Classical and Quantum Chaos:

Part III: Material available on www.nbi.dk/ChaosBook

version 10, Aug 26 2003 printed September 2, 2003
www.nbi.dk/ChaosBook/ChaosBook comments to: predrag@nbi.dk

Appendix A

A brief history of chaos

Laws of attribution

1. **Arnol'd's Law**: everything that is discovered is named after someone else (including Arnol'd's law)
2. **Berry's Law**: sometimes, the sequence of antecedents seems endless. So, nothing is discovered for the first time.
3. **Whiteheads's Law**: Everything of importance has been said before by someone who did not discover it.

M.V. Berry

A.1 Chaos is born

(R. Mainieri)

Classical mechanics has not stood still since Newton. The formalism that we use today was developed by Euler and Lagrange. By the end of the 1800's the three problems that would lead to the notion of chaotic dynamics were already known: the three-body problem, the ergodic hypothesis, and nonlinear oscillators.

A.1.1 Three-body problem

Trying to predict the motion of the Moon has preoccupied astronomers since antiquity. Accurate understanding of its motion was important for determining the longitude of ships while traversing open seas. Kepler's Rudolphine tables had been a great improvement over previous tables, and Kepler was justly proud of his achievements. He wrote in the introduction to the announcement of Kepler's third law, *Harmonice Mundi* (Linz, 1619) in a style that would not fly with the contemporary *Physical Review Letters* editors:

> What I prophesied two-and-twenty years ago, as soon as I discovered the five solids among the heavenly orbits – what I firmly believed long before I had seen Ptolemy's *Harmonics* – what I had promised my friends in the title of this book, which I named before I was sure of my discovery – what sixteen years ago, I urged as the thing to be sought – that for which I joined Tycho Brahé, for which I settled in Prague, for which I have devoted the best part of my life to astronomical contemplations, at length I have brought to light, and recognized its truth beyond my most sanguine expectations. It is not eighteen months since I got the first glimpse of light, three months since the dawn, very few days since the unveiled sun, most admirable to gaze upon, burst upon me. Nothing holds me; I will indulge my sacred fury; I will triumph over mankind by the honest confession that I have stolen the golden vases of the Egyptians to build up a tabernacle for my God far away from the confines of Egypt. If you forgive me, I rejoice; if you are angry, I can bear it; the die is cast, the book is written, to be read either now or in posterity, I care not which; it may well wait a century for a reader, as God has waited six thousand years for an observer.

Bernoulli used Newton's work on mechanics to derive the elliptic orbits of Kepler and set an example of how equations of motion could be solved by integrating. But the motion of the Moon is not well approximated by an ellipse with the Earth at a focus; at least the effects of the Sun have to be taken into account if one wants to reproduce the data the classical Greeks already possessed. To do that one has to consider the motion of three bodies: the Moon, the Earth, and the Sun. When the planets are replaced by point particles of arbitrary masses, the problem to be solved is known as the three-body problem. The three-body problem was also a model to another concern in astronomy. In the Newtonian model of the solar system it is possible for one of the planets to go from an elliptic orbit around the Sun to an orbit that escaped its domain or that plunged right into it. Knowing if any of the planets would do so became the problem of the stability of the solar system. A planet would not meet this terrible end if solar system consisted of two celestial bodies, but whether such fate could befall in the three-body case remained unclear.

After many failed attempts to solve the three-body problem, natural philosophers started to suspect that it was impossible to integrate. The usual technique for integrating problems was to find the conserved quantities, quantities that do not change with time and allow one to relate the momenta and positions different times. The first sign on the impossibility of integrating the three-body problem came from a result of Burns that showed that there were no conserved quantities that were polynomial in the momenta and positions. Burns' result did not preclude the possibility of more complicated conserved quantities. This problem was settled by Poincaré and Sundman in two very different ways.

In an attempt to promote the journal *Acta Mathematica*, Mittag-Leffler got the permission of the King Oscar II of Sweden and Norway to establish a mathematical competition. Several questions were posed (although the king would have preferred only one), and the prize of 2500 kroner would go to the best submission. One of the questions was formulated by Weierstrass:

> Given a system of arbitrary mass points that attract each other according to Newton's laws, under the assumption that no two points ever collide, try to find a representation of the coordinates of each point as a series in a variable that is some known function of time and for all of whose values the series converges uniformly.
>
> This problem, whose solution would considerably extend our understanding of the solar system, ...

Poincaré's submission won the prize. He showed that conserved quantities that were analytic in the momenta and positions could not exist. To show that he introduced methods that were very geometrical in spirit: the importance of phase flow, the role of periodic orbits and their cross sections, the homoclinic points.

The interesting thing about Poincaré's work was that it did not solve the problem posed. He did not find a function that would give the coordinates as a function of time for all times. He did not show that it was impossible either, but rather that it could not be done with the Bernoulli technique of finding a conserved quantity and trying to integrate. Integration would seem unlikely from Poincaré's prize-winning memoir, but it was accomplished by the Finnish-born Swedish mathematician Sundman. Sundman showed that to integrate the three-body problem one had to confront the two-body collisions. He did that by making them go away through a trick known as regularization of the collision manifold. The trick is not to expand the coordinates as a function of time t, but rather as a function of $\sqrt[3]{t}$. To solve the problem for all times he used a conformal map into a strip. This allowed Sundman to obtain a series expansion for the coordinates valid for all times, solving the problem that was proposed by Weirstrass in the King Oscar II's competition.

The Sundman's series are not used today to compute the trajectories of any three-body system. That is more simply accomplished by numerical methods or through series that, although divergent, produce better numerical results. The conformal map and the collision regularization mean that the series are effectively in the variable $1 - e^{-\sqrt[3]{t}}$. Quite rapidly this gets exponentially close to one, the radius of convergence of the series. Many terms, more terms than any one has ever wanted to compute, are needed to achieve numerical convergence. Though Sundman's work deserves better credit than it gets, it did not live up to Weirstrass's expectations, and the series solution did not "considerably extend our understanding of the solar system." The work that followed from Poincaré did.

A.1.2 Ergodic hypothesis

The second problem that played a key role in development of chaotic dynamics was the ergodic hypothesis of Boltzmann. Maxwell and Boltzmann had combined the mechanics of Newton with notions of probability in order to create statistical mechanics, deriving thermodynamics from the equations of mechanics. To evaluate the heat capacity of even a simple system, Boltzmann had to make a great simplifying assumption of ergodicity: that the dynamical system would visit every part of the phase space allowed by conservations law equally often. This hypothesis was extended to other averages used in statistical mechanics and was called the ergodic hypothesis. It was reformulated by Poincaré to say that a trajectory comes as close as desired to any phase space point.

Proving the ergodic hypothesis turned out to be very difficult. By the end of our own century it has only been shown true for a few systems and wrong for quite a few others. Early on, as a mathematical necessity, the proof of the hypothesis was broken down into two parts. First one would show that the mechanical system was ergodic (it would go near any point) and then one would show that it would go near each point equally often and regularly so that the computed averages made mathematical sense. Koopman took the first step in proving the ergodic hypothesis when he noticed that it was possible to reformulate it using the recently developed methods of Hilbert spaces. This was an important step that showed that it was possible to take a finite-dimensional nonlinear problem and reformulate it as a infinite-dimensional linear problem. This does not make the problem easier, but it does allow one to use a different set of mathematical tools on the problem. Shortly after Koopman started lecturing on his method, von Neumann proved a version of the ergodic hypothesis, giving it the status of a theorem. He proved that if the mechanical system was ergodic, then the computed averages would make sense. Soon afterwards Birkhoff published a much stronger version of the theorem.

A.1.3 Nonlinear oscillators

The third problem that was very influential in the development of the theory of chaotic dynamical systems was the work on the nonlinear oscillators. The problem is to construct mechanical models that would aid our understanding of physical systems. Lord Rayleigh came to the problem through his interest in understanding how musical instruments generate sound. In the first approximation one can construct a model of a musical instrument as a linear oscillator. But real instruments do not produce a simple tone forever as the linear oscillator does, so Lord Rayleigh modified this simple model by adding friction and more realistic models for the spring. By a clever use of negative friction he created two basic models for the musical instruments. These models have more than a pure tone and decay with time when not stroked. In his book *The Theory of Sound* Lord Rayleigh introduced a series of methods that would prove quite general, such as the notion of a limit cycle, a periodic motion a system goes to regardless of the initial conditions.

A.2 Chaos grows up

(R. Mainieri)

The theorems of von Neumann and Birkhoff on the ergodic hypothesis were published in 1912 and 1913. This line of enquiry developed in two directions. One direction took an abstract approach and considered dynamical systems as transformations of measurable spaces into themselves. Could we classify these transformations in a meaningful way? This lead Kolmogorov to the introduction of the concept of entropy for dynamical systems. With entropy as a dynamical invariant it became possible to classify a set of abstract dynamical systems known as the Bernoulli systems. The other line that developed from the ergodic hypothesis was in trying to find mechanical systems that are ergodic. An ergodic system could not have stable orbits, as these would break ergodicity. So in 1898 Hadamard published a paper with a playful title of "... billiards ...," where he showed that the motion of balls on surfaces of constant negative curvature is everywhere unstable. This dynamical system was to prove very useful and it was taken up by Birkhoff. Morse in 1923 showed that it was possible to enumerate the orbits of a ball on a surface of constant negative curvature. He did this by introducing a symbolic code to each orbit and showed that the number of possible codes grew exponentially with the length of the code. With contributions by Artin, Hedlund, and Hopf it was eventually proven that the motion of a ball on a surface of constant negative curvature was ergodic. The importance of this result escaped most physicists, one exception being Krylov, who understood that a physical billiard was a dynamical system on a surface of negative curvature, but with the curvature concentrated along the lines of collision. Sinai, who was the first to show that a physical billiard can be ergodic, knew Krylov's work well.

The work of Lord Rayleigh also received vigorous development. It prompted many experiments and some theoretical development by van der Pol, Duffing, and Hayashi. They found other systems in which the nonlinear oscillator played a role and classified the possible motions of these systems. This concreteness of experiments, and the possibility of analysis was too much of temptation for Mary Lucy Cartwright and J.E. Littlewood, who set out to prove that many of the structures conjectured by the experimentalists and theoretical physicists did indeed follow from the equations of motion. Birkhoff had found a "remarkable curve" in a two dimensional map; it appeared to be non-differentiable and it would be nice to see if a smooth flow could generate such a curve. The work of Cartwright and Littlewood lead to the work of Levinson, which in turn provided the basis for the horseshoe construction of Smale.

In Russia, Lyapunov paralleled the methods of Poincaré and initiated the strong Russian dynamical systems school. Andronov carried on with the study of nonlinear oscillators and in 1937 introduced together with Pontryagin the notion of coarse systems. They were formalizing the understanding garnered from the study of nonlinear oscillators, the understanding that many of the details on how these oscillators work do not affect the overall picture of the phase space: there will still be limit cycles if one changes the dissipation or spring force function by a little bit. And changing the system a little bit has the great advantage of eliminating exceptional cases in the mathematical analysis. Coarse systems were the concept that caught Smale's attention and enticed him to study dynamical systems.

A.3 Chaos with us

(R. Mainieri)

In the fall of 1961 Steven Smale was invited to Kiev where he met Arnol'd, Anosov, Sinai, and Novikov. He lectured there, and spent a lot of time with Anosov. He suggested a series of conjectures, most of which Anosov proved within a year. It was Anosov who showed that there are dynamical systems for which all points (as opposed to a non–wandering set) admit the hyperbolic structure, and it was in honor of this result that Smale named these systems Axiom-A. In Kiev Smale found a receptive audience that had been thinking about these problems. Smale's result catalyzed their thoughts and initiated a chain of developments that persisted into the

1970's.

Smale collected his results and their development in the 1967 review article on dynamical systems, entitled "Differentiable dynamical systems". There are many great ideas in this paper: the global foliation of invariant sets of the map into disjoint stable and unstable parts; the existence of a horseshoe and enumeration and ordering of all its orbits; the use of zeta functions to study dynamical systems. The emphasis of the paper is on the global properties of the dynamical system, on how to understand the topology of the orbits. Smale's account takes you from a local differential equation (in the form of vector fields) to the global topological description in terms of horseshoes.

The path traversed from ergodicity to entropy is a little more confusing. The general character of entropy was understood by Weiner, who seemed to have spoken to Shannon. In 1948 Shannon published his results on information theory, where he discusses the entropy of the shift transformation. Kolmogorov went far beyond and suggested a definition of the metric entropy of an area preserving transformation in order to classify Bernoulli shifts. The suggestion was taken by his student Sinai and the results published in 1959. In 1960 Rohlin connected these results to measure-theoretical notions of entropy. The next step was published in 1965 by Adler and Palis, and also Adler, Konheim, McAndrew; these papers showed that one could define the notion of topological entropy and use it as an invariant to classify continuous maps. In 1967 Anosov and Sinai applied the notion of entropy to the study of dynamical systems. It was in the context of studying the entropy associated to a dynamical system that Sinai introduced Markov partitions in 1968.

Markov partitions allow one to relate dynamical systems and statistical mechanics; this has been a very fruitful relationship. It adds measure notions to the topological framework laid down in Smale's paper. Markov partitions divide the phase space of the dynamical system into nice little boxes that map into each other. Each box is labeled by a code and the dynamics on the phase space maps the codes around, inducing a symbolic dynamics. From the number of boxes needed to cover all the space, Sinai was able to define the notion of entropy of a dynamical system. In 1970 Bowen came up independently with the same ideas, although there was presumably some flow of information back and forth before these papers got published. Bowen also introduced the important concept of shadowing of chaotic orbits. We do not know whether at this point the relations with statistical mechanics were clear to every one. They became explicit in the work of Ruelle. Ruelle understood that the topology of the orbits could be specified by a symbolic code, and that one could associate an "energy" to each orbit. The energies could be formally combined in a "partition function" to generate the invariant measure of the system.

After Smale, Sinai, Bowen, and Ruelle had laid the foundations of the statistical mechanics approach to chaotic systems, research turned to studying particular cases. The simplest case to consider is one-dimensional maps. The topology of the orbits for parabola-like maps was worked out in 1973 by Metropolis, Stein, and Stein. The more general one-dimensional case was worked out in 1976 by Milnor and Thurston in a widely circulated preprint, whose extended version eventually got published in 1988.

A lecture of Smale and the results of Metropolis, Stein, and Stein inspired Feigenbaum to study simple maps. This lead him to the discovery of the universality in quadratic maps and the application of ideas from field-theory to dynamical systems. Feigenbaum's work was the culmination in the study of one-dimensional systems; a complete analysis of a nontrivial transition to chaos. Feigenbaum introduced many new ideas into the field: the use of the renormalization group which lead him to introduce functional equations in the study of dynamical systems, the scaling function which completed the link between dynamical systems and statistical mechanics, and the use of presentation functions as the dynamics of scaling functions.

The work in more than one dimension progressed very slowly and is still far from completed. The first result in trying to understand the topology of the orbits in two dimensions (the equivalent of Metropolis, Stein, and Stein, or Milnor and Thurston's work) was obtained by Thurston. Around 1975 Thurston was giving lectures "On the geometry and dynamics of diffeomorphisms of surfaces". Thurston's techniques exposed in that lecture have not been applied in physics, but much of the classification that Thurston developed can be obtained from the notion of a "pruning front" developed independently by Cvitanović.

Once one develops an understanding for the topology of the orbits of a dynamical system, one needs to be able to compute its properties. Ruelle had already generalized the zeta function introduced by Artin and Mazur so that it could be used to compute the average value of observables. The difficulty with Ruelle's zeta function is that it does not converge very well. Starting out from Smale's observation that a chaotic dynamical system is dense with a set of periodic orbits, Cvitanović used these orbits as a skeleton on which to evaluate the averages of observables, and organized such calculations in terms of rapidly converging cycle expansions. This convergence is attained by using the shorter orbits used as a basis for shadowing the longer orbits.

This account is far from complete, but we hope that it will help get a sense of perspective on the field. It is not a fad and it will not die anytime soon.

Remark A.1 Notion of global foliations. For each paper cited in dynamical systems literature, there are many results that went into its development. As an example, take the notion of global foliations that we attribute to Smale. As far as we can trace the idea, it goes back to René Thom; local foliations were already used by Hadamard. Smale attended a seminar of Thom in 1958 or 1959. In that seminar Thom was explaining his notion of transversality. One of Thom's disciples introduced Smale to Brazilian mathematician Peixoto. Peixoto (who had learned the results of the Andronov-Pontryagin school from Lefschetz) was the closest Smale had ever come until then to the Andronov-Pontryagin school. It was from Peixoto that Smale learned about structural stability, a notion that got him enthusiastic about dynamical systems, as it blended well with his topological background. It was from discussions with Peixoto that Smale got the problems in dynamical systems that lead him to his 1960 paper on Morse inequalities. The next year Smale published his result on the hyperbolic structure of the nonwandering set. Smale was not the first to consider a hyperbolic point, Poincaré had already done that; but Smale was the first to introduce a global hyperbolic structure. By 1960 Smale was already lecturing on the horseshoe as a structurally stable dynamical system with an infinity of periodic points and promoting his global viewpoint.

(R. Mainieri)

Remark A.2 Levels of ergodicity. In the mid 1970's A. Katok and Ya.B. Pesin tried to use geometry to establish positive Lyapunov exponents. A. Katok and J.-M. Strelcyn carried out the program and developed a theory of general dynamical systems with singularities. They studied uniformly hyperbolic systems (as strong as Anosov's), but with sets of singularities. Under iterations a dense set of points hits the singularities. Even more important are the points that never hit the singularity set. In order to establish some control over how they approach the set, one looks at trajectories that apporach the set by some given ϵ^n, or faster.

Ya.G. Sinai, L. Bunimovich and Chernov studied the geometry of billiards in a very detailed way. A. Katok and Ya.B. Pesin's idea was much more robust. Look at the discontinuity set (geometry of it matters not at all), take an ϵ neighborhood around it. Given that the Lebesgue measure is ϵ^α and the stability grows not faster than $(\text{distance})^n$, A. Katok and J.-M. Strelcyn prove that the Lyapunov exponent is non-zero.

In mid 1980's Ya.B. Pesin studied the dissipative case. Now the problem has no invariant Lebesgue measure. Assuming uniform hyperbolicity, with singularities, and tying together Lebesgue measure and discontinutities, and given that the stability grows not faster than $(\text{distance})^n$, Ya.B. Pesin proved that the Lyapunov exponent is non-zero, and that SBR measure exists. He also proved that the Lorenz, Lozi and Byelikh attractors satisfy these conditions.

In the the systems were uniformly hyperbolic, all trouble was in differentials. For the Hénon attractor, already the differentials are nonhyperbolic. The points do not separate uniformly, but the analogue of the singularity set can be obtained by excizing the regions that do not separate. Hence there are 3 levels of ergodic systems:

1. Anosov flow
2. Anosov flow + singularity set
   - the Hamiltonian systems: general case A. Katok and J.-M. Strelcyn, billiards Ya.G. Sinai and L. Bunimovich.
   - the dissipative case: Ya.B. Pesin
3. Hénon
   - The first proof was given by M. Benedicks and L. Carleson [9.35].
   - A more readable proof is given in M. Benedicks and L.-S. Young [3.19]

(based on Ya.B. Pesin's comments)

A.3.1 Periodic orbit theory

The history of the periodic orbit theory is rich and curious, and the recent advances are to equal degree inspired by a century of separate development of three disparate subjects; 1. *classical chaotic dynamics*, initiated by Poincaré and put on its modern footing by Smale, Ruelle, and many others; 2. *quantum theory* initiated by Bohr, with the modern "chaotic" formulation by Gutzwiller; and 3. analytic *number theory* initiated by Riemann and formulated as a spectral problem by Selberg. Following totally different lines of reasoning and driven by very different motivations, the three separate roads all arrive at formally nearly identical *trace formulas*, *zeta functions* and *spectral determinants*.

That these topics should be related is far from obvious. Connection between dynamics and number theory arises from Selberg's observation that description of geodesic motion and wave mechanics on spaces of constant negative curvature is essentially a number-theoretic problem. *A posteriori*, one can say that zeta functions arise in both classical and quantum mechanics because in both the dynamical evolution can be described by the action of linear evolution (or transfer) operators on infinite-dimensional vector spaces. The spectra of these operators are given by the zeros of appropriate determinants. One way to evaluate determinants is to expand them in terms of traces, *log det = tr log*, and in this way the spectrum of an evolution operator becames related to its traces, that is, periodic orbits. A perhaps deeper way of restating this is to observe that the trace formulas perform the same service in all of the above problems; they relate the spectrum of lengths (local dynamics) to the spectrum of eigenvalues (global averages), and for nonlinear geometries they play a role analogous to that the Fourier transform plays for the circle.

A.4 Death of the Old Quantum Theory

> In 1913 Otto Stern and Max Theodor Felix von Laue went up for a walk up the Uetliberg. On the top they sat down and talked about physics. In particular they talked about the new atom model of Bohr. There and then they made the "Uetli Schwur": If that crazy model of Bohr turned out to be right, then they would leave physics. It did and they didn't.
> A. Pais, *Inward Bound: of Matter and Forces in the Physical World*

In an afternoon of May 1991 Dieter Wintgen is sitting in his office at the Niels Bohr Institute beaming with the unparalleled glee of a boy who has just committed a major mischief. The starting words of the manuscript he has just penned are

> The failure of the Copenhagen School to obtain a reasonable ...

34 years old at the time, Dieter was a scruffy kind of guy, always in sandals and holed out jeans, a left winger and a mountain climber, working around the clock with his students Gregor and Klaus to complete the work that Bohr himself would have loved to see done back in 1916: a "planetary" calculation of the helium spectrum.

Never mind that the "Copenhagen School" refers not to the old quantum theory, but to something else. The old quantum theory was no theory at all; it was a set of rules bringing some order to a set of phenomena which defied logic of classical theory. The electrons were supposed to describe planetary orbits around the nucleus; their wave aspects were yet to be discovered. The foundations seemed obscure, but Bohr's answer for the once-ionized helium to hydrogen ratio was correct to five significant figures and hard to ignore. The old quantum theory marched on, until by 1924 it reached an impasse: the helium spectrum and the Zeeman effect were its death knell.

Since the late 1890's it had been known that the helium spectrum consists of the orthohelium and parahelium lines. In 1915 Bohr suggested that the two kinds of helium lines might be associated with two distinct shapes of orbits (a suggestion that turned out to be wrong). In 1916 he got Kramers to work on the problem, and wrote to Rutherford: "I have used all my spare time in the last months to make a serious attempt to solve the problem of ordinary helium spectrum ...I think really that at last I have a clue to the problem." To other colleagues he wrote that "the theory was worked out in the fall of 1916" and of having obtained a "partial agreement with the measurements." Nevertheless, the Bohr-Sommerfeld theory, while by and large successful for hydrogen, was a disaster for neutral helium. Heroic efforts of the young generation, including Kramers and Heisenberg, were of no avail.

For a while Heisenberg thought that he had the ionization potential for helium, which he had obtained by a simple perturbative scheme. He wrote enthusiastic letters to Sommerfeld and was drawn into a collaboration with Max Born to compute the spectrum of helium using Born's systematic perturbative scheme. In first approximation, they reproduced the earlier calculations. The next level of corrections turned out to be larger than the computed effect. The concluding paragraph of Max Born's classic "Vorlesungen über Atommechanik" from 1925 sums it up in a somber tone:

> (...) the systematic application of the principles of the quantum theory (...) gives results in agreement with experiment only in those cases where the motion of a single electron is considered; it fails even in the treatment of the motion of the two electrons in the helium atom.
>
> This is not surprising, for the principles used are not really consistent. (...) A complete systematic transformation of the classical mechanics into a discontinuous mechanics is the goal towards which the quantum theory strives.

That year Heisenberg suffered a bout of hay fever, and the old quantum theory was dead. In 1926 he gave the first quantitative explanation of the helium spectrum. He used wave mechanics, electron spin and the Pauli exclusion principle, none of which belonged to the old quantum theory, and planetary orbits of electrons were cast away for nearly half a century.

Why did Pauli and Heisenberg fail with the helium atom? It was not the fault of the old quantum mechanics, but rather it reflected their lack of understanding of the subtleties of classical mechanics. Today we know what they missed in 1913-24: the role of conjugate points (topological indices) along classical trajectories was not accounted for, and they had no idea of the importance of periodic orbits in nonintegrable systems.

Since then the calculation for helium using the methods of the old quantum mechanics has been fixed. Leopold and Percival added the topological indices in 1980, and in 1991 Wintgen and collaborators orbits. Dieter had good reasons to gloat; while the rest of us were preparing to sharpen our pencils and supercomputers in order to approach the dreaded 3-body problem, they just went ahead and did it. What it took - and much else - is described in this book. One is also free to ponder what quantum theory would look like today if all this was worked out in 1917.

Remark A.3 Sources. This tale, aside from a few personal recollections, is in large part lifted from Abraham Pais' accounts of the demise of the old quantum theory [A.5, A.6], as well as Jammer's account [27.1]. The helium spectrum is taken up in chapter 29. In August 1994 Dieter Wintgen died in a climbing accident in the Swiss Alps.

References

[A.1] F. Diacu and P. Holmes, *Celestial Encounters, The Origins of Chaos and Stability* (Princeton Univ. Press, Princeton NJ 1996).

[A.2] T. Li and J. Yorke, *"Period 3 implies chaos"*, *Amer. Math. Monthly* **82**, 985 (1975).

[A.3] M. Born, *Vorlesungen über Atommechanik* (Springer, Berlin 1925). English translation: *The Mechanics of the Atom*, (F. Ungar Publishing Co., New York 1927).

[A.4] J. G. Leopold and I. Percival, *J. Phys. B*, **13**, 1037 (1980).

[A.5] A. Pais, *Inward Bound: of Matter and Forces in the Physical World* (Oxford Univ. Press, Oxford 1986).

[A.6] A. Pais, *Niels Bohr's Times, in Physics, Philosophy and Polity* (Oxford Univ. Press, Oxford 1991).

[A.7] D. Wintgen, K. Richter and G. Tanner, *CHAOS* **2**, 19 (1992).

Appendix B

Infinite-dimensional flows

Flows described by partial differential equations are considered infinite dimensional because if one writes them down as a set of ordinary differential equations (ODEs) then one needs an infinity of the ordinary kind to represent the dynamics of one equation of the partial kind (PDE). Even though the phase space is infinite dimensional, for many systems of physical interest the global attractor is finite dimensional. We illustrate how this works with a concrete example, the Kuramoto-Sivashinsky system.

B.0.1 Partial differential equations

First, a few words about partial differential equations in general. Many of the partial differential equations of mathematical physics can be written in the quasi-linear form

$$\partial_t u = Au + N(u)\,, \tag{B.1}$$

where u is a function (possibly a vector function) of the coordinate x and time t, A is a linear operator, usually containing the Laplacian and a few other derivatives of u, and $N(u)$ is the nonlinear part of the equation (terms like $u\partial_x u$ in (2.14)).

Not all equations are stated in the form (B.1), but they can easily be so transformed, just as the ordinary differential equations can be rewritten as first-order systems. We will illustrate the method with a variant of the D'Alambert's wave equation describing a plucked string:

$$\partial_{tt} y = \left(c + \frac{1}{2}\left(\partial_x y\right)^2\right)\partial_{xx} y\,. \tag{B.2}$$

Were the term $\partial_x y$ small, this equation would be just the ordinary wave equation. To rewrite the equation in the first order form (B.1), we need a field $u = (y, w)$ that is two-dimensional,

$$\partial_t \begin{bmatrix} y \\ w \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ c\partial_{xx} & 0 \end{bmatrix} \begin{bmatrix} y \\ w \end{bmatrix} + \begin{bmatrix} 0 \\ \partial_{xx} y (\partial_x y)^2/2 \end{bmatrix} . \qquad \text{(B.3)}$$

The $[2\times 2]$ matrix is the linear operator A and the vector on the far right is the nonlinear function $N(u)$. Unlike ordinary functions, differentiations are part of the function. The nonlinear part can also be expressed as a function on the infinite set of numbers that represent the field, as exemplified by the Kuramoto-Sivashinsky system (2.14).

The usual technique for solving the linear part is to use Fourier methods. Just as in the ordinary differential equation case, one can integrate the linear part of

☞ chapter 4.2

$$\partial_t u = Au \qquad \text{(B.4)}$$

to obtain

$$u(x,t) = e^{tA} u(x,0) \, . \qquad \text{(B.5)}$$

If u is expressed as Fourier series $\sum_k a_k \exp(ikx)$, as we will do for the Kuramoto-Shivashinsky system, then we can determine the action of e^{tA} on $u(x,0)$. This can be done because differentiations in A act rather simply on the exponentials. For example,

$$e^{t\partial_x} u(x,0) = e^{t\partial_x} \sum_k a_k e^{ikx} = \sum_k a_k \frac{(it)^k}{k!} e^{ikx} \, . \qquad \text{(B.6)}$$

Depending on the behavior of the linear part, one distinguishes three classes of partial differential equations: *diffusion, wave, and potential.* The classification relies on the solution by a Fourier series, as in (B.5). In mathematical literature these equations are also called *parabolic, hyperbolic* and *elliptic*. If the nonlinear part $N(u)$ is as big as the linear part, the classification is not a good indication of behavior, and one can encounter features of one class of equations while studying the others.

In *diffusion-type* equations the modes of high frequency tend to become smooth, and all initial conditions tend to an attractor, called the *inertial manifold*. The Kuramoto-Sivashinsky system studied below is of this type.

The solution being attracted to the inertial manifold does not mean that the amplitudes of all but a finite number of modes go to zero (alas were we so lucky), but that there is a finite set of modes that could be used to describe any solution of the inertial manifold. The only catch is that there is no simple way to discover what these inertial manifold modes might be.

In *wave-like* equations the high frequency modes do not die out and the solutions tend to be distributions. The equations can be solved by variations on the WKB idea: the wave-like equations can be approximated by the trajectories of the wave fronts.

☞ chapter 26

2.10
page 50

Elliptic equations have no time dependence and do not represent dynamical systems.

Appendix C

Stability of Hamiltonian flows

C.1 Symplectic invariance

(M.J. Feigenbaum and P. Cvitanović)

The symplectic structure of Hamilton's equations buys us much more than the incompressibility, or the phase space volume conservation alluded to in sect. 5.1. We assume you are at home with Hamiltonian formalism. If you would like to see the Hamilton's equations derived, Hamilton's original line of reasoning is retraced in sect. 26.1.1. The evolution equations for any p, q dependent quantity $Q = Q(p, q)$ are given by (7.33).

sect. 26.1.1

In terms of the Poisson brackets, the time evolution equation for $Q = Q(p, q)$ is given by (7.35). We now recast the symplectic condition (5.8) in a form convenient for using the symplectic constraints on $\mathbf{J}$. Writing $x(t) = x' = [p', q']$ and the Jacobian matrix and its inverse

$$\mathbf{J} = \begin{pmatrix} \frac{\partial q'}{\partial q} & \frac{\partial q'}{\partial p} \\ \frac{\partial p'}{\partial q} & \frac{\partial p'}{\partial p} \end{pmatrix}, \qquad \mathbf{J}^{-1} = \begin{pmatrix} \frac{\partial q}{\partial q'} & \frac{\partial q}{\partial p'} \\ \frac{\partial p}{\partial q'} & \frac{\partial p}{\partial p'} \end{pmatrix}, \tag{C.1}$$

we can spell out the symplectic invariance condition (5.8):

$$\begin{aligned} \frac{\partial q'_k}{\partial q_i}\frac{\partial p'_k}{\partial q_j} - \frac{\partial p'_k}{\partial q_i}\frac{\partial q'_k}{\partial q_j} &= 0 \\ \frac{\partial q'_k}{\partial p_i}\frac{\partial p'_k}{\partial p_j} - \frac{\partial p'_k}{\partial p_i}\frac{\partial q'_k}{\partial p_j} &= 0 \\ \frac{\partial q'_k}{\partial q_i}\frac{\partial p'_k}{\partial p_j} - \frac{\partial p'_k}{\partial q_i}\frac{\partial q'_k}{\partial p_j} &= \delta_{ij} \,. \end{aligned} \tag{C.2}$$

From (5.9) we obtain

$$\frac{\partial q_i}{\partial q'_j} = \frac{\partial p'_j}{\partial p_i}, \quad \frac{\partial p_i}{\partial p'_j} = \frac{\partial q'_j}{\partial q_i}, \quad \frac{\partial q_i}{\partial p'_j} = -\frac{\partial q'_j}{\partial p_i}, \quad \frac{\partial p_i}{\partial q'_j} = -\frac{\partial p'_j}{\partial q_i}. \tag{C.3}$$

Taken together, (C.3) and (C.2) imply that the flow conserves the $[p, q]$ Poisson brackets

$$\begin{aligned} [q_i, q_j] &= \frac{\partial q_i}{\partial p'_k}\frac{\partial q_j}{\partial q'_k} - \frac{\partial q_j}{\partial p'_k}\frac{\partial q_i}{\partial q'_k} = 0 \\ [p_i, p_j] &= 0\,, \qquad [p_i, q_j] = \delta_{ij}\,, \end{aligned} \tag{C.4}$$

that is, the transformations induced by a Hamiltonian flow are *canonical*, preserving the form of the equations of motion. The first two relations are symmetric under i, j interchange and yield $D(D-1)/2$ constraints each; the last relation yields D^2 constraints. Hence only $(2D)^2 - 2D(D-1)/2 - D^2 = 2D^2 + D$ elements of $\mathbf{J}$ are linearly independent, as it behooves group elements of the symplectic group $Sp(2D)$.

