}{T} \tag{4.1}$$

(this can be regarded as an expression of the energy-time uncertainty relation). The wavelength of the largest oscillation, which comes from the shortest orbit $T_{\min}$, is our outer scale $\varepsilon_{\max}$ [see eq. (3.4)]. Longer orbits give finer energy fluctuations (contrasting with the phase space oscillations, which for long orbits have a constant limiting wavelength). Thus, the orbits that contribute to the spectrum on the finest scale, corresponding to the mean level separation, eq. (3.4a), [compare eq. (3.18)] have period

$$T_{\max} = \frac{1}{h^{N-1}} \int \mathrm{d}x\, \delta[E - H(x)]\,. \tag{4.2}$$

Semiclassically, these are *very long orbits*. This means that any attempt to describe the spectrum on scales corresponding to a given number of mean level spacings – e.g., any study of level statistics (see section 4.3), which because it involves large numbers of excited states is inevitably semiclassical – requires a knowledge of the Gutzwiller terms, eq. (3.28), for large T. Alternatively stated, the need for this 'asymptotics of asymptotics' comes

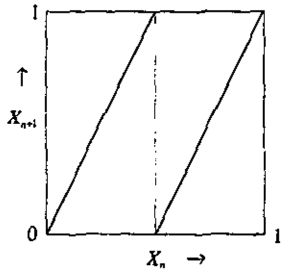

Figure 4.1.

from choosing the energy scale at a fixed number of mean level spacings because this requires continued magnification in energy as $\hbar \to 0$ [27].

Two aspects of the long orbits are important. The first is that for a chaotic system their number proliferates exponentially with period T, the law being [63],

$$\begin{aligned}\rho(T) &\equiv \frac{\text{number of orbits with periods } T \text{ to } T+\mathrm{d}T}{\mathrm{d}T} \\ &\approx \frac{\exp(h_\mathrm{T} T)}{T} \quad \text{as} \quad T \to \infty \end{aligned} \tag{4.3}$$

where h_T is the topological entropy (mixing rate).

I will not give a general proof of this law, but will illustrate it by calculating the orbit density for the 'doubling' map on the unit interval $0 \leqslant x \leqslant 1$, namely

$$x_{n+1} = 2x_n \bmod 1 \tag{4.4}$$

(fig. 4.1). This is uniformly hyperbolic, with invariant measure unity and entropy $h_\mathrm{T} = \log 2$. The evolution equation (4.4) is equivalent to moving the point one place to the right in the binary representation of x_n, so that the whole orbit can be read off from the binary digit string representing x_0. Periodic orbits are periodic digit strings (representing rational x_0). The number of digit strings with period n is 2^n. To get the number of distinct orbits, it is necessary to divide by n, because sequences of the same n digits but with different starting-points correspond to the same orbit. Thus

$$\text{number of orbits with period } n = \frac{2^n}{n} = \frac{\exp(n \log 2)}{n} = \frac{\exp(n h_\mathrm{T})}{n}. \tag{4.5}$$

Reinstating time (for example by letting n correspond to $T = n\tau$), we immediately obtain eq. (4.3). (In this argument, we have ignored the fact that some periodic orbits are repetitions of shorter ones. but these are a negligible fraction of the total as $n \to \infty$.) For a different illustration of eq. (4.3) (the Sinai billiard) see appendix H of [64].

The second important aspect is that the amplitudes of the long orbit contributions decay exponentially with T. We write Gutzwiller's series in the form

$$d(E:\varepsilon) \approx \bar{d}(E) + \frac{1}{\hbar}\sum_j A_j(E)\exp\left[-\frac{\varepsilon T_j(E)}{\hbar}\right]\exp\left\{\mathrm{i}\left[\frac{S_j(E)}{\hbar}+\gamma_j\right]\right\} \tag{4.6}$$

where $\bar{d}$ is the mean spectral density given by eq. (3.18) and the sum now includes negative as well as positive traversals, and the amplitude is [see eq. (3.28)]

$$A_j = \frac{T_j}{k_j\sqrt{\det(\mathbf{M}_j - \mathbf{I})}} . \tag{4.7}$$

Almost all long orbits are not repetitions of shorter ones. so we can set $|k_j| = 1$. Long orbits explore the whole energy surface uniformly, so the $N-1$ expanding eigenvalues λ_m (> 1) of their Poincaré map matrices $\mathbf{M}$ generate the entropy,

$$\begin{aligned}|\det(\mathbf{M}-\mathbf{I})| &= \left|\prod_1^{N-1}(\lambda_m - 1)\left(\frac{1}{\lambda_m}-1\right)\right|_{T\to\infty}\\ &\approx \exp\left(\sum_1^{N-1}\log\lambda_m\right) = \exp\left(h_{\mathrm{KS}}T\right) ,\end{aligned} \tag{4.8}$$

where h_{KS} is the Kolmogorov–Sinai entropy. Thus. on the average for orbits with period near $T \gg T_{\min}$, we can write the Gutzwiller amplitude as a function of T:

$$A(T) = T\exp\left(-\tfrac{1}{2}h_{\mathrm{KS}}T\right) . \tag{4.9}$$

If we assume that (after averaging over orbits) the two entropies are equal. then we can combine the proliferating number and decaying amplitudes into a sum rule for the Gutzwiller amplitudes:

$$\sum_j A_j^2\delta(T-T_j) \to A^2(T)\rho(T) = T . \tag{4.10}$$

This is a classical sum rule, because it does not involve $\hbar$. It was obtained by Hannay and Ozorio de Almeida [65, 66] as a direct consequence of the uniformity of exploration of phase space ('democracy') by long closed orbits, without invoking equality of entropies. For simplicity, I will give their argument as it applies to maps rather than flows. Let $\boldsymbol{X}_n(\boldsymbol{X}_0)$ denote the iterates of $\boldsymbol{X}_0$ under a mixing transformation of a phase space with volume Ω. The totality of many iterates of an initial δ-function will form a distribution that in a coarse-grained sense covers the space uniformly. Thus,

$$\frac{1}{N}\sum_{1}^{N}\delta[\boldsymbol{X}-\boldsymbol{X}_n(\boldsymbol{X}_0)]\to\frac{1}{\Omega}\quad\text{as}\quad N\to\infty\,. \tag{4.11}$$

This applies to any $\boldsymbol{X}$ and $\boldsymbol{X}_0$; in particular, we can choose $\boldsymbol{X}=\boldsymbol{X}_0$, in which case

$$\delta[\boldsymbol{X}-\boldsymbol{X}_n(\boldsymbol{X}_0)]=\sum_j\frac{\delta(\boldsymbol{X}_0-\boldsymbol{X}_{jn})}{\det(\mathsf{M}_j-\mathsf{I})}\,, \tag{4.12}$$

where the sum is over points $\boldsymbol{X}_{jn}$ of periodic orbits with period n and where M_j is the stability matrix $\partial\boldsymbol{X}_n/\partial\boldsymbol{X}_0$, evaluated at $\boldsymbol{X}_{jn}$. Now we substitute into eq. (4.11) and integrate over $\boldsymbol{X}_0$. Because the transformation is mixing, the totality of periodic points of any long period N covers the space uniformly. Therefore, we can omit the average over n in eq. (4.11), to get

$$N\sum_j\frac{1}{\det(\mathsf{M}_j-\mathsf{I})}\to 1\quad\text{as}\quad N\to\infty\,, \tag{4.13}$$

where now j labels periodic *orbits* with period N (each orbit has N periodic points – hence the extra factor N). Multiplying both sides by $N\tau$, where τ is the time for one iteration, and recalling eq. (4.7), we obtain the discrete-time version of eq. (4.10) [note that $N\tau=T$, and the extra factor $1/\tau$ corresponds to the δ-function in eq. (4.10)].

An important feature of the Hannay–Ozorio classical sum rule, eq. (4.10), for long orbits is its universality: the cancellation of exponentials representing decay and proliferation leaves a formula containing no trace of any particular dynamics. It depends only on the fact that the motion is chaotic. By contrast, the Gutzwiller amplitudes, eq. (4.7), for short orbits are not universal; these orbits do not participate in phase-space democracy, and their periods and stabilities are peculiar to them.

4.3. Energy-level statistics

As will be discussed later, it is difficult to get a semiclassical quantization condition for the energy levels E_n from the Gutzwiller series of eq. (4.6), or indeed by any other means – at least for the classically chaotic systems we are concentrating on here (for integrable systems, see [1,38,39,67]). One response to this inability to penetrate the details of quantum-to-classical correspondence has been to study the statistics of the levels rather than their individual energies. This is appropriate semiclassically because as $\hbar \to 0$ infinitely many levels crowd into any small fixed energy range.

The simplest statistic is the mean level density, which as we have seen is given by eq. (3.18). This gives a useful first view of the spectrum, particularly when supplemented by smooth (i.e., analytic in $\hbar$) corrections where these are known, as is the case for billiards [55] (for which the corrected mean density has become a standard means to check the completeness of computed sets of levels [64,68,69]).

Much more interesting, however, are the statistics describing *fluctuations* about the mean level density. These are conveniently studied by multiplying the E_n in the energy range of interest by the mean level density, to obtain a set of scaled levels e_n whose mean density is unity. Sometimes this magnification (whose strength is of order $\hbar^{-N}$) is called 'unfolding the spectrum'. The remarkable result of many numerical studies is that the statistics of the e_n, i.e., the fluctuation statistics, fall into universality classes, depending only on broad features of the classical dynamics. If the motion is integrable, the e_n are Poisson-distributed [67]. If the motion is chaotic, the statistics of the e_n are those of the eigenvalues of *random matrices* [70] chosen from ensembles [26] reflecting the presence or absence of time-reversal symmetry [68] or the Bosonic or Fermionic nature of the system [71]. There is some subtlety about this question of symmetry [72], and interesting complications can arise if the system is partly integrable and partly chaotic [73,74]. Here we seek to understand the simplest case, where the classical motion is wholly chaotic and there is no symmetry at all. Then the appropriate random-matrix theory is that of the Gaussian unitary ensemble (GUE), consisting of complex Hermitian matrices whose elements are independently Gauss-distributed apart from the constraint that the ensemble is invariant under unitary transformations in Hilbert space.

For reasons already explained, spectral fluctuations are inevitably semiclassical. They are described by the nonanalytic corrections, arising from the closed orbits, in the Gutzwiller series [eq. (4.6)]. Recalling the discussion following eq. (4.2), we expect spectral statistics involving finite numbers of the e_n to depend on the long orbits. We know, however, that these orbits display universality, in the form of the classical sum rule (4.10). Therefore, it is not surprising that the quantum level statistics exhibit some reflection of this classical universality. On the other hand, the classical universality does not hold for short orbits, so by the duality embodied in eq. (4.1) we expect the quantum universality to break down for level statistics involving e_n so distant as to correspond to energy ranges of order $\hbar$ (rather than $\hbar^N$ as for nearby e_n).

Elsewhere [75], I have given the detailed explanation of this short-range universality and its breakdown at long range, in the context of a particular statistic, the spectral rigidity of Dyson and Mehta [76]. However, the theory works for any statistic that is bilinear in the spectral density, and here I choose to illustrate it with the number variance $\Sigma^2(L)$. This is defined as

$$\Sigma^2(L) = \left\langle [n(L) - L]^2 \right\rangle , \tag{4.14}$$

where $n(L)$ is the number of levels e_n in an e-interval of length L (in which the expected number is of course L), and the brackets $\langle \ldots \rangle$ denote an average over many such e-intervals near the energy E of interest. This statistic is easy to compute from a sequence of eigenvalues: the longer the sequence, the larger is the value of L that can be studied. We expect breakdown of universality when $L > L_{\max}$, where $L_{\max}$ corresponds to the energy range $\varepsilon_{\max}$ of eq. (3.4), i.e.,

$$L_{\max} = \frac{\varepsilon_{\max}}{\varepsilon_{\min}} = \frac{\bar{d}h}{T_{\min}} \sim \frac{1}{\hbar^{N-1}} . \tag{4.15}$$

To calculate Σ^2 we first express $n(L)$ in terms of the e spectral density (whose mean value is unity). If the centre of the e-interval is e_c, then

$$n(L) - L = \int_{e_c - L/2}^{e_c + L/2} \mathrm{d}e \, [d(e) - 1] . \tag{4.16}$$

At this point, we could substitute the Gutzwiller series (4.6), but it is convenient to make a further transformation and introduce the *form factor*

$$K(\tau) \equiv \int_{-\infty}^{\infty} \mathrm{d}\xi \exp(2\pi \mathrm{i} \xi \tau) \left\langle \left[d\left(e - \tfrac{1}{2}\xi\right) - 1\right] \left[d\left(e + \tfrac{1}{2}\xi\right) - 1\right] \right\rangle . \tag{4.17}$$

Fourier inversion and substitution into eqs. (4.14) and (4.16) leads to

$$\Sigma^2(L) = \frac{2}{\pi^2} \int_0^{\infty} \mathrm{d}\tau \, \frac{K(\tau)}{\tau^2} \sin^2(\pi L \tau) . \tag{4.18}$$

Now, we shall obtain a semiclassical representation of $K(\tau)$. The first step is to substitute the semiclassical series (4.6) into each factor of the definition (4.17) and use

$$d(e) = \frac{d(E)}{\bar{d}(E)}. \tag{4.19}$$

The mean level density cancels, leaving only the periodic orbit sum. Each term involves the action, which for the purposes of integration in eq. (4.17) we expand as

$$S\left[E\left(e \pm \tfrac{1}{2}\xi\right)\right] \approx S(E) \pm \frac{\xi T(E)}{2\bar{d}}. \tag{4.20}$$

A short calculation then gives the semiclassical spectral form factor as the double sum over orbits

$$K(\tau) \sim \frac{1}{h\bar{d}} \left\langle \sum_i \sum_j A_i A_j \exp\left[\frac{\mathrm{i}}{\hbar}(S_i - S_j)\right] \delta[T - \tfrac{1}{2}(T_i + T_j)] \right\rangle, \tag{4.21}$$

where T is a scaled version of the variable τ, with the dimensions of time:

$$T \equiv \tau h \bar{d}. \tag{4.22}$$

A natural approximation now is to implement the energy average $\langle \ldots \rangle$ by replacing the double sum in eq. (4.21) by its diagonal terms $i = j$, on the grounds that the energy dependence of $S_i - S_j$ will wash out the off-diagonal terms by phase incoherence. This gives

$$K(\tau) \sim K_\mathrm{D}(\tau) = \frac{1}{h\bar{d}} \sum_j A_j^2 \delta(T - T_j). \tag{4.23}$$

Thus we see that $K(t)$ has δ-spikes at the periods of closed orbits. From eq. (4.22), the first spike, corresponding to $T_{\min}$, occurs at a very small τ-value, of order $\hbar^{N-1}$. Larger τ correspond to long orbits, for which we can use the classical sum rule (4.10), to obtain

$$K_\mathrm{D}(\tau) \approx \tau, \quad \text{if} \quad \tau \gg \frac{T_{\min}}{hd}. \tag{4.24}$$

Plausible though this random-phase argument is, it cannot hold for arbitrarily large τ, one reason being that it would cause the form factor (4.18)

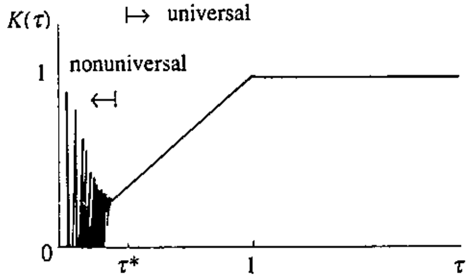

Figure 4.2.

to diverge. The replacement of K by K_D fails because the exponential proliferation of very long orbits means that in the double sum (4.21) there are pairs of different orbits with energy E whose action difference is much less than $\hbar$, and these are not washed out by incoherence. An argument I wish to postpone until section 5 shows that the double sum obeys the *semiclassical sum rule*

$$K(\tau) \to 1 \quad \text{if} \quad \tau \gg 1 . \tag{4.25}$$

Note that according to eq. (4.2) $\tau = 1$ represents a very long period T that corresponds, through the duality (4.1), to the mean level spacing.

These results still leave the transition between eqs. (4.24) and (4.25) unresolved. The simplest choice [75] is that the transition is abrupt. Then, we can synthesize the three limiting results just obtained into

$$\begin{aligned} K(\tau) &\approx \frac{1}{h\bar{d}} \sum_j A_j^2 \delta(T - T_j) \quad \text{if} \quad \tau < \tau^* , \quad \text{where} \quad \frac{T_{\min}}{h\bar{d}} \ll \tau^* \ll 1 . \\ &\approx \tau \quad \text{if} \quad \tau^* < \tau < 1 , \\ &\approx 1 \quad \text{if} \quad \tau > 1. \end{aligned} \tag{4.26}$$

Figure 4.2 is a sketch of the graph of this formula. The last two members, for $\tau > \tau^*$, constitute exactly the form factor of the GUE [77]. In a sense this is the simplest semiclassical limit of $K(\tau)$, because as $\hbar \to 0$ we can let $\tau^* \to 0$. However, the importance of the tiny range $0 < \tau < \tau^*$ will soon become apparent.