We have now succeeded in making the full set of constraints explicit - as we shall see in appendix D, this will enable us to implement dynamics in such a way that the symplectic invariance will be automatically preserved.

Consider the symplectic product of two infinitesimal vectors

$$\begin{aligned} (\delta x, \delta\hat{x}) &= \delta x^T \omega \delta\hat{x} = \delta p_i \delta\hat{q}_i - \delta q_i \delta\hat{p}_i \\ &= \sum_{i=1}^{D} \{\text{oriented area in the } (p_i, q_i) \text{ plane}\}\,. \end{aligned} \tag{C.5}$$

Time t later we have

$$(\delta x', \delta\hat{x}') = \delta x^T \mathbf{J}^T \omega \mathbf{J} \delta\hat{x} = \delta x^T \omega \delta\hat{x}\,.$$

This has the following geometrical meaning. We imagine there is a reference phase space point. We then define two other points infinitesimally close so that the vectors δx and $\delta\hat{x}$ describe their displacements relative to the reference point. Under the dynamics, the three points are mapped to three new points which are still infinitesimally close to one another. The meaning of the above expression is that the symplectic area spanned by the three final points is the same as that spanned by the inital points. The integral (Stokes theorem) version of this infinitesimal area invariance states that for

Hamiltonian flows the D oriented areas $\mathcal{V}_i$ bounded by D loops $\Omega\mathcal{V}_i$, one per each (p_i, q_i) plane, are separately conserved:

$$\int_{\mathcal{V}} dp \wedge dq = \oint_{\Omega\mathcal{V}} p \cdot dq = \text{invariant}\,. \tag{C.6}$$

C.2 Monodromy matrix for Hamiltonian flows

(G. Tanner)

It is not the Jacobian matrix of the flow, but the *monodromy* matrix, which enters the trace formula. This matrix gives the time dependence of a displacement perpendicular to the flow on the energy manifold. Indeed, we discover some trivial parts in the Jacobian matrix $\mathbf{J}$. An initial displacement in the direction of the flow $x = \omega\nabla H(x)$ transfers according to $\delta x(t) = x_t(t)\delta t$ with δt time independent. The projection of any displacement on δx on $\nabla H(x)$ is constant, i.e. $\nabla H(x(t))\delta x(t) = \delta E$. We get the equations of motion for the monodromy matrix directly choosing a suitable local coordinate system on the orbit $x(t)$ in form of the (non singular) transformation $\mathbf{U}(x(t))$:

$$\tilde{\mathbf{J}}(x(t)) = \mathbf{U}^{-1}(x(t))\,\mathbf{J}(x(t))\,\mathbf{U}(x(0)) \tag{C.7}$$

These lead to

$$\begin{array}{rrcl} & \dot{\tilde{\mathbf{J}}} & = & \tilde{\mathbf{L}}\,\tilde{\mathbf{J}} \\ \text{with} & \tilde{\mathbf{L}} & = & \mathbf{U}^{-1}(\mathbf{L}\mathbf{U} - \dot{\mathbf{U}}) \end{array} \tag{C.8}$$

Note that the properties a) – c) are only fulfilled for $\tilde{\mathbf{J}}$ and $\tilde{\mathbf{L}}$, if $\mathbf{U}$ itself is symplectic.
Choosing $x_E = \nabla H(t)/|\nabla H(t)|^2$ and x_t as local coordinates uncovers the two trivial eigenvalues 1 of the transformed matrix in (C.7) at any time t. Setting $\mathbf{U} = (x_t^T, x_E^T, x_1^T, \ldots, x_{2d-2}^T)$ gives

$$\tilde{\mathbf{J}} = \left(\begin{array}{cccccc} 1 & * & * & \ldots & * \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & * & & & \\ \vdots & \vdots & & \mathbf{m} & \\ 0 & * & & & \end{array}\right); \qquad \tilde{\mathbf{L}} = \left(\begin{array}{cccccc} 0 & * & * & \ldots & * \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & * & & & \\ \vdots & \vdots & & \mathbf{l} & \\ 0 & * & & & \end{array}\right), \tag{C.9}$$

The matrix $\mathbf{m}$ is now the monodromy matrix and the equation of motion are given by

$$\dot{\mathbf{m}} = \mathbf{l}\,\mathbf{m}. \tag{C.10}$$

The vectors $x_1, \dots, x_{2d-2}$ must span the space perpendicular to the flow on the energy manifold.

For a system with two degrees of freedom, the matrix $\mathbf{U(t)}$ can be written down explicitly, i.e.

$$\mathbf{U}(t) = (x_t, x_1, x_E, x_2) = \left(\begin{array}{cccc} \dot{x} & -\dot{y} & -\dot{u}/q^2 & -\dot{v}/q^2 \\ \dot{y} & \dot{x} & -\dot{v}/q^2 & \dot{u}/q^2 \\ \dot{u} & \dot{v} & \dot{x}/q^2 & -\dot{y}/q^2 \\ \dot{v} & -\dot{u} & \dot{y}/q^2 & \dot{x}/q^2 \end{array}\right) \tag{C.11}$$

with $x^T = (x, y; u, v)$ and $q = |\nabla H| = |\dot{x}|$. The matrix $\mathbf{U}$ is non singular and symplectic at every phase space point x (except the equilibrium points $\dot{x} = 0$). The matrix elements for $\mathbf{l}$ are given (C.13). One distinguishes 4 classes of eigenvalues of $\mathbf{m}$.

- *stable* or *elliptic*, if $\Lambda = e^{\pm i\pi\nu}$ and $\nu \in]0, 1[$.
- *marginal*, if $\Lambda = \pm 1$.
- *hyperbolic*, *inverse hyperbolic*, if $\Lambda = e^{\pm\lambda}$, $\Lambda = -e^{\pm\lambda}$; $\lambda > 0$ is called the Lyapunov exponent of the periodic orbit.
- *loxodromic*, if $\Lambda = e^{\pm u \pm i\Psi}$ with u and Ψ real. This is the most general case possible only in systems with 3 or more degree of freedoms.

For 2 degrees of freedom, i.e. $\mathbf{m}$ is a (2×2) matrix, the eigenvalues are determined by

$$\lambda = \frac{\mathrm{Tr}(\mathbf{m}) \pm \sqrt{\mathrm{Tr}(\mathbf{m})^2 - 4}}{2}, \tag{C.12}$$

i.e. $\mathrm{Tr}(\mathbf{m}) = 2$ separates stable and unstable behavior.

The $\mathbf{l}$ matrix elements for the local transformation (C.11) are

$$\tilde{\mathbf{l}}_{11} = \frac{1}{q}[(h_x^2 - h_y^2 - h_u^2 + h_v^2)(h_{xu} - h_{yv}) + 2(h_x h_y - h_u h_v)(h_{xv} + h_{yu})$$

$$
\begin{aligned}
& -(h_x h_u + h_y h_v)(h_{xx} + h_{yy} - h_{uu} - h_{vv})] \\
\tilde{\mathbf{l}}_{12} &= \frac{1}{q^2}[(h_x^2 + h_v^2)(h_{yy} + h_{uu}) + (h_y^2 + h_u^2)(h_{xx} + h_{vv}) \\
& -2(h_x h_u + h_y h_v)(h_{xu} + h_{yv}) - 2(h_x h_y - h_u h_v)(h_{xy} - h_{uv})] \\
\tilde{\mathbf{l}}_{21} &= -(h_x^2 + h_y^2)(h_{uu} + h_{vv}) - (h_u^2 + h_v^2)(h_{xx} + h_{yy}) \\
& +2(h_x h_u - h_y h_v)(h_{xu} - h_{yv}) + 2(h_x h_v + h_y h_u)(h_{xv} + h_{yu}) \\
\tilde{\mathbf{l}}_{22} &= -\tilde{\mathbf{l}}_{11}, \qquad \text{(C.13)}
\end{aligned}
$$

with h_i, h_{ij} is the derivative of the Hamiltonian H with respect to the phase space coordinates and $q = |\nabla H|^2$.

Remark C.1 The sign convention of the Poisson bracket. The Poisson bracket is antisymmetric in its arguments and there is a freedom to define it with the opposite sign convention. When such freedoms exist in physics, it is certain that both conventions are in use and this is no exception. In several texts you will see the right hand side of (7.34) defined as $[B, A]$ so that (7.35) is $\frac{dQ}{dt} = [Q, H]$. As long as one is consistent, there should be no problem.

Remark C.2 The sign convention of ω. The overall sign of ω, the symplectic invariant in (5.4), is set by the convention that the Hamilton's principal function (for energy conserving flows) is given by $R(q, q', t) = \int_q^{q'} p_i dq_i - Et$. With this sign convention the action along a classical path is minimal, and the kinetic energy of a free particle is positive.

Appendix D

Implementing evolution

The large body of accrued wisdom on the subject of flows called fluid dynamics is about physicaly real flows of media with (typically) continuous densities. On the other hand, the flows in state spaces of complicated systems are not only abstract, but may lead to fractal or other forms of complication in the densities. The latter frequently require more abstract tools that we develop in this text. To sharpen our intuition about those, it might be helpful to outline the more tangible fluid dynamical vision.

D.1 Material invariants

> I'm a material girl in a material world.
> Madonna, *Material Girl*

(E.A. Spiegel and P. Cvitanović)

We consider first the simplest property of a flow called a *material invariant* by fluid dynamicists, who have no choice but to ponder such matters. A material invariant $I(x)$ is a property attached to each point x that is preserved by the flow, $I(x) = I(f^t(x))$; for example, at this point the fluid is green. As $I(x) = I(f^t(x))$ is invariant, its total time derivative vanishes, $\dot{I}(x) = 0$. Written in terms of partial derivatives this is the *conservation equation* for the material invariant

$$\partial_t I + v_i \partial_i I = 0 \,. \tag{D.1}$$

Let the *density* of representative points be $\rho(x,t)$. To each representative point x we assign the value $I(x)$ of some physical property of a system.

The manner in which the flow redistributes $I(x)$ is governed by a partial differential equation whose form is relatively simple because the representative points are neither created nor destroyed. This conservation property is expressed in the integral statement

$$\partial_t \int_{\mathcal{V}} dx\, \rho I = - \int_{\partial \mathcal{V}} d\sigma\, n_i v_i \rho I \,, \tag{D.2}$$

where $\mathcal{V}$ is an arbitrary volume in $\mathcal{M}$, $\partial\mathcal{V}$ is its surface and $\mathbf{n}$ is its outward normal. Repeated indices are summed over throughout and ∂_t is the partial derivative with respect to time.

We may use the divergence theorem to turn the surface integral into a volume integral and obtain

$$\int_{\mathcal{V}} \left[\partial_t(\rho I) + \partial_i(v_i \rho I)\right] dx = 0 \,,$$

where ∂_i is the partial derivative operator with respect to x_i. Since the integration is over an arbitrary volume, we conclude that

$$\partial_t(\rho I) + \partial_i(\rho I v_i) = 0 \,. \tag{D.3}$$

The choice $I \equiv 1$ yields the *continuity equation* for the density:

$$\partial_t \rho + \partial_i(\rho v_i) = 0 \,. \tag{D.4}$$

If the density itself is a material invariant, combining (D.1) and (D.4) we conclude that $\partial_i v_i = 0$ and $\mathbf{J}^t(x_0) = 1$. An example of such incompressible flow is the Hamiltonian flow of sect. 5.1.1. For incompressible flows the continuity equation (D.4) becomes a statement of conservation of the phase-space volume (see sect. 5.1.1), or the Liouville theorem

$$\partial_t \rho + v_i \partial_i \rho = 0 \,. \tag{D.5}$$

D.2 Koopmania

The way in which time evolution acts on densities may be rephrased in the language of functional analysis, by introducing the *Koopman operator*, whose action on a phase space function $a(x)$ is to replace it by its downstream value time t later, $a(x) \to a(x(t))$ evaluated at the trajectory point $x(t)$:

$$\mathcal{K}^t a(x) \;=\; a(f^t(x))\,. \tag{D.6}$$

Observable $a(x)$ has no explicit time dependence; all time dependence is carried in its evaluation at $x(t)$ rather than at $x = x(0)$.

Suppose we are starting with an initial density of representative points $\rho(x)$: then the average value of $a(x)$ evolves as

$$\langle a \rangle(t) \;=\; \frac{1}{|\rho_{\mathcal{M}}|}\int_{\mathcal{M}} dx\, a(f^t(x))\rho(x) \;=\; \frac{1}{|\rho_{\mathcal{M}}|}\int_{\mathcal{M}} dx\;\left[\mathcal{K}^t a(x)\right]\rho(x)\,.$$

An alternative point of view (analogous to the shift from the Heisenberg to the Schrödinger picture in quantum mechanics) is to push dynamical effects into the density. In contrast to the Koopman operator which advances the trajectory by time t, the Perron-Frobenius operator (7.10) depends on the trajectory point time t in the past, so the Perron-Frobenius operator is the adjoint of the Koopman operator

$$\int_{\mathcal{M}} dx\;\left[\mathcal{K}^t a(x)\right]\,\rho(x) = \int_{\mathcal{M}} dx\, a(x)\;\left[\mathcal{L}^t \rho(x)\right]\,. \tag{D.7}$$

Checking this is an easy change of variables exercise. For finite dimensional deterministic invertible flows the Koopman operator (D.6) is simply the inverse of the Perron-Frobenius operator (7.6), so in what follows we shall not distinguish the two. However, for infinite dimensional flows contracting forward in time and for stochastic flows such inverses do not exist, and there you need to be more careful.

☞ sect. 2.4.2

The family of Koopman's operators $\left\{\mathcal{K}^t\right\}_{t\in\mathbb{R}_+}$ forms a semigroup parametrized by time

(a) $\mathcal{K}^0 = I$

(b) $\mathcal{K}^t\mathcal{K}^{t'} = \mathcal{K}^{t+t'} \qquad t, t' \geq 0 \qquad$ (semigroup property) ,

with the *generator* of the semigroup, the generator of infinitesimal time translations defined by

$$\mathcal{A} \;=\; \lim_{t\to 0^+}\frac{1}{t}\left(\mathcal{K}^t - I\right)\,.$$

(If the flow is finite-dimensional and invertible, $\mathcal{A}$ is a generator of a group). The explicit form of $\mathcal{A}$ follows from expanding dynamical evolution up to first order, as in (2.4):

$$\mathcal{A}a(x) \;=\; \lim_{t\to 0^+} \frac{1}{t}\left(a(f^t(x)) - a(x)\right) \;=\; v_i(x)\partial_i a(x)\,. \tag{D.8}$$

Of course, that is nothing but the definition of the time derivative, so the equation of motion for $a(x)$ is

$$\left(\frac{d}{dt} - \mathcal{A}\right) a(x) \;=\; 0\,. \tag{D.9}$$

☞ appendix D.3

The finite time Koopman operator (D.6) can be formally expressed by exponentiating the time evolution generator $\mathcal{A}$ as

$$\mathcal{K}^t = e^{t\mathcal{A}}\,. \tag{D.10}$$

✎ 7.10 page 123

The generator $\mathcal{A}$ looks very much like the generator of translations. Indeed, for a constant velocity field dynamical evolution is nothing but a translation by time × velocity:

✎ 7.11 page 123

$$e^{tv\frac{\partial}{\partial x}} a(x) = a(x+tv)\,. \tag{D.11}$$

As we will not need to implement a computational formula for general $e^{t\mathcal{A}}$ in what follows, we relegate making sense of such operators to appendix D.3. Here we limit ourselves to a brief remark about the notion of "spectrum" of a linear operator.

☞ appendix D.3

The Koopman operator $\mathcal{K}$ acts multiplicatively in time, so it is reasonable to suppose that there exist constants $M > 0$, $\beta \geq 0$ such that $||\mathcal{K}^t|| \leq Me^{t\beta}$ for all $t \geq 0$. What does that mean? The operator norm is define in the same spirit in which we defined the matrix norms in sect. K.2: We are assuming that no value of $\mathcal{K}^t\rho(x)$ grows faster than exponentially for any choice of function $\rho(x)$, so that the fastest possible growth can be bounded by $e^{t\beta}$, a reasonable expectation in the light of the simplest example studied so far, the exact escape rate (7.14). If that is so, multiplying $\mathcal{K}^t$ by $e^{-t\beta}$ we construct a new operator $e^{-t\beta}\mathcal{K}^t = e^{t(\mathcal{A}-\beta)}$ which decays exponentially for large t, $||e^{t(\mathcal{A}-\beta)}|| \leq M$. We say that $e^{-t\beta}\mathcal{K}^t$ is an element of a *bounded* semigroup with generator $\mathcal{A} - \beta I$. Given this bound, it follows by the Laplace transform

$$\int_0^\infty dt\, e^{-st}\mathcal{K}^t = \frac{1}{s-\mathcal{A}}\,, \qquad \mathrm{Re}\, s > \beta\,, \tag{D.12}$$

sect. K.2

that the *resolvent* operator $(s-\mathcal{A})^{-1}$ is bounded ("resolvent" = able to cause separation into constituents)

$$\left\|\frac{1}{s-\mathcal{A}}\right\| \leq \int_0^\infty dt\, e^{-st} M e^{t\beta} = \frac{M}{s-\beta}\,.$$

If one is interested in the spectrum of $\mathcal{K}$, as we will be, the resolvent operator is a natural object to study. The main lesson of this brief aside is that for the continuous time flows the Laplace transform is the tool that brings down the generator in (7.29) into the resolvent form (7.31) and enables us to study its spectrum.

D.3 Implementing evolution

(R. Artuso and P. Cvitanović)

We now come back to the semigroup of operators $\mathcal{K}^t$. We have introduced the generator of the semigroup (7.26) as

$$\mathcal{A} = \left.\frac{d}{dt}\mathcal{K}^t\right|_{t=0}.$$

If we now take the derivative at arbitrary times we get

$$\begin{aligned}\left(\frac{d}{dt}\mathcal{K}^t\psi\right)(x) &= \lim_{\eta\to 0}\frac{\psi(f^{t+\eta}(x))-\psi(f^t(x))}{\eta}\\ &= v_i(f^t(x))\left.\frac{\partial}{\partial\tilde{x}_i}\psi(\tilde{x})\right|_{\tilde{x}=f^t(x)}\\ &= \left(\mathcal{K}^t\mathcal{A}\psi\right)(x)\end{aligned}$$

which can be formally integrated like an ordinary differential equation yielding

7.10
page 123

$$\mathcal{K}^t = e^{t\mathcal{A}}\,. \tag{D.13}$$

This guarantees that the Laplace transform manipulations in sect. 7.4 are correct. Though the formal expression of the semigroup (D.13) is quite simple one has to take care in implementing its action. If we express the exponential through the power series

$$\mathcal{K}^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathcal{A}^k \,, \tag{D.14}$$

we encounter the problem that the infinitesimal generator (7.26) contains non-commuting pieces, that is, there are i, j combinations for which the commutator does not satisfy

$$\left[\frac{\partial}{\partial x_i}, v_j(x) \right] = 0 \,.$$

To derive a more useful representation, we follow the strategy used for finite-dimensional matrix operators in sects. 4.2 and 4.3 and use the semigroup property to write

$$\mathcal{K}^t = \prod_{m=1}^{t/\delta\tau} \mathcal{K}^{\delta\tau}$$

as the starting point for a discretized approximation to the continuous time dynamics, with time step $\delta\tau$. Omitting terms from the second order onwards in the expansion of $\mathcal{K}^{\delta\tau}$ yields an error of order $O(\delta\tau^2)$. This might be acceptable if the time step $\delta\tau$ is sufficiently small. In practice we write the Euler product

$$\mathcal{K}^t = \prod_{m=1}^{t/\delta\tau} \left(1 + \delta\tau \mathcal{A}_{(m)} \right) + O(\delta\tau^2) \tag{D.15}$$

where

$$\left(\mathcal{A}_{(m)} \psi \right)(x) = v_i(f^{m\delta\tau}(x)) \left. \frac{\partial \psi}{\partial \tilde{x}_i} \right|_{\tilde{x} = f^{m\delta\tau}(x)}$$

As far as the x dependence is concerned, $e^{\delta\tau \mathcal{A}_i}$ acts as

$$e^{\delta\tau \mathcal{A}_i} \left\{ \begin{array}{c} x_1 \\ \cdot \\ x_i \\ x_d \end{array} \right\} \to \left\{ \begin{array}{c} x_1 \\ \cdot \\ x_i + \delta\tau v_i(x) \\ x_d \end{array} \right\} . \tag{D.16}$$

2.6 page 49

We see that the product form (D.15) of the operator is nothing else but a prescription for finite time step integration of the equations of motion - in this case the simplest Euler type integrator which advances the trajectory by $\delta\tau\times$velocity at each time step.

D.3.1 A symplectic integrator

The procedure we described above is only a starting point for more sophisticated approximations. As an example on how to get a sharper bound on the error term consider the Hamiltonian flow $\mathcal{A} = \mathcal{B} + \mathcal{C}$, $\mathcal{B} = p_i \frac{\partial}{\partial q_i}$, $\mathcal{C} = -\partial_i V(q) \frac{\partial}{\partial p_i}$. Clearly the potential and the kinetic parts do not commute. We make sense of the formal solution (D.15) by spliting it into infinitesimal steps and keeping terms up to $\delta\tau^2$ in

D.2 page 636

$$\mathcal{K}^{\delta\tau} = \hat{\mathcal{K}}^{\delta\tau} + \frac{1}{24}(\delta\tau)^3[\mathcal{B} + 2\mathcal{C}, [\mathcal{B}, \mathcal{C}]] + \cdots \,, \tag{D.17}$$

where

$$\hat{\mathcal{K}}^{\delta\tau} = e^{\frac{1}{2}\delta\tau\mathcal{B}} e^{\delta\tau\mathcal{C}} e^{\frac{1}{2}\delta\tau\mathcal{B}} \,. \tag{D.18}$$

The approximate infinitesimal Liouville operator $\hat{\mathcal{K}}^{\delta\tau}$ is of the form that now generates evolution as a sequence of mappings induced by (7.30), a free flight by $\frac{1}{2}\delta\tau\mathcal{B}$, scattering by $\delta\tau\partial V(q')$, followed again by $\frac{1}{2}\delta\tau\mathcal{B}$ free flight:

$$\begin{array}{rclcl}
e^{\frac{1}{2}\delta\tau\mathcal{B}} \begin{Bmatrix} q \\ p \end{Bmatrix} & \rightarrow & \begin{Bmatrix} q' \\ p' \end{Bmatrix} & = & \begin{Bmatrix} q - \frac{\delta\tau}{2} p \\ p \end{Bmatrix} \\
e^{\delta\tau\mathcal{C}} \begin{Bmatrix} q' \\ p' \end{Bmatrix} & \rightarrow & \begin{Bmatrix} q'' \\ p'' \end{Bmatrix} & = & \begin{Bmatrix} q' \\ p' + \delta\tau\partial V(q') \end{Bmatrix} \\
e^{\frac{1}{2}\delta\tau\mathcal{B}} \begin{Bmatrix} q'' \\ p'' \end{Bmatrix} & \rightarrow & \begin{Bmatrix} q''' \\ p''' \end{Bmatrix} & = & \begin{Bmatrix} q' - \frac{\delta\tau}{2} p'' \\ p'' \end{Bmatrix}
\end{array} \tag{D.19}$$

Collecting the terms we obtain an integration rule for this type of symplectic flow which is better than the straight Euler integration (D.16) as it is accurate up to order $\delta\tau^2$:

$$\begin{aligned}
q_{n+1} &= q_n - \delta\tau\, p_n - \frac{(\delta\tau)^2}{2}\partial V\,(q_n - \delta\tau p_n/2) \\
p_{n+1} &= p_n + \delta\tau\partial V\,(q_n - \delta\tau p_n/2)
\end{aligned} \tag{D.20}$$

The Jacobian matrix of one integration step is given by

$$\mathbf{J} = \begin{pmatrix} 1 & -\delta\tau/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \delta\tau\partial V(q') & 1 \end{pmatrix} \begin{pmatrix} 1 & -\delta\tau/2 \\ 0 & 1 \end{pmatrix} . \tag{D.21}$$

Note that the billiard flow (5.22) is an example of such symplectic integrator. In that case the free flight is interupted by instantaneous wall reflections, and can be integrated out.

Commentary

Remark D.1 Koopman operators. The "Heisenberg picture" in dynamical system theory has been introduced by Koopman refs. [D.2, D.3], see also ref. [D.1]. Inspired by the contemporary advances in quantum mechanics, Koopman [D.2] observed in 1931 that $\mathcal{K}^t$ is unitary on $L^2(\mu)$ Hilbert spaces. The Liouville/Koopman operator is the classical analogue of the quantum evolution operator — the kernel of $\mathcal{L}^t(y, x)$ introduced in (7.16) (see also sect. 8.2) is the analogue of the Green's function. The relation between the spectrum of the Koopman operator and classical ergodicity was formalized by von Neumann [D.3]. We shall not use Hilbert spaces here and the operators that we shall study *will not* be unitary. For a discussion of the relation between the Perron-Frobenius operators and the Koopman operators for finite dimensional deterministic invertible flows, infinite dimensional contracting flows, and stochastic flows, see Lasota-Mackey [D.1] and Gaspard [1.4].

Remark D.2 Symplectic integration. The reviews [D.6] and [D.7] offer a good starting point for exploring the symplectic integrators literature.

References

[D.1] A. Lasota and M.C. Mackey, *Chaos, Fractals and Noise* (Springer, New York 1994).

[D.2] B.O. Koopman, *Proc. Nat. Acad. Sci. USA* **17**, 315 (1931).

[D.3] J. von Neumann, *Ann. Math.* **33**, 587 (1932).

[D.4] B.A. Shadwick, J.C. Bowman, and P.J. Morrison, *Exactly Conservative Integrators*, `chao-dyn/9507012`, Submitted to *SIAM J. Sci. Comput.*

[D.5] D.J.D. Earn, *Symplectic integration without roundoff error*, `astro-ph/9408024`.

[D.6] P.J. Channell and C. Scovel, *Nonlinearity* **3**, 231 (1990).

[D.7] J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian problems* (Chapman and Hall, London, 1994).

Exercises

Exercise D.1 Symplectic volume preservation. *Check that the sequence of mappings (D.19) is volume preserving,* $\det \hat{U} = 1$.

Exercise D.2 Noncommutativity. *Check that the commutators in (D.17) are not vanishing by showing that*

$$[\mathcal{B}, \mathcal{C}] = -p \left(V'' \frac{\partial}{\partial p} - V' \frac{\partial}{\partial q} \right) .$$

Exercise D.3 Symplectic leapfrog integrator. *Implement (D.20) for 2-dimensional Hamiltonian flows; compare it with Runge-Kutta integrator by integrating trajectories in some (chaotic) Hamiltonian flow.*

Appendix E

Symbolic dynamics techniques

The kneading theory for unimodal mappings is developed in sect. E.1. The prime factorization for dynamical itineraries of sect. E.2 illustrates the sense in which prime cycles are "prime" - the product structure of zeta functions is a consequence of the unique factorization property of symbol sequences.

E.1 Topological zeta functions for infinite subshifts

(P. Dahlqvist)

The Markov graph methods outlined in chapter 9 are well suited for symbolic dynamics of finite subshift type. A sequence of well defined rules leads to the answer, the topological zeta function, which turns out to be a polynomial. For infinite subshifts one would have to go through an infinite sequence of graph constructions and it is of course very difficult to make any asymptotic statements about the outcome. Luckily, for some simple systems the goal can be reached by much simpler means. This is the case for unimodal maps.

We will restrict our attention to the topological zeta function for unimodal maps with one external parameter $f_\Lambda(x) = \Lambda g(x)$. As usual, symbolic dynamics is introduced by mapping a time series $\ldots x_{i-1}x_ix_{i+1}\ldots$ onto a sequence of symbols $\ldots s_{i-1}s_is_{i+1}\ldots$ where

$$\begin{aligned} s_i = 0 &\quad x_i < x_c \\ s_i = C &\quad x_i = x_c \\ s_i = 1 &\quad x_i > x_c \end{aligned} \tag{E.1}$$

$I(C)$	$\zeta_{top}^{-1}(z)/(1-z)$	$I(C)$	$\zeta_{top}^{-1}(z)/(1-z)$
$1C$		$1001C$	
$101C$		$100111C$	
$1011101C$		$10011C$	
$H^\infty(1)$	$\prod_{n=0}^{\infty}(1-z^{2^n})$	$100110C$	
$10111C$		$100C$	
$1011111C$		$100010C$	
101^∞	$(1-2z^2)/(1+z)$	$10001C$	
$10111111C$		$100011C$	
$101111C$		$1000C$	
$1011C$		$100001C$	
$101101C$		$10000C$	
$10C$	$(1-z-z^2)$	$100000C$	
$10010C$		10^∞	$(1-2z)/(1-z)$
$100101C$			

Table E.1: All ordered kneading sequences up to length seven, as well as some longer kneading sequences. Harmonic extension $H^\infty(1)$ is defined below.

and x_c is the critical point of the map (that is maximum of g). In addition to the usual binary alphabet we have added a symbol C for the critical point. The kneading sequence K_Λ is the itinerary of the critical point. The crucial observation is that no periodic orbit can have a topological coordinate (see sect. E.1.1) beyond that of the kneading sequence. The kneading sequence thus inserts a border in the list of periodic orbits (ordered according to maximal topological coordinate), cycles up to this limit are allowed, all beyond are pruned. All unimodal maps (obeying some further constraints) with the same kneading sequence thus have the same set of periodic orbitsand the same topological zeta function. The topological coordinate of the kneading sequence increases with increasing Λ.

The kneading sequence can be of one of three types

1. It maps to the critical point again, after n iterations. If so, we adopt the convention to terminate the kneading sequence with a C, and refer to the kneading sequence as finite.
2. Preperiodic, that is it is infinite but with a periodic tail.
3. Aperiodic.

As an archetype unimodal map we will choose the *tent map*

$$x \mapsto f(x) = \begin{cases} \Lambda x & x \in [0,1/2] \\ \Lambda(1-x) & x \in (1/2,1] \end{cases} , \tag{E.2}$$

where the parameter $\Lambda \in (1,2]$. The topological entropy is $h = \log \Lambda$. This follows from the fact any trajectory of the map is bounded, the escape rate is strictly zero, and so the dynamical zeta function

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right) = \prod_p \left(1 - \left(\frac{z}{\Lambda}\right)^{n_p}\right) = 1/\zeta_{\mathrm{top}}(z/\Lambda)$$

has its leading zero at $z = 1$.

The set of periodic points of the tent map is countable. A consequence of this fact is that the set of parameter values for which the kneading sequence is periodic or preperiodic are countable and thus of measure zero and consequently *the kneading sequence is aperiodic for almost all* Λ. For general unimodal maps the corresponding statement is that the kneading sequence is aperiodic for almost all topological entropies.

For a given periodic kneading sequence of period n, $\underline{K}_\Lambda = PC = s_1 s_2 \ldots s_{n-1} C$ there is a simple expansion for the topological zeta function. Then the expanded zeta function is a polynomial of degree n

$$1/\zeta_{\mathrm{top}}(z) = \prod_p (1 - z_p^n) = (1 - z) \sum_{i=0}^{n-1} a_i z^i \ , \qquad a_i = \prod_{j=1}^{i} (-1)^{s_j} \quad \text{(E.3)}$$

and $a_0 = 1$.

Aperiodic and preperiodic kneading sequences are accounted for by simply replacing n by ∞.

Example. Consider as an example the kneading sequence $K_\Lambda = 10C$. From (E.3) we get the topological zeta function $1/\zeta_{top}(z) = (1 - z)(1 - z - z^2)$, see table E.1. This can also be realized by redefining the alphabet. The only forbidden subsequence is 100. All allowed periodic orbits, except $\overline{0}$, can can be built from a alphabet with letters $\underline{10}$ and $\underline{1}$. We write this alphabet as $\{\underline{10}, \underline{1}; \overline{0}\}$, yielding the topological zeta function $1/\zeta_{top}(z) = (1 - z)(1 - z - z^2)$. The leading zero is the inverse golden mean $z_0 = (\sqrt{5} - 1)/2$.

Example. As another example we consider the preperiodic kneading sequence $K_\Lambda = 101^\infty$. From (E.3) we get the topological zeta function $1/\zeta_{top}(z) = (1 - z)(1 - 2z^2)/(1 + z)$, see table E.1. This can again be realized by redefining the alphabet. There are now an infinite number of forbidden subsequences, namely $101^{2n}0$ where $n \geq 0$. These pruning rules are respected by the alphabet $\{\underline{01^{2n+1}}; \overline{1}, \overline{0}\}$, yielding the topological zeta function above. The pole in the zeta function $\zeta^{-1}_{top}(z)$ is a consequence of the infinite alphabet.

An important consequence of (E.3) is that the sequence $\{a_i\}$ has a periodic tail if and only if the kneading sequence has one (however, their period may differ by a factor of two). We know already that the kneading sequence is aperiodic for almost all Λ.