For the number variance, evaluation of eq. (4.18) with this form factor gives

$$\Sigma^2(L) \approx \frac{1}{\pi^2}\left[\log(2\pi L) - \mathrm{Ci}(2\pi L) - 2\pi L\,\mathrm{Si}(2\pi L) + \pi^2 L - \cos(2\pi L) + 1 + \gamma_{\mathrm{E}}\right] + \frac{1}{\pi^2}\left[2\sum_{T_j < h\bar{d}\tau^*} \frac{A_j^2}{T_j^2}\sin^2\left(\frac{LT_j}{2\hbar\bar{d}}\right) + \mathrm{Ci}(2\pi L\tau^*) - \log(2\pi L\tau^*) - \gamma_{\mathrm{E}}\right], \tag{4.27}$$

where Ci and Si are the sine and cosine integrals [19] and γ_{E} is the Euler constant. The value of Σ^2 does not depend on the cutoff τ^*. provided this lies in the range specified in eq. (4.26). The terms in the first set of braces dominate if $L \ll L_{\max}$ [given by eq. (4.15)]. They constitute precisely the number variance of the GUE. Here. we have obtained this quantum universal formula by purely semiclassical arguments: random-matrix theory played no part. The terms in the second set of brackets are not universal. for they depend on the individual short-period orbits; they become important when $L > L_{\max}$ (a more accurate value for the transition is $L_{\max}/2\pi$), which we earlier obtained as the limit of universality by uncertainty-principle duality arguments. Analyzing eq. (4.27) we find three regimes of L. i.e..

$$\begin{aligned}\Sigma^2(L) &\approx L \quad \text{if} \quad L \leqslant 1\\ &\approx \frac{1}{\pi^2}\left[\log(2\pi L) + 1 + \gamma_{\mathrm{E}}\right] \quad \text{if} \quad 1 \ll L \ll L_{\max}\\ &\approx \frac{1}{\pi^2}\left(\sum_{T_j < h\bar{d}\tau^*} \frac{A_j^2}{T_j^2} + 1 - \log\tau^* + \text{oscillations}\right) \quad \text{if} \quad L \gg L_{\max}\,.\end{aligned} \tag{4.28}$$

Figure 4.3 is a sketch of $\Sigma^2(L)$, showing these three regimes. Again it is true that the simplest semiclassical limit of the function $\Sigma^2(L)$ is the random-matrix formula, because $L_{\max} \to \infty$ as $\hbar \to 0$. But our eq. (4.27) has a wider range of validity; it is a uniform approximation for $\Sigma^2(L)$, valid also in the nonuniversal regime $L > L_{\max}$, corresponding to energy ranges of order $\hbar$ (i.e.. classically small but large compared with the mean level spacing $\hbar^N$).

Nobody has yet tested the variance formula (4.27), or its analogue for spectral rigidity [75], over the full range of L, universal and nonuniversal, for any individual quantum system with a classical counterpart. To do a proper

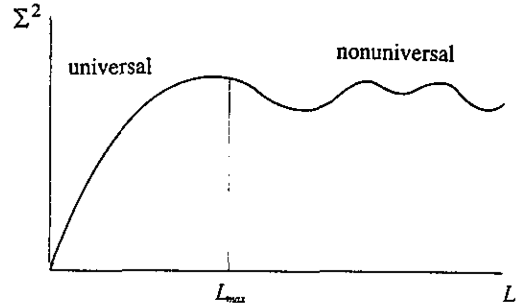

Figure 4.3.

job would require several thousands of levels. which are still not available for an appropriate system (for reasons explained in [75]. the stadium billiard is not a very suitable system). This is a technological barrier which I expect will be overcome soon. There have. however. been many confirmations [78. 68,79] in the random-matrix range $L \ll L_{\max}$. which require fewer levels. The full formula has. however. been tested on a (so far) mock quantum system – the Riemann zeros – as will be explained in the next section. The analogous formula for classically *integrable* systems [75] is. however, easy to test – because it is easy to get long level sequences – and has been tested. on the rectangle billiard. for which numerical calculations by Casati, Chirikov and Guarneri [99] first showed long-range deviation from universality (in this case of the Poisson type) and inspired the theory just outlined.

As I have said. semiclassical techniques lead to formulae for all spectral statistics that are bilinear in the level density. Other statistics, involving higher powers of the density. are proving more resistant to analysis (although the broad approach seems clear enough). even though we know some higher-order semiclassical sum rules analogous to eq. (4.25). Most difficult is the statistic that is most familiar (because it is easy to compute from a set of eigenvalues), namely the distribution $P(S)$ of spacings between neighbouring levels. This depends on level density correlations of all orders, and its semiclassical study – if possible at all [80] – would require very fine control of the very long orbits.

A completely different approach [64, 1, 81] yields the form of $P(S)$ as $S \to 0$ (i.e., S^2 when there are no symmetries, S for Bosons with time-reversal or similar [72] symmetry, S^4 for Fermions with time-reversal symmetry [71], and S^0 when the classical system is integrable [67]). This is based on the degeneracy structure of typical families of quantum systems with the appropriate symmetry, and, as well as making it possible to understand the short-range $P(S)$. led eventually to the discovery of the geometric phase [82].

Yet another approach to spectral statistics. attractive in its generality, was pioneered by Dyson [83], revived in the context of quantum chaology by Pechukas [84], and further developed by Yukawa [85] and Nakamura and Lakshmanan [86]. It considers the spectrum not of a single Hamiltonian but of a one-parameter family of Hamiltonians. Usually the parameter is taken to be the coefficient of a perturbation in the Hamiltonian, that switches on chaos. I have argued [81] that it is more appropriate to choose $\hbar$. Wilkinson [87] makes the stimulating suggestion of a parameter that explores the cloud of quantum Hamiltonians obtained by applying different quantization prescriptions to the given classical Hamiltonian (the size of the cloud is of order $\hbar^2$).

This parameter is regarded as a fictitious time variable t, and evolution equations are obtained for the levels $E_n(t)$ and certain matrix elements $M_{ij}(t)$ [84,85] or the states $|n(t)\rangle$ themselves [86]. The aim is to understand spectral statistics in terms of the statistical mechanics of this set of quantities. It was shown, however, in a remarkable development [86], that this infinite set of equations is completely integrable, making it difficult to see how the levels reach thermal equilibrium, which would correspond to random-matrix theory [83,85]. The integrability may be another manifestation of the absence of quantum chaos (section 2); although the details are surprising [86], the fact of integrability is perhaps no more surprising than the integrability of the set of coupled harmonic oscillators obtained by discretizing Schrödinger's equation on a spatial grid.

I think it will be very hard to make a proper theory of spectral statistics in terms of the dynamical system consisting of the $E_n(t)$ and $M_{ij}(t)$ or $|n(t)\rangle$. The aim seems to be to provide a justification for random-matrix theory [83-85], or to argue against the applicability of random-matrix theory [86]. In the latter argument, t is regarded as the strength of a perturbation away from classical integrability, and the integrability of the spectrum equations is invoked to support the claim that finding the spectrum for finite t is in some sense trivial. But the actual exploitation of the integrability involves operating on the initial spectrum with a time-ordered infinite matrix exponential, a process apparently equivalent in difficulty to diagonalizing the finite-t Hamiltonian in a basis of the $t = 0$ Hamiltonian.

In any case, it is obvious that quantum Hamiltonians derived from classical ones are not random matrices but have a structure that encodes the orbits of the particular classical system that generated them. This is revealed by the long-range spectral statistics which, as we have seen, are not universal. The ingredient lacking in the 'level motion' approach is semiclassical asymptotics. This is essential not only to the understanding of statistics over long as well as short ranges but also at the more elementary level of distinguishing the statistics that correspond to chaotic and non-chaotic classical motion. Alternatively stated, there is no obvious connection between the integrability of the spectrum dynamics in fictitious time (i.e., as a parameter varies) and the integrability or chaos of the underlying classical dynamics in real time.

4.4. Application(?) to the Riemann zeros

The Riemann zeta function [88] is defined as a Dirichlet series, or Euler product over primes, namely

$$\zeta(z) \equiv \sum_{1}^{\infty} \frac{1}{n^z} = \prod_{p} \frac{1}{1-p^{-z}} \quad \mathrm{Re}\, z > 1\,. \tag{4.29}$$

and by analytic continuation elsewhere in the z plane. The celebrated Riemann hypothesis states that its complex zeros all have real part $\frac{1}{2}$, so that the quantities E_n defined by

$$\zeta(\tfrac{1}{2} + \mathrm{i}E_n) = 0\,. \tag{4.30}$$

are all real. A variety of evidence, described elsewhere [75, 89] strongly suggests that the E_n are eigenvalues of a quantum Hamiltonian obtained by quantizing a classical system without time-reversal symmetry whose orbits are chaotic. We still have no idea what that system is (but an increasing knowledge of what it is not).

Here I will concentrate on the ζ analogue of the Gutzwiller sum over periodic orbits. To obtain this, we first note that whilst moving up just to the right of the critical line $\mathrm{Re}\, z = \frac{1}{2}$, the logarithm of $\zeta(z)$ increases by $\mathrm{i}\pi$ at every zero E_n. Thus, the density of zeros is

$$d(E) - \bar{d}(E) = \frac{\mathrm{d}}{\mathrm{d}E}\frac{1}{\pi}\,\mathrm{Im}\log\zeta(\tfrac{1}{2} + \mathrm{i}E) \tag{4.31}$$

(the subtraction of the mean density is required by the behaviour of ζ as $\mathrm{Re}\, z \to +\infty$). Now we substitute the Euler product (4.29), postponing discussion of convergence difficulties to section 5, and obtain

$$d(E) - \bar{d}(E) \approx -\frac{1}{2\pi}\sum_{p}\sum_{k=-\infty}^{\infty}{}' \log p \exp(-\tfrac{1}{2}|k|\log p)\exp(\mathrm{i}Ek\log p)\,, \tag{4.32}$$

where the prime denotes the absence of the term $k = 0$. The mean density is known [88] to be

$$\bar{d}(E) = \frac{1}{2\pi}\log\left(\frac{E}{2\pi}\right). \tag{4.33}$$

In the series (4.32) there is of course no $\hbar$, so in making the analogy with the Gutzwiller series (4.6) (with zero smoothing ε), the semiclassical limit $\hbar \to 0$ is replaced by $E \to \infty$, a situation familiar in other scaling systems (e.g., billiards). The analogy begins by interpreting eq. (4.32) as a sum over primitive periodic orbits, labelled by primes p, and repetitions k (positive and negative). The action of the orbit (k, p) is (setting $\hbar = 1$),

$$S(E) = Ek\log p. \tag{4.34}$$

Therefore, the period of this orbit is

$$T(E) = \frac{\mathrm{d}S}{\mathrm{d}E} = k\log p. \tag{4.35}$$

Thus we identify the amplitude A in eq. (4.6) as

$$A(E) = \log p \exp(-\tfrac{1}{2}|k|\log p) = \frac{T}{k\sqrt{\exp(|k|\log p)}}. \tag{4.36}$$

The analogy with eq. (4.6) is very close: A contains the factor T/k and although the determinantal factor looks slightly different, there is a transformation [89] that makes the resemblance closer.