The analytic structure of the function represented by the infinite series $\sum a_i z_i$ with unity as radius of convergence, depends on whether the tail of $\{a_i\}$ is periodic or not. If the period of the tail is N we can write

$$1/\zeta_{top}(z) = p(z) + q(z)(1 + z^N + z^{2N} \ldots) = p(z) + \frac{q(z)}{1 - z^N} \quad ,$$

for some polynomials $p(z)$ and $q(z)$. The result is a set of poles spread out along the unit circle. This applies to the preperiodic case. An aperiodic sequence of coefficients would formally correspond to infinite N and it is natural to assume that the singularities will fill the unit circle. There is indeed a theorem ensuring that this is the case [9.37], provided the a_i's can only take on a finite number of values. The unit circle becomes a *natural boundary*, already apparent in a finite polynomial approximations to the topological zeta function, as in fig. 10.4. A function with a natural boundary lacks an analytic continuation outside it.

To conclude: The topological zeta function $1/\zeta_{top}$ for unimodal maps has the unit circle as a natural boundary for almost all topological entropies and for the tent map (E.2), for almost all Λ.

Let us now focus on the relation between the analytic structure of the topological zeta function and the number of periodic orbits, or rather (10.6), the number N_n of fixed points of $f^n(x)$. The trace formula is (see sect. 10.4)

$$N_n = \operatorname{tr} T^n = \frac{1}{2\pi i} \oint_{\gamma_r} dz\, z^{-n} \frac{d}{dz} \log \zeta_{top}^{-1}$$

where γ_r is a (circular) contour encircling the origin $z = 0$ in clockwise direction. Residue calculus turns this into a sum over zeros z_0 and poles z_p of ζ_{top}^{-1}

$$N_n = \sum_{z_0 : r < |z_0| < R} z_0^{-n} - \sum_{z_p : r < |z_p| < R} z_0^{-n} + \frac{1}{2\pi i} \oint_{\gamma_R} dz\, z^{-n} \frac{d}{dz} \log \zeta_{top}^{-1}$$

and a contribution from a large circle γ_R. For meromorphic topological zeta functions one may let $R \to \infty$ with vanishing contribution from γ_R, and N_n will be a sum of exponentials.

The leading zero is associated with the topological entropy, as discussed in chapter 10.

We have also seen that for preperiodic kneading there will be poles on the unit circle.

To appreciate the role of natural boundaries we will consider a (very) special example. Cascades of period doublings is a central concept for the description of unimodal maps. This motivates a close study of the function

$$\Xi(z) = \prod_{n=0}^{\infty} (1 - z^{2^n}) \quad . \tag{E.4}$$

This function will appear again when we derive (E.3).

The expansion of $\Xi(z)$ begins as $\Xi(z) = 1 - z - z^2 + z^3 - z^4 + z^5 \ldots$. The radius of convergence is obviously unity. The simple rule governing the expansion will effectively prohibit any periodicity among the coefficients making the unit circle a natural boundary.

It is easy to see that $\Xi(z) = 0$ if $z = \exp(2\pi m/2^n)$ for any integer m and n. (Strictly speaking we mean that $\Xi(z) \to 0$ when $z \to \exp(2\pi m/2^n)$ from inside). Consequently, zeros are dense on the unit circle. One can also show that singular points are dense on the unit circle, for instance $|\Xi(z)| \to \infty$ when $z \to \exp(2\pi m/3^n)$ for any integer m and n.

As an example, the topological zeta function at the accumulation point of the first Feigenbaum cascade is $\zeta_{top}^{-1}(z) = (1 - z)\Xi(z)$. Then $N_n = 2^{l+1}$ if $n = 2^l$, otherwise $N_n = 0$. The growth rate in the number of cycles is anything but exponential. It is clear that N_n cannot be a sum of exponentials, the contour γ_R cannot be pushed away to infinity, R is restricted to $R \leq 1$ and N_n is entirely determined by $\int_{\gamma_R}$ which picks up its contribution from the natural boundary.

We have so far studied the analytic structure for some special cases and we know that the unit circle is a natural boundary for almost all Λ. But how does it look out there in the complex plane for some typical parameter values? To explore that we will imagine a journey from the origin $z = 0$ out towards the unit circle. While traveling we let the parameter Λ change slowly. The trip will have a distinct science fiction flavor. The first zero we encounter is the one connected to the topological entropy. Obviously it moves smoothly and slowly. When we move outward to the unit circle we encounter zeros in increasing densities. The closer to the unit circle they are, the wilder and stranger they move. They move from and back to the horizon, where they are created and destroyed through bizarre bifurcations. For some special values of the parameter the unit circle suddenly gets transparent and and we get (infinitely) short glimpses of another world beyond the horizon.

We end this section by deriving eqs (E.5) and (E.6). The impenetrable prose is hopefully explained by the accompanying tables.

We know one thing from chapter 9, namely for that finite kneading sequence of length n the topological polynomial is of degree n. The graph contains a node which is connected to itself only via the symbol 0. This implies that a factor $(1-z)$ may be factored out and $\zeta_{top}(z) = (1-z)\sum_{i=0}^{n-1} a_i z^i$. The problem is to find the coefficients a_i.

periodic orbits	finite kneading sequences
$\overline{P1} = A^\infty(P)$	
	PC
$\overline{P0}$	
	$P0PC$
$\overline{P0P1}$	
	$P0P1P0PC$
$\downarrow$	$\downarrow$
$H^\infty(P)$	$H^\infty(P)$

Table E.2: Relation between periodic orbits and finite kneading sequences in a harmonic cascade. The string P is assumed to contain an odd number of 1's.

The ordered list of (finite) kneading sequences table E.1 and the ordered list of periodic orbits (on maximal form) are intimately related. In table E.2 we indicate how they are nested during a period doubling cascade. Every finite kneading sequence PC is bracketed by two periodic orbits, $\overline{P1}$ and $\overline{P0}$. We have $\overline{P1} < PC < \overline{P0}$ if P contains an odd number of 1's, and $\overline{P0} < PC < \overline{P1}$ otherwise. From now on we will assume that P contains an odd number of 1's. The other case can be worked out in complete analogy. The first and second harmonic of PC are displayed in table E.2. The periodic orbit $\overline{P1}$ (and the corresponding infinite kneading sequence) is sometimes referred to as the antiharmonic extension of PC (denoted $A^\infty(P)$) and the accumulation point of the cascade is called the harmonic extension of PC [9.14] (denoted $H^\infty(P)$).

A central result is the fact that a period doubling cascade of PC is not interfered by any other sequence. Another way to express this is that a kneading sequence PC and its harmonic are adjacent in the list of kneading sequences to any order.

Table E.3 illustrates another central result in the combinatorics of kneading sequences. We suppose that P_1C and P_2C are neighbors in the list of order 5 (meaning that the shortest finite kneading sequence $P'C$ between P_1C and P_2C is longer than 5.) The important result is that P' (of length

$I(C)$			$\zeta_{top}^{-1}(z)/(1-z)$
P_1	$=$	$100C$	$1-z-z^2-z^3$
$H^\infty(P_1)$	$=$	$10001001100\ldots$	$1-z-z^2-z^3-z^4+z^5+z^6+z^7-z^8\ldots$
P'	$=$	$10001C$	$1-z-z^2-z^3-z^4+z^5$
$A^\infty(P_2)$	$=$	$1000110001\ldots$	$1-z-z^2-z^3-z^4+z^5-z^6-z^7-z^8\ldots$
P_2	$=$	$1000C$	$1-z-z^2-z^3-z^4$

Table E.3: Example of a step in the iterative construction of the list of kneading sequences PC.

$n'=6$) has to coincide with the first $n'-1$ letters of both $H^\infty(P_1)$ and $A^\infty(P_2)$. This is exemplified in the left column of table E.3. This fact makes it possible to generate the list of kneading sequences in an iterative way.

The zeta function at the accumulation point $H^\infty(P_1)$ is

$$\zeta_{P_1}^{-1}(z)\Xi(z^{n_1}) \quad , \tag{E.5}$$

and just before $A^\infty(P_2)$

$$\zeta_{P_2}^{-1}(z)/(1-z^{n_2}) \quad . \tag{E.6}$$

A short calculation shows that this is exactly what one would obtain by applying (E.3) to the antiharmonic and harmonic extensions directly, provided that it applies to $\zeta_{P_1}^{-1}(z)$ and $\zeta_{P_2}^{-1}(z)$. This is the key observation.

Recall now the product representation of the zeta function $\zeta^{-1}=\prod_p(1-z^{n_p})$. We will now make use of the fact that the zeta function associated with $P'C$ is a polynomial of order n'. There is no periodic orbit of length shorter than $n'+1$ between $H^\infty(P_1)$ and $A^\infty(P_2)$. It thus follows that the coefficients of this polynomial coincides with those of (E.5) and (E.6), see Table E.3. We can thus conclude that our rule can be applied directly to $P'C$.

This can be used as an induction step in proving that the rule can be applied to every finite and infinite kneading sequences.

Remark E.1 How to prove things. The explicit relation between the kneading sequence and the coefficients of the topological zeta function is not commonly seen in the literature. The result can proven by combining some theorems of Milnor and Thurston [9.16]. That approach is hardly instructive in the present context. Our derivation was inspired by Metropolis, Stein and Stein classical paper [9.14]. For further detail, consult [9.36].

E.1.1 Periodic orbits of unimodal maps

A *periodic point* (or a *cycle* point) x_i belonging to a cycle of period n is a real solution of

$$f^n(x_i) = f(f(\dots f(x_i)\dots)) = x_i\,, \quad i = 0, 1, 2, \dots, n-1 \tag{E.7}$$

The nth iterate of a unimodal map crosses the diagonal at most 2^n times. Similarly, the backward and the forward Smale horseshoes intersect at most 2^n times, and therefore there will be 2^n or fewer periodic points of length n. A cycle of length n corresponds to an infinite repetition of a length n symbol string, customarily indicated by a line over the string:

$$S = (s_1 s_2 s_3 \dots s_n)^\infty = \overline{s_1 s_2 s_3 \dots s_n}\,.$$

If $\overline{s_1 s_2 \dots s_n}$ is the symbol string associated with x_0, its cyclic permutation $\overline{s_k s_{k+1} \dots s_n s_1 \dots s_{k-1}}$ corresponds to the point x_{k-1} in the same cycle. A cycle p is called *prime* if its itinerary S cannot be written as a repetition of a shorter block S'.

Each cycle yields n rational values of γ. The repeating string $s_1, s_2, \dots s_n$ contains an odd number "1"s, the string of well ordered symbols $w_1 w_2 \dots w_n$ has to be of the double length before it repeats itself. The value γ is a geometrical sum which we can write as the finite sum

$$\gamma(\overline{s_1 s_2 \dots s_n}) = \frac{2^{2n}}{2^{2n}-1} \sum_{t=1}^{2n} w_t/2^t$$

Using this we can calculate the $\hat{\gamma}(S)$ for all short cycles.

Here we give explicit formulas for the topological coordinate of a periodic point, given its itinerary. For the purpose of what follows it is convenient to compactify the itineraries by replacing the binary alphabet $s_i = \{0, 1\}$ by the infinite alphabet

$$\{a_1, a_2, a_3, a_4, \cdots; \overline{0}\} = \{1, 10, 100, 1000, \dots; \overline{0}\}\,. \tag{E.8}$$

In this notation the itinerary $S = a_i a_j a_k a_l \cdots$ and the corresponding topological coordinate (9.13) are related by $\gamma(S) = .1^i 0^j 1^k 0^l \cdots$. For example:

$$\begin{array}{rclcl} S & = & 111011101001000\dots & = & a_1 a_1 a_2 a_1 a_1 a_2 a_3 a_4 \dots \\ \gamma(S) & = & .101101001110000\dots & = & .1^1 0^1 1^2 0^1 1^1 0^2 1^3 0^4 \dots \end{array}$$

Cycle points whose itineraries start with $w_1 = w_2 = \ldots = w_i = 0$, $w_{i+1} = 1$ remain on the left branch of the tent map for i iterations, and satisfy $\gamma(0 \ldots 0S) = \gamma(S)/2^i$.

A *periodic point* (or a *cycle* point) x_i belonging to a cycle of period n is a real solution of

$$f^n(x_i) = f(f(\ldots f(x_i)\ldots)) = x_i\,, \quad i = 0, 1, 2, \ldots, n-1\,. \tag{E.9}$$

The nth iterate of a unimodal map has at most 2^n monotone segments, and therefore there will be 2^n or fewer periodic points of length n. A periodic orbit of length n corresponds to an infinite repetition of a length n symbol string, customarily indicated by a line over the string:

$$S = (s_1 s_2 s_3 \ldots s_n)^\infty = \overline{s_1 s_2 s_3 \ldots s_n}\,.$$

As all itineraries are infinite, we shall adopt convention that a finite string itinerary $S = s_1 s_2 s_3 \ldots s_n$ stands for infinite repetition of a finite block, and routinely omit the overline. If $\overline{s_1 s_2 \ldots s_n}$ is the symbol string associated with x_0, its cyclic permutation $\overline{s_k s_{k+1} \ldots s_n s_1 \ldots s_{k-1}}$ corresponds to the point x_{k-1} in the same cycle. A periodic orbit p is called *prime* if its itinerary S cannot be written as a repetition of a shorter block S'.

Periodic points correspond to rational values of γ, but we have to distinguish *even* and *odd* cycles. The even (odd) cycles contain even (odd) number of a_i in the repeating block, with periodic points given by

$$\gamma(a_i a_j \cdots a_k a_\ell) = \begin{cases} \frac{2^n}{2^n-1} .1^i 0^j \cdots 1^k & \text{even} \\ \frac{1}{2^n+1}\left(1 + 2^n \times .1^i 0^j \cdots 1^\ell\right) & \text{odd} \end{cases}, \tag{E.10}$$

where $n = i + j + \cdots + k + \ell$ is the cycle period. The maximal value cycle point is given by the cyclic permutation of S with the largest a_i as the first symbol, followed by the smallest available a_j as the next symbol, and so on. For example:

$$\begin{array}{lllllllll} \hat\gamma(1) &=& \gamma(a_1) &=& .10101\ldots &=& .\overline{10} &=& 2/3 \\ \hat\gamma(10) &=& \gamma(a_2) &=& .1^2 0^2 \ldots &=& .\overline{1100} &=& 4/5 \\ \hat\gamma(100) &=& \gamma(a_3) &=& .1^3 0^3 \ldots &=& .\overline{111000} &=& 8/9 \\ \hat\gamma(101) &=& \gamma(a_2 a_1) &=& .1^2 0^1 \ldots &=& .\overline{110} &=& 6/7 \end{array}$$

An example of a cycle where only the third symbol determines the maximal value cycle point is

$$\hat\gamma(1101110) = \gamma(a_2 a_1 a_2 a_1 a_1) = .\overline{11011010010010} = 100/129\,.$$

Maximal values of all cycles up to length 5 are given in table!?

E.2 Prime factorization for dynamical itineraries

The Möbius function is not only a number-theoretic function, but can be used to manipulate ordered sets of noncommuting objects such as symbol strings. Let $\mathcal{P} = \{p_1, p_2, p_3, \cdots\}$ be an ordered set of *prime* strings, and

$$\mathcal{N} = \{n\} = \left\{ p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_j^{k_j} \right\} ,$$

$j \in \mathbb{N}$, $k_i \in \mathbb{Z}_+$, be the set of all strings n obtained by the ordered concatenation of the "primes" p_i. By construction, every string n has a unique prime factorization. We say that a string has a divisor d if it contains d as a substring, and define the string division n/d as n with the substring d deleted. Now we can do things like this: defining $t_n := t_{p_1}^{k_1} t_{p_2}^{k_2} \cdots t_{p_j}^{k_j}$ we can write the inverse dynamical zeta function (15.2) as

$$\prod_p (1 - t_p) = \sum_n \mu(n) t_n ,$$

and, if we care (we do in the case of the Riemann zeta function), the dynamical zeta function as .

$$\prod_p \frac{1}{1 - t_p} = \sum_n t_n \tag{E.11}$$

A striking aspect of this formula is its resemblance to the factorization of natural numbers into primes: the relation of the cycle expansion (E.11) to the product over prime cycles is analogous to the Riemann zeta (exercise c) represented as a sum over natural numbers vs. its Euler product representation.

We now implement this factorization explicitly by decomposing recursively binary strings into ordered concatenations of prime strings. There are 2 strings of length 1, both prime: $p_1 = 0$, $p_2 = 1$. There are 4 strings of length 2: 00, 01, 11, 10. The first three are ordered concatenations of primes: $00 = p_1^2$, $01 = p_1 p_2$, $11 = p_2^2$; by ordered concatenations we mean that $p_1 p_2$ is legal, but $p_2 p_1$ is not. The remaining string is the only prime of length 2, $p_3 = 10$. Proceeding by discarding the strings which are concatenations of shorter primes $p_1^{k_1} p_2^{k_2} \cdots p_j^{k_j}$, with primes lexically ordered, we generate the standard list of primes, in agreement with table 9.2: 0, 1, 10, 101, 100, 1000, 1001, 1011, 10000, 10001, 10010, 10011, 10110, 10111, 100000, 100001, 100010, 100011, 100110, 100111, 101100, 101110, 101111, ... This factorization is illustrated in table E.4.

factors	string
p_1	0
p_2	1
p_1^2	00
p_1p_2	01
p_2^2	11
p_3	10
p_1^3	000
$p_1^2p_2$	001
$p_1p_2^2$	011
p_2^3	111
p_1p_3	010
p_2p_3	110
p_4	100
p_5	101

factors	string
p_1^4	0000
$p_1^3p_2$	0001
$p_1^2p_2^2$	0011
$p_1p_2^3$	0111
p_2^4	1111
$p_1^2p_3$	0010
$p_1p_2p_3$	0110
$p_2^2p_3$	1110
p_3^2	1010
p_1p_4	0100
p_2p_4	1100
p_1p_5	0101
p_2p_5	1101
p_6	1000
p_7	1001
p_8	1011

factors	string
p_1^5	00000
$p_1^4p_2$	00001
$p_1^3p_2^2$	00011
$p_1^2p_2^3$	00111
$p_1p_2^4$	01111
p_2^5	11111
$p_1^3p_3$	00010
$p_1^2p_2p_3$	00110
$p_1p_2^2p_3$	01110
$p_2^3p_3$	11110
$p_1p_3^2$	01010
$p_2p_3^2$	11010
$p_1^2p_4$	00100
$p_1p_2p_4$	01100
$p_2^2p_4$	11100
p_3p_4	10100

factors	string
$p_1^2p_5$	00101
$p_1p_2p_5$	01101
$p_2^2p_5$	11101
p_3p_5	10101
p_1p_6	01000
p_2p_6	11000
p_1p_7	01001
p_2p_7	11001
p_1p_8	01011
p_2p_8	11011
p_9	10000
p_{10}	10001
p_{11}	10010
p_{12}	10011
p_{13}	10110
p_{14}	10111

Table E.4: Factorization of all periodic points strings up to length 5 into ordered concatenations $p_1^{k_1}p_2^{k_2}\cdots p_n^{k_n}$ of prime strings $p_1 = 0$, $p_2 = 1$, $p_3 = 10$, $p_4 = 100$, . . . , $p_{14} = 10111$.

E.2.1 Prime factorization for spectral determinants

Following sect. E.2, the spectral determinant cycle expansions is obtained by expanding F as a multinomial in prime cycle weights t_p

$$F = \prod_p \sum_{k=0}^{\infty} C_{p^k} t_p^k = \sum_{k_1 k_2 k_3 \cdots = 0}^{\infty} \tau_{p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots} \tag{E.12}$$

where the sum goes over all pseudocycles. In the above we have defined

$$\tau_{p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots} = \prod_{i=1}^{\infty} C_{{p_i}^{k_i}} t_{p_i}^{k_i} \ . \tag{E.13}$$

C page 234

A striking aspect of the spectral determinant cycle expansion is its resemblance to the factorization of natural numbers into primes: as we already noted in sect. E.2, the relation of the cycle expansion (E.12) to the product formula (12.9) is analogous to the Riemann zeta represented as a sum over natural numbers vs. its Euler product representation.

This is somewhat unexpected, as the cycle weights factorize exactly with respect to r repetitions of a prime cycle, $t_{pp\ldots p} = t_p^r$, but only approximately (*shadowing*) with respect to subdividing a string into prime substrings, $t_{p_1p_2} \approx t_{p_1}t_{p_2}$.

The coefficients C_{p^k} have a simple form only in 1-d, given by the Euler formula (13.26). In higher dimensions C_{p^k} can be evaluated by expanding (12.9), $F(z) = \prod_p F_p$, where

$$F_p = 1 - \left(\sum_{r=1}^{\infty} \frac{t_p^r}{r d_{p,r}} \right) + \frac{1}{2} \left(\sum_{r=1}^{\infty} \frac{t_p^r}{r d_{p,r}} \right)^2 - \ldots .$$

Expanding and recollecting terms, and suppressing the p cycle label for the moment, we obtain

$$\begin{aligned} F_p &= \sum_{r=1}^{\infty} C_k t^k, \quad C_k = (-)^k c_k / D_k, \\ D_k &= \prod_{r=1}^{k} d_r = \prod_{a=1}^{d} \prod_{r=1}^{k} (1 - u_a^r) \end{aligned} \tag{E.14}$$

where evaluation of c_k requires a certain amount of not too luminous algebra:

$$\begin{aligned} c_0 &= 1 \\ c_1 &= 1 \\ c_2 &= \frac{1}{2} \left(\frac{d_2}{d_1} - d_1 \right) = \frac{1}{2} \left(\prod_{a=1}^{d} (1 + u_a) - \prod_{a=1}^{d} (1 - u_a) \right) \\ c_3 &= \frac{1}{3!} \left(\frac{d_2 d_3}{d_1^2} + 2 d_1 d_2 - 3 d_3 \right) \\ &= \frac{1}{6} \left(\prod_{a=1}^{d} (1 + 2u_a + 2u_a^2 + u_a^3) \right. \\ &\quad \left. + 2 \prod_{a=1}^{d} (1 - u_a - u_a^2 + u_a^3) - 3 \prod_{a=1}^{d} (1 - u_a^3) \right) \end{aligned}$$

etc.. For example, for a general 2-dimensional map we have

$$F_p = 1 - \frac{1}{D_1} t + \frac{u_1 + u_2}{D_2} t^2 - \frac{u_1 u_2 (1 + u_1)(1 + u_2) + u_1^3 + u_2^3}{D_3} t^3 + \ldots . \tag{E.15}$$

We discuss the convergence of such cycle expansions in sect. J.4.

With $\tau_{p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}}$ defined as above, the prime factorization of symbol strings is unique in the sense that *each symbol string can be written as a* *unique concatenation of prime strings,* up to a convention on ordering of primes. This factorization is a nontrivial example of the utility of generalized Möbius inversion, sect. E.2.

How is the factorization of sect. E.2 used in practice? Suppose we have computed (or perhaps even measured in an experiment) all prime cycles up to length n, that is we have a list of t_p's and the corresponding Jacobian matrix eigenvalues $\Lambda_{p,1}, \Lambda_{p,2}, \ldots \Lambda_{p,d}$. A cycle expansion of the Selberg product is obtained by generating all strings in order of increasing length j allowed by the symbolic dynamics and constructing the multinomial

$$F = \sum_n \tau_n \tag{E.16}$$

where $n = s_1 s_2 \cdots s_j$, s_i range over the alphabet, in the present case $\{0, 1\}$. Factorizing every string $n = s_1 s_2 \cdots s_j = p_1^{k_1} p_2^{k_2} \cdots p_j^{k_j}$ as in table E.4, and substituting $\tau_{p_1^{k_1} p_2^{k_2} \cdots}$ we obtain a multinomial approximation to F. For example, $\tau_{001001010101} = \tau_{\underline{001}\,\underline{001}\,\underline{01}\,\underline{01}\,\underline{01}} = \tau_{001^2} \tau_{01^3}$, and τ_{01^3}, τ_{001^2} are known functions of the corresponding cycle eigenvalues. The zeros of F can now be easily determined by standard numerical methods. The fact that as far as the symbolic dynamics is concerned, the cycle expansion of a Selberg product is simply an average over all symbolic strings makes Selberg products rather pretty.

To be more explicit, we illustrate the above by expressing binary strings as concatenations of prime factors. We start by computing N_n, the number of terms in the expansion (E.12) of the total cycle length n. Setting $C_{p^k} t_p^k = z^{n_p k}$ in (E.12), we obtain

$$\sum_{n=0}^{\infty} N_n z^n = \prod_p \sum_{k=0}^{\infty} z^{n_p k} = \frac{1}{\prod_p (1 - z^{n_p})} \,.$$

So the generating function for the number of terms in the Selberg product is the topological zeta function. For the complete binary dynamics we have $N_n = 2^n$ contributing terms of length n:

$$\zeta_{top} = \frac{1}{\prod_p (1 - z^{n_p})} = \frac{1}{1 - 2z} = \sum_{n=0}^{\infty} 2^n z^n$$

Hence the number of distinct terms in the expansion (E.12) is the same as the number of binary strings, and conversely, the set of binary strings of length n suffices to label all terms of the total cycle length n in the expansion (E.12).

Appendix F

Counting itineraries

F.1 Counting curvatures

One consequence of the finitness of topological polynomials is that the contributions to curvatures at every order are even in number, half with positive and half with negative sign. For instance, for complete binary labelling (15.5),

$$\begin{aligned} c_4 &= -t_{0001} - t_{0011} - t_{0111} - t_0 t_{01} t_1 \\ &\quad + t_0 t_{001} + t_0 t_{011} + t_{001} t_1 + t_{011} t_1 \,. \end{aligned} \tag{F.1}$$

We see that 2^3 terms contribute to c_4, and exactly half of them appear with a negative sign - hence if all binary strings are admissible, this term vanishes in the counting expression.

F.2
page 653

Such counting rules arise from the identity

$$\prod_p (1 + t_p) = \prod_p \frac{1 - {t_p}^2}{1 - t_p} \,. \tag{F.2}$$

Substituting $t_p = z^{n_p}$ and using (10.15) we obtain for unrestricted symbol dynamics with N letters

$$\prod_p^{\infty} (1 + z^{n_p}) = \frac{1 - Nz^2}{1 - Nz} = 1 + Nz + \sum_{k=2}^{\infty} z^k \left(N^k - N^{k-1} \right)$$

The z^n coefficient in the above expansion is the number of terms contributing to c_n curvature, so we find that for a complete symbolic dynamics of N symbols and $n > 1$, the number of terms contributing to c_n is $(N-1)N^{k-1}$ (of which half carry a minus sign).

F.4 page 654

We find that for complete symbolic dynamics of N symbols and $n > 1$, the number of terms contributing to c_n is $(N - 1)N^{n-1}$. So, superficially, not much is gained by going from periodic orbits trace sums which get N^n contributions of n to the curvature expansions with $N^n(1-1/N)$. However, the point is not the number of the terms, but the cancellations between them.

Exercises

Exercise F.1 Lefschetz zeta function. *Elucidate the relation betveen the topological zeta function and the Lefschetz zeta function.*

Exercise F.2 Counting the 3-disk pinball counterterms. *Verify that the number of terms in the 3-disk pinball curvature expansion (15.31) is given by*

$$\prod_p (1+t_p) = \frac{1-3z^4-2z^6}{1-3z^2-2z^3} = 1+3z^2+2z^3+\frac{z^4(6+12z+2z^2)}{1-3z^2-2z^3}$$
$$= 1+3z^2+2z^3+6z^4+12z^5+20z^6+48z^7+84z^8+184z^9+ \quad \text{(F..3)}$$

This means that, for example, c_6 has a total of 20 terms, in agreement with the explicit 3-disk cycle expansion (15.32).

Exercise F.3 Cycle expansion denominators**. *Prove that the denominator of c_k is indeed D_k, as asserted (E.14).*

Exercise F.4 Counting subsets of cycles. *The techniques developed above can be generalized to counting subsets of cycles. Consider the simplest example of a dynamical system with a complete binary tree, a repeller map (9.10) with two straight branches, which we label 0 and 1. Every cycle weight for such map factorizes, with a factor t_0 for each 0, and factor t_1 for each 1 in its symbol string. The transition matrix traces (10.5) collapse to $tr(T^k) = (t_0+t_1)^k$, and $1/\zeta$ is simply*

$$\prod_p (1-t_p) = 1-t_0-t_1 \quad \text{(F.4)}$$

Substituting into the identity

$$\prod_p (1+t_p) = \prod_p \frac{1-{t_p}^2}{1-t_p}$$

we obtain

$$\prod_p (1+t_p) = \frac{1-t_0^2-t_1^2}{1-t_0-t_1} = 1+t_0+t_1+\frac{2t_0t_1}{1-t_0-t_1}$$
$$= 1+t_0+t_1+\sum_{n=2}^{\infty}\sum_{k=1}^{n-1} 2\binom{n-2}{k-1} t_0^k t_1^{n-k}. \quad \text{(F.5)}$$

Hence for $n \geq 2$ the number of terms in the expansion ?! with k 0's and $n-k$ 1's in their symbol sequences is $2\binom{n-2}{k-1}$. This is the degeneracy of distinct cycle *eigenvalues in fig.?!; for systems with non-uniform hyperbolicity this degeneracy is lifted (see fig. ?!).*

In order to count the number of prime cycles in each such subset we denote with $M_{n,k} \quad (n = 1, 2, \ldots\, ; \, k = \{0, 1\}$ *for* $n = 1$; $k = 1, \ldots, n-1$ *for* $n \geq 2)$ *the number of prime* n*-cycles whose labels contain* k *zeros, use binomial string counting and Möbius inversion and obtain*

$$\begin{aligned} M_{1,0} &= M_{1,1} = 1 \\ nM_{n,k} &= \sum_{m \left| \frac{n}{k} \right.} \mu(m) \binom{n/m}{k/m}, \quad n \geq 2\,, k = 1, \ldots, n-1 \end{aligned}$$

where the sum is over all m *which divide both* n *and* k.

Appendix G

Finding cycles

(C. Chandre)

G.1 Newton-Raphson method

G.1.1 Contraction rate

Consider a d-dimensional map $x' = f(x)$ with an unstable fixed point x_*. The Newton-Raphson algorithm is obtained by iterating the following map

$$x' = g(x) = x - (J(x) - 1)^{-1} (f(x) - x) .$$

The linearization of g near x_* leads to

$$x_* + \epsilon' = x_* + \epsilon - (J(x_*) - \mathbf{1})^{-1} (f(x_*) + J(x_*)\epsilon - x_* - \epsilon) + O\left(\|\epsilon\|^2\right) ,$$

where $\epsilon = x - x_*$. Therefore,

$$x' - x_* = O\left((x - x_*)^2\right) .$$

After n steps and if the initial guess x_0 is close to x_*, the error decreases super-exponentially

$$g^n(x_0) - x_* = O\left((x_0 - x_*)^{2^n}\right) .$$

G.1.2 Computation of the inverse

The Newton-Raphson method for finding n-cycles of d-dimensional mappings using the multi-shooting method reduces to the following equation

$$\begin{pmatrix} \mathbf{1} & & & -Df(x_n) \\ -Df(x_1) & \mathbf{1} & & \\ & \cdots & \mathbf{1} & \\ & & -Df(x_{n-1}) & \mathbf{1} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \cdots \\ \delta_n \end{pmatrix} = - \begin{pmatrix} F_1 \\ F_2 \\ \cdots \\ F_n \end{pmatrix}, \tag{G.1}$$

where $Df(x)$ is the $[d \times d]$ Jacobian matrix of the map evaluated at the point x, and $\delta_m = x'_m - x_m$ and $F_m = x_m - f(x_{m-1})$ are d-dimensional vectors. By some starightforward algebra, the vectors δ_m are expressed as functions of the vectors F_m:

$$\delta_m = -\sum_{k=1}^{m} \beta_{k,m-1} F_k - \beta_{1,m-1} \left(\mathbf{1} - \beta_{1,n}\right)^{-1} \left(\sum_{k=1}^{n} \beta_{k,n} F_k \right), \tag{G.2}$$

for $m = 1, \ldots, n$, where $\beta_{k,m} = Df(x_m)Df(x_{m-1}) \cdots Df(x_k)$ for $k < m$ and $\beta_{k,m} = \mathbf{1}$ for $k \geq m$. Therefore, finding n-cycles by a Newton-Raphson method with multiple shooting requires the inversing of a $[d \times d]$ matrix $\mathbf{1} - Df(x_n)Df(x_{n-1}) \cdots Df(x_1)$.