Long ago, Selberg, inspired by the earlier suggestion that there might be a dynamical system underlying the Riemann zeros and frustrated by his inability to find it, devised his celebrated trace formula [90–92]. This is a formula closely resembling eq. (4.32), giving exactly the density of eigenvalues of the Laplace–Beltrami operator on compact surfaces of constant negative curvature. A natural interpretation [92] is as the quantization of the dynamical system whose orbits are the geodesics, here chaotic. Selberg's formula is an anticipation of Gutzwiller's for this special case, where it is exact rather than asymptotic. As a model for quantum chaology the Riemann zeros appear to be better than the eigenvalues of these surfaces, because all evidence points to the fact that the Riemann E_n mimic a generic quantum system, whereas the high symmetry of the compact surfaces (still not completely understood) makes the Selberg eigenvalues nongeneric. For example, their short-range statistics are not those of random-matrix theory

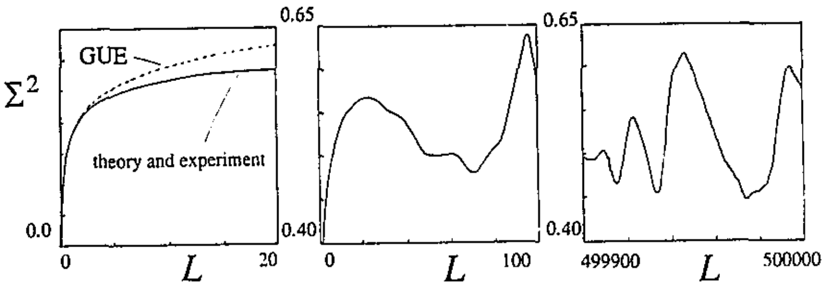

Figure 4.4. After [97], with kind permission of A.M. Odlyzko.

[93,94]. In this respect, the Selberg system resembles the quantized Arnold cat map [37], which is also classically chaotic and can be understood in considerable analytical detail but whose eigenvalues have special statistics [34, 35], unlike those of random-matrix theory.

The analogy between the Gutzwiller formula (4.6) and its Riemann look-alike (4.32) was greatly strengthened by the discovery [75] that the Riemann amplitudes (4.36) for the long 'orbits' (i.e., the high primes) satisfy exactly the classical sum rule (4.10), or, what is (as we now know) equivalent, that the form factor $K(\tau)$ of the zeros satisfies the random-matrix formula $K(\tau) = \tau$ in the appropriate regime [cf. the second equality in eq. (4.26)], which was earlier discovered by Montgomery [95].

It is natural to test the analogy by numerical experiment. This has been done by Odlyzko, who studied 10^5 Riemann zeros, starting at the 10^{12}th [96] and more recently 175×10^6 zeros, starting at the 10^{20}th [97]. Such extreme heights are necessary because ζ approaches its asymptotics very slowly. The results show that statistics of near-neighbouring zeros, such as $P(S)$, are accurately given by the GUE of random-matrix theory, but long-range statistics show the expected breakdown of universality. In [98] I derived the number variance $\Sigma^2(L)$, given by the semiclassical formula (4.27) with the amplitudes of eq. (4.36), and found excellent agreement with Odlyzko's data for zeros near the 10^{12}th. Figure 4.4 (taken from [97]) shows the comparison for zeros near the 10^{20}th, for three ranges of L: in the random-matrix regime, in the transition zone (near $L_{\max}/2\pi = \log(E/2\pi)/(2\pi \log 2) = 9.72$), and far into the nonuniversal regime. On these graphs the experimental curves, calculated from Odlyzko's zeros, and the 'theoretical' curves, calculated from the semiclassical formula, are indistinguishable. This lends powerful support to the picture of $\zeta(z)$ as a model for quantum chaology as well as confirming our understanding of spectral fluctuations (at least for this and other bilinear statistics) in the universal and nonuniversal regimes.

Odlyzko's computations were of epic proportions: 1000 hours of Cray X-MP time, generating 2000 Mb of data. By contrast, evaluation of the semiclassical formula for $\Sigma^2(L)$ takes only a few seconds, illustrating the power and usefulness of asymptotics.

5. Divergence and resurgence

5.1. Making the periodic-orbit sum converge

Without smoothing, i.e., with $\varepsilon = 0$, Gutzwiller's sum (4.6) could converge at best only conditionally, in order to reproduce the δ-functions in the unsmoothed spectral density [eq. (3.5)]. In fact, it does not converge in any sense. To see this, we use the exponential decay law (4.9) together with the exponential proliferation law (4.3) to obtain the estimate

$$A\rho \approx \exp(\tfrac{1}{2}h_{\rm KS}T) \quad \text{as} \quad T \to \infty\,. \tag{5.1}$$

Thus we have an oscillatory series whose terms have, effectively, exponentially increasing amplitudes, reflecting the dominance of the diminishing A's by the proliferating number of orbits. (It is correct to scale the sum over j into an integration over T, because for long orbits action S is proportional to T, so this scaling makes the distribution of phases asymptotically uniform.) The same divergence occurs in the special case of the Selberg trace formula [92] and in the Riemann series (4.32) (for which $h_{\rm KS} = 1$).

An obvious remedy is to reinstate the exponential smoothing ε. However, convergence demands a finite ε, namely [cf. eqs. (4.6) and (5.1)],

$$\varepsilon_{\rm conv} = \hbar h_{\rm KS} \tag{5.2}$$

[for $\zeta(z)$ this reproduces the validity condition $\operatorname{Re} z > 1$ in eq. (4.29)]. This has the fatal defect of blurring $d(E)$ on a scale much larger than the mean level spacing $\varepsilon_{\rm min} \sim \hbar^N$, thereby preventing semiclassical determination of the individual levels – which would be the main aim of summing the series over orbits.

The word 'semiclassical' is important here. It means that the levels which cannot be reconstructed are the high ones. In favourable cases some low levels can be found from the early terms of even the unsmoothed series. In this way, Gutzwiller [100] found several levels of the classically chaotic anisotropic Kepler system, and I [75] reproduced the first few Riemann zeros. Here. the meaning of 'low' depends on the context: for $\zeta(z)$ it means the first $(4\pi - 1)\exp(4\pi) = 33166712$ zeros! It is a curious reflection of the failure of 'asymptotics of asymptotics' that a semiclassical theory is capable of generating low levels but not high ones.

Faced with the failure of exponential smoothing. we must try other means to get the series to converge. Two remarkably simple possibilities are suggested by exact results for $\zeta(z)$. The first. by Guinand ([101]. see also [102]). is to regularize eq. (4.32) by taking the limit of the sum truncated at a period T^*, after subtracting the integral over the limiting form (5.1). In terms of the spectral density and after a little reduction. Guinand's result can be written

$$\begin{aligned} d(E) =& \left\{ \frac{1}{2\pi} \log\left(\frac{E}{2\pi}\right) + \frac{1}{2\pi}\left[\frac{\mathrm{d}}{\mathrm{d}E}\arg\Gamma(\tfrac{1}{2}+\mathrm{i}E) - \log E\right] \right. \\ & \left. - \frac{\cosh \pi E}{4(1+\sinh^2 \pi E)} \right\} \\ & - \frac{1}{\pi} \lim_{T^*\to\infty} \left\{ \sum_p \sum_{k=1}^{k\log p<T^*} \frac{\log p}{p^{k/2}} \cos(Ek\log p) - \frac{2\exp(\frac{1}{2}T^*)}{1+4E^2} \right. \\ & \left. \times [2E\sin(ET^*) - \cos(ET^*)] \right\} \end{aligned} \tag{5.3}$$

The first set of braces contains an exact representation of the mean density of zeros. and the second set contains the regularized sum. Actually. this result depends [101] on the vanishing of

$$\Sigma \equiv \lim_{T^*\to\infty} \frac{\displaystyle\sum_p \sum_{k=1}^{k\log p<T^*} \frac{\log p}{p^{k/2}} - 2\exp(\tfrac{1}{2}T^*)}{T^*} \tag{5.4}$$

a question apparently analytically difficult and for which my numerical explorations are inconclusive. If S does not vanish. an extra term must be added to eq. (5.3).

Guinand's formula suggests that a similar regularization might work for the general Gutzwiller formula (4.6). To subtract the integral over periods T using the amplitude rule (5.1), we have to know how the phases behave. Here we have the result [65]

$$S \approx \frac{NT}{\dfrac{\mathrm{d}}{\mathrm{d}E}\log\displaystyle\int \mathrm{d}x\ \Theta[E - H(x)]} \equiv \mathcal{E}T \quad \text{for} \quad T \gg T_{\min}\,. \tag{5.5}$$

where $\mathcal{E}$ is an energy associated with the classical energy surface. This leads to the (unexplored) hope that the following replacement might improve the convergence of eq. (4.6).

$$\sum_j A_j \exp\left[\mathrm{i}\left(\frac{S_j}{\hbar}+\gamma_j\right)\right] \to$$
$$\sum_{T_j<T^*} A_j \exp\left[\mathrm{i}\left(\frac{S_j}{\hbar}+\gamma_j\right)\right] - \frac{\exp(\frac{1}{2}h_{\mathrm{KS}}T^*)\exp\left[\mathrm{i}\left(\frac{\mathcal{E}T^*}{\hbar}+\gamma^*\right)\right]}{\frac{1}{2}h_{\mathrm{KS}}+\mathrm{i}\frac{\mathcal{E}}{\hbar}}. \tag{5.6}$$

Here γ^* is the (large?) focal phase corresponding to orbits with periods near T^*.

The second exact result for $\zeta(z)$, from Delsarte [103], is in terms of Gaussian rather than exponential smoothing. Delsarte defines the Gaussian-smoothed density of Riemann zeros,

$$d_{\mathrm{G}}(E;\varepsilon) \equiv \frac{1}{\varepsilon}\int_{-\infty}^{\infty} \mathrm{d}E' \ \exp\left[-\frac{\pi}{\varepsilon^2}(E-E')^2\right] d(E'). \tag{5.7}$$

and shows that

$$\begin{aligned} d_{\mathrm{G}}(E;\varepsilon) =& \frac{2}{\varepsilon}\exp\left[\frac{\pi}{\varepsilon^2}(\tfrac{1}{4}-E)^2\right]\cos\left(\frac{\pi E}{\varepsilon^2}\right) \\ &+ \frac{1}{2\pi\varepsilon}\int_{-\infty}^{\infty} \mathrm{d}E' \ \exp\left[-\frac{\pi}{\varepsilon^2}(E-E')^2\right] \\ &\times \left[\mathrm{Re}\,\frac{\Gamma'(\frac{1}{2}+\mathrm{i}E')}{\Gamma(\frac{1}{2}+\mathrm{i}E')} - \log(2\pi) - \frac{\pi}{2\cosh(\pi E')}\right] \\ &- \frac{1}{\pi}\sum_p\sum_{k=1}\frac{\log p}{p^{k/2}}\cos(Ek\log p)\exp\left[-\frac{\varepsilon^2(k\log p)^2}{4\pi}\right] \end{aligned} \tag{5.8}$$

The advantage of Gaussian smoothing is that it guarantees convergence of the sum for arbitrarily small ε and so, in principle, permits the recovery of the Riemann zeros from the primes.

Steiner [104] has independently emphasized the virtues of Gaussian smoothing in the context of Selberg's trace formula. It is important because it shows how, in principle, the quantum eigenvalues are generated exactly by the closed orbits of this system. It is true in general that Gaussian smoothing makes the Gutzwiller sum converge. Instead of eq. (4.6) we have

$$d_G(E:\varepsilon) \approx \bar{d}_G(E) + \frac{1}{\hbar}\sum_j A_j(E)\exp\left[-\frac{\varepsilon^2 T_j^2(E)}{4\pi\hbar^2}\right] \times \exp\left\{\mathrm{i}\left[\frac{S_j(E)}{\hbar} + \gamma_j\right]\right\} . \tag{5.9}$$

whose convergence is obvious from eq. (5.1). Because the Gutzwiller formula is in general asymptotic rather than exact, it is likely that the convergence for $\varepsilon \ll \varepsilon_{\min}$ is to something other than the desired string of δ-functions situated on the quantum levels. It is far from obvious what this is. Probably it is not even a string of δ-functions at all, but a disorderly oscillation getting ever noisier as $\varepsilon \to 0$.

Suppose, however, that this pessimism proves unjustified, and somebody proves that one of the preceding regularizations, suggested by $\zeta(z)$ or by the Selberg formula, or some other regularization, makes the Gutzwiller sum converge to a set of δ-functions semiclassically close to the quantum energies. This would certainly be substantial progress, but would hardly advance the problem of actually determining the levels. The reason (already discussed in [1]) is that in order to resolve the δ-functions it is necessary to include at least all the orbits whose oscillatory contributions to eq. (4.6) have wavelengths (4.1) greater than the spacing $\varepsilon_{\min}$. These are all the orbits with periods less than $T_{\max}$ given by eq. (4.2). The difficulty is that this is an *exponentially large number*: it is of order $\exp(1/\hbar^{N-1})$ - yet another semiclassical nonanalyticity.

Thus, in the semiclassical limit the sum over periodic orbits gets worse much faster than more conventional techniques for locating eigenvalues, based on diagonalizing matrices whose size increases as an inverse power of $\hbar$. As already mentioned, this paradoxical behaviour is the result of taking the long-time limit of what started life as the legitimate semiclassical approximation (3.16) for the propagator. Voros [121] gives a transparent account of the pitfalls in a literal interpretation of eq. (4.6), and Keating and Berry [122] discuss a particular (integrable) example in detail.

It is likely, however, that the periodic orbits do encode the quantum levels in some way, more subtle than a literal interpretation of eq. (4.6) or one of its possible regularizations. The problem is to find the key to unlock the periodic-orbit sum and enable its secrets to be decoded. Unless this can be done, we cannot claim to understand the mechanism of quantization for chaotic systems.

In the remainder of these lectures I will explore with the aid of three examples one approach to decoding information contained in the tails of series, especially divergent ones. This is the idea of resurgence – a term invented by Écalle [105,106] in purely mathematical contexts that I confess to finding hard to follow. In resurgence, the tail of a series is decoded to give not merely a small numerical correction to the leading terms but qualitatively new phenomena. All my examples have been published separately, but here I want to emphasize that they illustrate the same idea – something I realized only recently.

5.2. Resurgence 1: exponentially small waves

This first example has nothing to do with chaos, but serves to introduce resurgence in a familiar context. Consider the semiclassical limit of the one-dimensional Schrödinger equation in the complex plane z. This is

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\Psi(z) = \frac{R^2(z)}{\hbar^2}\Psi(z)\,, \tag{5.10}$$

where for a particle of mass m and energy E in a potential $V(z)$,

$$R^2(z) = 2m[V(z) - E]\,. \tag{5.11}$$

For $\hbar \to 0$ the WKB, or phase-integral, method can be applied, to give Ψ approximately as a linear combination of two exponentials [107,108,15], i.e., as two 'waves'. As is well-known, the simplest form of the method fails at the classical turning points z^*, at which $V(z^*) = E$. In what follows, let z^* denote the turning point closest to the z at which Ψ is being studied. Then a convenient way to write the two approximate solutions is

$$\Psi_\pm(z) = \frac{\exp\left[\pm\frac{1}{\hbar}\int_{z^*}^{z} \mathrm{d}z\; R(z)\right]}{R^{1/2}(z)}\,. \tag{5.12}$$

On the simplest view, Ψ can be expressed approximately as a linear combination of Ψ_+ and Ψ_- with constant multipliers $b_\pm$. This would be an exact representation if R were constant. But, as has been known (but not well-known) since 1857, when R varies it is wrong, even as a lowest approximation, because of the Stokes phenomenon [109]. This refers to rapid changes in the multiplier b_- of the small exponential across *Stokes lines*, where the integral in eq. (5.12) is positive real, so that the disparity between the two exponentials is greatest. The change in b_- is i times the multiplier b_+ of the dominant exponential. The Stokes phenomenon is unavoidable if Ψ is to be correctly approximated globally in terms of Ψ_+ and Ψ_-.

Dingle [110] has explained the Stokes phenomenon as resurgence of the tail of the asymptotic series that converts Ψ_+ in eq. (5.12) from an approximate to an exact solution. Write the exact Ψ as

$$\Psi(z) = \frac{\exp\left[\frac{1}{\hbar}\int_{z^*}^{z} \mathrm{d}z\, R(z)\right]}{R^{1/2}(z)} \sum_{r=0}^{\infty} Y_r(z)\,, \tag{5.13}$$

where the Y_r are defined by substitution into eq. (5.10) and the stipulations $Y_0 = 1$ and $Y_r \propto \hbar^r$. This semiclassical expansion apparently contains just a single exponential – the dominant one. But the expansion is divergent, so eq. (5.13) represents an encoding of Ψ, whose interpretation in conventional terms, as a numerically well defined function, requires a decoding of the tail of the series. Out of the tail will be born the small exponential.

Underlying this resurgence is a universal formula, discovered by Dingle [110] for the late terms $r \gg 1$ of the semiclassical expansion, i.e., for the asymptotics of the asymptotics,

$$Y_r(z) \approx \frac{(r-1)\,!}{2\pi F^r} \quad \text{if} \quad r \gg 1\,, \tag{5.14}$$

where F, the 'singulant', denotes the difference between the two exponents, i.e.,

$$F(z) \equiv \frac{2}{\hbar}\int_{z^*}^{z} \mathrm{d}z\, R(z)\,. \tag{5.15}$$

(Actually Dingle goes much farther, and develops an asymptotic expansion for the late Y_r, of which eq. (5.14) is just the first term (see also [125]), but we shall not require this.) It is clear from eq. (5.14) that the semiclassical series behaves in the typical manner of an asymptotic expansion: for small $\hbar$ – i.e., large F – the terms get rapidly smaller at first but eventually increase factorially. The least term is the integer r^* closest to $|F|$.