G.2 Hybrid Newton-Raphson / relaxation method

Consider a d-dimensional map $x' = f(x)$ with an unstable fixed point x_*. The transformed map is the following one:

$$x' = g(x) = x + \gamma C(f(x) - x),$$

where $\gamma > 0$ and C is a $d \times d$ invertible constant matrix. We notice that x_* is also a fixed point of g. Consider the stability matrix at the fixed point x_*

$$A_g = \left. \frac{dg}{dx} \right|_{x=x_*} = 1 + \gamma C(A_f - 1).$$

The matrix C is constructed such that the eigenvalues of A_g are of modulus less than one. Assume that A_f is diagonalizable: In the basis of diagonalization, the matrix writes:

$$\tilde{A}_g = 1 + \gamma \tilde{C}(\tilde{A}_f - 1),$$

where $\tilde{A}_f$ is diagonal with elements μ_i. We restrict the set of matrices $\tilde{C}$ to diagonal matrices with $\tilde{C}_{ii} = \epsilon_i$ where $\epsilon_i = \pm 1$. Thus $\tilde{A}_g$ is diagonal with eigenvalues $\gamma_i = 1 + \gamma\epsilon_i(\mu_i - 1)$. The choice of γ and ϵ_i is such that $|\gamma_i| < 1$. It is easy to see that if $\mathrm{Re}(\mu_i) < 1$ one has to choose $\epsilon_i = 1$, and if $\mathrm{Re}(\mu_i) > 1$, $\epsilon_i = -1$. If λ is chosen such that

$$0 < \gamma < \min_{i=1,\ldots,d} \frac{2|\mathrm{Re}(\mu_i) - 1|}{|\mu_i - 1|^2},$$

all the eigenvalues of A_g have modulus less that one. The contraction rate at the fixed point for the map g is then $\max_i |1 + \gamma\epsilon_i(\mu_i - 1)|$. We notice that if $\mathrm{Re}(\mu_i) = 1$, it is not possible to stabilize x_* by the set of matrices γC.

From the construction of C, we see that 2^d choices of matrices are possible. For example, for two-dimensional systems, these matrices are

$$C \in \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right\}.$$

For 2-dimensional dissipative maps, the eigenvalues satisfy $\mathrm{Re}(\mu_1)\mathrm{Re}(\mu_2) \leq \det Df < 1$. The case $(\mathrm{Re}(\mu_1) > 1, \mathrm{Re}(\mu_2) > 1)$ which is stabilized by $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ has to be discarded. The minimal set is reduced to three matrices.

Appendix H

Applications

Man who says it cannot be done should not interrupt man doing it.
Sayings of Vattay Gábor

In this appendix we show that the multidimensional Lyapunov exponents and relaxation exponents (dynamo rates) of vector fields can be expressed in terms of leading eigenvalues of appropriate evolution operators.

H.1 Evolution operator for Lyapunov exponents

Lyapunov exponents were introduced and computed for 1-d maps in sect. 8.3.2. For higher-dimensional flows only the Jacobian matrices are multiplicative, not individual eigenvalues, and the construction of the evolution operator for evaluation of the Lyapunov spectra requires the extension of evolution equations to the flow in the tangent space. We now develop the requisite theory.

Here we construct a multiplicative evolution operator (H.4) whose spectral determinant (H.8) yields the leading Lyapunov exponent of a d-dimensional flow (and is entire for Axiom A flows).

The key idea is to extending the dynamical system by the tangent space of the flow, suggested by the standard numerical methods for evaluation of Lyapunov exponents: start at x_0 with an initial infinitesimal tangent space vector $\eta(0) \in \mathbf{T}\mathcal{M}_x$, and let the flow transport it along the trajectory $x(t) = f^t(x_0)$.

The dynamics in the $(x, \eta) \in U \times TU_x$ space is governed by the system of equations of variations [6.1]:

$$\dot{x} = \mathbf{v}(x)\,, \quad \dot{\eta} = \mathbf{D}\mathbf{v}(x)\eta\,.$$

Here $\mathbf{D}\mathbf{v}(x)$ is the derivative matrix of the flow. We write the solution as

$$x(t) = f^t(x_0)\,, \quad \eta(t) = \mathbf{J}^t(x_0)\cdot\eta_0\,, \tag{H.1}$$

with the tangent space vector η transported by the stability matrix $\mathbf{J}^t(x_0) = \partial x(t)/\partial x_0$.

As explained in sect. 4.1, the growth rate of this vector is multiplicative along the trajectory and can be represented as $\eta(t) = |\eta(t)|/|\eta(0)|\mathbf{u}(t)$ where $\mathbf{u}(t)$ is a "unit" vector in some norm $||.||$. For asymptotic times and for almost every initial $(x_0, \eta(0))$, this factor converges to the leading eigenvalue of the linearized stability matrix of the flow.

We implement this multiplicative evaluation of stability eigenvalues by adjoining the d-dimensional transverse tangent space $\eta \in \mathbf{T}\mathcal{M}_x$; $\eta(x)\mathbf{v}(x) = 0$ to the $(d{+}1)$-dimensional dynamical evolution space $x \in \mathcal{M} \subset \mathbb{R}^{d+1}$. In order to determine the length of the vector η we introduce a homogeneous differentiable scalar function $g(\eta) = ||\eta||$. It has the property $g(\Lambda\eta) = |\Lambda|g(\eta)$ for any Λ. An example is the projection of a vector to its dth component

$$g\begin{pmatrix}\eta_1\\ \eta_2\\ \cdots\\ \eta_d\end{pmatrix} = |\eta_d|\,.$$

Any vector $\eta \in TU_x$ can now be represented by the product $\eta = \Lambda\mathbf{u}$, where $\mathbf{u}$ is a "unit" vector in the sense that its norm is $||\mathbf{u}|| = 1$, and the factor

$$\Lambda^t(x_0, \mathbf{u}_0) = g(\eta(t)) = g(\mathbf{J}^t(x_0)\cdot\mathbf{u}_0) \tag{H.2}$$

is the multiplicative "stretching" factor.

Unlike the leading eigenvalue of the Jacobian the stretching factor is multiplicative along the trajectory:

$$\Lambda^{t'+t}(x_0, \mathbf{u}_0) = \Lambda^{t'}(x(t), \mathbf{u}(t))\,\Lambda^t(x_0, \mathbf{u}_0).$$

H.1
page 671

The $\mathbf{u}$ evolution constrained to $E\mathbf{T}_{g,x}$, the space of unit transverse tangent vectors, is given by rescaling of (H.1):

$$\mathbf{u}' = R^t(x,\mathbf{u}) = \frac{1}{\Lambda^t(x,\mathbf{u})}\mathbf{J}^t(x)\cdot\mathbf{u}\,. \tag{H.3}$$

Eqs. (H.1), (H.2) and (H.3) enable us to define a *multiplicative* evolution operator on the extended space $U \times E\mathbf{T}_{g,x}$

$$\mathcal{L}^t(x',\mathbf{u}';x,\mathbf{u}) = \delta\big(x' - f^t(x)\big)\,\frac{\delta\big(\mathbf{u}' - R^t(x,\mathbf{u})\big)}{|\Lambda^t(x,\mathbf{u})|^{\beta-1}}\,, \tag{H.4}$$

where β is a variable.

To evaluate the expectation value of $\log|\Lambda^t(x,\mathbf{u})|$ which is the Lyapunov exponent we again have to take the proper derivative of the leading eigenvalue of (H.4). In order to derive the trace formula for the operator (H.4) we need to evaluate $\mathrm{Tr}\,\mathcal{L}^t = \int dx d\mathbf{u}\,\mathcal{L}^t(\mathbf{u},x;\mathbf{u},x)$. The $\int dx$ integral yields a weighted sum over prime periodic orbits p and their repetitions r:

$$\begin{aligned}
\mathrm{Tr}\,\mathcal{L}^t &= \sum_p T_p \sum_{r=1}^{\infty} \frac{\delta(t - rT_p)}{|\det\left(1 - \mathbf{J}_p^r\right)|}\Delta_{p,r}, \\
\Delta_{p,r} &= \int_g d\mathbf{u}\,\frac{\delta\big(\mathbf{u} - R^{T_p r}(x_p,\mathbf{u})\big)}{|\Lambda^{T_p r}(x_p,\mathbf{u})|^{\beta-1}}\,,
\end{aligned} \tag{H.5}$$

where $\mathbf{J}_p$ is the prime cycle p transverse stability matrix. As we shall see below, $\Delta_{p,r}$ is intrinsic to cycle p, and independent of any particular cycle point x_p.

We note next that if the trajectory $f^t(x)$ is periodic with period T, the tangent space contains d periodic solutions

$$\mathbf{e}_i(x(T+t)) = \mathbf{e}_i(x(t))\,, \quad i = 1,...,d,$$

corresponding to the d unit eigenvectors $\{\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_d\}$ of the transverse stability matrix, with "stretching" factors (H.2) given by its eigenvalues

$$\mathbf{J}_p(x)\cdot\mathbf{e}_i(x) = \Lambda_{p,i}\,\mathbf{e}_i(x)\,, \quad i = 1,...,d. \qquad \text{(no summation on } i)$$

The $\int d\mathbf{u}$ integral in (H.5) picks up contributions from these periodic solutions. In order to compute the stability of the ith eigendirection solution, it is convenient to expand the variation around the eigenvector $\mathbf{e}_i$ in the stability matrix eigenbasis $\delta\mathbf{u} = \sum \delta u_\ell \, \mathbf{e}_\ell$. The variation of the map (H.3) at a complete period $t = T$ is then given by

$$\begin{aligned} \delta R^T(\mathbf{e}_i) &= \frac{\mathbf{J}\cdot\delta\mathbf{u}}{g(\mathbf{J}\cdot\mathbf{e}_i)} - \frac{\mathbf{J}\cdot\mathbf{e}_i}{g(\mathbf{J}\cdot\mathbf{e}_i)^2}\left(\frac{\partial g(\mathbf{e}_i)}{\partial\mathbf{u}}\cdot\mathbf{J}\cdot\delta\mathbf{u}\right) \\ &= \sum_{k\neq i}\frac{\Lambda_{p,k}}{\Lambda_{p,i}}\left(\mathbf{e}_k - \mathbf{e}_i\frac{\partial g(\mathbf{e}_i)}{\partial u_k}\right)\delta u_k \,. \end{aligned} \tag{H.6}$$

The δu_i component does not contribute to this sum since $g(\mathbf{e}_i + du_i\mathbf{e}_i) = 1 + du_i$ implies $\partial g(\mathbf{e}_i)/\partial u_i = 1$. Indeed, infinitesimal variations $\delta\mathbf{u}$ must satisfy

$$g(\mathbf{u}+\delta\mathbf{u}) = g(\mathbf{u}) = 1 \quad \Longrightarrow \quad \sum_{\ell=1}^{d}\delta u_\ell\frac{\partial g(\mathbf{u})}{\partial u_\ell} = 0\,,$$

so the allowed variations are of form

$$\delta\mathbf{u} = \sum_{k\neq i}\left(\mathbf{e}_k - \mathbf{e}_i\frac{\partial g(\mathbf{e}_i)}{\partial u_k}\right)c_k\,, \quad |c_k| \ll 1\,,$$

and in the neighborhood of the $\mathbf{e}_i$ eigenvector the $\int d\mathbf{u}$ integral can be expressed as

$$\int_g d\mathbf{u} = \int\prod_{k\neq i}dc_k\,.$$

Inserting these variations into the $\int d\mathbf{u}$ integral we obtain

$$\begin{aligned} \int_g d\mathbf{u} \quad & \delta\big(\mathbf{e}_i + \delta\mathbf{u} - R^T(\mathbf{e}_i) - \delta R^T(\mathbf{e}_i) + \ldots\big) \\ = \quad & \int\prod_{k\neq i}dc_k\,\delta((1-\Lambda_k/\Lambda_i)c_k + \ldots) \\ & = \prod_{k\neq i}\frac{1}{|1-\Lambda_k/\Lambda_i|}\,, \end{aligned}$$

and the $\int d\mathbf{u}$ trace (H.5) becomes

$$\Delta_{p,r} = \sum_{i=1}^{d} \frac{1}{|\Lambda_{p,i}^r|^{\beta-1}} \prod_{k\neq i} \frac{1}{|1-\Lambda_{p,k}^r/\Lambda_{p,i}^r|}. \tag{H.7}$$

The corresponding spectral determinant is obtained by observing that the Laplace transform of the trace (11.19) is a logarithmic derivative $\mathrm{Tr}\,\mathcal{L}(s) = -\frac{d}{ds}\log F(s)$ of the spectral determinant:

$$F(\beta,s) = \exp\left(-\sum_{p,r} \frac{e^{sT_p r}}{r\,|\det(1-\mathbf{J}_p^r)|}\Delta_{p,r}(\beta)\right). \tag{H.8}$$

This determinant is the central result of this section. Its zeros correspond to the eigenvalues of the evolution operator (H.4), and can be evaluated by the cycle expansion methods.

The leading zero of (H.8) is called "pressure" (or free energy)

$$P(\beta) = s_0(\beta). \tag{H.9}$$

The average Lyapunov exponent is then given by the first derivative of the pressure at $\beta = 1$:

$$\overline{\lambda} = P'(1). \tag{H.10}$$

The simplest application of (H.8) is to 2-dimensional hyperbolic Hamiltonian maps. The stability eigenvalues are related by $\Lambda_1 = 1/\Lambda_2 = \Lambda$, and the spectral determinant is given by

$$\begin{aligned} F(\beta,z) &= \exp\left(-\sum_{p,r}\frac{z^{rn_p}}{r\,|\Lambda_p^r|}\frac{1}{(1-1/\Lambda_p^r)^2}\Delta_{p,r}(\beta)\right) \\ \Delta_{p,r}(\beta) &= \frac{|\Lambda_p^r|^{1-\beta}}{1-1/\Lambda_p^{2r}} + \frac{|\Lambda_p^r|^{\beta-3}}{1-1/\Lambda_p^{2r}}. \end{aligned} \tag{H.11}$$

The dynamics (H.3) can be restricted to a u unit eigenvector neighborhood corresponding to the largest eigenvalue of the Jacobi matrix. On this neighborhood the largest eigenvalue of the Jacobi matrix is the only fixed point, and the spectral determinant obtained by keeping only the largest term the $\Delta_{p,r}$ sum in (H.7) is also entire.

In case of maps it is practical to introduce the logarithm of the leading zero and to call it "pressure"

$$P(\beta) = \log z_0(\beta). \tag{H.12}$$

The average of the Lyapunov exponent of the map is then given by the first derivative of the pressure at $\beta = 1$:

$$\overline{\lambda} = P'(1). \tag{H.13}$$

By factorizing the determinant (H.11) into products of zeta functions we can conclude that the leading zero of the (H.4) can also be recovered from the leading zeta function

$$1/\zeta_0(\beta, z) = \exp\left(-\sum_{p,r} \frac{z^{rn_p}}{r|\Lambda_p^r|^\beta}\right). \tag{H.14}$$

This zeta function plays a key role in thermodynamic applications as we will will see in Chapter 17.

H.2 Advection of vector fields by chaotic flows

Fluid motions can move embedded vector fields around. An example is the magnetic field of the Sun which is "frozen" in the fluid motion. A passively evolving vector field $\mathbf{V}$ is governed by an equation of the form

$$\partial_t \mathbf{V} + \mathbf{u} \cdot \nabla \mathbf{V} - \mathbf{V} \cdot \nabla \mathbf{u} = 0, \tag{H.15}$$

where $\mathbf{u}(x, t)$ represents the velocity field of the fluid. The strength of the vector field can grow or decay during its time evolution. The amplification of the vector field in such a process is called the "dynamo effect". In a strongly chaotic fluid motion we can characterize the asymptotic behavior of the field with an exponent

$$\mathbf{V}(x, t) \sim \mathbf{V}(x) e^{\nu t}, \tag{H.16}$$

where ν is called the fast dynamo rate. The goal of this section is to show that periodic orbit theory can be developed for such a highly non-trivial system as well.

We can write the solution of (H.15) formally, as shown by Cauchy. Let $\mathbf{x}(t, \mathbf{a})$ be the position of the fluid particle that was at the point $\mathbf{a}$ at $t = 0$. Then the field evolves according to

$$\mathbf{V}(\mathbf{x}, t) = \mathbf{J}(\mathbf{a}, t)\mathbf{V}(\mathbf{a}, 0) \quad , \tag{H.17}$$

where $\mathbf{J}(\mathbf{a}, t) = \partial(\mathbf{x})/\partial(\mathbf{a})$ is the Jacobian matrix of the transformation that moves the fluid into itself $\mathbf{x} = \mathbf{x}(\mathbf{a}, t)$.

We write $\mathbf{x} = f^t(\mathbf{a})$, where f^t is the flow that maps the initial positions of the fluid particles into their positions at time t. Its inverse, $\mathbf{a} = f^{-t}(\mathbf{x})$, maps particles at time t and position $\mathbf{x}$ back to their initial positions. Then we can write (H.17)

$$V_i(\mathbf{x}, t) = \sum_j \int d^3\mathbf{a}\, \mathcal{L}^t_{ij}(\mathbf{x}, \mathbf{a}) V_j(\mathbf{a}, 0) \quad , \tag{H.18}$$

with

$$\mathcal{L}^t_{ij}(\mathbf{x}, \mathbf{a}) = \delta(\mathbf{a} - f^{-t}(\mathbf{x})) \frac{\partial x_i}{\partial a_j} \quad . \tag{H.19}$$

For large times, the effect of $\mathcal{L}^t$ is dominated by its leading eigenvalue, $e^{\nu_0 t}$ with $Re(\nu_0) > Re(\nu_i)$, $i = 1, 2, 3, ...$ In this way the transfer operator furnishes the fast dynamo rate, $\nu := \nu_0$.

The trace of the transfer operator is the sum over all periodic orbit contributions, with each cycle weighted by its intrinsic stability

$$\mathbf{Tr}\mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{\mathrm{tr}\, \mathbf{J}_p^r}{\left|\det\, \left(\mathbf{1} - \mathbf{J}_p^{-r}\right)\right|} \delta(t - rT_p). \tag{H.20}$$

We can construct the corresponding spectral determinant as usual

$$F(s) = \exp\left[- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{\mathrm{tr}\, \mathbf{J}_p^r}{\left|\det\, \left(\mathbf{1} - \mathbf{J}_p^{-r}\right)\right|} e^{srT_p} \right] \quad . \tag{H.21}$$

Note that in this formuli we have omitted a term arising from the Jacobian transformation along the orbit which would give $1+\mathrm{tr}\ \mathbf{J}_p^r$ in the numerator rather than just the trace of $\mathbf{J}_p^r$. Since the extra term corresponds to advection along the orbit, and this does not evolve the magnetic field, we have chosen to ignore it. It is also interesting to note that the negative powers of the Jacobian occur in the denominator, since we have f^{-t} in (H.19).

In order to simplify $F(s)$, we factor the denominator cycle stability determinants into products of expanding and contracting eigenvalues. For a 3-dimensional fluid flow with cycles possessing one expanding eigenvalue Λ_p (with $|\Lambda_p| > 1$), and one contracting eigenvalue λ_p (with $|\lambda_p| < 1$) the determinant may be expanded as follows:

$$\left|\det\left(\mathbf{1}-\mathbf{J}_p^{-r}\right)\right|^{-1} = |(1-\Lambda_p^{-r})(1-\lambda_p^{-r})|^{-1} = |\lambda_p|^r \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \Lambda_p^{-jr}\lambda_p^{kr} \quad . \tag{H.22}$$

With this decomposition we can rewrite the exponent in (H.21) as

$$\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{(\lambda_p^r + \Lambda_p^r)e^{srT_p}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^{-r}\right)\right|} = \sum_p \sum_{j,k=0}^{\infty}\sum_{r=1}^{\infty} \frac{1}{r}\left(|\lambda_p|\Lambda_p^{-j}\lambda_p^k e^{sT_p}\right)^r (\lambda_p^r+\Lambda_p^r) \quad , \tag{H.23}$$

which has the form of the expansion of a logarithm:

$$\sum_p \sum_{j,k} \left[\log\left(1 - e^{sT_p}|\lambda_p|\Lambda_p^{1-j}\lambda_p^k\right) + \log\left(1 - e^{sT_p}|\lambda_p|\Lambda_p^{-j}\lambda_p^{1+k}\right)\right] \quad . \tag{H.24}$$

The spectral determinant is therefore of the form,

$$F(s) = F_e(s)F_c(s) \quad , \tag{H.25}$$

where

$$F_e(s) = \prod_p \prod_{j,k=0}^{\infty} \left(1 - t_p^{(jk)}\Lambda_p\right) \quad , \tag{H.26}$$

$$F_c(s) = \prod_p \prod_{j,k=0}^{\infty} \left(1 - t_p^{(jk)}\lambda_p\right) , \tag{H.27}$$

with

$$t_p^{(jk)} = e^{sT_p} |\lambda_p| \frac{\lambda_p^k}{\Lambda_p^j} \quad . \tag{H.28}$$

The two factors present in $F(s)$ correspond to the expanding and contracting exponents. (Had we not neglected a term in (H.21), there would be a third factor corresponding to the translation.)

For 2-d Hamiltonian volume preserving systems, $\lambda = 1/\Lambda$ and (H.26) reduces to

$$F_e(s) = \prod_p \prod_{k=0}^{\infty} \left(1 - \frac{t_p}{\Lambda_p^{k-1}}\right)^{k+1} , \quad t_p = \frac{e^{sT_p}}{\mid \Lambda_p \mid} \quad . \tag{H.29}$$

With $\sigma_p = \Lambda_p/|\Lambda_p|$, the Hamiltonian zeta function (the $j = k = 0$ part of the product (H.27)) is given by

$$1/\zeta_{dyn}(s) = \prod_p \left(1 - \sigma_p e^{sT_p}\right) . \tag{H.30}$$

This is a curious formula — the zeta function depends only on the return times, not on the eigenvalues of the cycles. Furthermore, the identity,

$$\frac{\Lambda + 1/\Lambda}{|(1-\Lambda)(1-1/\Lambda)|} = \sigma + \frac{2}{|(1-\Lambda)(1-1/\Lambda)|} \quad ,$$

when substituted into (H.25), leads to a relation between the vector and scalar advection spectral determinants:

$$F_{dyn}(s) = F_0^2(s)/\zeta_{dyn}(s) . \tag{H.31}$$

The spectral determinants in this equation are entire for hyperbolic (axiom A) systems, since both of them correspond to multiplicative operators.

In the case of a flow governed by a map, we can adapt the formulas (H.29) and (H.30) for the dynamo determinants by simply making the substitution

$$z^{n_p} = e^{sT_p} \quad , \tag{H.32}$$

where n_p is the integer order of the cycle. Then we find the spectral determinant $F_e(z)$ given by equation (H.29) but with

$$t_p = \frac{z^{n_p}}{|\Lambda_p|} \tag{H.33}$$

for the weights, and

$$1/\zeta_{dyn}(z) = \Pi_p \left(1 - \sigma_p z^{n_p}\right) \tag{H.34}$$

for the zeta-function

For *maps* with finite Markov partition the inverse zeta function (H.34) reduces to a polynomial for z since curvature terms in the cycle expansion vanish. For example, for maps with complete binary partition, and with the fixed point stabilities of opposite signs, the cycle expansion reduces to

$$1/\zeta_{dyn}(s) = 1. \tag{H.35}$$

For such *maps* the dynamo spectral determinant is simply the square of the scalar advection spectral determinant, and therefore all its zeros are double. In other words, for flows governed by such discrete maps, the fast dynamo rate equals the scalar advection rate.

In contrast, for three-dimensional *flows*, the dynamo effect is distinct from the scalar advection. For example, for flows with finite symbolic dynamical grammars, (H.31) implies that the dynamo zeta function is a ratio of two entire determinants:

$$1/\zeta_{dyn}(s) = F_{dyn}(s)/F_0^2(s) \,. \tag{H.36}$$

This relation implies that for *flows* the zeta function has double poles at the zeros of the scalar advection spectral determinant, with zeros of the dynamo spectral determinant no longer coinciding with the zeros of the scalar advection spectral determinant; Usually the leading zero of the dynamo spectral determinant is larger than the scalar advection rate, and the rate of decay of the magnetic field is no longer governed by the scalar advection.

H.2 page 671

Commentary

Remark H.1 Dynamo zeta. The dynamo zeta (H.34) has been introduced by Aurell and Gilbert [H.3] and reviewed in ref. [H.5]. Our exposition follows ref. [H.4].

References

[H.1] J. H. Hannay and A. M. Ozorio de Almeida, *J. Phys.* **A 17**, 3429, (1984).

[H.2] Ya.B. Pesin, *Uspekhi Mat. Nauk* **32,** 55 (1977), [*Russian Math. Surveys* **32,** 55 (1977)].

[H.3] E. Aurell and A. Gilbert, *Geophys. & Astrophys. Fluid Dynamics* (1993).

[H.4] N.J. Balmforth, P. Cvitanović, G.R. Ierley, E.A. Spiegel and G. Vattay, *Stochastic Processes in Astrophysics*, *Annals of New York Academy of Sciences* **706**, 148 (1993).

[H.5] S. Childress and A.D. Gilbert *Stretch, Twist, Fold: The Fast Dynamo*, *Lecture Notes in Physics* **37** (Springer Verlag, Berlin 1995).

[H.6] J. Balatoni and A. Renyi, *Publi. Math. Inst. Hung. Acad. Sci.* **1**, 9 (1956); english translation **1**, 588 (Akademia Budapest, 1976).

[H.7] R. Benzi, G. Paladin, G. Parisi and A. Vulpiani, *J. Phys.* **A17**, 3521 (1984).

[H.8] Even though the thermodynamic formalism is of older vintage (we refer the reader to ref. [1.16] for a comprehensive overview), we adhere here to the notational conventions of ref. [H.9], more frequently employed in the physics literature.

[H.9] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia and B.I. Shraiman, *Phys. Rev.* **A107**, 1141 (1986).

[H.10] P. Grassberger, *Phys. Lett.* **97A**, 227 (1983); **107A**, 101 (1985); H.G.E. Hentschel and I. Procaccia, *Physica* **8D**, 435 (1983); R. Benzi, G. Paladin, G. Parisi and A. Vulpiani, em J. Phys. **A17**, 3521 (1984).

[H.11] C. Grebogi, E. Ott and J.A. Yorke, *Phys. Rev.* **A 36**, 3522 (1987).

[H.12] C. Grebogi, E. Ott and J. Yorke, *Physica* **D 7**, 181 (1983).

[H.13] C. Grebogi, E. Ott and J.A. Yorke, *Phys. Rev.* **A36**, 3522 (1987).

[H.14] C. Grebogi, E. Ott and J. Yorke, *Phys. Rev. A* **37,** 1711 (1988).

[H.15] E. Ott, C. Grebogi and J.A. Yorke, *Phys. Lett. A* **135**, 343 (1989).

[H.16] P. Grassberger and I. Procaccia, *Phys. Rev. A* **31**, 1872 (1985).

[H.17] C. Shannon, *Bell System Technical Journal,* **27**, 379 (1948).

[H.18] H. Fujisaka, *Progr. Theor. Phys* **70**, 1264 (1983).

[H.19] A. Politi, R. Badii and P. Grassberger, *J. Phys. A* **15**, L763 (1988); P. Grassberger, R. Badii and A. Politi, *J. Stat. Phys.* **51**, 135 (1988).

[H.20] M. Barnsley, *Fractals Everywhere*, (Academic Press, New York 1988).

[H.21] J.D. Crawford and J.R. Cary, *Physica* **D6**, 223 (1983)

[H.22] P. Collet and S. Isola, *Commun.Math.Phys.* **139**, 551 (1991)

[H.23] F. Christiansen, S. Isola, G. Paladin and H.H. Rugh, *J. Phys.* **A 23**, L1301 (1990).

[H.24] A.S. Pikovsky, unpublished.

[H.25] C. Beck, "Higher correlation functions of chaotic dynamical systems - a graph theoretical approach", *Nonlinearity* **4**, 1131 (1991); to be published.

[H.26] The 4-point correlation function is given in ref. [H.25].

[H.27] G. Hackenbroich and F. von Oppen, "Semiclassical theory of transport in antidot lattices", *Z. Phys.* **B 97**, 157 (1995).

[H.28] M.J. Feigenbaum, *J. Stat. Phys.* **21**, 669 (1979); reprinted in ref. [15.5].

[H.29] P. Szépfalusy, T. Tél, A. Csordás and Z. Kovács, *Phys. Rev.* **A 36**, 3525 (1987).

Exercises

Exercise H.1 Stretching factor. *Prove the multiplicative property of the stretching factor (H.2). Why should we extend the phase space with the tangent space?*

Exercise H.2 Dynamo rate. *Suppose that the fluid dynamics is highly dissipative and can be well approximated by the piecewise linear map*

$$f(x) = \begin{cases} 1 + ax & \textit{if} \quad x < 0, \\ 1 - bx & \textit{if} \quad x > 0, \end{cases} \tag{H.37}$$

on an appropriate surface of section ($a, b > 2$). Suppose also that the return time is constant T_a for $x < 0$ and T_b for $x > 0$. Show that the dynamo zeta is

$$1/\zeta_{dyn}(s) = 1 - e^{sT_a} + e^{sT_b}. \tag{H.38}$$

Show also that the escape rate is the leading zero of

$$1/\zeta_0(s) = 1 - e^{sT_a}/a - e^{sT_b}/b. \tag{H.39}$$

Calculate the dynamo and the escape rates analytically if $b = a^2$ and $T_b = 2T_a$. Do the calculation for the case when you reverse the signs of the slopes of the map. What is the difference?

Appendix I

Discrete symmetries

I.1 Preliminaries and definitions

(P. Cvitanović and A. Wirzba)

In the following we will define what we mean by the concepts *group*, *representation*, *symmetry of a dynamical system*, and *invariance*.

Group axioms. First, we define a group in abstract terms: A group G is a set of elements $g_1, g_2, g_3, \ldots$ for which a law of *composition* or *group multiplication* is given such that the product $g_2 \circ g_1$ (which we will also just abbreviate as $g_2 g_1$) of any two elements satisfies the following conditions:

1. If $g_1, g_2 \in G$, then $g_2 \circ g_1 \in G$.
2. The group multiplication is associative: $g_3 \circ (g_2 \circ g_1) = (g_3 \circ g_2) \circ g_1$.
3. The group G contains an element e called *identity* such that $g \circ e = e \circ g = g$ for every element $g \in G$.
4. For every element $g \in G$, there exists an unique element $h \in G$ such that $h \circ g = g \circ h = e$. The element h is called *inverse* of g, and is denoted by $h = g^{-1}$.

A *finite* group is a group with a finite number of elements

$$G = \{e, g_2, \ldots, g_{|G|}\} ,$$

where $|G|$, the number of elements, will be referred to as *order* of the group.

Matrix group on vector space. We will now apply these abstract group definitions to the set of $[d \times d]$-dimensional non-singular matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$ acting in a d-dimensional vector space $V \in \mathbb{R}^d$, that is the product of matrices $\mathbf{A}$ and $\mathbf{B}$ gives the single matrix $\mathbf{C}$, such that

$$\mathbf{C}v = \mathbf{B}(\mathbf{A}v) \in V, \qquad \forall v \in V. \tag{I.1}$$

The identity of the group is the unit matrix $\mathbf{1}$ which leaves all vectors in V unchanged. Every matrix in the group has a unique inverse.

Linear representation of a group. Let us now map the abstract group G *homeomorphically* on a group of matrices $\mathsf{D}(G)$ in the vector space V, that isin such a way that the group properties, especially the group multiplication, are preserved:

1. Any $g \in G$ is mapped to a matrix $\mathbf{D}(g) \in \mathsf{D}(G)$.

2. The group product $g_2 \circ g_1 \in G$ is mapped onto the matrix product $\mathbf{D}(g_2 \circ g_1) = \mathbf{D}(g_2)\mathbf{D}(g_1)$.

3. The associativity is preserved: $\mathbf{D}(g_3 \circ (g_2 \circ g_1)) = \mathbf{D}(g_3)(\mathbf{D}(g_2)\mathbf{D}(g_1)) = (\mathbf{D}(g_3)(\mathbf{D}(g_2))\mathbf{D}(g_1)$.

4. The identity element $e \in G$ is mapped onto the unit matrix $\mathsf{D}(e) = \mathbf{1}$ and the inverse element $g^{-1} \in G$ is mapped onto the inverse matrix $\mathbf{D}(g^{-1}) = [\mathbf{D}(g)]^{-1} \equiv \mathbf{D}^{-\mathbf{1}}(g)$.

We call the so defined matrix group $\mathsf{D}(G)$ a linear or matrix *representation* of the group G in the *representation space* V. Note that the matrix operation on a vector is by definition linear. We use the specification *linear* in order to discriminate the matrix representations from other operator representations that do not have to be linear, in general. Throughout this appendix we only consider linear representations.

If the dimensionality of V is d, we say the representation is an *d-dimensional representation* or has the *degree d*. The matrices $\mathbf{D}(g) \in \mathsf{D}(G)$ are non-singular $[d \times d]$ matrices, which we will also just abbreviate as $\mathbf{g}$, that is$x' = \mathbf{g}x$ corresponds to the normal matrix operation $x'_i = \sum_{j=1}^{d} (\mathbf{g})_{ij} x_j = \sum_{j=1}^{d} g_{ij} x_j$.

Character of a representation. The character of $\chi_\alpha(g)$ of an d-dimensional representation $\mathbf{D}(g)$ of the group element g of a discrete group G is defined as trace

$$\chi_\alpha(g) = \sum_{i=1}^{n} \mathbf{D}_{ii}(g) \equiv \mathrm{tr}\, \mathbf{D}(g) \,.$$

Note especially that $\chi(e) = n$, since $\mathbf{D}_{ij}(e) = \delta_{ij}$ for $1 \leq i, j \leq n$.

Faithful representations. If the homomorphismus mapping G on $\mathsf{D}(G)$ becomes an isomorphism, the representation is said to be *faithful*. In this case the order of the group of matrices $D(G)$ is equal to the order $|G|$ of the group. In general, however, there will be several elements of G that will be mapped on the unit matrix $\mathbf{D}(e) = \mathbf{1}$. This property can be used to define a subgroup $H \subset G$ of the group G consisting of all elements $h \in G$ that are mapped to the unit matrix of a given representation. Then the cosidered representation is a faithful representation of the *factor group* G/H.

Equivalent representations. From this remarks it should be clear that the representation of a group is by no means unique. If the basis in the d-dimensional vector space V is changed, the matrices $\mathbf{D}(g)$ have to be replaced by their transformations $\mathbf{D}'(g)$. In this case, however, the new matrices $\mathbf{D}'(g)$ and the old matrices $\mathbf{D}(g)$ are related by an equivalence transformation through a non-singular matrix $\mathbf{C}$

$$\mathbf{D}'(g) = \mathbf{C}\, \mathbf{D}(g)\, \mathbf{C}^{-1} \,.$$

Thus, the group of matrices $\mathbf{D}'(g)$ form an equivalent representation $\mathsf{D}'(G)$ to the representation $\mathsf{D}(G)$ of the group G. The equivalent representations have the same structure, although the matrices look different. Because of the cylic nature of the trace and because equivalent representations have the same dimension, the character of equivalent representations is the same

$$\chi(g) = \sum_{i=1}^{n} \mathbf{D}'_{ii}(g) = \mathrm{tr}\, \left(\mathbf{D}'(g)\right) = \mathrm{tr}\, \left(\mathbf{C}\mathbf{D}(g)\mathbf{C}^{-1}\right) \,.$$

Regular representation of a group. The *regular* representation of a group is a special representation that is defined as follows: If we define the elements of a finite group as $g_1, g_2, \ldots, g_{|G|}$, the multiplying from the left by any element g_ν permutes the $g_1, g_2, \ldots, g_{|G|}$ among themselves. We can represent the element g_ν by the permutations of the $|G|$ "coordinates" $g_1, g_2, \ldots, g_{|G|}$. Thus for $i, j = 1, \ldots, |G|$, we define the *regular representation*

$$\mathbf{D}_{ij}(g_\nu) = \begin{cases} \delta_{jl_i} & \text{if } g_\nu g_i = g_{l_i} \text{ with } l_i = 1, \ldots, |G|\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

In this regular representation the diagonal elements of all matrices are zero except for the element g_{ν_0} with $g_{\nu_0} g_i = g_i$, that isfor the identity element e. So in the regular representation the character is given by

$$\chi(g) = \begin{cases} 1 & \text{for} \quad g = e\,, \\ 0 & \text{for} \quad g \neq e\,. \end{cases}$$

Passive and active coordinate transformations. We have to discriminate between *active* and *passive* coordinate transformations. An active (coordinate) transformation corresponds to an non-singular $d \times d$ matrix that actively shifts/changes the vector $x \in \mathcal{M}$

$$x \to \mathbf{T}x.$$

The corresponding passive coordinate transformation changes the coordinate system with respect to which the vector $f(x) \in \mathcal{M}$ is measured. Thus it is given by

$$f(x) \to \mathbf{T}^{-1} f(x) = f(\mathbf{T}^{-1} x).$$

Note that the combination of an passive and active coordinate transformation results to the identity

$$f(x) = \mathbf{T}^{-1} f(\mathbf{T}x)\,.$$

On the other hand, the evolution operator $\mathcal{L}(x, y)$ satisfies the following identity

$$\mathcal{L}(x, y) = \left| \det \left(\frac{\partial \mathbf{T}x}{\partial x} \right) \right| \mathcal{L}(\mathbf{T}x, \mathbf{T}y) = |\det \mathbf{T}|\, \mathcal{L}(\mathbf{T}x, \mathbf{T}y).$$

Note the appearance of $\det \mathbf{T}$ instead of $\det \mathbf{T}^{-1}$ and therefore the contravariant transformation property of $\mathcal{L}(x, y)$ in correspondence to maps

$f(x)$. If the coordinate transformation $\mathbf{T}$ belongs to the linear non-singular representation of a *discrete* (that isfinite) symmetry group G, then $|\det \mathbf{T}| = 1$, since for any element g of a finite group G, where exists a number m such that

$$g^n \equiv \underbrace{g \circ g \circ \ldots \circ g}_{m \text{ times}} = e.$$

Thus $\mathbf{T}$ corresponds to the mth root of $\mathbf{1}$ and the modulus of its determinant is unity.

Symmetry of dynamical system. A dynamical system $(\mathcal{M}, f)$ is invariant under a discrete symmetry group $G = \{e, g_2, \ldots, g_{|G|}\}$, if the map $f : \mathcal{M} \to \mathcal{M}$ (or the continous flow f^t) from the d-dimensional manifold $\mathcal{M}$ into itself (with d finite) is invariant

$$f(\mathbf{g}x) = \mathbf{g}f(x)$$

for *any* coordinate $x \in \mathcal{M}$ and *any* finite non-singular linear representation (that isa non-singular $d \times d$ matrix) $\mathbf{g}$ of any element $g \in G$. So a *symmetry for a dynamical system* $(\mathcal{M}, f)$ has to satisfy the two conditions

$$\begin{array}{ll} 1) & \mathbf{g}x \in \mathcal{M} \quad \forall x \in \mathcal{M} \text{ and } \forall g \in G\,, \\ 2) & [\mathbf{D}(g), f] = 0 \quad \forall f : \mathcal{M} \to \mathcal{M} \text{ and } \forall g \in G\,. \end{array}$$

Group integration. Note the following laws

$$\frac{1}{|G|} \sum_{g \in G} = 1$$

and therefore

$$\frac{1}{|G|} \sum_{g \in G} \mathbf{D}(g_i) = \mathbf{D}(g_{i_0}), \quad i_0 \text{ fixed}\,.$$

However,

$$\frac{1}{|G|} \sum_{g \in G} \mathbf{D}(g) = \mathbf{0},$$

where $\mathbf{0}$ is the zero matrix of same dimension as the representations $\mathbf{D}(g) \in D(G)$. In particular,

$$\frac{1}{|G|}\sum_{g\in G}\chi_\alpha(g) = \frac{1}{|G|}\sum_{g\in G}\sum_{i=1}^{d_\alpha} D(g)_{ii} = 0.$$

Furthermore, if we consider all non-equilavent irreducible representations of a group G, then the quantities $D^{(\alpha)}_{ij}(g)$ for fixed α, i and j

Orthonormalitity of characters. But what can we say about

$$\frac{1}{|G|}\sum_{g\in G}\chi_\alpha(hg)\chi_\alpha(g^{-1}k^{-1}) \quad \text{with } h,k\in G \text{ fixed ?}$$

Note the following relation

$$\delta_{ab}\delta_{cd} = \frac{1}{n}\delta_{ad}\delta_{cb} + \left(\delta_{ab}\delta_{cd} - \frac{1}{n}\delta_{ad}\delta_{cb}\right).$$

Projection operators. The projection operator onto the α irreducible subspace of dimension d_α is given by

$$\mathbf{P}_\alpha = \frac{d_\alpha}{|G|}\sum_{g\in G}\chi_\alpha(g)\mathbf{g}^{-1}.$$

Note that $\mathbf{P}_\alpha$ is a $[d\times d]$-dimensional matrix as the representation $\mathbf{g}$.

Irreducible subspaces of the evolution operator.

$$\mathcal{L} = \sum_\alpha \mathrm{tr}\,\mathcal{L}_\alpha$$

with

$$\mathcal{L}_\alpha(y,x) = \frac{d_\alpha}{|G|}\sum_{g\in G}\chi_\alpha(g)\mathcal{L}(\mathbf{g}^{-1}y,x),$$

where the prefactor d_α reflects the fact that a d_α-dimensional representation occurs d_α times.

I.2 C_{4v} factorization

If an N-disk arrangement has C_N symmetry, and the disk visitation sequence is given by disk labels $\{\epsilon_1\epsilon_2\epsilon_3\ldots\}$, only the relative increments $\rho_i = \epsilon_{i+1} - \epsilon_i \bmod N$ matter. Symmetries under reflections across axes increase the group to C_{Nv} and add relations between symbols: $\{\epsilon_i\}$ and $\{N-\epsilon_i\}$ differ only by a reflection. As a consequence of this reflection increments become decrements until the next reflection and vice versa. Consider four equal disks placed on the vertices of a square (fig. I.1a). The symmetry group consists of the identity $\mathbf{e}$, the two reflections σ_x, σ_y across x, y axes, the two diagonal reflections σ_{13}, σ_{24}, and the three rotations C_4, C_2 and C_4^3 by angles $\pi/2$, π and $3\pi/2$. We start by exploiting the C_4 subgroup symmetry in order to replace the absolute labels $\epsilon_i \in \{1,2,3,4\}$ by relative increments $\rho_i \in \{1,2,3\}$. By reflection across diagonals, an increment by 3 is equivalent to an increment by 1 and a reflection; this new symbol will be called $\underline{1}$. Our convention will be to first perform the increment and then to change the orientation due to the reflection. As an example, consider the fundamental domain cycle 112. Taking the disk $1 \to$ disk 2 segment as the starting segment, this symbol string is mapped into the disk visitation sequence $1_{+1}2_{+1}3_{+2}1\ldots = \overline{123}$, where the subscript indicates the increments (or decrements) between neighboring symbols; the period of the cycle $\overline{112}$ is thus 3 in both the fundamental domain and the full space. Similarly, the cycle $\overline{\underline{1}12}$ will be mapped into $1_{+1}2_{-1}1_{-2}3_{-1}2_{+1}3_{+2}1 = \overline{121323}$ (note that the fundamental domain symbol $\underline{1}$ corresponds to a flip in orientation after the second and fifth symbols); this time the period in the full space is twice that of the fundamental domain. In particular, the fundamental domain fixed points correspond to the following 4-disk cycles:

4-disk		reduced
12	$\leftrightarrow$	$\underline{1}$
1234	$\leftrightarrow$	1
13	$\leftrightarrow$	2

Conversions for all periodic orbits of reduced symbol period less than 5 are listed in table I.1.

This symbolic dynamics is closely related to the group-theoretic structure of the dynamics: the global 4-disk trajectory can be generated by mapping the fundamental domain trajectories onto the full 4-disk space by the accumulated product of the C_{4v} group elements $g_1 = C$, $g_2 = C^2$, $g_{\underline{1}} = \sigma_{diag}C = \sigma_{axis}$, where C is a rotation by $\pi/2$. In the $\overline{\underline{1}12}$ example worked out above, this yields $g_{\underline{1}12} = g_2 g_1 g_{\underline{1}} = C^2 C\sigma_{axis} = \sigma_{diag}$, listed in the last column of table I.1. Our convention is to multiply group elements in

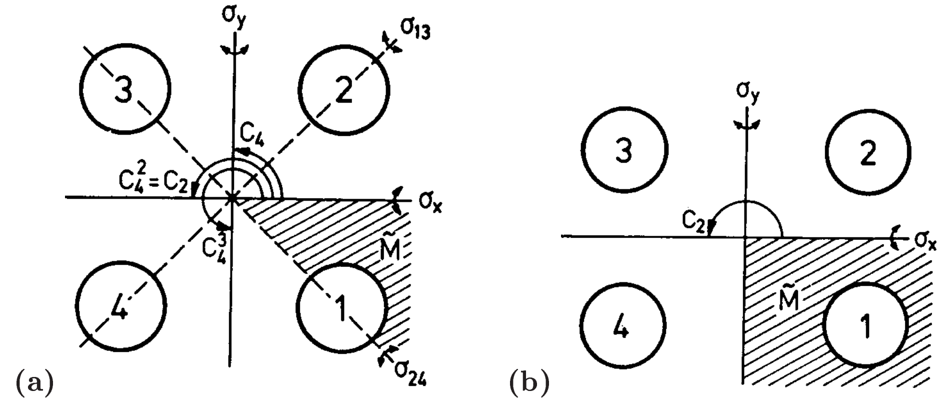


Figure I.1: (a) The symmetries of four disks on a square. (b) The symmetries of four disks on a rectangle. The fundamental domains are indicated by the shaded wedges.

$\tilde{p}$	p	$\mathbf{h}_{\tilde{p}}$
0	1 2	σ_x
1	1 2 3 4	C_4
2	1 3	C_2, σ_{13}
01	12 14	σ_{24}
02	12 43	σ_y
12	12 41 34 23	C_4^3
001	121 232 343 414	C_4
002	121 343	C_2
011	121 434	σ_y
012	121 323	σ_{13}
021	124 324	σ_{13}
022	124 213	σ_x
112	123	e
122	124 231 342 413	C_4

$\tilde{p}$	p	$\mathbf{h}_{\tilde{p}}$
0001	1212 1414	σ_{24}
0002	1212 4343	σ_y
0011	1212 3434	C_2
0012	1212 4141 3434 2323	C_4^3
0021 (a)	1213 4142 3431 2324	C_4^3
0022	1213	e
0102 (a)	1214 2321 3432 4143	C_4
0111	1214 3234	σ_{13}
0112 (b)	1214 2123	σ_x
0121 (b)	1213 2124	σ_x
0122	1213 1413	σ_{24}
0211	1243 2134	σ_x
0212	1243 1423	σ_{24}
0221	1242 1424	σ_{24}
0222	1242 4313	σ_y
1112	1234 2341 3412 4123	C_4
1122	1231 3413	C_2
1222	1242 4131 3424 2313	C_4^3

Table I.1: C_{4v} correspondence between the ternary fundamental domain prime cycles $\tilde{p}$ and the full 4-disk {1,2,3,4} labelled cycles p, together with the C_{4v} transformation that maps the end point of the $\tilde{p}$ cycle into an irreducible segment of the p cycle. For typographical convenience, the symbol $\underline{1}$ of sect. I.2 has been replaced by 0, so that the ternary alphabet is $\{0, 1, 2\}$. The degeneracy of the p cycle is $m_p = 8n_{\tilde{p}}/n_p$. Orbit $\overline{2}$ is the sole boundary orbit, invariant both under a rotation by π and a reflection across a diagonal. The two pairs of cycles marked by (a) and (b) are related by time reversal, but cannot be mapped into each other by C_{4v} transformations.

the reverse order with respect to the symbol sequence. We need these group elements for our next step, the dynamical zeta function factorizations.

The C_{4v} group has four one-dimensional representations, either symmetric (A_1) or antisymmetric (A_2) under both types of reflections, or symmetric under one and antisymmetric under the other (B_1, B_2), and a degenerate pair of two-dimensional representations E. Substituting the C_{4v} characters

C_{4v}	A_1	A_2	B_1	B_2	E
e	1	1	1	1	2
C_2	1	1	1	1	-2
C_4, C_4^3	1	1	-1	-1	0
σ_{axes}	1	-1	1	-1	0
σ_{diag}	1	-1	-1	1	0

into (19.15) we obtain:

$$
\begin{array}{rllllllll}
h_{\tilde{p}} & & & A_1 & A_2 & B_1 & B_2 & E \\
e\!: & (1-t_{\tilde{p}})^8 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}})^4 \\
C_2\!: & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}})^4 \\
C_4, C_4^3\!: & (1-t_{\tilde{p}}^4)^2 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}^2)^2 \\
\sigma_{axes}\!: & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}^2)^2 \\
\sigma_{diag}\!: & (1-t_{\tilde{p}}^2)^4 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}^2)^2
\end{array}
$$

The possible irreducible segment group elements $\mathbf{h}_{\tilde{p}}$ are listed in the first column; σ_{axes} denotes a reflection across either the x-axis or the y-axis, and σ_{diag} denotes a reflection across a diagonal (see fig. I.1a). In addition, degenerate pairs of boundary orbits can run along the symmetry lines in the full space, with the fundamental domain group theory weights $\mathbf{h}_p = (C_2 + \sigma_x)/2$ (axes) and $\mathbf{h}_p = (C_2 + \sigma_{13})/2$ (diagonals) respectively:

$$
\begin{array}{rlll}
 & & & A_1 \quad\; A_2 \quad\;\; B_1 \quad\; B_2 \quad\;\; E \\
\text{axes:} & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1+t_{\tilde{p}})^2 \\
\text{diagonals:} & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1-0t_{\tilde{p}})(1-t_{\tilde{p}})(1+t_{\tilde{p}})^2
\end{array}
\tag{I.2}
$$

(we have assumed that $t_{\tilde{p}}$ does not change sign under reflections across symmetry axes). For the 4-disk arrangement considered here only the diagonal orbits $\overline{13}$, $\overline{24}$ occur; they correspond to the $\overline{2}$ fixed point in the fundamental domain.

The A_1 subspace in C_{4v} cycle expansion is given by

$$
1/\zeta_{A_1} = (1-t_0)(1-t_1)(1-t_2)(1-t_{01})(1-t_{02})(1-t_{12})
$$

$$
\begin{aligned}
& (1-t_{001})(1-t_{002})(1-t_{011})(1-t_{012})(1-t_{021})(1-t_{022})(1-t_{112}) \\
& (1-t_{122})(1-t_{0001})(1-t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\ldots \\
= \; & 1-t_0-t_1-t_2-(t_{01}-t_0t_1)-(t_{02}-t_0t_2)-(t_{12}-t_1t_2) \\
& -(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
& -(t_{022}-t_2t_{02})-(t_{112}-t_1t_{12})-(t_{122}-t_2t_{12}) \\
& -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\ldots \qquad \text{(I.3)}
\end{aligned}
$$

(for typographical convenience, $\underline{1}$ is replaced by 0 in the remainder of this section). For one-dimensional representations, the characters can be read off the symbol strings: $\chi_{A_2}(\mathbf{h}_{\tilde{\mathbf{p}}})=(-1)^{n_0}$, $\chi_{B_1}(\mathbf{h}_{\tilde{\mathbf{p}}})=(-1)^{n_1}$, $\chi_{B_2}(\mathbf{h}_{\tilde{\mathbf{p}}})=(-1)^{n_0+n_1}$, where n_0 and n_1 are the number of times symbols 0, 1 appear in the $\tilde{p}$ symbol string. For B_2 all t_p with an odd total number of 0's and 1's change sign:

$$
\begin{aligned}
1/\zeta_{B_2} = \; & (1+t_0)(1+t_1)(1-t_2)(1-t_{01})(1+t_{02})(1+t_{12}) \\
& (1+t_{001})(1-t_{002})(1+t_{011})(1-t_{012})(1-t_{021})(1+t_{022})(1-t_{112}) \\
& (1+t_{122})(1-t_{0001})(1+t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\ldots \\
= \; & 1+t_0+t_1-t_2-(t_{01}-t_0t_1)+(t_{02}-t_0t_2)+(t_{12}-t_1t_2) \\
& +(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
& +(t_{022}-t_2t_{02})-(t_{112}-t_1t_{12})+(t_{122}-t_2t_{12}) \\
& -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\ldots \qquad \text{(I.4)}
\end{aligned}
$$

The form of the remaining cycle expansions depends crucially on the special role played by the boundary orbits: by (I.2) the orbit t_2 does not contribute to A_2 and B_1,

$$
\begin{aligned}
1/\zeta_{A_2} = \; & (1+t_0)(1-t_1)(1+t_{01})(1+t_{02})(1-t_{12}) \\
& (1-t_{001})(1-t_{002})(1+t_{011})(1+t_{012})(1+t_{021})(1+t_{022})(1-t_{112}) \\
& (1-t_{122})(1+t_{0001})(1+t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\ldots \\
= \; & 1+t_0-t_1+(t_{01}-t_0t_1)+t_{02}-t_{12} \\
& -(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
& +t_{022}-t_{122}-(t_{112}-t_1t_{12})+(t_{012}+t_{021}-t_0t_{12}-t_1t_{02})\ldots \qquad \text{(I.5)}
\end{aligned}
$$

and

$$
\begin{aligned}
1/\zeta_{B_1} = \; & (1-t_0)(1+t_1)(1+t_{01})(1-t_{02})(1+t_{12}) \\
& (1+t_{001})(1-t_{002})(1-t_{011})(1+t_{012})(1+t_{021})(1-t_{022})(1-t_{112}) \\
& (1+t_{122})(1+t_{0001})(1-t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\ldots \\
= \; & 1-t_0+t_1+(t_{01}-t_0t_1)-t_{02}+t_{12} \\
& +(t_{001}-t_0t_{01})-(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
& -t_{022}+t_{122}-(t_{112}-t_1t_{12})+(t_{012}+t_{021}-t_0t_{12}-t_1t_{02})\ldots \qquad \text{(I.6)}
\end{aligned}
$$

In the above we have assumed that t_2 does not change sign under C_{4v} reflections. For the mixed-symmetry subspace E the curvature expansion is given by

$$\begin{aligned}
1/\zeta_E \;=\;& 1 + t_2 + (-{t_0}^2 + {t_1}^2) + (2t_{002} - t_2{t_0}^2 - 2t_{112} + t_2{t_1}^2) \\
&+(2t_{0011} - 2t_{0022} + 2t_2t_{002} - {t_{01}}^2 - {t_{02}}^2 + 2t_{1122} - 2t_2t_{112} \\
&+{t_{12}}^2 - {t_0}^2{t_1}^2) + (2t_{00002} - 2t_{00112} + 2t_2t_{0011} - 2t_{00121} - 2t_{00211} \\
&+2t_{00222} - 2t_2t_{0022} + 2t_{01012} + 2t_{01021} - 2t_{01102} - t_2{t_{01}}^2 + 2t_{02022} \\
&-t_2{t_{02}}^2 + 2t_{11112} - 2t_{11222} + 2t_2t_{1122} - 2t_{12122} + t_2{t_{12}}^2 - t_2{t_0}^2{t_1}^2 \\
&+2t_{002}(-{t_0}^2 + {t_1}^2) - 2t_{112}(-{t_0}^2 + {t_1}^2))
\end{aligned} \tag{I.7}$$

A quick test of the $\zeta = \zeta_{A_1}\zeta_{A_2}\zeta_{B_1}\zeta_{B_2}\zeta_E^2$ factorization is afforded by the topological polynomial; substituting $t_p = z^{n_p}$ into the expansion yields

$$1/\zeta_{A_1} = 1 - 3z \ , \quad 1/\zeta_{A_2} = 1/\zeta_{B_1} = 1 \ , \quad 1/\zeta_{B_2} = 1/\zeta_E = 1 + z \ ,$$

in agreement with (10.42).

15.9
page 303

Remark I.1 Labelling conventions While there is a variety of labelling conventions [20.15, 19.13] for the reduced C_{4v} dynamics, we prefer the one introduced here because of its close relation to the group-theoretic structure of the dynamics: the global 4-disk trajectory can be generated by mapping the fundamental domain trajectories onto the full 4-disk space by the accumulated product of the C_{4v} group elements.

I.3 C_{2v} factorization

An arrangement of four identical disks on the vertices of a rectangle has C_{2v} symmetry (fig. I.1b). C_{2v} consists of $\{e, \sigma_x, \sigma_y, C_2\}$, *i.e.*, the reflections across the symmetry axes and a rotation by π.

This system affords a rather easy visualization of the conversion of a 4-disk dynamics into a fundamental domain symbolic dynamics. An orbit leaving the fundamental domain through one of the axis may be folded back by a reflection on that axis; with these symmetry operations $g_0 = \sigma_x$ and $g_1 = \sigma_y$ we associate labels 1 and 0, respectively. Orbits going to the diagonally opposed disk cross the boundaries of the fundamental domain twice; the product of these two reflections is just $C_2 = \sigma_x\sigma_y$, to

$\tilde{p}$	p	g
0	1 4	σ_y
1	1 2	σ_x
2	1 3	C_2
01	14 32	C_2
02	14 23	σ_x
12	12 43	σ_y
001	141 232	σ_x
002	141 323	C_2
011	143 412	σ_y
012	143	e
021	142	e
022	142 413	σ_y
112	121 343	C_2
122	124 213	σ_x

$\tilde{p}$	p	g
0001	1414 3232	C_2
0002	1414 2323	σ_x
0011	1412	e
0012	1412 4143	σ_y
0021	1413 4142	σ_y
0022	1413	e
0102	1432 4123	σ_y
0111	1434 3212	C_2
0112	1434 2343	σ_x
0121	1431 2342	σ_x
0122	1431 3213	C_2
0211	1421 2312	σ_x
0212	1421 3243	C_2
0221	1424 3242	C_2
0222	1424 2313	σ_x
1112	1212 4343	σ_y
1122	1213	e
1222	1242 4313	σ_y

Table I.2: C_{2v} correspondence between the ternary $\{0, 1, 2\}$ fundamental domain prime cycles $\tilde{p}$ and the full 4-disk {1,2,3,4} cycles p, together with the C_{2v} transformation that maps the end point of the $\tilde{p}$ cycle into an irreducible segment of the p cycle. The degeneracy of the p cycle is $m_p = 4n_{\tilde{p}}/n_p$. Note that the 012 and 021 cycles are related by time reversal, but cannot be mapped into each other by C_{2v} transformations. The full space orbit listed here is generated from the symmetry reduced code by the rules given in sect. I.3, starting from disk 1.

which we assign the label 2. For example, a ternary string $0\,0\,1\,0\,2\,0\,1\ldots$ is converted into 12143123..., and the associated group-theory weight is given by $\ldots g_1 g_0 g_2 g_0 g_1 g_0 g_0$.

Short ternary cycles and the corresponding 4-disk cycles are listed in table I.2. Note that already at length three there is a pair of cycles ($012 = 143$ and $021 = 142$) related by time reversal, but *not* by any C_{2v} symmetries.

The above is the complete description of the symbolic dynamics for 4 sufficiently separated equal disks placed at corners of a rectangle. However, if the fundamental domain requires further partitioning, the ternary description is insufficient. For example, in the stadium billiard fundamental domain one has to distinguish between bounces off the straight and the curved sections of the billiard wall; in that case five symbols suffice for constructing the covering symbolic dynamics.

The group C_{2v} has four one-dimensional representations, distinguished by their behavior under axis reflections. The A_1 representation is symmetric with respect to both reflections; the A_2 representation is antisymmetric with respect to both. The B_1 and B_2 representations are symmetric under one and antisymmetric under the other reflection. The character table is

C_{2v}	A_1	A_2	B_1	B_2
e	1	1	1	1
C_2	1	1	-1	-1
σ_x	1	-1	1	-1
σ_y	1	-1	-1	1

Substituted into the factorized determinant (19.14), the contributions of periodic orbits split as follows

$$
\begin{array}{rlcllll}
g_{\tilde{p}} & & & A_1 & A_2 & B_1 & B_2 \\
e\colon & (1-t_{\tilde{p}})^4 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) \\
C_2\colon & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1-t_{\tilde{p}}) \\
\sigma_x\colon & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) \\
\sigma_y\colon & (1-t_{\tilde{p}}^2)^2 & = & (1-t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1+t_{\tilde{p}}) & (1-t_{\tilde{p}})
\end{array}
$$

Cycle expansions follow by substituting cycles and their group theory factors from table I.2. For A_1 all characters are $+1$, and the corresponding cycle expansion is given in (I.3). Similarly, the totally antisymmetric subspace factorization A_2 is given by (I.4), the B_2 factorization of C_{4v}. For B_1 all t_p with an odd total number of 0's and 2's change sign:

$$
\begin{aligned}
1/\zeta_{B_1} &= (1+t_0)(1-t_1)(1+t_2)(1+t_{01})(1-t_{02})(1+t_{12}) \\
&\quad (1-t_{001})(1+t_{002})(1+t_{011})(1-t_{012})(1-t_{021})(1+t_{022})(1+t_{112}) \\
&\quad (1-t_{122})(1+t_{0001})(1-t_{0002})(1-t_{0011})(1+t_{0012})(1+t_{0021})\dots \\
&= 1+t_0-t_1+t_2+(t_{01}-t_0t_1)-(t_{02}-t_0t_2)+(t_{12}-t_1t_2) \\
&\quad -(t_{001}-t_0t_{01})+(t_{002}-t_0t_{02})+(t_{011}-t_1t_{01}) \\
&\quad +(t_{022}-t_2t_{02})+(t_{112}-t_1t_{12})-(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\dots
\end{aligned} \tag{I.8}
$$

For B_2 all t_p with an odd total number of 1's and 2's change sign:

$$
\begin{aligned}
1/\zeta_{B_2} &= (1-t_0)(1+t_1)(1+t_2)(1+t_{01})(1+t_{02})(1-t_{12}) \\
&\quad (1+t_{001})(1+t_{002})(1-t_{011})(1-t_{012})(1-t_{021})(1-t_{022})(1+t_{112}) \\
&\quad (1+t_{122})(1+t_{0001})(1+t_{0002})(1-t_{0011})(1-t_{0012})(1-t_{0021})\dots \\
&= 1-t_0+t_1+t_2+(t_{01}-t_0t_1)+(t_{02}-t_0t_2)-(t_{12}-t_1t_2) \\
&\quad +(t_{001}-t_0t_{01})+(t_{002}-t_0t_{02})-(t_{011}-t_1t_{01}) \\
&\quad -(t_{022}-t_2t_{02})+(t_{112}-t_1t_{12})+(t_{122}-t_2t_{12}) \\
&\quad -(t_{012}+t_{021}+t_0t_1t_2-t_0t_{12}-t_1t_{02}-t_2t_{01})\dots
\end{aligned} \tag{I.9}
$$

Note that all of the above cycle expansions group long orbits together with their pseudoorbit shadows, so that the shadowing arguments for convergence still apply.

The topological polynomial factorizes as

$$\frac{1}{\zeta_{A_1}} = 1 - 3z \quad , \quad \frac{1}{\zeta_{A_2}} = \frac{1}{\zeta_{B_1}} = \frac{1}{\zeta_{B_2}} = 1 + z,$$

consistent with the 4-disk factorization (10.42).

Remark I.2 C_{2v} symmetry C_{2v} is the symmetry of several systems studied in the literature, such as the stadium billiard [3.17], and the 2-dimensional anisotropic Kepler potential [27.18].

I.4 Hénon map symmetries

We note here a few simple symmetries of the Hénon map (3.10). For $b \neq 0$ the Hénon map is reversible: the backward iteration of (3.11) is given by

$$x_{n-1} = -\frac{1}{b}(1 - ax_n^2 - x_{n+1}) . \tag{I.10}$$

Hence the time reversal amounts to $b \to 1/b$, $a \to a/b^2$ symmetry in the parameter plane, together with $x \to -x/b$ in the coordinate plane, and there is no need to explore the (a, b) parameter plane outside the strip $b \in \{-1, 1\}$. For $b = -1$ the map is orientation and area preserving (see (17.1) below),

$$x_{n-1} = 1 - ax_n^2 - x_{n+1} , \tag{I.11}$$

the backward and the forward iteration are the same, and the non–wandering set is symmetric across the $x_{n+1} = x_n$ diagonal. This is one of the simplest models of a Poincaré return map for a Hamiltonian flow. For the orientation reversing $b = 1$ case we have

$$x_{n-1} = 1 - ax_n^2 + x_{n+1} , \tag{I.12}$$

and the non–wandering set is symmetric across the $x_{n+1} = -x_n$ diagonal.

I.5 Symmetries of the symbol square

• advanced section •

Depending on the type of dynamical system, the symbol square might have a variety of symmetries. Under the time reversal

$$\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots \to \cdots s_3s_2s_1.s_0s_{-1}s_{-2}\cdots$$

the points in the symbol square for an orientation preserving map are symmetric across the diagonal $\gamma = \delta$, and for the orientation reversing case they are symmetric with respect to the $\gamma = 1 - \delta$ diagonal. Consequently the periodic orbits appear either in dual pairs $p = s_1s_2s_3\ldots s_n$, $\overline{p} = s_ns_{n-1}s_{n-2}\ldots s_1$, or are self-dual under time reversal, $S_p = S_{\overline{p}}$. For the orientation preserving case a self-dual cycle of odd period has at least one point on the symmetry diagonal. In particular, all fixed points lie on the symmetry diagonal. Determination of such symmetry lines can be of considerable practical utility, as it reduces some of the periodic orbit searches to 1-dimensional searches.

Remark I.3 Symmetries of the symbol square. For a more detailed discussion of the symbolic dynamics symmetries, see refs. [3.4, 9.38].

Appendix J

Convergence of spectral determinants

J.1 Curvature expansions: geometric picture

If you has some experience with numerical estimates of fractal dimensions, you will note that the numerical convergence of cycle expansions for systems such as the 3-disk game of pinball, table 15.2, is very impressive; only three input numbers (the two fixed points $\overline{0}$, $\overline{1}$ and the 2-cycle $\overline{10}$) already yield the escape rate to 4 significant digits! We have omitted an infinity of unstable cycles; so why does approximating the dynamics by a finite number of cycles work so well?

Looking at the cycle expansions simply as sums of unrelated contributions is not specially encouraging: the cycle expansion (15.2) is not absolutely convergent in the sense of Dirichlet series of sect. 15.5, so what one makes of it depends on the way the terms are arranged.

The simplest estimate of the error introduced by approximating smooth flow by periodic orbits is to think of the approximation as a tessalation of a smooth curve by piecewise linear tiles, fig. 1.9.

J.1.1 Tessalation of a smooth flow by cycles

One of the early high accuracy computations of π was due to Euler. Euler computed the circumference of the circee of unit radius by inscribing into it a regular polygon with N sides; the error of such computation is proportional to $1-\cos(2\pi/N) \propto N^{-2}$. In a periodic orbit tessalation of a smooth flow, we cover the phase space by e^{hn} tiles at the nth level of resolution, where h is the topological entropy, the growth rate of the number of tiles. Hence we expect the error in approximating a smooth flow by e^{hn} linear segments to be exponentially small, of order $N^{-2} \propto e^{-2hn}$.