With the aid of eq. (5.14) we split Ψ into the series (5.13) down to the least term, plus the tail:

$$\Psi(z) \approx \frac{\exp\left[\frac{1}{\hbar}\int_{z^*}^{z} \mathrm{d}z\, R(z)\right]}{R^{1/2}(z)} \sum_{r=0}^{r^*} Y_r(z)$$
$$+ \frac{\exp\left[-\frac{1}{\hbar}\int_{z^*}^{z} \mathrm{d}z\, R(z)\right]}{R^{1/2}(z)} \left[\exp(F) \sum_{r=r^*+1}^{\infty} \frac{(r-1)\,!}{2\pi F^r}\right]. \tag{5.16}$$

This has been cast into the form of two exponentials, with $b_+ = 1$ and b_- given by the term in brackets, suitably decoded. The decoding can be achieved by Borel summation, that is by replacing the factorial by its integral representation and performing the sum (after interchanging summation and integration). Dingle shows that the resulting b_- jumps by i across a Stokes line, on which F is positive real.

I have been able to go further, and determine the precise manner in which the change occurs [111,112]. This requires asymptotic approximation of the integral generated by the Borel method (asymptotics of asymptotics of asymptotics), and yields the following result, whose remarkable simplicity depends on the fact that eq. (5.16) is truncated at the least term:

$$b_- \approx \text{constant} + \tfrac{1}{2} i b_+ \operatorname{Erf}\left(\frac{\operatorname{Im} F}{\sqrt{2 \operatorname{Re} F}}\right). \tag{5.17}$$

in which Erf denotes the error function [19]. The constant is arbitrary, reflecting the freedom to choose any linear combination of exponential solutions to eq. (5.10) at a given point z: in the Borel method the arbitrariness enters through a choice of integration contour near a pole. From the definition (5.15) it now follows that the jump ib_- in the subdominant multiplier across the Stokes line, that is as $\operatorname{Im} F$ passes zero, is essentially completed over a z interval whose size is of order $\hbar^{1/2}$, which can be regarded as the 'width' of the Stokes line. I have confirmed the correctness of eq. (5.17) by numerical computation, which shows the smooth switching-on of the small exponential when the dominant asymptotic series is subtracted from the exactly-calculated function.

The universality of the resurgence (5.17) extends beyond the context described here, of second-order differential equations in the complex plane. It also applies, for example, to exponentials generated by the method of steepest descent (applied, e.g., to the diffraction catastrophes of short-wave optics [113]), where Stokes lines (and their higher-dimensional generalizations [114]) can occur in spaces of real – and therefore physically accessible – variables, as well as complex ones. I have given a detailed discussion of the application to the Schrödinger equation [115], and Voros [116] earlier employed similar ideas in a very different way. What we learn from this example is that the resurgence of a divergent series can indeed yield qualitatively new results. It gives physics as well as mathematics, because the small exponentials whose birth is being described correspond (for example) to weak reflected waves, small changes in adiabatic 'invariants', and evanescent diffracted waves beyond caustics.

5.3. Resurgence 2: Riemann–Siegel

The method used in practice to compute the Riemann zeros (e.g., by Odlyzko [96,97]) is an asymptotic representation of $\zeta(z)$ which, unlike the Gutzwiller look-alike formulae (4.32), is valid on the critical line $\mathrm{Re}\, z = \frac{1}{2} + \mathrm{i}E$, where (on the Riemann hypothesis and apparently in reality) the zeros lie. This is the Riemann–Siegel formula [88]. Its dominant part is a finite series of real terms. Here, I will show how to obtain this by resurgence, in a derivation considerably simpler than the usual one [88].

To begin, we write the Dirichlet series in the definition (4.29) on the critical line:

$$\zeta(\tfrac{1}{2} + \mathrm{i}E) = \sum_{n=1}^{\infty} \frac{\exp(-\mathrm{i}E \log n)}{n^{1/2}} . \tag{5.18}$$

This is divergent. By transforming to the uniformizing phase variable $\log n$ it can be seen that the divergence is exponential - not as bad as the factorial divergences in the last section, but still requiring decoding to get a useful representation of $\zeta(z)$ on the line.

Again we divide the sum into a finite part and a tail:

$$\zeta(\tfrac{1}{2} + \mathrm{i}E) = \sum_{n=1}^{n^*} \frac{\exp(-\mathrm{i}E \log n)}{n^{1/2}} + \sum_{n=n^*+1}^{\infty} \frac{\exp(-\mathrm{i}E \log n)}{n^{1/2}} . \tag{5.19}$$

To interpret the tail, we employ Poisson rather than Borel summation. This technique [117] transforms a sum into another sum through the identity

$$\sum_{n=a}^{b} f(n) = \sum_{m=-\infty}^{\infty} \int_{a-1/2}^{b+1/2} \mathrm{d}x \; f(x) \exp(2\pi \mathrm{i} m x) . \tag{5.20}$$

where $f(x)$ is any function interpolating $f(n)$. Applied to eq. (5.19) it gives the tail as

$$\sum_{n=n^*+1}^{\infty} \frac{\exp(-\mathrm{i}E \log n)}{n^{1/2}} = \sum_{m=-\infty}^{\infty} \int_{n^*+1/2}^{\infty} \mathrm{d}x \frac{\exp[\mathrm{i}(2\pi m x - E \log x)]}{x^{1/2}} . \tag{5.21}$$

The next step is to approximate the integrals on the right-hand side by the method of stationary phase – a procedure ubiquitous in applications of Poisson summation. The only contributions come from those m for which the phase has a stationary point in the range of integration. Each integral has a single stationary point, at

$$x = \frac{E}{2\pi m}. \tag{5.22}$$

so that the contributing m lie in the range $1 \leqslant m \leqslant E/(2\pi n^*)$. Evaluating these contributions, we obtain for the tail the approximation

$$\sum_{n=n^*+1}^{\infty} \frac{\exp(-iE\log n)}{n^{1/2}} \approx \exp[-2\pi i\mathcal{N}(E)] \sum_{m=1}^{E/2\pi n^*} \frac{\exp(iE\log m)}{m^{1/2}}. \tag{5.23}$$

where

$$\mathcal{N}(E) = \frac{E}{2\pi}\log\frac{E}{2\pi e} + \frac{7}{8}. \tag{5.24}$$

Now we note a remarkable fact. Apart from an overall phase shift, the terms in the tail are complex conjugates of the terms in the finite part of the ζ sum (5.19). Moreover, if we choose the splitting point at

$$n^* = \left[\sqrt{\frac{E}{2\pi}}\right]. \tag{5.25}$$

where the brackets denote 'integer part', the transformed tail contains the same number of terms as the finite part. Thus we obtain

$$\zeta(\tfrac{1}{2} + iE) \approx 2\exp[-\pi i\mathcal{N}(E)] \sum_{m=1}^{[\sqrt{E/2\pi}]} \frac{\cos[E\log m - \pi\mathcal{N}(E)]}{m^{1/2}}. \tag{5.26}$$

Resurgence has transformed the divergent series (5.18), with no obvious zeros, into a finite sum which is real precisely where we expect the zeros, i.e. real E.

What we have obtained differs from the 'official' Riemann–Siegel formula in two respects, neither of which is important in the large-E limit we are interested in. First, our simple technique has failed to capture the Riemann–Siegel correction terms, the first of which is of order $E^{-3/4}$. Second, instead of $\mathcal{N}(E)$ the usual formula contains complex gamma functions (cf. sections 5.3 and 5.8), equivalent to our expressions through Stirling's formula for large E.

As a representation of $\zeta(z)$ from which zeros can be determined, eq. (5.26) is vastly superior to the Gutzwiller look-alike (4.32) or its exact regularizations (5.3) and (5.8). It requires neither truncation [like eq. (4.2)] nor limiting procedures [like eqs. (5.3) and (5.8)]. And the most optimistic truncation of eq. (4.2) requires the square of the number of terms in eq. (5.26). Moreover, even the first term $m = 1$ of eq. (5.26) generates approximate zeros with the correct density (4.33), whose integral $\mathcal{N}(E)$ [eq. (5.24)] counts the number of states with energy less than E. complete with the correct constant [88].

A natural question is whether this resurgence of the long orbits (i.e., large n) into a compact and usable formula has any analogue for chaotic dynamical systems. I have discussed this elsewhere [89] without coming to a firm conclusion. The generalization I envisage would be a sum formula for the spectral determinant $\det(E - \hat{H})$ (in some appropriate representation), rather than the spectral density, which is the derivative of the logarithm of this quantity. One conjecture along these lines is presented in [123].

Note the differences between this Riemann resurgence and the Stokes resurgence in the previous section. There. a factorially divergent series engendered a single term. exponentially small relative to the originals. Here. an exponentially divergent series engendered a whole series of terms, equal in magnitude to the originals.