J.1.2 Shadowing and convergence of curvature expansions

We have shown in chapter 10 that if the symbolic dynamics is defined by a finite grammar, a finite number of cycles, let us say the first k terms in the cycle expansion are necessary to correctly count the pieces of the Cantor set generated by the dynamical system.

They are composed of products of non–intersecting loops on the Markov graph, see (10.13). We refer to this set of non–intersecting loops as the *fundamental* cycles of the strange set. It is only after these terms have been included that the cycle expansion is expected to converge smoothly, that is only for $n > k$ are the curvatures c_n in (9.2??) a measure of the variation of the quality of a linearized covering of the dynamical Cantor set by the length n cycles, and expected to fall off rapidly with n.

The rate of fall-off of the cycle expansion coefficients can be estimated by observing that for subshifts of finite type the contributions from longer orbits in curvature expansions such as (15.5) can always be grouped into shadowing combinations of pseudo-cycles. For example, a cycle with itinerary $\overline{ab}= s_1 s_2 \cdots s_n$ will appear in combination of form

$$1/\zeta = 1 - \cdots - (t_{ab} - t_a t_b) - \cdots ,$$

with $\overline{ab}$ shadowed by cycle $\overline{a}$ followed by cycle $\overline{b}$, where $a = s_1 s_2 \cdots s_m$, $b = s_{m+1} \cdots s_{n-1} s_n$, and s_k labels the Markov partition $\mathcal{M}_{s_k}$ (9.4) that the trajectory traverses at the kth return. If the two trajectories coincide in the first m symbols, at the mth return to a Poincaré section they can land anywhere in the phase space $\mathcal{M}$

$$\left| f^{T_a}(x_a) - f^{T_{a\ldots}}(x_{a\ldots}) \right| \approx 1 ,$$

where we have assumed that the $\mathcal{M}$ is compact, and that the maximal possible separation across $\mathcal{M}$ is $O(1)$. Here x_a is a point on the $\overline{a}$ cycle of period T_a, and $x_{a\ldots}$ is a nearby point whose trajectory tracks the cycle $\overline{a}$ for the first m Poincaré section returns completed at the time $T_{a\ldots}$. An estimate of the maximal separation of the initial points of the two neighboring trajectories is achieved by Taylor expanding around $x_{a\ldots} = x_{\overline{a}} + \delta x_{a\ldots}$

$$f^{T_a}(x_{\overline{a}}) - f^{T_{a\ldots}}(x_{a\ldots}) \approx \frac{\partial f^{T_a}(x_{\overline{a}})}{\partial x} \cdot \delta x_{a\ldots} = \mathbf{J}_a \cdot \delta x_{a\ldots} ,$$

hence the hyperbolicity of the flow forces the initial points of neighboring trajectories that track each other for at least m consecutive symbols to lie exponentially close

$$|\delta x_{a\dots}| \propto \frac{1}{|\Lambda_a|} .$$

Similarly, for any observable (8.1) integrated along the two nearby trajectories

$$A^{T_{a\cdots}}(x_{a\dots}) \approx A^{T_a}(x_{\overline{a}}) + \left.\frac{\partial A^{T_a}}{\partial x}\right|_{x=x_{\overline{a}}} \cdot \delta x_{a\dots} ,$$

so

$$\left|A^{T_{a\cdots}}(x_{a\dots}) - A^{T_a}(x_{\overline{a}})\right| \propto \frac{T_a \mathrm{Const}}{|\Lambda_a|} ,$$

As the time of return is itself an integral along the trajectory, return times of nearby trajectories are exponentially close

$$|T_{a\dots} - T_a| \propto \frac{T_a \mathrm{Const}}{|\Lambda_a|} ,$$

and so are the trajectory stabilities

$$\left|A^{T_{a\cdots}}(x_{a\dots}) - A^{T_a}(x_{\overline{a}})\right| \propto \frac{T_a \mathrm{Const}}{|\Lambda_a|} ,$$

Substituting t_{ab} one finds

$$\frac{t_{ab} - t_a t_b}{t_{ab}} = 1 - e^{-s(T_a+T_b-T_{ab})} \left|\frac{\Lambda_a \Lambda_b}{\Lambda_{ab}}\right| .$$

Since with increasing m segments of $\overline{ab}$ come closer to $\overline{a}$, the differences in action and the ratio of the eigenvalues converge exponentially with the eigenvalue of the orbit $\overline{a}$,

$$T_a + T_b - T_{ab} \approx \mathrm{Const} \times \Lambda_a^{-j} , \quad |\Lambda_a \Lambda_b / \Lambda_{ab}| \approx \exp(-\mathrm{Const}/\Lambda_{ab})$$

n	$t_{ab} - t_a t_b$	$T_{ab} - (T_a + T_b)$	$\log \left[\frac{\Lambda_a \Lambda_b}{\Lambda_{ab}}\right]$	$ab - a \cdot b$
2	-5.23465150784×10^4	4.85802927371×10^2	-6.3×10^2	01-0·1
3	-7.96028600139×10^6	5.21713101432×10^3	-9.8×10^3	001-0·01
4	-1.03326529874×10^7	5.29858199419×10^4	-1.3×10^3	0001-0·001
5	-1.27481522016×10^9	5.35513574697×10^5	-1.6×10^4	00001-0·0001
6	-1.52544704823×10^{11}	5.40999882625×10^6	-1.8×10^5	000001-0·00001
2	-5.23465150784×10^4	4.85802927371×10^2	-6.3×10^2	01-0·1
3	5.30414752996×10^6	-3.67093656690×10^3	7.7×10^3	011-01·1
4	-5.40934261680×10^8	3.14925761316×10^4	-9.2×10^4	0111-011·1
5	4.99129508833×10^{10}	-2.67292822795×10^5	1.0×10^4	01111-0111·1
6	-4.39246000586×10^{12}	2.27087116266×10^6	-1.0×10^5	011111-01111·1

Table J.1: Demonstration of shadowing in curvature combinations of cycle weights of form $t_{ab} - t_a t_b$, the 3-disk fundamental domain cycles at $R : d = 6$, table 25.3. The ratio $\Lambda_a \Lambda_b / \Lambda_{ab}$ is approaching unity exponentially fast.

Expanding the exponentials one thus finds that this term in the cycle expansion is of the order of

$$t_{a^j b} - t_a t_{a^{j-1} b} \approx \text{Const} \times t_{a^j b} \Lambda_a^{-j} \,. \tag{J.1}$$

Even though the number of terms in a cycle expansion grows exponentially, the shadowing cancellations improve the convergence by an exponential factor compared to trace formulas, and extend the radius of convergence of the periodic orbit sums. Table J.1 shows some examples of such compensations between long cycles and their pseudo-cycle shadows.

It is crucial that the curvature expansion is grouped (and truncated) by topologically related cycles and pseudo-cycles; truncations that ignore topology, such as inclusion of all cycles with $T_p < T_{max}$, will contain orbits unmatched by shadowed orbits, and exhibit a mediocre convergence compared with the curvature expansions.

Note that the existence of a pole at $z = 1/c$ implies that the cycle expansions have a finite radius of convergence, and that analytic continuations will be required for extraction of the non-leading zeros of $1/\zeta$. Preferably, one should work with cycle expansions of Selberg products, as discussed in sect. 15.1.3.

J.1.3 No shadowing, poorer convergence

Conversely, if the dynamics is not of a finite subshift type, there is no finite topological polynomial, there are no "curvature" corrections, and the convergence of the cycle expansions will be poor.

J.2 On importance of pruning

If the grammar is not finite and there is no finite topological polynomial, there will be no "curvature" expansions, and the convergence will be poor. That is the generic case, and one strategy for dealing with it is to find a good sequence of approximate but finite grammars; for each approximate grammar cycle expansions yield exponentially accurate eigenvalues, with successive approximate grammars converging toward the desired infinite grammar system.

When the dynamical system's symbolic dynamics does not have a finite grammar, and we are not able to arrange its cycle expansion into curvature combinations (15.5), the series is truncated as in sect. 15.4, by including all pseudo-cycles such that $|\Lambda_{p_1} \cdots \Lambda_{p_k}| \leq |\Lambda_P|$, where P is the most unstable prime cycle included into truncation. The truncation error should then be of order $O(e^{hT_P} T_P / |\Lambda_P|)$, with h the topological entropy, and e^{hT_P} roughly the number of pseudo-cycles of stability $\approx |\Lambda_P|$. In this case the cycle averaging formulas do not converge significantly better than the approximations such as the trace formula (16.18).

Numerical results (see for example the plots of the accuracy of the cycle expansion truncations for the Hénon map in ref. [15.3]) indicate that the truncation error of most averages tracks closely the fluctuations due to the irregular growth in the number of cycles. It is not known whether one can exploit the sum rules such as the mass flow conservation (16.11) to improve the accuracy of dynamical averaging.

J.3 Ma-the-matical caveats

> "Lo duca e io per quel cammino ascoso intrammo a ritornar nel chiaro monde; e sanza cura aver d'alcun riposa
> salimmo sù, el primo e io secondo, tanto ch'i' vidi de le cose belle che porta 'l ciel, per un perutgio tondo."
> Dante

The periodic orbit theory is learned in stages. At first glance, it seems totally impenetrable. After basic exercises are gone through, it seems totally trivial; all that seems to be at stake are elementary manipulations with traces, determinants, derivatives. But if start thinking about you will get a more and more uncomfortable feeling that from the mathematical point of view, this is a perilous enterprise indeed. In chapter 13 we shall explain which parts of this enterprise are really solid; here you give a fortaste of what objections a mathematician might rise.

Birkhoff's 1931 ergodic theorem states that the time average (8.4) exists almost everywhere, and, if the flow is ergodic, it implies that $\langle a(x)\rangle = \langle a\rangle$ is a constant for almost all x. The problem is that the above cycle averaging formulas implicitly rely on ergodic hypothesis: they are strictly correct only if the dynamical system is locally hyperbolic and globally mixing. If one takes a β derivative of both sides

$$\rho_\beta(y) e^{ts(\beta)} = \int_{\mathcal{M}} dx\, \delta(y - f^t(x)) e^{\beta\cdot A^t(x)} \rho_\beta(x)\,,$$

and integrates over y

$$\begin{aligned}\int_{\mathcal{M}} dy\ \left.\frac{\partial}{\partial\beta}\rho_\beta(y)\right|_{\beta=0} \quad + \quad t\left.\frac{\partial s}{\partial\beta}\right|_{\beta=0} \int_{\mathcal{M}} dy\, \rho_0(y) \ = \\ \int_{\mathcal{M}} dx\, A^t(x)\rho_0(x) + \int_{\mathcal{M}} dx\ \left.\frac{\partial}{\partial\beta}\rho_\beta(x)\right|_{\beta=0}\,,\end{aligned}$$

one obtains in the long time limit

$$\left.\frac{\partial s}{\partial\beta}\right|_{\beta=0} = \int_{\mathcal{M}} dy\, \rho_0(x)\, \langle a(x)\rangle\ . \tag{J.2}$$

This is the expectation value (8.12) only if the time average (8.4) equals the space average (8.9), $\langle a(x)\rangle = \langle a\rangle$, for all x except a subset $x \in \mathcal{M}$ of zero measure; if the phase space is foliated into non-communicating subspaces $\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2$ of finite measure such that $f^t(\mathcal{M}_1) \cap \mathcal{M}_2 = \emptyset$ for all t, this fails. In other words, we have tacitly assumed metric indecomposability or transitivity. We have also glossed over the nature of the "phase space" $\mathcal{M}$. For example, if the dynamical system is open, such as the 3-disk game of pinball, $\mathcal{M}$ in the expectation value integral (8.18) is a Cantor set, the closure of the union of all periodic orbits. Alternatively, $\mathcal{M}$ can be considered continuous, but then the measure ρ_0 in (J.2) is highly singular. The beauty of the periodic orbit theory is that instead of using an arbitrary coordinatization of $\mathcal{M}$ it partitions the phase space by the intrinsic topology of the dynamical flow and builds the correct measure from cycle invariants, the stability eigenvalues of periodic orbits.

Were we to restrict the applications of the formalism only to systems which have been rigorously proven to be ergodic, we might as well fold up the shop right now. For example, even for something as simple as the

8.1
page 143

Hénon mapping we do not know whether the asymptotic time attractor is strange or periodic. Physics applications require a more pragmatic attitude. In the cycle expansions approach we construct the invariant set of the given dynamical system as a closure of the union of periodic orbits, and investigate how robust are the averages computed on this set. This turns out to depend very much on the observable being averaged over; dynamical averages exhibit "phase transitions", and the above cycle averaging formulas apply in the "hyperbolic phase" where the average is dominated by exponentially many exponentially small contributions, but fail in a phase dominated by few marginally stable orbits. Here the noise - always present, no matter how weak - helps us by erasing an infinity of small traps that the deterministic dynamics might fall into.

Still, in spite of all the caveats, periodic orbit theory is a beautiful theory, and the cycle averaging formulas are the most elegant and powerful tool available today for evaluation of dynamical averages for low dimensional chaotic deterministic systems.

J.4 Estimate of the nth cumulant

An immediate consequence of the exponential spacing of the eigenvalues is that the convergence of the Selberg product expansion (E.12) as function of the topological cycle length, $F(z) = \sum_n C_n z^n$, is faster than exponential. Consider a d–dimensional map for which all Jacobian matrix eigenvalues are equal: $u_p = \Lambda_{p,1} = \Lambda_{p,2} = \cdots = \Lambda_{p,d}$. The stability eigenvalues are generally not isotropic; however, to obtain qualitative bounds on the spectrum, we replace all stability eigenvalues with the least expanding one. In this case the p cycle contribution to the product (12.9) reduces to

$$\begin{aligned}
F_p(z) &= \prod_{k_1\cdots k_d=0}^{\infty}\left(1 - t_p u_p^{k_1+k_2+\cdots+k_d}\right) \\
&= \prod_{k=0}^{\infty}\left(1 - t_p u_p^{k}\right)^{m_k}; \qquad m_k = \binom{d-1+k}{d-1} = \frac{(k+d-1)!}{k!(d-1)!} \\
&= \prod_{k=0}^{\infty}\sum_{\ell=0}^{m_k}\binom{m_k}{\ell}\left(-u_p^k t_p\right)^{\ell}
\end{aligned} \tag{J.3}$$

In one dimension the expansion can be given in closed form (13.26), and the coefficients C_k in (E.12) are given by

$$\tau_{p^k} = (-1)^k \frac{u_p^{\frac{k(k-1)}{2}}}{\prod_{j=1}^{k}(1-u_p^j)} t_p^k \,. \tag{J.4}$$

Hence the coefficients in the $F(z) = \sum_n C_n z^n$ expansion of the spectral determinant (15.8) fall off faster than exponentially, as $|C_n| \approx u^{n(n-1)/2}$. In contrast, the cycle expansions of dynamical zeta functions fall of "only" exponentially; in numerical applications, the difference is dramatic.

In higher dimensions the expansions are not quite as compact. The leading power of u and its coefficient are easily evaluated by use of binomial expansions (J.3) of the $(1+tu^k)^{m_k}$ factors. More precisely, the leading u^n terms in t^k coefficients are of form

$$\begin{aligned}
\prod_{k=0}^{\infty}(1+tu^k)^{m_k} &= \ldots + u^{m_1+2m_2+\ldots+jm_j}\, t^{1+m_1+m_2+\ldots+m_j} + \ldots \\
&= \ldots + \left(u^{\frac{m_d}{d+1}} t\right)^{\binom{d+m}{m}} + \ldots \approx \ldots + u^{\frac{\sqrt[d]{d!}}{(d-1)!} n^{\frac{d+1}{d}}}\, t^n + \ldots
\end{aligned}$$

Hence the coefficients in the $F(z)$ expansion fall off faster than exponentially, as $u^{n^{1+1/d}}$. The Selberg products are entire functions in any dimension, provided that the symbolic dynamics is a finite subshift, and all cycle eigenvalues are sufficiently bounded away from 1.

The case of particular interest in many applications are the 2-d Hamiltonian mappings; their symplectic structure implies that $u_p = \Lambda_{p,1} = 1/\Lambda_{p,2}$, and the Selberg product (12.23) In this case the expansion corresponding to (13.26) is given by (13.27) and the coefficients fall off asymptotically as $C_n \approx u^{n^{3/2}}$.

Appendix K

Infinite dimensional operators

(A. Wirzba)

This appendix taken from ref. [K.1] summarizes the definitions and properties for trace-class and Hilbert-Schmidt matrices, the determinants over infinite dimensional matrices and possible regularization schemes for matrices or operators which are not of trace-class.

K.1 Matrix-valued functions

(P. Cvitanović)

As a preliminary we summarize some of the properties of functions of finite-dimensional matrices.

The derivative of a matrix is a matrix with elements

$$A'(x) = \frac{dA(x)}{dx}\,, \qquad A'_{ij}(x) = \frac{d}{dx}A_{ij}(x)\,. \tag{K.1}$$

Derivatives of products of matrices are evaluated by the chain rule

$$\frac{d}{dx}(A\mathbf{B}) = \frac{dA}{dx}\mathbf{B} + A\frac{d\mathbf{B}}{dx}\,. \tag{K.2}$$

A matrix and its derivative matrix in general do not commute

$$\frac{d}{dx}A^2 = \frac{dA}{dx}A + A\frac{dA}{dx}\,. \tag{K.3}$$

The derivative of the inverse of a matrix, follows from $\frac{d}{dx}(AA^{-1}) = 0$:

$$\frac{d}{dx}A^{-1} = -\frac{1}{A}\frac{dA}{dx}\frac{1}{A}. \tag{K.4}$$

A function of a single variable that can be expressed in terms of additions and multiplications generalizes to a matrix-valued function by replacing the variable by the matrix.

In particular, the exponential of a constant matrix can be defined either by its series expansion, or as a limit of an infinite product:

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!}A^k \,, \qquad A^0 = \mathbf{1} \tag{K.5}$$

$$= \lim_{N\to\infty}\left(\mathbf{1} + \frac{1}{N}A\right)^N \tag{K.6}$$

The first equation follows from the second one by the binomial theorem, so these indeed are equivalent definitions. That the terms of order $O(N^{-2})$ or smaller do not matter follows from the bound

$$\left(1 + \frac{x-\epsilon}{N}\right)^N < \left(1 + \frac{x+\delta x_N}{N}\right)^N < \left(1 + \frac{x+\epsilon}{N}\right)^N ,$$

where $|\delta x_N| < \epsilon$. If $\lim \delta x_N \to 0$ as $N \to \infty$, the extra terms do not contribute.

Consider now the determinant

$$\det\left(e^A\right) = \lim_{N\to\infty}\left(\det\,(\mathbf{1} + A/N)\right)^N .$$

To the leading order in $1/N$

$$\det\,(\mathbf{1} + A/N) = 1 + \frac{1}{N}\mathrm{tr}\,A + O(N^{-2}) .$$

hence

$$\det e^A = \lim_{N\to\infty}\left(1 + \frac{1}{N}\mathrm{tr}\,A + O(N^{-2})\right)^N = e^{\mathrm{tr}\,A} \tag{K.7}$$

Due to non-commutativity of matrices, generalization of a function of several variables to a function is not as straightforward. Expression involving several matrices depend on their commutation relations. For example, the commutator expansion

$$e^{t\mathbf{A}}\mathbf{B}e^{-t\mathbf{A}} = \mathbf{B}+t[\mathbf{A},\mathbf{B}]+\frac{t^2}{2}[\mathbf{A},[\mathbf{A},\mathbf{B}]]+\frac{t^3}{3!}[\mathbf{A},[\mathbf{A},[\mathbf{A},\mathbf{B}]]]+\cdots \tag{K.8}$$

sometimes used to establish the equivalence of the Heisenberg and Schrödinger pictures of quantum mechanics follows by recursive evaluation of t derivaties

$$\frac{d}{dt}\left(e^{t\mathbf{A}}\mathbf{B}e^{-t\mathbf{A}}\right) = e^{t\mathbf{A}}[\mathbf{A},\mathbf{B}]e^{-t\mathbf{A}}\,.$$

Manipulations of such ilk yield

$$e^{(\mathbf{A}+\mathbf{B})/N} = e^{\mathbf{A}/N}e^{\mathbf{B}/N} - \frac{1}{2N^2}[\mathbf{A},\mathbf{B}] + O(N^{-3})\,,$$

and the Trotter product formula: if $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{A} = \mathbf{B}+\mathbf{C}$ are matrices, then

$$e^{\mathbf{A}} = \lim_{N\to\infty}\left(e^{\mathbf{B}/N}e^{\mathbf{C}/N}\right)^N \tag{K.9}$$

K.2 Operator norms

(R. Mainieri and P. Cvitanović)

The limit used in the above definition involves matrices - operators in vector spaces - rather than numbers, and its convergence can be checked using tools familiar from calculus. We briefly review those tools here, as throughout the text we will have to consider many different operators and how they converge.

The $n \to \infty$ convergence of partial products

$$\mathbf{E}_n = \prod_{0\le m<n}\left(\mathbf{1}+\frac{t}{m}A\right)$$

can be verified using the Cauchy criterion, which states that the sequence $\{\mathbf{E}_n\}$ converges if the differences $\|\mathbf{E}_k - \mathbf{E}_j\| \to 0$ as $k, j \to \infty$. To make sense of this we need to define a sensible norm $\|\cdots\|$. Norm of a matrix is based on the Euclidean norm for a vector: the idea is to assign to a matrix $\mathbf{M}$ a norm that is the largest possible change it can cause to the length of a unit vector $\hat{n}$:

$$\|\mathbf{M}\| = \sup_{\hat{n}} \|\mathbf{M}\hat{n}\|, \qquad \|\hat{n}\| = 1. \tag{K.10}$$

We say that $\|\cdot\|$ is the operator norm induced by the vector norm $\|\cdot\|$. Constructing a norm for a finite-dimensional matrix is easy, but had $\mathbf{M}$ been an operator in an infinite-dimensional space, we would also have to specify the space $\hat{n}$ belongs to. In the finite-dimensional case, the sum of the absolute values of the components of a vector is also a norm; the induced operator norm for a matrix $\mathbf{M}$ with components M_{ij} in that case can be defined by

$$\|\mathbf{M}\| = \max_i \sum_j |M_{ij}|. \tag{K.11}$$

The operator norm (K.11) and the vector norm (K.10) are only rarely distinguished by different notation, a bit of notational laziness that we shall uphold.

Now that we have learned how to make sense out of norms of operators, we can check that

$$\|e^{tA}\| \le e^{t\|A\|}. \tag{K.12}$$

2.8
page 49

As $\|A\|$ is a number, the norm of e^{tA} is finite and therefore well defined. In particular, the exponential of a matrix is well defined for all values of t, and the linear differential equation (4.7) has a solution for all times.

K.3 Trace class and Hilbert-Schmidt class

This section is mainly an extract from ref. [K.9]. Refs. [K.7, K.10, K.11, K.14] should be consulted for more details and proofs. The trace class and Hilbert-Schmidt property will be defined here for linear, in general non-hermitian operators $\mathbf{A} \in \mathcal{L}(\mathcal{H})$: $\mathcal{H} \to \mathcal{H}$ (where $\mathcal{H}$ is a separable Hilbert space). The transcription to matrix elements (used in the prior chapters) is simply $a_{ij} = \langle \phi_i, \mathbf{A}\phi_j \rangle$ where $\{\phi_n\}$ is an orthonormal basis of $\mathcal{H}$ and $\langle\ ,\ \rangle$ is the inner product in $\mathcal{H}$ (see sect. K.5 where the theory of *von Koch matrices* of ref. [K.12] is discussed). So, the trace is the generalization of the usual notion of the sum of the diagonal elements of a matrix; but because infinite sums are involved, not all operators will have a trace:

Definition:

(a) An operator $\mathbf{A}$ is called **trace class**, $\mathbf{A} \in \mathcal{J}_1$, if and only if, for every orthonormal basis, $\{\phi_n\}$:

$$\sum_n |\langle \phi_n, \mathbf{A}\phi_n \rangle| < \infty. \tag{K.13}$$

The family of all trace class operators is denoted by $\mathcal{J}_1$.

(b) An operator $\mathbf{A}$ is called **Hilbert-Schmidt**, $\mathbf{A} \in \mathcal{J}_2$, if and only if, for every orthonormal basis, $\{\phi_n\}$:

$$\sum_n \|\mathbf{A}\phi_n\|^2 < \infty. \tag{K.14}$$

The family of all Hilbert-Schmidt operators is denoted by $\mathcal{J}_2$.

Bounded operators are dual to trace class operators. They satisfy the the following condition: $|\langle \psi, B\phi \rangle| \le C\|\psi\|\|\phi\|$ with $C < \infty$ and $\psi, \phi \in \mathcal{H}$. If they have eigenvalues, these are bounded too. The family of bounded operators is denoted by $\mathcal{B}(\mathcal{H})$ with the norm $\|B\| = \sup_{\phi \neq 0} \frac{\|\mathbf{B}\phi\|}{\|\phi\|}$ for $\phi \in \mathcal{H}$. Examples for bounded operators are unitary operators and especially the unit matrix. In fact, every bounded operator can be written as linear combination of four unitary operators.

A bounded operator $\mathbf{C}$ is *compact*, if it is the norm limit of finite rank operators.

An operator $\mathbf{A}$ is called *positive*, $\mathbf{A} \ge 0$, if $\langle \mathbf{A}\phi, \phi \rangle \ge 0\ \forall \phi \in \mathcal{H}$. Notice that $\mathbf{A}^\dagger\mathbf{A} \ge 0$. We define $|\mathbf{A}| = \sqrt{\mathbf{A}^\dagger\mathbf{A}}$.

The most important properties of the trace and Hilbert-Schmidt classes are summarized in (see refs. [K.7, K.9]):

(a) $\mathcal{J}_1$ and $\mathcal{J}_2$ are $*$ideals., i.e., they are vector spaces closed under scalar multiplication, sums, adjoints, and multiplication with bounded operators.

(b) $\mathbf{A} \in \mathcal{J}_1$ if and only if $\mathbf{A} = \mathbf{BC}$ with $\mathbf{B}, \mathbf{C} \in \mathcal{J}_2$.

(c) $\mathcal{J}_1 \subset \mathcal{J}_2 \subset$ Compact operators.

(d) For any operator $\mathbf{A}$, we have $\mathbf{A} \in \mathcal{J}_2$ if $\sum_n \|\mathbf{A}\phi_n\|^2 < \infty$ for a single basis.
For any operator $\mathbf{A} \geq 0$ we have $\mathbf{A} \in \mathcal{J}_1$ if $\sum_n |\langle \phi_n, \mathbf{A}\phi_n \rangle| < \infty$ for a single basis.

(e) If $\mathbf{A} \in \mathcal{J}_1$, $\mathrm{Tr}(\mathbf{A}) = \sum \langle \phi_n, \mathbf{A}\phi_n \rangle$ is independent of the basis used.

(f) Tr is linear and obeys $\mathrm{Tr}(\mathbf{A}^\dagger) = \overline{\mathrm{Tr}(\mathbf{A})}$; $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$ if either $\mathbf{A} \in \mathcal{J}_1$ and $\mathbf{B}$ bounded, $\mathbf{A}$ bounded and $\mathbf{B} \in \mathcal{J}_1$ or both $\mathbf{A}, \mathbf{B} \in \mathcal{J}_2$.

(g) $\mathcal{J}_2$ endowed with the inner product $\langle \mathbf{A}, \mathbf{B} \rangle_2 = \mathrm{Tr}(\mathbf{A}^\dagger \mathbf{B})$ is a Hilbert space. If $\|\mathbf{A}\|_2 = [\,\mathrm{Tr}(\mathbf{A}^\dagger \mathbf{A})\,]^{\frac{1}{2}}$, then $\|\mathbf{A}\|_2 \geq \|\mathbf{A}\|$ and $\mathcal{J}_2$ is the $\| \; \|_2$-closure of the *finite* rank operators.

(h) $\mathcal{J}_1$ endowed with the norm $\|\mathbf{A}\|_1 = \mathrm{Tr}(\sqrt{\mathbf{A}^\dagger \mathbf{A}})$ is a Banach space. $\|\mathbf{A}\|_1 \geq \|\mathbf{A}\|_2 \geq \|\mathbf{A}\|$ and $\mathcal{J}_1$ is the $\| \; \|_1$-norm closure of the *finite* rank operators. The dual space of $\mathcal{J}_1$ is $\mathcal{B}(\mathcal{H})$, the family of bounded operators with the duality $\langle \mathbf{B}, \mathbf{A} \rangle = \mathrm{Tr}(\mathbf{BA})$.

(i) If $\mathbf{A}, \mathbf{B} \in \mathcal{J}_2$, then $\|\mathbf{AB}\|_1 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$. If $\mathbf{A} \in \mathcal{J}_2$ and $\mathbf{B} \in \mathcal{B}(\mathcal{H})$, then $\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|$. If $\mathbf{A} \in \mathcal{J}_1$ and $\mathbf{B} \in \mathcal{B}(\mathcal{H})$, then $\|\mathbf{AB}\|_1 \leq \|\mathbf{A}\|_1 \|\mathbf{B}\|$.

Note the most important property for proving that an operator is trace class is the decomposition **(b)** into two Hilbert-Schmidt ones, as the Hilbert-Schmidt property can easily be verified in one single orthonormal basis (see **(d)**). Property **(e)** ensures then that the trace is the same in any basis. Properties **(a)** and **(f)** show that trace class operators behave in complete analogy to finite rank operators. The proof whether a matrix is trace-class (or Hilbert-Schmidt) or not simplifies enormously for diagonal matrices, as then the second part of property **(d)** is directly applicable: just the moduli of the eigenvalues (or – in case of Hilbert-Schmidt – the squares of the eigenvalues) have to be summed up in order to answer that question. A good strategy in checking the trace-class character of a general matrix $\mathbf{A}$ is therefore the decomposition of that matrix into two matrices $\mathbf{B}$ and $\mathbf{C}$ where one, say $\mathbf{C}$, should be chosen to be diagonal and either just barely of Hilbert-Schmidt character leaving enough freedom for its partner $\mathbf{B}$ or of trace-class character such that one only has to show the boundedness for $\mathbf{B}$.

K.4 Determinants of trace class operators

This section is mainly based on refs. [K.8, K.10] which should be consulted for more details and proofs. See also refs. [K.11, K.14].

Pre-definitions (Alternating algebra and Fock spaces):
Given a Hilbert space $\mathcal{H}$, $\otimes^n\mathcal{H}$ is defined as the vector space of multi-linear functionals on $\mathcal{H}$ with $\phi_1 \otimes \cdots \otimes \phi_n \in \otimes^n\mathcal{H}$ in case $\phi_1, \ldots, \phi_n \in \mathcal{H}$. $\bigwedge^n(\mathcal{H})$ is defined as the subspace of $\otimes^n\mathcal{H}$ spanned by the wedge-product

$$\phi_1 \wedge \cdots \wedge \phi_n = \frac{1}{\sqrt{n!}} \sum_{\pi \in \mathcal{P}_n} \epsilon(\pi)[\phi_{\pi(1)} \otimes \cdots \otimes \phi_{\pi(n)}] \tag{K.15}$$

where $\mathcal{P}_n$ is the group of all permutations of n letters and $\epsilon(\pi) = \pm 1$ depending on whether π is an even or odd permutation, respectively. The inner product in $\bigwedge^n(\mathcal{H})$ is given by

$$(\phi_1 \wedge \cdots \wedge \phi_n, \eta_1 \wedge \cdots \wedge \eta_n) = \det\{(\phi_i, \eta_j)\} \tag{K.16}$$

where $\det\{a_{ij}\} = \sum_{\pi \in \mathcal{P}_n} \epsilon(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)}$. $\bigwedge^n(\mathbf{A})$ is defined as functor (a functor satisfies $\bigwedge^n(\mathbf{AB}) = \bigwedge^n(\mathbf{A}) \bigwedge^n(\mathbf{B})$) on $\bigwedge^n(\mathcal{H})$ with

$$\bigwedge^n(\mathbf{A})\,(\phi_1 \wedge \cdots \wedge \phi_n) = \mathbf{A}\phi_1 \wedge \cdots \wedge \mathbf{A}\phi_n\ . \tag{K.17}$$

When $n = 0$, $\bigwedge^n(\mathcal{H})$ is defined to be C and $\bigwedge^n(\mathbf{A})$ as 1: $C \to C$.

Properties: If $\mathbf{A}$ trace class, i.e., $\mathbf{A} \in \mathcal{J}_1$, then for any k, $\bigwedge^k(\mathbf{A})$ is trace class, and for any orthonormal basis $\{\phi_n\}$ the cumulant

$$\mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right) = \sum_{i_1 < \cdots < i_k} ((\phi_{i_1} \wedge \cdots \wedge \phi_{i_k}), (\mathbf{A}\phi_{i_1} \wedge \cdots \wedge \mathbf{A}\phi_{i_k})) < \infty \tag{K.18}$$

is independent of the basis (with the understanding that $\mathrm{Tr}\bigwedge^0(\mathbf{A}) \equiv 1$).

Definition: Let $\mathbf{A} \in \mathcal{J}_1$, then $\mathbf{det(1+A)}$ is defined as

$$\det(\mathbf{1} + \mathbf{A}) = \sum_{k=0}^{\infty} \mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right) \tag{K.19}$$

Properties:

Let $\mathbf{A}$ be a linear operator on a separable Hilbert space $\mathcal{H}$ and $\{\phi_j\}_1^\infty$ an orthonormal basis.

(a) $\sum_{k=0}^{\infty} \mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right)$ converges for each $\mathbf{A} \in \mathcal{J}_1$.

(b) $|\det(\mathbf{1}+\mathbf{A})| \leq \prod_{j=1}^{\infty}(1+\mu_j(\mathbf{A}))$ where $\mu_j(\mathbf{A})$ are the *singular* values of $\mathbf{A}$, i.e., the eigenvalues of $|\mathbf{A}| = \sqrt{\mathbf{A}^\dagger\mathbf{A}}$.

(c) $|\det(\mathbf{1}+\mathbf{A})| \leq \exp(\|\mathbf{A}\|_1)$.

(d) For any $\mathbf{A}_1, \ldots, \mathbf{A}_n \in \mathcal{J}_1$, $\langle z_1, \ldots, z_n\rangle \mapsto \det\left(\mathbf{1} + \sum_{i=1}^n z_i\mathbf{A}_i\right)$ is an entire analytic function.

(e) If $\mathbf{A}, \mathbf{B} \in \mathcal{J}_1$, then

$$\begin{aligned} \det(\mathbf{1}+\mathbf{A})\det(\mathbf{1}+\mathbf{B}) &= \det\left(\mathbf{1}+\mathbf{A}+\mathbf{B}+\mathbf{AB}\right) \\ &= \det\left((\mathbf{1}+\mathbf{A})(\mathbf{1}+\mathbf{B})\right) \\ &= \det\left((\mathbf{1}+\mathbf{B})(\mathbf{1}+\mathbf{A})\right) . \end{aligned} \tag{K.20}$$

If $\mathbf{A} \in \mathcal{J}_1$ and $\mathbf{U}$ unitary, then

$$\det\left(\mathbf{U}^{-1}(\mathbf{1}+\mathbf{A})\mathbf{U}\right) = \det\left(\mathbf{1}+\mathbf{U}^{-1}\mathbf{A}\mathbf{U}\right) = \det(\mathbf{1}+\mathbf{A}) . \tag{K.21}$$

(f) If $\mathbf{A} \in \mathcal{J}_1$, then $(\mathbf{1}+\mathbf{A})$ is invertible if and only if $\det(\mathbf{1}+\mathbf{A}) \neq 0$.

(g) If $\lambda \neq 0$ is an n-times degenerate eigenvalue of $\mathbf{A} \in \mathcal{J}_1$, then $\det(\mathbf{1}+z\mathbf{A})$ has a zero of order n at $z = -1/\lambda$.

(h) For any ϵ, there is a $C_\epsilon(\mathbf{A})$, depending on $\mathbf{A} \in \mathcal{J}_1$, so that $|\det(\mathbf{1}+z\mathbf{A})| \leq C_\epsilon(\mathbf{A})\exp(\epsilon|z|)$.

(i) For any $\mathbf{A} \in \mathcal{J}_1$,

$$\det(\mathbf{1}+\mathbf{A}) = \prod_{j=1}^{N(\mathbf{A})}(1+\lambda_j(\mathbf{A})) \tag{K.22}$$

where here and in the following $\{\lambda_j(\mathbf{A})\}_{j=1}^{N(\mathbf{A})}$ are the eigenvalues of $\mathbf{A}$ counted with algebraic multiplicity .

(j) *Lidskii's theorem:* For any $\mathbf{A} \in \mathcal{J}_1$,

$$\mathrm{Tr}(\mathbf{A}) = \sum_{j=1}^{N(\mathbf{A})} \lambda_j(\mathbf{A}) < \infty . \tag{K.23}$$

(k) If $\mathbf{A} \in \mathcal{J}_1$, then

$$\begin{aligned} \mathrm{Tr}\left(\bigwedge^k(\mathbf{A})\right) &= \sum_{j=1}^{N\left(\bigwedge^k(\mathbf{A})\right)} \lambda_j\left(\bigwedge^k(\mathbf{A})\right) \\ &= \sum_{1\leq j_1<\cdots<j_k\leq N(\mathbf{A})} \lambda_{j_1}(\mathbf{A})\cdots\lambda_{j_k}(\mathbf{A}) < \infty. \end{aligned}$$

(l) If $\mathbf{A} \in \mathcal{J}_1$, then

$$\det(1 + z\mathbf{A}) = \sum_{k=0}^{\infty} z^k \sum_{1 \le j_1 < \cdots < j_k \le N(\mathbf{A})} \lambda_{j_1}(\mathbf{A}) \cdots \lambda_{j_k}(\mathbf{A}) < \infty \tag{K.24}$$

(m) If $\mathbf{A} \in \mathcal{J}_1$, then for $|z|$ small (that is $|z| \max|\lambda_j(\mathbf{A})| < 1$) the series $\sum_{k=1}^{\infty} z^k \mathrm{Tr}\left((-\mathbf{A})^k\right)/k$ converges and

$$\begin{aligned} \det(1 + z\mathbf{A}) &= \exp\left(-\sum_{k=1}^{\infty} \frac{z^k}{k} \mathrm{Tr}\left((-\mathbf{A})^k\right)\right) \\ &= \exp\left(\mathrm{Tr}\ln(\mathbf{1} + z\mathbf{A})\right) . \end{aligned} \tag{K.25}$$

(n) *The Plemelj-Smithies formula:* Define $\alpha_m(\mathbf{A})$ for $\mathbf{A} \in \mathcal{J}_1$ by

$$\det(\mathbf{1} + z\mathbf{A}) = \sum_{m=0}^{\infty} z^m \frac{\alpha_m(\mathbf{A})}{m!} . \tag{K.26}$$

Then $\alpha_m(\mathbf{A})$ is given by the $m \times m$ determinant:

$$\alpha_m(\mathbf{A}) = \begin{vmatrix} \mathrm{Tr}(\mathbf{A}) & m-1 & 0 & \cdots & 0 \\ \mathrm{Tr}(\mathbf{A}^2) & \mathrm{Tr}(\mathbf{A}) & m-2 & \cdots & 0 \\ \mathrm{Tr}(\mathbf{A}^3) & \mathrm{Tr}(\mathbf{A}^2) & \mathrm{Tr}(\mathbf{A}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 1 \\ \mathrm{Tr}(\mathbf{A}^m) & \mathrm{Tr}(\mathbf{A}^{(m-1)}) & \mathrm{Tr}(\mathbf{A}^{(m-2)}) & \cdots & \mathrm{Tr}(\mathbf{A}) \end{vmatrix} \tag{K.27}$$

with the understanding that $\alpha_0(\mathbf{A}) \equiv 1$ and $\alpha_1(\mathbf{A}) \equiv \mathrm{Tr}(\mathbf{A})$. Thus the cumulants $c_m(\mathbf{A}) \equiv \alpha_m(\mathbf{A})/m!$ satisfy the following recursion relation

$$\begin{aligned} c_m(\mathbf{A}) &= \frac{1}{m} \sum_{k=1}^{m} (-1)^{k+1} c_{m-k}(\mathbf{A})\, \mathrm{Tr}(\mathbf{A}^k) \quad \text{for } m \ge 1 \\ c_0(\mathbf{A}) &\equiv 1 . \end{aligned} \tag{K.28}$$

Note that in the context of quantum mechanics formula (K.26) is the quantum analog to the curvature expansion of the semiclassical zeta function with $\mathrm{Tr}(\mathbf{A}^m)$ corresponding to the sum of all periodic orbits (prime and also repeated ones) of *total* topological length m, that is let $c_m(\text{s.c.})$ denote the m^{th} curvature term, then the curvature expansion of the semiclassical zeta function is given by the recursion relation

$$\begin{aligned} c_m(\text{s.c.}) &= \frac{1}{m} \sum_{k=1}^{m} (-1)^{k+m+1} c_{m-k}(\text{s.c.}) \sum_{\substack{p;r>0 \\ \text{with } [p]r=k}} [p] \frac{t_p(k)^r}{1 - \left(\frac{1}{\Lambda_p}\right)^r} \quad \text{for } m \ge 1 \\ c_0(\text{s.c.}) &\equiv 1 . \end{aligned} \tag{K.29}$$

In fact, in the cumulant expansion (K.26) as well as in the curvature expansion there are large cancellations involved. Let us order – without lost of generality – the eigenvalues of the operator $\mathbf{A} \in \mathcal{J}_1$ as follows:

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{i-1}| \geq |\lambda_i| \geq |\lambda_{i+1}| \geq \cdots$$

(This is always possible because of $\sum_{i=1}^{N(\mathbf{A})} |\lambda_i| < \infty$.) Then, in the standard (Plemelj-Smithies) cumulant evaluation of the determinant, eq. (K.26), we have enormous cancellations of big numbers, e.g. at the k^{th} cumulant order ($k > 3$), all the intrinsically large 'numbers' λ_1^k, $\lambda_1^{k-1}\lambda_2$, ..., $\lambda_1^{k-2}\lambda_2\lambda_3$, ... and many more have to cancel out exactly until only $\sum_{1\leq j_1<\cdots<j_k\leq N(\mathbf{A})} \lambda_{j_1}\cdots\lambda_{j_k}$ is finally left over. Algebraically, the fact that there are these large cancellations is of course of no importance. However, if the determinant is calculated numerically, the big cancellations might spoil the result or even the convergence. Now, the curvature expansion of the semiclassical zeta function, as it is known today, *is* the semi-classical approximation to the curvature expansion (unfortunately) in the Plemelj-Smithies form. As the exact quantum mechanical result is approximated semi-classically, the errors introduced in the approximation might lead to big effects as they are done with respect to large quantities which eventually cancel out and not – as it would be of course better – with respect to the small surviving cumulants. Thus it would be very desirable to have a semi-classical analog to the reduced cumulant expansion (K.24) or even to (K.22) directly. It might not be possible to find a direct semi-classical analog for the individual eigenvalues λ_j. Thus the direct construction of the semi-classical equivalent to (K.22) is rather unlikely. However, in order to have a semi-classical "cumulant" summation without large cancellations – see (K.24) – it would be just sufficient to find the semi-classical analog of each complete cumulant (K.24) and not of the single eigenvalues. Whether this will eventually be possible is still an open question.

K.5 Von Koch matrices

Implicitly, many of the above properties are based on the theory of von Koch matrices [K.11, K.12, K.13]: An infinite matrix $\mathbf{1}-\mathbf{A} = \|\delta_{jk} - a_{jk}\|_1^\infty$, consisting of complex numbers, is called a matrix with an *absolutely convergent determinant*, if the series $\sum |a_{j_1k_1}a_{j_2k_2}\cdots a_{j_n,k_n}|$ converges, where the sum extends over all pairs of systems of indices $(j_1, j_2, \cdots, j_n)$ and $(k_1, k_2, \cdots, k_n)$ which differ from each other only by a permutation, and $j_1 < j_2 < \cdots j_n$ $(n = 1, 2, \cdots)$. Then the limit

$$\lim_{n\to\infty} \det\|\delta_{jk} - a_{jk}\|_1^n = \det(\mathbf{1} - \mathbf{A})$$

exists and is called the determinant of the matrix $\mathbf{1} - \mathbf{A}$. It can be represented in the form

$$\det(\mathbf{1} - \mathbf{A}) = 1 - \sum_{j=1}^{\infty} a_{jj} + \frac{1}{2!} \sum_{j,k=1}^{\infty} \begin{vmatrix} a_{jj} & a_{jk} \\ a_{kj} & a_{kk} \end{vmatrix} - \frac{1}{3!} \sum_{j,k,m=1}^{\infty} \begin{vmatrix} a_{jj} & a_{jk} & a_{jm} \\ a_{kj} & a_{kk} & a_{km} \\ a_{mj} & a_{mk} & a_{mm} \end{vmatrix} + \cdots ,$$

where the series on the r.h.s. will remain convergent even if the numbers a_{jk} $(j, k = 1, 2, \cdots)$ are replaced by their moduli and if all the terms obtained by expanding the determinants are taken with the plus sign. The matrix $\mathbf{1} - \mathbf{A}$ is called *von Koch matrix*, if both conditions

$$\sum_{j=1}^{\infty} |a_{jj}| \;<\; \infty , \tag{K.30}$$

$$\sum_{j,k=1}^{\infty} |a_{jk}|^2 \;<\; \infty \tag{K.31}$$

are fulfilled. Then the following holds (see ref. [K.11, K.13]): (1) Every von Koch matrix has an absolutely convergent determinant. If the elements of a von Koch matrix are functions of some parameter μ ($a_{jk} = a_{jk}(\mu)$, $j, k = 1, 2, \cdots$) and both series in the defining condition converge uniformly in the domain of the parameter μ, then as $n \to \infty$ the determinant $\det\|\delta_{jk} - a_{jk}(\mu)\|_1^n$ tends to the determinant $\det(\mathbf{1} + \mathbf{A}(\mu))$ uniformly with respect to μ, over the domain of μ. (2) If the matrices $\mathbf{1} - \mathbf{A}$ and $\mathbf{1} - \mathbf{B}$ are von Koch matrices, then their product $\mathbf{1} - \mathbf{C} = (\mathbf{1} - \mathbf{A})(\mathbf{1} - \mathbf{B})$ is a von Koch matrix, and

$$\det(\mathbf{1} - \mathbf{C}) = \det(\mathbf{1} - \mathbf{A}) \det(\mathbf{1} - \mathbf{B}) \ . \tag{K.32}$$

Note that every trace-class matrix $\mathbf{A} \in \mathcal{J}_1$ is also a von Koch matrix (and that any matrix satisfying condition (K.31) is Hilbert-Schmidt and vice versa). The inverse implication, however, is not true: von Koch matrices are not automatically trace-class. The caveat is that the definition of von Koch matrices is basis-dependent, whereas the trace-class property is basis-*independent*. As the traces involve infinite sums, the basis-independence is not at all trivial. An example for an infinite matrix which is von Koch, but not trace-class is the following:

$$\mathbf{A}_{ij} = \begin{cases} 2/j & \text{for} \quad i - j = -1 \quad \text{and} \quad j \text{ even} , \\ 2/i & \text{for} \quad i - j = +1 \quad \text{and} \quad i \text{ even} , \\ 0 & \text{else} , \end{cases}$$

i.e.,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1/2 & 0 & 0 & \cdots \\ 0 & 0 & 1/2 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1/3 & \ddots \\ 0 & 0 & 0 & 0 & 1/3 & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix} . \tag{K.33}$$

Obviously, condition (K.30) is fulfilled by definition. Secondly, condition (K.31) is satisfied as $\sum_{n=1}^{\infty} 2/n^2 < \infty$. However, the sum over the moduli of the eigenvalues is just twice the harmonic series $\sum_{n=1}^{\infty} 1/n$ which does not converge. The matrix (K.33) violates the trace-class definition (K.13), as in its eigenbasis the sum over the moduli of its diagonal elements is infinite. Thus the *absolute* convergence is traded for a *conditional* convergence, since the sum over the eigenvalues themselves can be arranged to still be zero, if the eigenvalues with the same modulus are summed first. Absolute convergence is of course essential, if sums have to be rearranged or exchanged. Thus, the trace-class property is indispensable for any controlled unitary transformation of an infinite determinant, as then there will be necessarily a change of basis and in general also a re-ordering of the corresponding traces. Therefore the claim that *a Hilbert-Schmidt operator with a vanishing trace is automatically trace-class* is false. In general, such an operator has to be regularized in addition (see next chapter).

K.6 Regularization

Many interesting operators are not of trace class (although they might be in some $\mathcal{J}_p$ with $p > 1$ - an operator A is in $\mathcal{J}_p$ iff $\mathrm{Tr}|A|^p < \infty$ in any orthonormal basis). In order to compute determinants of such operators, an extension of the cumulant expansion is needed which in fact corresponds to a regularization procedure [K.8, K.10]:
E.g. let $\mathbf{A} \in \mathcal{J}_p$ with $p \leq n$. Define

$$R_n(z\mathbf{A}) = (\mathbf{1} + z\mathbf{A}) \exp\left(\sum_{k=1}^{n-1} \frac{(-z)^k}{k} \mathbf{A}^k \right) - \mathbf{1} \tag{K.34}$$

as the regulated version of the operator $z\mathbf{A}$. Then the regulated operator $R_n(z\mathbf{A})$ is trace class, i.e., $R_n(z\mathbf{A}) \in \mathcal{J}_1$. Define now $\det_n(\mathbf{1} + z\mathbf{A}) =$

$\det(\mathbf{1} + R_n(z\mathbf{A}))$. Then the regulated determinant

$$\det{}_n(\mathbf{1} + z\mathbf{A}) = \prod_{j=1}^{N(z\mathbf{A})} \left[(1 + z\lambda_j(\mathbf{A})) \exp\left(\sum_{k=1}^{n-1} \frac{(-z\lambda_j(\mathbf{A}))^k}{k} \right) \right] < \infty \tag{K.35}$$

exists and is finite. The corresponding Plemelj-Smithies formula now reads [K.10]:

$$\det{}_n(\mathbf{1} + z\mathbf{A}) = \sum_{m=0}^{\infty} z^m \frac{\alpha_m^{(n)}(\mathbf{A})}{m!} \,. \tag{K.36}$$

with $\alpha_m^{(n)}(\mathbf{A})$ given by the $m \times m$ determinant:

$$\alpha_m^{(n)}(\mathbf{A}) = \begin{vmatrix} \sigma_1^{(n)} & m-1 & 0 & \cdots & 0 \\ \sigma_2^{(n)} & \sigma_1^{(n)} & m-2 & \cdots & 0 \\ \sigma_3^{(n)} & \sigma_2^{(n)} & \sigma_1^{(n)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 1 \\ \sigma_m^{(n)} & \sigma_{m-1}^{(n)} & \sigma_{m-2}^{(n)} & \cdots & \sigma_1^{(n)} \end{vmatrix} \tag{K.37}$$

where

$$\sigma_k^{(n)} = \begin{cases} \mathrm{Tr}(\mathbf{A}^k) & k \geq n \\ 0 & k \leq n-1 \end{cases}$$

As Simon [K.10] says simply, the beauty of (K.37) is that we get $\det_n(\mathbf{1} + \mathbf{A})$ from the standard Plemelj-Smithies formula (K.26) by simply setting $\mathrm{Tr}(\mathbf{A})$, $\mathrm{Tr}(\mathbf{A}^2)$, …, $\mathrm{Tr}(\mathbf{A}^{n-1})$ to zero.

See also ref. [K.15] where $\{\lambda_j\}$ are the eigenvalues of an elliptic (pseudo)-differential operator $\mathbf{H}$ of order m on a compact or bounded manifold of dimension d, $0 < \lambda_0 \leq \lambda_1 \leq \cdots$ and $\lambda_k \uparrow +\infty$. and the Fredholm determinant

$$\Delta(\lambda) = \prod_{k=0}^{\infty} \left(1 - \frac{\lambda}{\lambda_k} \right) \tag{K.38}$$

is regulated in the case $\mu \equiv d/m > 1$ as Weierstrass product

$$\Delta(\lambda) = \prod_{k=0}^{\infty} \left[\left(1 - \frac{\lambda}{\lambda_k} \right) \exp\left(\frac{\lambda}{\lambda_k} + \frac{\lambda^2}{2\lambda_k^2} + \cdots + \frac{\lambda^{[\mu]}}{[\mu]\lambda_k^{[\mu]}} \right) \right] \tag{K.39}$$

where $[\mu]$ denotes the integer part of μ. This is, see ref. [K.15], the unique entire function of order μ having zeros at $\{\lambda_k\}$ and subject to the normalization conditions

$$\ln\Delta(0) = \frac{d}{d\lambda}\ln\Delta(0) = \cdots = \frac{d^{[\mu]}}{d\lambda^{[\mu]}}\ln\Delta(0) = 0\ . \tag{K.40}$$

Clearly eq. (K.39) is the same as (K.35); one just has to identify $z = -\lambda$, $\mathbf{A} = 1/\mathbf{H}$ and $n-1 = [\mu]$. An example is the regularization of the spectral determinant

$$\Delta(E) = \det\left[(E - \mathbf{H})\right] \tag{K.41}$$

which – as it stands – would only make sense for a finite dimensional basis (or finite dimensional matrices). In ref. [K.16] the regulated spectral determinant for the example of the hyperbola billiard in two dimensions (thus $d = 2$, $m = 2$ and hence $\mu = 1$) is given as

$$\Delta(E) = \det\ \left[(E - \mathbf{H})\Omega(E, \mathbf{H})\right] \tag{K.42}$$

where

$$\Omega(E, \mathbf{H}) = -\mathbf{H}^{-1}e^{E\mathbf{H}^{-1}} \tag{K.43}$$

such that the spectral determinant in the eigenbasis of $\mathbf{H}$ (with eigenvalues $E_n \neq 0$) reads

$$\Delta(E) = \prod_n \left(1 - \frac{E}{E_n}\right) e^{E/E_n} < \infty\ . \tag{K.44}$$

Note that $\mathbf{H}^{-1}$ is for this example of Hilbert-Schmidt character.

References

[K.1] A. Wirzba, *Quantum Mechanics and Semiclassics of Hyperbolic n-Disk Scattering*, Habilitationsschrift, Technische Universität, Germany, 1997, HAB, chao-dyn/9712015, *Physics Reports* in press.

[K.2] A. Grothendieck, *"La théorie de Fredholm", Bull. Soc. Math. France*, **84**, 319 (1956).

[K.3] A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, Amer. Meth. Soc. **16**, Providence R. I. (1955).

[K.4] C.A. Tracy and H. Widom, CHECK THIS!: *Fredholm Determinants, Differential Equations and Matrix Models*, hep-th/9306042.

[K.5] M.G. Krein, *On the Trace Formula in Perturbation Theory* Mat.. Sborn. (N.S.) **33** (1953) 597-626; *Perturbation Determinants and Formula for Traces of Unitary and Self-adjoint Operators* Sov. Math.-Dokl. **3** (1962) 707-710. M.S. Birman and M.G. Krein, *On the Theory of Wave Operators and Scattering Operators*, Sov. Math.-Dokl. **3** (1962) 740-744.

[K.6] J. Friedel, Nuovo Cim. Suppl. **7** (1958) 287-301.

[K.7] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. I: *Functional Analysis*, Chap. VI, Academic Press (New York), 1972.

[K.8] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. IV: *Analysis of Operators*, Chap. XIII.17, Academic Press (New York), 1976.

[K.9] B. Simon, *Quantum Mechanics for Hamiltonians defined as Quadratic Forms*, Princeton Series in Physics, 1971, Appendix.

[K.10] B. Simon, *Notes on Infinite Determinants of Hilbert Space Operators*, Adv. Math. **24** (1977) 244-273.

[K.11] I.C. Gohberg and M.G. Krein, *Introduction to the theory of linear nonselfadjoint operators*, Translations of Mathematical Monographs **18**, Amer. Math. Soc. (1969).

[K.12] H. von Koch, *Sur quelques points de la théorie des déterminants infinis*, Acta. Math. **24** (1900) 89-122; *Sur la convergence des déterminants infinis*, Rend. Circ. Mat. Palermo **28** (1909) 255-266.

[K.13] E. Hille and J.D. Tamarkin, *On the characteristic values of linear integral equations*, Acta Math. **57** (1931) 1-76.

[K.14] T. Kato, *Perturbation Theory of Linear Operators* (Springer, New York, 1966), Chap. X, § 1.3 and § 1.4.

[K.15] A. Voros, *Spectral Functions, Special Functions and the Selberg Zeta Function*, *Comm. Math Phys.* **110**, 439 (1987).

[K.16] J.P. Keating and M. Sieber, *Calculation of Spectral Determinants*, preprint 1994.

Appendix L

Statistical mechanics recycled

(R. Mainieri)

A spin system with long-range interactions can be converted into a chaotic dynamical system that is differentiable and low-dimensional. The thermodynamic limit quantities of the spin system are then equivalent to long time averages of the dynamical system. In this way the spin system averages can be recast as the cycle expansions. If the resulting dynamical system is analytic, the convergence to the thermodynamic limit is faster than with the standard transfer matrix techniques.

L.1 The thermodynamic limit

There are two motivations to recycle statistical mechanics: one gets better control over the thermodynamic limit and one gets detailed information on how one is converging to it. From this information, most other quantities of physical interst can be computed.

In statistical mechanics one computes the averages of observables. These are functions that return a number for every state of the system; they are an abstraction of the process of measuring the pressure or temperature of a gas. The average of an observable is computed in the thermodynamic limit — the limit of system with an arbitrarily large number of particles. The thermodynamic limit is an essential step in the computation of averages, as it is only then that one observes the bulk properties of matter.

Without the thermodynamic limit many of the thermodynamic properties of matter could not be derived within the framework of statistical mechanics. There would be no extensive quantities, no equivalence of ensembles, and no phase transitions. From experiments it is known that certain quantities are extensive, that is, they are proportional to the size of the system. This is not true for an interacting set of particles. If two systems interacting via pairwise potentials are brought close together, work will be required to join them, and the final total energy will not be the sum of the energies of each of the parts. To avoid the conflict between the experiments and the theory of Hamiltonian systems, one needs systems with an infinite number of particles. In the canonical ensemble the probability of a state is given by the Boltzman factor which does not impose the conservation of energy; in the microcanonical ensemble energy is conserved but the Boltzmann factor is no longer exact. The equality between the ensembles only appears in the limit of the number of particles going to infinity at constant density. The phase transitions are interpreted as points of non-analyticity of the free energy in the thermodynamic limit. For a finite system the partition function cannot have a zero as a function of the inverse temperature β, as it is a finite sum of positive terms.

The thermodynamic limit is also of central importance in the study of field theories. A field theory can be first defined on a lattice and then the lattice spacing is taken to zero as the correlation length is kept fixed. This continuum limit corresponds to the thermodynamic limit. In lattice spacing units the correlation length is going to infinity, and the interacting field theory can be thought of as a statistical mechanics model at a phase transition.

For general systems the convergence towards the thermodynamic limit is slow. If the thermodynamic limit exists for an interaction, the convergence of the free energy per unit volume f is as an inverse power in the linear dimension of the system.

$$f(\beta) \to \frac{1}{n} \tag{L.1}$$

where n is proportional to $V^{1/d}$, with V the volume of the d-dimensional system. Much better results can be obtained if the system can be described by a transfer matrix. A transfer matrix is concocted so that the trace of its nth power is exactly the partition function of the system with one of the dimensions proportional to n. When the system is described by a transfer matrix then the convergence is exponential,

$$f(\beta) \to e^{-\alpha n} \tag{L.2}$$

and may only be faster than that if all long-range correlations of the system are zero — that is, when there are no interactions. The coefficient α depends only on the inverse correlation length of the system.

One of the difficulties in using the transfer matrix techniques is that they seem at first limited to systems with finite range interactions. Phase transitions can happen only when the interaction is long range. One can try to approximate the long range interaction with a series of finite range interactions that have an ever increasing range. The problem with this approach is that in a formally defined transfer matrix, not all the eigenvalues of the matrix correspond to eigenvalues of the system (in the sense that the rate of decay of correlations is not the ratio of eigenvalues).

Knowledge of the correlations used in conjunction with finite size scaling to obtain accurate estimates of the parameters of systems with phase transitions. (Accurate critical exponents are obtained by series expansions or transfer matrices, and seldomly by renormalization group arguments or Monte Carlo.) In a phase transition the coefficient α of the exponential convergence goes to zero and the convergence to the thermodynamic limit is power-law.

The computation of the partition function is an example of a functional integral. For most interactions these integrals are ill-defined and require some form of normalization. In the spin models case the functional integral is very simple, as "space" has only two points and only "time" being infinite has to be dealt with. The same problem occurs in the computation of the trace of transfer matrices of systems with infinite range interactions. If one tries to compute the partition function Z_n

$$Z_n = \operatorname{tr} T^n$$

when T is an infinite matrix, the result may be infinite for any n. This is not to say that Z_n is infinite, but that the relation between the trace of an operator and the partition function breaks down. We could try regularizing the expression, but as we shall see below, that is not necessary, as there is a better physical solution to this problem.

What will described here solves both of these problems in a limited context: it regularizes the transfer operator in a physically meaningful way, and as a a consequence, it allows for the faster than exponential convergence to the thermodynamic limit and complete determination of the spectrum. The steps to achieve this are:

- Redefine the transfer operator so that there are no limits involved except for the thermodynamic limit.
- Note that the divergences of this operator come from the fact that it acts on a very large space. All that is needed is the smallest subspace containing the eigenvector corresponding to the largest eigenvalue (the Gibbs state).

- Rewrite all observables as depending on a local effective field. The eigenvector is like that, and the operator restricted to this space is trace-class.
- Compute the spectrum of the transfer operator and observe the magic.

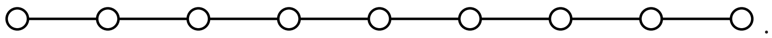
L.2 Ising models

The Ising model is a simple model to study the cooperative effects of many small interacting magnetic dipoles. The dipoles are placed on a lattice and their interaction is greatly simplified. There can also be a field that includes the effects of an external magnetic field and the average effect of the dipoles among themselves. We will define a general class of Ising models (also called spin systems) where the dipoles can be in one of many possible states and the interactions extend beyond the nearest neighboring sites of the lattice. But before we extend the Ising model, we will examine the simplest model in that class.

L.2.1 Ising model

One of the simplest models in statistical mechanics is the Ising model. One imagines that one has a one-dimensional lattice with small magnets at each site that can point either up or down.

Each little magnet interacts only with its neighbors. If they both point in the same direction, then they contribute an energy $-J$ to the total energy of the system; and if they point in opposite directions, then they contribute $+J$. The signs are chsen so that they prefer to be aligned. Let us suppose that we have n small magnets arranged in a line: A line is drawn between two sites to indicate that there is an interaction between the small magnets that are located on that site



(L.3)

(This figure can be thought of as a graph, with sites being vertices and interacting magnets indicated by edges.) To each of the sites we associate a variable, that we call a spin, that can be in either of two states: up ($\uparrow$) or down ($\downarrow$). This represents the two states of the small magnet on that site, and in general we will use the notation Σ_0 to represent the set of possible values of a spin at any site; all sites assume the same set of values. A configuration consists of assigning a value to the spin at each site; a typical configuration is

$$\begin{array}{ccccccccccccccccc} \uparrow & & \uparrow & & \downarrow & & \uparrow & & \downarrow & & \uparrow & & \downarrow & & \downarrow & & \uparrow \\ \circ & \text{—} & \circ & \text{—} & \circ & \text{—} & \circ & \text{—} & \circ & \text{—} & \circ & \text{—} & \circ & \text{—} & \circ & \text{—} & \circ \end{array} \quad . \tag{L.4}$$

The set of all configurations for a lattice with n sites is called Ω_0^n and is formed by the Cartesian product $\Omega_0 \times \Omega_0 \cdots \times \Omega_0$, the product repeated n times. Each configuration $\sigma \in \Omega^n$ is a string of n spins

$$\sigma = \{\sigma_0, \sigma_1, \ldots \sigma_n\} , \tag{L.5}$$

In the example configuration (L.4) there are two pairs of spins that have the same orientation and six that have the opposite orientation. Therefore the total energy H of the configuration is $J \times 6 - J \times 2 = 4J$. In general we can associate an energy H to every configuration

$$H(\sigma) = \sum_i J\delta(\sigma_i, \sigma_{i+1}) , \tag{L.6}$$

where

$$\delta(\sigma_1, \sigma_2) = \left\{ \begin{array}{ll} +1 & \text{if } \sigma_1 = \sigma_2 \\ -1 & \text{if } \sigma_1 \neq \sigma_2 \end{array} \right. . \tag{L.7}$$

One of the problems that was avoided when computing the energy was what to do at the boundaries of the one-dimensional chain. Notice that as written, (L.6) requires the interaction of spin n with spin $n+1$. In the absence of phase transitions the boundaries do not matter much to the thermodynamic limit and we will connect the first site to the last, implementing periodic boundary conditions.

Thermodynamic quantities are computed from the partition function $Z^{(n)}$ as the size n of the system becomes very large. For example, the free energy per site f at inverse temperature β is given by

$$-\beta f(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln Z^{(n)} . \tag{L.8}$$

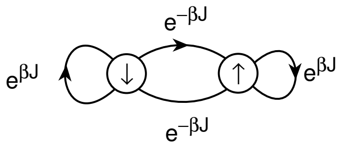
The partition function $Z^{(n)}$ is computed by a sum that runs over all the possible configurations on the one-dimensional chain. Each configuration contributes with its Gibbs factor $\exp(-\beta H(\sigma))$ and the partition function $Z^{(n)}$ is

$$Z^{(n)}(\beta) = \sum_{\sigma \in \Omega_0^n} e^{-\beta H(\sigma)} . \tag{L.9}$$

The partition function can be computed using transfer matrices. This is a method that generalizes to other models. At first, it is a little mysterious that matrices show up in the study of a sum. To see where they come from, we can try and build a configuration on the lattice site by site. The frst thing to do is to expand out the sum for the energy of the configuration

$$Z^{(n)}(\beta) = \sum_{\sigma \in \Omega^n} e^{\beta J \delta(\sigma_1,\sigma_2)} e^{\beta J \delta(\sigma_2,\sigma_3)} \cdots e^{\beta J \delta(\sigma_n,\sigma_1)} . \tag{L.10}$$

Let us use the configuration in (L.4). The first site is $\sigma_1 =\uparrow$. As the second site is $\uparrow$, we know that the first term in (L.10) is a term $e^{\beta J}$. The third spin is $\downarrow$, so the second term in (L.10) is $e^{-\beta J}$. If the third spin had been $\uparrow$, then the term would have been $e^{\beta J}$ but it would not depend on the value of the first spin σ_1. This means that the configuration can be built site by site and that to compute the Gibbs factor for the configuration just requires knowing the last spin added. We can then think of the configuration as being a weighted random walk where each step of the walk contributes according to the last spin added. The random walk take place on the Markov graph



.