5.4. Resurgence 3: bootstrapping the closed orbits

This final example of resurgence is different again. Imagine that the Gutzwiller series (4.6) is convergent. or is made to be convergent by one of the regularizations suggested in section 5.2. Then as pointed out already the convergence can be at best conditional because it has to generate the δ-functions of the spectral density. Consider the mechanism of convergence. Each finite periodic orbit contributes a smooth oscillation to $d(E)$. Therefore, δ-singularities can emerge only from infinitely long orbits, i.e., from the tail of the series: the only role of the finite orbits is to cancel the smooth background between the δ-functions. In determining the positions of the levels, then, any number of the finite orbits can be omitted. But once the levels have been determined, their average density can be calculated, and this is the first term $\bar{d}$ of the series. Thus must the tail of Gutzwiller's series determine its head. a resurgence for which we can adopt the felicitous term 'analytic bootstrap' invented by Voros [116] in another context.

The bootstrap implies severe constraints on the amplitudes and phases from the closed orbits. in order that they add to give δ-functions with the right density. One consequence is the semiclassical sum rule (4.25) whose long-deferred derivation I will give now. Originally [75] I employed the exponential smoothing, as in eq. (4.6). Here I will obtain the same result with Gaussian smoothing, i.e., with eq. (5.9).

We can square the spectral density (3.5) and ignore the terms from the overlaps of the smoothed δ-functions provided we choose $\varepsilon \ll \varepsilon_{\min}$. Thus,

$$d^2(E,\varepsilon) = \sum_n \delta_\varepsilon^2(E - E_n)\,. \tag{5.27}$$

With Gaussian smoothing, i.e.,

$$\delta_\varepsilon(x) \equiv \frac{1}{\varepsilon}\exp\left(-\frac{\pi x^2}{\varepsilon^2}\right)\,, \tag{5.28}$$

we have

$$\delta(x) = \lim_{\varepsilon\to 0} \varepsilon\sqrt{2}\delta_\varepsilon^2(x)\,, \tag{5.29}$$

and hence,

$$d(E) = \lim_{\varepsilon\to 0} \varepsilon\sqrt{2}d^2(E,\varepsilon)\,. \tag{5.30}$$

Now we substitute the series (5.9). Because of the multiplication by ε, only the closed-orbit terms survive, i.e., the term $\bar{d}$ does not contribute on the right. Averaging over an energy range exceeding $\varepsilon_{\max}$, we obtain

$$\bar{d} = \lim_{\varepsilon\to 0} \frac{\varepsilon\sqrt{2}}{h^2}\left\langle \sum_i \sum_j A_i A_j \exp\left(\mathrm{i}\frac{S_i - S_j}{\hbar}\right)\exp\left[-\frac{\varepsilon^2(T_i^2 + T_j^2)}{4\pi\hbar^2}\right]\right\rangle\,. \tag{5.31}$$

This shows explicitly how the resurgence of the long orbits determines the mean density. To get the desired sum rule we make the approximation

$$T_i^2 + T_j^2 \approx \tfrac{1}{2}(T_i + T_j)^2 \tag{5.32}$$

(justified by the smallness of the omitted term $(T_i - T_j)/2$ in comparison with the actions in the oscillatory exponential). Then, we can express this in terms of the semiclassical form factor $K(\tau)$ given by eq. (4.21):

$$\bar{d} \approx \lim_{\varepsilon\to 0} \varepsilon\sqrt{2}\bar{d}^2 \int_{-\infty}^{\infty} \mathrm{d}\tau\ K(\tau)\exp(-2\pi\varepsilon^2\tau^2\bar{d}^2)\,. \tag{5.33}$$

The sum rule (4.25) follows immediately if we disregard possible long-range oscillations in $K(\tau)$.

By taking higher powers of smoothed δ-functions. the same resurgence principle can be used to make higher-order semiclassical sum rules. I have given one set of these rules [75] and Verbaarschot [118] has given another.

Note how this resurgence differs from the previous two. Operating on a conditionally convergent rather than a divergent series. it has not generated new terms but *reproduced* the term at the beginning of the series. I have to state. however. that the direct derivation of eq. (4.25) from the double sum (4.21) (or otherwise) still eludes us.

6. Last words

I had intended to end by expressing astonishment at the power of the semiclassical way of thinking about these quantum phenomena. where one gets understanding about the real world of finite $\hbar$ by imagining the asymptotic limit $\hbar \to 0$. Then. I discovered that Ulam [119] said it already:

> "..why are asymptotic theorems so much simpler than finite approximations? Infinity does not correspond to the popular image. It is a guiding light. a star that draws us to finite ways of thinking. God knows why."

References

[1] M. V. Berry. Semiclassical Mechanics of Regular and Irregular Motion. in: Chaotic Behaviour of Deterministic systems. Les Houches Lectures XXXVI. eds.. G. Iooss. R. H. G. Helleman and R. Stora (North-Holland. Amsterdam. 1983) pp171–271.

[2] M. V. Berry. Regular and Irregular Motion. in: Topics in Nonlinear Mechanics. ed.. S. Jorna. (Am. Inst. Phys. Conf. Proc. 46. 1978) 16–120.

[3] B. Eckhardt. Phys. Rep. 163 (1987) 205–297.

[4] J. Ford. in: The New Physics, ed., P. C. W. Davies (Cambridge University Press. 1989) 348–372.

[5] G. Casati. ed.. Chaotic Behavior in Quantum Systems (London. Plenum. 1985).

[6] T. H. Seligman and H. Nishioka. Quantum Chaos and Statistical Nuclear Physics. Lecture Notes in Physics No. 263 (Springer. 1986).

[7] E. R. Pike and S. Sarkar, Quantum Measurement and Chaos. NATO ASI Series B, Vol. 161 (London, Plenum, 1987).

[8] A. M. Ozorio de Almeida, Hamiltonian Systems: Chaos and Quantization (Cambridge University Press, 1988).

[9] M. Tabor. Chaos and Integrability in Nonlinear Dynamics (New York. Wiley. 1989).

[10] U. Frisch. in ref. [1] (1983) 665–704.

[11] B. B. Mandelbrot. The Fractal Geometry of Nature (Freeman, San Francisco. 1982).
[12] R. C. Tolman. The Principles of Statistical Mechanics (Oxford University Press. 1938).
[13] K. G. Wilson. Rev. Mod. Phys. 47 (1975) 773–840.
[14] B. Russell and A.N. Whitehead. Principia Mathematica (Cambridge University Press. 2nd ed.. 1927) Vol. 2. p. 83. Proposition *110.643.
[15] M. V. Berry and K. E. Mount. Rep. Prog. Phys. 35 (1972) 315-397.
[16] G. B. Airy. Trans. Cambridge Phil. Soc. 6 (1838) 379–403.
[17] M. V. Berry and C. Upstill. Progress in Optics, 18 (1980) 257–346.
[18] Yu. A. Kravtsov. Progress in Optics 26 (1988) 227–348.
[19] M. Abramowitz and I. A. Stegun. Handbook of Mathematical Functions (National Bureau of Standards. Washington, 1964).
[20] A. N. Varchenko. Funkt. Anal. i Prilozhen. (Moscow) 10 No. 3 13-38. (Funct. Anal. Appl. 10 (1976) 175-196).
[21] M. V. Berry. J. Phys. A 10 (1977) 2061-2081.
[22] M. V. Berry. in: Wave Propagation and Scattering. ed. B. J. Uscinski (Oxford. Clarendon Press. 1986) 11-35.
[23] M. V. Berry and J. Goldberg. Nonlinearity 1 (1988) 1–26.
[24] M. V. Berry. Physica D 33 (1988) 26-33.
[25] G. Casati. B. V. Chirikov. J. Ford and F. M. Izraelev. in: Stochastic Behavior in Classical and Quantum Hamiltonian Systems. eds.. G. Casati. J. Ford. Lecture Notes in Physics 93 (Springer. 1979) pp334 352.
[26] O. Bohigas and M.-J. Giannoni. Chaotic Motion and Random-matrix Theories. in: Mathematical and Computational Methods in Nuclear Physics. eds.. J. S. Dehesa. J. M. G. Gomez and A. Polls, Lecture Notes in Physics 209 (Springer-Verlag. New York. 1984) pp1 99.
[27] M. V. Berry. Proc. R. Soc. London A413 (1987) 183 198.
[28] E. J. Heller. Phys. Rev. Lett. 53 (1984) 1515-1518.
[29] E. J. Heller. in ref. [6] (1986) 162-181.
[30] E. B. Bogomolny. Physica D 31 (1988) 169-189.
[31] M. V. Berry. Proc. Roy. Soc. Lond. A423 (1989) 219-231.
[32] M. V. Berry. Physica Scripta 40 (1989) 335–336.
[33] J. P. Keating. Semiclassical Properties of the Cat Maps. Ph.D. Thesis. Department of Physics. (University of Bristol. U.K.. 1989).
[34] J. P. Keating. in: Number Theory and Physics. eds.. J.-M. Luck. P. Moussa and M. Waldschmidt. Proceedings in Physics 47 (Springer. 1990) pp302–310.
[35] J. P. Keating. Nonlinearity 4 (1991) 277–307 and 309–341.
[36] V. I. Arnold and A. Avez. Ergodic Problems of Classical Mechanics (Benjamin. 1968).
[37] J. H. Hannay and M. V. Berry. Physica 1D (1980) 267–290.
[38] I. C. Percival. Adv. Chem. Phys. 36 (1977) 1–61.
[39] V. P. Maslov and M. V. Fedoriuk. Semiclassical Approximation in Quantum Mechanics (D. Reidel, Dordrecht. 1981).
[40] M. V. Berry. J. Phys. A 10 (1977) 2083–2091.
[41] A. Voros. in: Stochastic Behavior in Classical and Quantum Hamiltonian Systems. eds.. G. Casati and J. Ford. Lecture Notes in Physics 93 (Springer, Berlin. 1979) 326–333.
[42] A. I. Shnirelman. Usp. Mat. Nauk. 29 No. 6 (1974) 181–182.

[43] Y. Colin de Verdière. Commun. Math. Phys. 102 (1985) 497–502.
[44] E. P. Wigner. Phys. Rev. 40 (1932) 749–759.
[45] H. J. Groenewold. Physica 12 (1946) 405–460.
[46] T. Takabayashi. Proc. Theor. Phys. Jap. 11 (1954) 341–373.
[47] J. E. Moyal. Proc. Cambridge Phil. Soc. 45 (1949) 99–124.
[48] G. A. Baker Jr. Phys. Rev. 109 (1958) 2198–2206.
[49] K. Takahashi. J. Phys. Soc. Jap. 55 (1986) 762–779.
[50] P. Leboeuf and M. Saraceno, J. Phys. A 23 (1990) 1745–1764.
[51] J. H. Van Vleck. Proc. Nat. Acad. Sci. 14 (1928) 178–188.
[52] V. I. Arnol'd. Mathematical Methods of Classical Mechanics (Springer. New York. 1978).
[53] M. V. Berry. Philos. Trans. R. Soc. London A287 (1977) 237–271.
[54] A. M. Ozorio de Almeida and J. H. Hannay. Ann. Phys. (NY) 138 (1982) 115–154.
[55] H. P. Baltes and E. R. Hilf. Spectra of Finite Systems (B-I Wissenschaftsverlag. Mannheim. 1976).
[56] N. L. Balazs. Physica 102A (1980) 236–254.
[57] M. C. Gutzwiller. J. Math. Phys. 12 (1971) 343–358.
[58] R. Balian and C. Bloch. Ann. Phys. (NY) 69 (1972) 76–160.
[59] B. Eckhardt. J. Ford and F. Vivaldi. Physica 13D (1984) 339–356.
[60] J. Ford. G. Mantica and G. H. Ristow. Physica D 50 (1991) 493–520.
[61] S. W. McDonald and A. N. Kaufman. Phys. Rev. A37 (1988) 3067–3086.
[62] M. Shapiro and G. Goelman. Phys. Rev. Lett. 53 (1984) 1714–1717.
[63] Ya. G. Sinai. Introduction to Ergodic Theory (Princeton University Press. 1976).
[64] M. V. Berry. Ann. Phys. (NY) 131 (1981) 163–216.
[65] J. H. Hannay and A. M. Ozorio de Almeida. J. Phys. A 17 (1984) 3429–3440.
[66] J. H. Hannay. in ref. [5] (1985) 141–147.
[67] M. V. Berry and M. Tabor. Proc. R. Soc. London A356 (1977) 375–394.
[68] M. V. Berry and M. Robnik. J. Phys. A 19 (1986) 649–668.
[69] M. V. Berry. J. Phys. A 20 (1987) 2389–2403.
[70] C. E. Porter. Statistical Theories of Spectra: Fluctuations (Academic Press. New York. 1965).
[71] R. Scharf. B. Dietz. M. Kus, F. Haake and M. V. Berry. Europhys. Lett. 5 (1988) 383–389.
[72] M. Robnik and M. V. Berry. J. Phys. A 19 (1986) 669–682.
[73] M. Robnik. J. Phys. A 17 (1984) 1049–1074.
[74] H.-D. Meyer. E. Haller. H. Köppel and L. S. Cederbaum. J. Phys. A 17 (1984) L831–836.
[75] M. V. Berry. Proc. R. Soc. London A400 (1985) 229–251.
[76] F. J. Dyson and M. L. Mehta, J. Math. Phys. 4 (1963) 701–712.
[77] M. L. Mehta. Random Matrices and the Statistical Theory of Energy Levels (Academic Press. New York and London, 1967)
[78] T. H. Seligman. J. J. M. Verbaarschot and M. R. Zirnbauer, J. Phys. A 18 (1985) 2751–2770.
[79] M. V. Berry and R. J. Mondragon, Proc. R. Soc. London A412 (1987) 53–74.
[80] M. Robnik. in ref. [6] (1986) 120–130.
[81] M. V. Berry, in ref. [5] (1985) 123–140.
[82] A. Shapere and F. Wilczek, Geometric Phases in Physics (World Scientific. 1989).
[83] F. J. Dyson. J. Math. Phys. 3 (1962) 1191–1198.

[84] P. Pechukas. Phys. Rev. Lett. 51 (1983) 943–946.
[85] T. Yukawa. Phys. Rev. Lett. 54 (1985) 1883–1886.
[86] K. Nakamura and M. Lakshmanan. Phys. Rev. Lett. 57 (1986) 1661–1664.
[87] M. Wilkinson. J. Phys. A 21 (1988) 1173–1190.
[88] H. M. Edwards. Riemann's Zeta Function (Academic Press. New York and London. 1974).
[89] M. V. Berry. in ref. [6] (1986) 1–17.
[90] H. P. McKean. Comm. Pure App. Math. 25 (1972) 225–246.
[91] D. A. Hejhal. Duke Math. J. 43 (1976) 441–482.
[92] N. L. Balazs and A. Voros. Phys. Rep. 143 (1986) 109–240.
[93] C. Schmit (1986) private communication. and these proceedings.
[94] R. Aurich and F. Steiner. Physica D 39 (1989) 169–193.
[95] H. L. Montgomery. Proc. Symp. Pure Math. 24 (1973) 181–193.
[96] A. M. Odlyzko. Math. of Comp.. Vol. 48. No. 177 (1987). 273–308.
[97] A. M. Odlyzko. The 10^{20}th Zero of the Riemann Zeta Function and 175 million of its Neighbours (to be published. 1991).
[98] M. V. Berry. Nonlinearity 1 (1988) 399–407.
[99] G. Casati. B. V. Chirikov and I. Guarneri. Phys. Rev. Lett. 54 (1985) 1350–1353.
[100] M. C. Gutzwiller. Physica 5D (1982) 183–207.
[101] A. P. Guinand. Proc. London Math. Soc. 50 (1948) 107–119.
[102] A. P. Guinand. Proc. London Math. Soc. 51 (1950) 401–414.
[103] J. Delsarte. J. Anal. Math. (Jerusalem) 17 (1966) 419–431.
[104] R. Aurich. M. Sieber and F. Steiner. Phys. Rev. Lett. 61 (1988) 483–487.
[105] J. Écalle. Les Fonctions Résurgentes (Publ. Math. Université de Paris-Sud. Orsay. 1981) in 3 volumes.
[106] J. Écalle. Cinq Applications des Fonctions Résurgentes (Preprint 84T62. Orsay. 1984).
[107] J. Heading. An Introduction to Phase-Integral Methods (Methuen. London. 1962).
[108] N. Fröman and P. O. Fröman. JWKB Approximation: Contributions to the Theory (North-Holland. Amsterdam. 1965).
[109] G. G. Stokes. Trans. Cambridge Phil. Soc. 10 (1864) 106–128. reprinted in Mathematical and Physical papers by the late Sir George Gabriel Stokes (Cambridge University Press. 1904) Vol. IV. 77–109.
[110] R. B. Dingle. Asymptotic Expansions: Their Derivation and Interpretation (Academic Press. New York and London. 1973).
[111] M. V. Berry. Proc. R. Soc. London A422 (1989) 7–21.
[112] M. V. Berry. Publications Mathématiques IHES (Paris) 68 (1989) 211–221.
[113] F. J. Wright. J. Phys. A 13 (1980) 2913–2928.
[114] M. V. Berry and C. J. Howls. Nonlinearity 3 (1990) 281–291.
[115] M. V. Berry. Proc. R. Soc. London A 427 (1990) 265–280.
[116] A. Voros. Ann. Inst. H. Poincaré 39 A (1983) 211–338.
[117] M. J. Lighthill, Introduction to Fourier Analysis and Generalized Functions (Cambridge University Press. 1958).
[118] J. J. Verbaarschot. J. Phys. A 20 (1987) 5589–5605.
[119] N. G. Cooper. ed.. From Cardinals to Chaos: Reflections on the Life and Legacy of Stanislaw Ulam (Cambridge University Press. 1989) 311.
[120] A. Voros. Ann. Inst. H. Poincaré 24 A (1976) 31–90.
[121] A. Voros. J. Phys. A 21 (1988) 685–692.

[122] J. P. Keating and M. V. Berry, J. Phys. A 20 (1987) L1139–L1141.
[123] M. V. Berry and J. P. Keating, J. Phys. A 23 (1990) 4839–4849.
[124] M. V. Berry, in: Proc. 5th Yukawa Symposium, eds., Y. Abe, H. Horiuchi and K. Matsuyanagi (Springer, 1991) in press.
[125] M. V. Berry and C. J. Howls, Proc. R. Soc. London A 430 (1990) 653-668.