Choose one of the two sites as a starting point. Walk along any allowed edge making your choices randomly and keep track of the accumulated weight as you perform the n steps. To implement the periodic boundary conditions make sure that you return to the starting node of the Markov graph. If the walk is carried out in all possible 2^n ways then the sum of all the weights is the partition function. To perform the sum we consider the matrix

$$T(\beta) = \begin{bmatrix} e^{\beta J} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J} \end{bmatrix} . \tag{L.11}$$

As in chapter 9 the sum of all closed walks is given by the trace of powers of the matrix. These powers can easily be re-expressed in terms of the two eigenvalues λ_1 and λ_2 of the transfer matrix:

$$Z^{(n)}(\beta) = \mathrm{tr}\, T^n(\beta) = \lambda_1(\beta)^n + \lambda_2(\beta)^n \,. \tag{L.12}$$

L.2.2 Averages of observables

Averages of observables can be re-expressed in terms of the eigenvectors of the transfer matrix. Alternatively, one can introduce a modified transfer matrix and compute the averages through derivatives. Sounds familiar?

L.2.3 General spin models

The more general version of the Ising model — the spin models — will be defined on a regular lattice, $\mathbb{Z}^D$. At each lattice site there will be a spin variable that can assumes a finite number of states identified by the set Ω_0.

The transfer operator $\mathcal{T}$ was introduced by Kramers and Wannier [L.12] to study the Ising model on a strip and concocted so that the trace of its nth power is the partition function Z_n of system when one of its dimensions is n. The method can be generalized to deal with any finite-range interaction. If the range of the interaction is L, then $\mathcal{T}$ is a matrix of size $2^L \times 2^L$. The longer the range, the larger the matrix.

L.3 Fisher droplet model

In a series of articles [L.20], Fisher introduced the droplet model. It is a model for a system containing two phases: gas and liquid. At high temperatures, the typical state of the system consists of droplets of all sizes floating in the gas phase. As the temperature is lowered, the droplets coalesce, forming larger droplets, until at the transition temperature, all droplets form one large one. This is a first order phase transition.

Although Fisher formulated the model in three-dimensions, the analytic solution of the model shows that it is equivalent to a one-dimensional lattice gas model with long range interactions. Here we will show how the model can be solved for an arbitrary interaction, as the solution only depends on the asymptotic behavior of the interaction.

The interest of the model for the study of cycle expansions is its relation to intermittency. By having an interaction that behaves asymptotically as the scaling function for intermittency, one expects that the analytic structure (poles and cuts) will be same.

Fisher used the droplet model to study a first order phase transition [L.20]. Gallavotti [L.21] used it to show that the zeta functions cannot in general be extended to a meromorphic functions of the entire complex plane. The droplet model has also been used in dynamical systems to explain features of mode locking, see Artuso [L.22]. In computing the zeta function for the droplet model we will discover that at low temperatures the cycle expansion has a limited radius of convergence, but it is possible to factorize the expansion into the product of two functions, each of them with a better understood radius of convergence.

L.3.1 Solution

The droplet model is a one-dimensional lattice gas where each site can have two states: empty or occupied. We will represent the empty state by 0 and the occupied state by 1. The configurations of the model in this notation are then strings of zeros and ones. Each configuration can be viewed as groups of contiguous ones separated by one or more zeros. The contiguous ones represent the droplets in the model. The droplets do not interact with each other, but the individual particles within each droplet do.

To determine the thermodynamics of the system we must assign an energy to every configuration. At very high temperatures we would expect a gaseous phase where there are many small droplets, and as we decrease the temperature the droplets would be expected to coalesce into larger ones until at some point there is a phase transition and the configuration is dominated by one large drop. To construct a solvable model and yet one with a phase transition we need long range interaction among all the particles of a droplet. One choice is to assign a fixed energy θ_n for the interactions of the particles of a cluster of size n. In a given droplet one has to consider all the possible clusters formed by contiguous particles. Consider for example the configuration 0111010. It has two droplets, one of size three and another of size one. The droplet of size one has only one cluster of size one and therefore contributes to the energy of the configuration with θ_1. The cluster of size three has one cluster of size three, two clusters of size two, and three clusters of size one; each cluster contributing a θ_n term to the energy. The total energy of the configuration is then

$$H(0111010) = 4\theta_1 + 2\theta_2 + 1\theta_3 . \tag{L.13}$$

If there where more zeros around the droplets in the above configuration the energy would still be the same. The interaction of one site with the others is assumed to be finite, even in the ground state consisting of a single droplet, so there is a restriction on the sum of the cluster energies given by

$$a = \sum_{n>0} \theta_n < \infty \,. \tag{L.14}$$

The configuration with all zeros does not contribute to the energy.

Once we specify the function θ_n we can computed the energy of any configuration, and from that determine the thermodynamics. Here we will evaluate the cycle expansion for the model by first computing the generating function

$$G(z,\beta) = \sum_{n>0} z^n \frac{Z_n(\beta)}{n} \tag{L.15}$$

and then considering its exponential, the cycle expansion. Each partition function Z_n must be evaluated with periodic boundary conditions. So if we were computing Z_3 we must consider all eight binary sequences of three bits, and when computing the energy of a configuration, say 011, we should determine the energy per three sites of the long chain

$$\ldots 011011011011 \ldots$$

In this case the energy would be $\theta_2 + 2\theta_1$. If instead of 011 we had considered one of its rotated shifts, 110 or 101, the energy of the configuration would have been the same. To compute the partition function we only need to consider one of the configurations and multiply by the length of the configuration to obtain the contribution of all its rotated shifts. The factor $1/n$ in the generating function cancels this multiplicative factor. This reduction will not hold if the configuration has a symmetry, as for example 0101 which has only two rotated shift configurations. To compensate this we replace the $1/n$ factor by a symmetry factor $1/s(b)$ for each configuration b. The evaluation of G is now reduced to summing over all configurations that are not rotated shift equivalent, and we call these the basic configurations and the set of all of them B. We now need to evaluate

$$G(z,\beta) = \sum_{b \in B} \frac{z^{|b|}}{s(b)} e^{-\beta H(b)} \,. \tag{L.16}$$

The notation $|\cdot|$ represents the cardinality of the set.

Any basic configuration can be built by considering the set of droplets that form it. The smallest building block has size two, as we must also put a zero next to the one so that when two different blocks get put next to each other they do not coalesce. The first few building blocks are

$$
\begin{array}{ll}
\text{size} & \text{droplets} \\
2 & 01 \\
3 & 001 \quad 011 \\
4 & 0001 \quad 0011 \quad 0111
\end{array}
\tag{L.17}
$$

Each droplet of size n contributes with energy

$$W_n = \sum_{1 \leq k \leq n} (n-k+1)\theta_k \,. \tag{L.18}$$

So if we consider the sum

$$\sum_{n \geq 1} \frac{1}{n} \Big(z^2 e^{-\beta H(01)} + z^3 (e^{-\beta H(001)} + e^{-\beta H(011)}) + \\ + z^4 (e^{-\beta H(0001)} + e^{-\beta H(0011)} + e^{-\beta H(0111)}) + \cdots \Big)^n \tag{L.19}$$

then the power in n will generate all the configurations that are made from many droplets, while the z will keep track of the size of the configuration. The factor $1/n$ is there to avoid the over-counting, as we only want the basic configurations and not its rotated shifts. The $1/n$ factor also gives the correct symmetry factor in the case the configuration has a symmetry. The sum can be simplified by noticing that it is a logarithmic series

$$-\ln \Big(1 - (z^2 e^{-\beta W_1} + z^3 (e^{-\beta W_1} + e^{-\beta W_2}) + \cdots \Big) \,, \tag{L.20}$$

where the $H(b)$ factors have been evaluated in terms of the droplet energies W_n. A proof of the equality of (L.19) and (L.20) can be given , but we there was not enough space on the margin to write it down. The series that is subtracted from one can be written as a product of two series and the logarithm written as

$$-\ln \Big(1 - (z^1 + z^2 + z^3 + \cdots)(z e^{-\beta W_1} + z^2 e^{-\beta W_2} + \cdots) \Big) \tag{L.21}$$

The product of the two series can be directly interpreted as the generating function for sequences of droplets. The first series adds one or more zeros to a configuration and the second series add a droplet.

There is a whole class of configurations that is not included in the above sum: the configurations formed from a single droplet and the vacuum configuration. The vacuum is the easiest, as it has zero energy it only contributes a z. The sum of all the null configurations of all sizes is

$$\sum_{n>0} \frac{z^n}{n} . \tag{L.22}$$

The factor $1/n$ is here because the original G had them and the null configurations have no rotated shifts. The single droplet configurations also do not have rotated shifts so their sum is

$$\sum_{n>0} \frac{z^n e^{-\beta H(\overbrace{11\ldots 11}^{n})}}{n} . \tag{L.23}$$

Because there are no zeros in the above configuration clusters of all size exist and the energy of the configuration is $n \sum \theta_k$ which we denote by na.

From the three sums (L.21), (L.22), and (L.23) we can evaluate the generating function G to be

$$G(z,\beta) = -\ln(1-z) - \ln(1-ze^{-\beta a}) - \ln(1 - \frac{z}{1-z} \sum_{n\geq 1} z^n e^{-\beta W_n}) . \tag{L.24}$$

The cycle expansion $\zeta^{-1}(z,\beta)$ is given by the exponential of the generating function e^{-G} and we obtain

$$\zeta^{-1}(z,\beta) = (1-ze^{-\beta a})(1 - z(1 + \sum_{n\geq 1} z^n e^{-\beta W_n})) \tag{L.25}$$

To pursue this model further we need to have some assumptions about the interaction strengths θ_n. We will assume that the interaction strength decreases with the inverse square of the size of the cluster, that is, $\theta_n = -1/n^2$. With this we can estimate that the energy of a droplet of size n is asymptotically

$$W_n \sim -n + \ln n + \mathcal{O}(\frac{1}{n}) . \tag{L.26}$$

If the power chosen for the polynomially decaying interaction had been other than inverse square we would still have the droplet term proportional to n, but there would be no logarithmic term, and the $\mathcal{O}$ term would be of a different power. The term proportional to n survives even if the interactions falls off exponentially, and in this case the correction is exponentially small in the asymptotic formula. To simplify the calculations we are going to assume that the droplet energies are exactly

$$W_n = -n + \ln n \tag{L.27}$$

in a system of units where the dimensional constants are one. To evaluate the cycle expansion (L.25) we need to evaluate the constant a, the sum of all the θ_n. One can write a recursion for the θ_n

$$\theta_n = W_n - \sum_{1 \leq k < n} (n - k + 1)\theta_k \tag{L.28}$$

and with an initial choice for θ_1 evaluate all the others. It can be verified that independent of the choice of θ_1 the constant a is equal to the number that multiplies the n term in (L.27). In the units used

$$a = -1\,. \tag{L.29}$$

For the choice of droplet energy (L.27) the sum in the cycle expansion can be expressed in terms of a special function: the Lerch transcendental ϕ_L. It is defined by

$$\phi_L(z, s, c) = \sum_{n \geq 0} \frac{z^n}{(n + c)^s}\,, \tag{L.30}$$

excluding from the sum any term that has a zero denominator. The Lerch function converges for $|z| < 1$. The series can be analytically continued to the complex plane and it will have a branch point at $z = 1$ with a cut chosen along the positive real axis. In terms of Lerch transcendental function we can write the cycle expansion (L.25) using (L.27) as

$$\zeta^{-1}(z, \beta) = \left(1 - ze^{\beta}\right)\left(1 - z(1 + \phi_L(ze^{\beta}, \beta, 1))\right) \tag{L.31}$$

This serves as an example of a zeta function that cannot be extended to a meromorphic function of the complex plane as one could conjecture.

The thermodynamics for the droplet model comes from the smallest root of (L.31). The root can come from any of the two factors. For large value of β (low temperatures) the smallest root is determined from the $(1 - ze^{\beta})$ factor, which gave the contribution of a single large drop. For small β (large temperatures) the root is determined by the zero of the other factor, and it corresponds to the contribution from the gas phase of the droplet model. The transition occurs when the smallest root of each of the factors become numerically equal. This determines the critical temperature β_c through the equation

$$1 - e^{-\beta_c}(1 + \zeta_R(\beta_c)) = 0 \tag{L.32}$$

which can be solved numerically. One finds that $\beta_c = 1.40495$. The phase transition occurs because the roots from two different factors get swapped in their roles as the smallest root. This in general leads to a first order phase transition. For large β the Lerch transcendental is being evaluated at the branch point, and therefore the cycle expansion cannot be an analytic function at low temperatures. For large temperatures the smallest root is within the radius of convergence of the series for the Lerch transcendental, and the cycle expansion has a domain of analyticity containing the smallest root.

As we approach the phase transition point as a function of β the smallest root and the branch point get closer together until at exactly the phase transition they collide. This is a sufficient condition for the existence of a first order phase transitions. In the literature of zeta functions [L.23] there have been speculations on how to characterize a phase transition within the formalism. The solution of the Fisher droplet model suggests that for first order phase transitions the factorized cycle expansion will have its smallest root within the radius of convergence of one of the series except at the phase transition when the root collides with a singularity. This does not seem to be the case for second order phase transitions.

The analyticity of the cycle expansion can be restored if we consider separate cycle expansions for each of the phases of the system. If we separate the two terms of ζ^{-1} in (L.31), each of them is an analytic function and contains the smallest root within the radius of convergence of the series for the relevant β values.

L.4 Scaling functions

> "Clouds are not spheres, mountains are not cones, coastlines are not circles and bark is not smooth, nor does lightning travel in straight line."
>
> B.B. Mandelbrot

There is a relation between general spin models and dynamical system. If one thinks of the boxes of the Markov partition of a hyperbolic system as the states of a spin system, then computing averages in the dynamical system is carrying out a sum over all possible states. One can even construct the natural measure of the dynamical system from a translational invariant "interaction function" call the scaling function.

There are many routes that lead to an explanation of what a scaling function is and how to compute it. The shortest is by breaking away from the historical development and considering first the presentation function of a fractal. The presentation function is a simple chaotic dynamical system (hyperbolic, unlike the circle map) that generates the fractal and is closely related to the definition of fractals of Hutchinson [L.24] and the iterated dynamical systems introduced by Barnsley and collaborators [H.20]. From the presentation function one can derive the scaling function, but we will not do it in the most elegant fashion, rather we will develop the formalism in a form that is directly applicable to the experimental data.

In the upper part of fig. L.1 we have the successive steps of the construction similar to the middle third Cantor set. The construction is done in levels, each level being formed by a collection of segments. From one level to the next, each "parent" segment produces smaller "children" segments by removing the middle section. As the construction proceeds, the segments better approximate the Cantor set. In the figure not all the segments are the same size, some are larger and some are smaller, as is the case with multifractals. In the middle third Cantor set, the ratio between a segment and the one it was generated from is exactly $1/3$, but in the case shown in the figure the ratios differ from $1/3$. If we went through the last level of the construction and made a plot of the segment number and its ratio to its parent segment we would have a scaling function, as indicated in the figure. A function giving the ratios in the construction of a fractal is the basic idea for a scaling function. Much of the formalism that we will introduce is to be able to give precise names to every segments and to arrange the "lineage" of segments so that the children segments have the correct parent. If we do not take these precautions, the scaling function would be a "wild function", varying rapidly and not approximated easily by simple functions.

To describe the formalism we will use a variation on the quadratic map that appears in the theory of period doubling. This is because the combinatorial manipulations are much simpler for this map than they are for the circle map. The scaling function will be described for a one dimensional map F as shown in fig. L.2. Drawn is the map

$$F(x) = 5x(1-x) \tag{L.33}$$

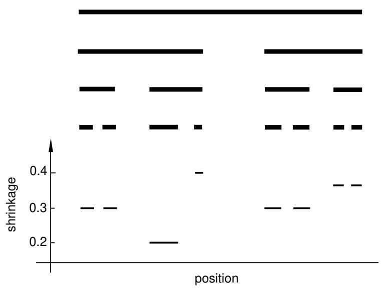


Figure L.1: Construction of the steps of the scaling function from a Cantor set. From one level to the next in the construction of the Cantor set the covers are shrunk, each parent segment into two children segments. The shrinkage of the last level of the construction is plotted and by removing the gaps one has an approximation to the scaling function of the Cantor set.

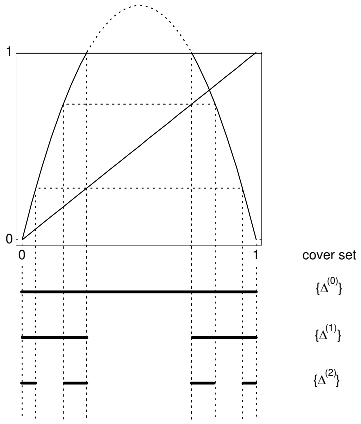


Figure L.2: A Cantor set presentation function. The Cantor set is the set of all points that under iteration do not leave the interval $[0, 1]$. This set can be found by backwards iterating the gap between the two branches of the map. The dotted lines can be used to find these backward images. At each step of the construction one is left with a set of segments that form a cover of the Cantor set.

restricted to the unit interval. We will see that this map is also a presentation function.

It has two branches separated by a gap: one over the left portion of the unit interval and one over the right. If we choose a point x at random in the unit interval and iterate it under the action of the map F, (L.33), it will hop between the branches and eventually get mapped to minus infinity. An orbit point is guaranteed to go to minus infinity if it lands in the gap. The hopping of the point defines the orbit of the initial point x: $x \mapsto x_1 \mapsto x_2 \mapsto \cdots$. For each orbit of the map F we can associate a symbolic code. The code for this map is formed from 0s and 1s and is found from the orbit by associating a 0 if $x_t < 1/2$ and a 1 if $x_t > 1/2$, with $t = 0, 1, 2, \ldots$.

Most initial points will end up in the gap region between the two branches. We then say that the orbit point has escaped the unit interval. The points that do not escape form a Cantor set $\mathcal{C}$ (or Cantor dust) and remain trapped in the unit interval for all iterations. In the process of describing all the points that do not escape, the map F can be used as a presentation of the Cantor set $\mathcal{C}$, and has been called a presentation function by Feigenbaum [21.13].

How does the map F "present" the Cantor set? The presentation is done in steps. First we determine the points that do not escape the unit interval in one iteration of the map. These are the points that are not part of the gap. These points determine two segments, which are an approximation to the Cantor set. In the next step we determine the points that do not escape in two iterations. These are the points that get mapped into the gap in one iteration, as in the next iteration they will escape; these points form the two segments $\Delta_0^{(1)}$ and $\Delta_1^{(1)}$ at level 1 in fig. L.2. The processes can be continued for any number of iterations. If we observe carefully what is being done, we discover that at each step the pre-images of the gap (backward iterates) are being removed from the unit interval. As the map has two branches, every point in the gap has two pre-images, and therefore the whole gap has two pre-images in the form of two smaller gaps. To generate all the gaps in the Cantor set one just has to iterate the gap backwards. Each iteration of the gap defines a set of segments, with the nth iterate defining the segments $\Delta_k^{(n)}$ at level n. For this map there will be 2^n segments at level n, with the first few drawn in fig. L.2. As $n \to \infty$ the segments that remain for at least n iterates converge to the Cantor set $\mathcal{C}$.

The segments at one level form a cover for the Cantor set and it is from a cover that all the invariant information about the set is extracted (the cover generated from the backward iterates of the gap form a Markov partition for the map as a dynamical system). The segments $\{\Delta_k^{(n)}\}$ at level n are a refinement of the cover formed by segments at level $n-1$. From successive covers we can compute the trajectory scaling function, the spectrum of scalings $f(\alpha)$, and the generalized dimensions.

To define the scaling function we must give labels (names) to the segments. The labels are chosen so that the definition of the scaling function allows for simple approximations. As each segment is generated from an inverse image of the unit interval, we will consider the inverse of the presentation function F. Because F does not have a unique inverse, we have to consider restrictions of F. Its restriction to the first half of the segment, from 0 to 1/2, has a unique inverse, which we will call F_0^{-1}, and its restriction to the second half, from 1/2 to 1, also has a unique inverse, which we will call F_1^{-1}. For example, the segment labeled $\Delta^{(2)}(0,1)$ in fig. L.2 is formed from the inverse image of the unit interval by mapping $\Delta^{(0)}$, the unit interval, with F_1^{-1} and then F_0^{-1}, so that the segment

$$\Delta^{(2)}(0,1) = F_0^{-1}\left(F_1^{-1}\left(\Delta^{(0)}\right)\right) . \tag{L.34}$$

The mapping of the unit interval into a smaller interval is what determines its label. The sequence of the labels of the inverse maps is the label of the segment:

$$\Delta^{(n)}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = F_{\epsilon_1}^{-1} \circ F_{\epsilon_2}^{-1} \circ \cdots F_{\epsilon_n}^{-1}\left(\Delta^{(0)}\right) .$$

The scaling function is formed from a set of ratios of segments length. We use $|\cdot|$ around a segment $\Delta^{(n)}(\epsilon)$ to denote its size (length), and define

$$\sigma^{(n)}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = \frac{|\Delta^{(n)}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)|}{|\Delta^{(n-1)}(\epsilon_2, \ldots, \epsilon_n)|}\,.$$

We can then arrange the ratios $\sigma^{(n)}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ next to each other as piecewise constant segments in increasing order of their binary label $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ so that the collection of steps scan the unit interval. As $n \to \infty$ this collection of steps will converge to the scaling function.

L.5 Geometrization

The $\mathcal{L}$ operator is a generalization of the transfer matrix. It gets more by considering less of the matrix: instead of considering the whole matrix it is possible to consider just one of the rows of the matrix. The $\mathcal{L}$ operator also makes explicit the vector space in which it acts: that of the observable functions. Observables are functions that to each configuration of the system associate a number: the energy, the average magnetization, the correlation between two sites. It is in the average of observables that one is interested in. Like the transfer matrix, the $\mathcal{L}$ operator considers only semi-infinite systems, that is, only the part of the interaction between spins to the right is taken into account. This may sound un-symmetric, but it is a simple way to count each interaction only once, even in cases where the interaction includes three or more spin couplings. To define the $\mathcal{L}$ operator one needs the interaction energy between one spin and all the rest to its right, which is given by the function ϕ. The $\mathcal{L}$ operators defined as

$$\mathcal{L}g(\sigma) = \sum_{\sigma_0 \in \Omega_0} g(\sigma_0\sigma) e^{-\beta\phi(\sigma_0\sigma)}\,.$$

To each possible value in Ω_0 that the spin σ_0 can assume, an average of the observable g is computed weighed by the Boltzmann factor $e^{-\beta\phi}$. The formal relations that stem from this definition are its relation to the free energy when applied to the observable ι that returns one for any configuration:

$$-\beta f(\beta) = \lim_{n\to\infty} \frac{1}{n} \ln \|\mathcal{L}^n \iota\|$$

and the thermodynamic average of an observable

$$\langle g \rangle = \lim_{n\to\infty} \frac{\|\mathcal{L}^n g\|}{\|\mathcal{L}^n \iota\|}\,.$$

Both relations hold for almost all configurations. These relations are part of theorem of Ruelle that enlarges the domain of the Perron-Frobenius theorem and sharpens its results. The theorem shows that just as the transfer matrix, the largest eigenvalue of the $\mathcal{L}$ operator is related to the free-energy of the spin system. It also hows that there is a formula for the eigenvector related to the largest eigenvalue. This eigenvector $|\rho\rangle$ (or the corresponding one for the adjoint $\mathcal{L}^*$ of $\mathcal{L}$) is the Gibbs state of the system. From it all averages of interest in statistical mechanics can be computed from the formula

$$\langle g\rangle = \langle\rho|g|\rho\rangle \,.$$

The Gibbs state can be expressed in an explicit form in terms of the interactions, but it is of little computational value as it involves the Gibbs state for a related spin system. Even then it does have an enormous theoretical value. Later we will see how the formula can be used to manipulate the space of observables into a more convenient space.

The geometrization of a spin system converts the shift dynamics (necessary to define the Ruelle operator) into a smooth dynamics. This is equivalent to the mathematical problem in ergodic theory of finding a smooth embedding for a given Bernoulli map.

The basic idea for the dynamics is to establish the a set of maps F_{σ_k} such that

$$F_{\sigma_k}(0) = 0$$

and

$$F_{\sigma_1} \circ F_{\sigma_2} \circ \cdots \circ F_{\sigma_n}(0) = \phi(+, \sigma_1, \sigma_2, \ldots, \sigma_n, -, -, \ldots) \,.$$

This is a formal relation that expresses how the interaction is to be converted into a dynamical systems. In most examples F_{σ_k} is a collection of maps from a subset of R^D to itself.

If the interaction is complicated, then the dimension of the set of maps may be infinite. If the resulting dynamical system is infinite have we gained anything from the transformation? The gain in this case is not in terms of added speed of convergence to the thermodynamic limit, but in the fact that the Ruelle operator is of trace-class and all eigenvalues are related to the spin system and not artifacts of the computation.

The construction of the higher dimensional system is done by borrowing the phase space reconstruction technique from dynamical systems. Phase space reconstruction can be done in several ways: by using delay coordinates, by using derivatives of the position, or by considering the value of several independent observables of the system. All these may be used in the construction of the equivalent dynamics. Just as in the study of dynamical systems, the exact method does not matter for the determination of the thermodynamics ($f(\alpha)$ spectra, generalized dimension), also in the construction of the equivalent dynamics the exact choice of observable does not matter.

We will only consider configurations for the half line. This is bescause for translational invariant interactions the thermodynamic limit on half line is the same as in the whole line. One can prove this by considering the difference in a thermodynamic average in the line and in the semiline and compare the two as the size of the system goes to infinity.

When the interactions are long range in principle one has to specify the boundary conditions to be able to compute the interaction energy of a configuration in a finite box. If there are no phase transitions for the interaction, then which boundary conditions are chosen is irrelevant in the thermodynamic limit. When computing quantities with the transfer matrix, the long rrange interaction is truncated at some finite range and the truncated interaction is then use to evaluate the transfer matrix. With the Ruelle operator the interaction is never truncated, and the boundary must be specified.

The interaction $\phi(\sigma)$ is any function that returns a number on a configuration. In general it is formed from pairwise spin interactions

$$\phi(\sigma) = \sum_{n>0} \delta_{\sigma_0,\sigma_n} J(n)$$

with different choices of $J(n)$ leading to differnt models. If $J(n) = 1$ only if $n = 1$ and) otherwise, then one has the nearest neighbor Ising model. If $J(n) = n^{-2}$, then one has the inverse square model relevant in the study of the Kondo problem.

Let us say that each site of the lattice can assume two values $+, -$ and the set of all possible configurations of the semiline is the set Ω. Then an observable g is a function from the set of configurations Ω to the reals. Each configuration is indexed by the integers from 0 up, and it is useful to think of the configuration as a string of spins. One can append a spin η_0 to its begining, $\eta \vee \sigma$, in which case η is at site 0, ω_0 at site 1, and so on.

The Ruelle operator $\mathcal{L}$ is defined as

$$\mathcal{L}g(\eta) = \sum_{\omega_0 \in \Omega_0} g(\omega_0 \vee \eta) e^{-\beta\phi(\omega_0 \vee \eta)} \,.$$

This is a positive and bounded operator over the space of bounded observables. There is a generalization of the Perron-Frobenius theorem by Ruelle that establishes that the largest eigenvalue of $\mathcal{L}$ is isolated from the rest of the spectrum and gives the thermodynamics of the spin system just as the largest eigenvalue of the transfer matrix does. Ruelle alos gave a formula for the eigenvector related to the largest eigenvalue.

The difficulty with it is that the relation between the partition function and the trace of its nth power, $\mathrm{tr}\,\mathcal{L}^n = Z_n$ no longer holds. The reason is that the trace of the Ruelle operator is ill-defined, it is infinite.

We now introduce a special set of observables $\{x_1(\sigma), \ldots, x_1(\sigma)\}$. The idea is to choose the observables in such a way that from their values on a particular configuration σ the configuration can be reconstructed. We also introduce the interaction observables h_{σ_0}

To geometrize spin systems, the interactions are assumed to be translationally invariant. The spins σ_k will only assume a finite number of values. For simplicity, we will take the interaction ϕ among the spins to depend only on pairwise interactions,

$$\phi(\sigma) = \phi(\sigma_0, \sigma_1, \sigma_2, \ldots) = J_0\sigma_0 + \sum_{n>0} \delta_{\sigma_0,\sigma_n} J_1(n)\,, \tag{L.35}$$

and limit σ_k to be in $\{+, -\}$. For the one-dimensional Ising model, J_0 is the external magnetic field and $J_1(n) = 1$ if $n = 1$ and 0 otherwise. For an exponentially decaying interaction $J_1(n) = e^{-\alpha n}$. Two- and three-dimensional models can be considered in this framework. For example, a strip of spins of $L \times \infty$ with helical boundary conditions is modeled by the potential $J_1(n) = \delta_{n,1} + \delta_{n,L}$.

The transfer operator $\mathcal{T}$ was introduced by Kramers and Wannier [L.12] to study the Ising model on a strip and concocted so that the trace of its nth power is the partition function Z_n of system when one of its dimensions is n. The method can be generalized to deal with any finite-range interaction. If the range of the interaction is L, then $\mathcal{T}$ is a matrix of size $2^L \times 2^L$. The longer the range, the larger the matrix. When the range of the interaction is infinite one has to define the $\mathcal{T}$ operator by its action on an observable g. Just as the observables in quantum mechanics, g is a function that associates a number to every state (configuration of spins). The energy density and the average magnetization are examples of observables. From this equivalent definition one can recover the usual transfer matrix by making all quantities finite range. For a semi-infinite configuration $\sigma = \{\sigma_0, \sigma_1, \ldots\}$:

$$\mathcal{T} g(\sigma) = g(+ \vee \sigma) e^{-\beta\phi(+\vee\sigma)} + g(- \vee \sigma) e^{-\beta\phi(-\vee\sigma)}\,. \tag{L.36}$$

By $+\vee\sigma$ we mean the configuration obtained by prepending $+$ to the beginning of σ resulting in the configuration $\{+,\sigma_0,\sigma_1,\ldots\}$. When the range becomes infinite, $\mathrm{tr}\,\mathcal{T}^n$ is infinite and there is no longer a connection between the trace and the partition function for a system of size n (this is a case where matrices give the wrong intuition). Ruelle [L.13] generalized the Perron-Frobenius theorem and showed that even in the case of infinite range interactions the largest eigenvalue of the $\mathcal{T}$ operator is related to the free-energy of the spin system and the corresponding eigenvector is related to the Gibbs state. By applying $\mathcal{T}$ to the constant observable u, which returns 1 for any configuration, the free energy per site f is computed as

$$-\beta f(\beta)=\lim_{n\to\infty}\frac{1}{n}\ln\|\mathcal{T}^n u\|\,. \tag{L.37}$$

To construct a smooth dynamical system that reproduces the properties of $\mathcal{T}$, one uses the phase space reconstruction technique of Packard *et al.* [L.6] and Takens [L.7], and introduces a vector of state observables $x(\sigma)=\{x_1(\sigma),\ldots,x_D(\sigma)\}$. To avoid complicated notation we will limit the discussion to the example $x(\sigma)=\{x_+(\sigma),x_-(\sigma)\}$, with $x_+(\sigma)=\phi(+\vee\sigma)$ and $x_-(\sigma)=\phi(-\vee\sigma)$; the more general case is similar and used in a later example. The observables are restricted to those g for which, for all configurations σ, there exist an analytic function G such that $G(x_1(\sigma),\ldots,x_D(\sigma))=g(\sigma)$. This at first seems a severe restriction as it may exclude the eigenvector corresponding to the Gibbs state. It can be checked that this is not the case by using the formula given by Ruelle [L.14] for this eigenvector. A simple example where this formalism can be carried out is for the interaction $\phi(\sigma)$ with pairwise exponentially decaying potential $J_1(n)=a^n$ (with $|a|<1$). In this case $\phi(\sigma)=\sum_{n>0}\delta_{\sigma_0,\sigma_n}a^n$ and the state observables are $x_+(\sigma)=\sum_{n>0}\delta_{+,\sigma_n}a^n$ and $x_-(\sigma)=\sum_{n>0}\delta_{-,\sigma_n}a^n$. In this case the observable x_+ gives the energy of $+$ spin at the origin, and x_- the energy of a $-$ spin.

Using the observables x_+ and x_-, the transfer operator can be re-expressed as

$$\mathcal{T}G\left(x(\sigma)\right)=\sum_{\eta\in\{+,-\}}G\left(x_+\left(\eta\vee\sigma\right),x_-\left(\eta\vee\sigma\right)\right)e^{-\beta x_\eta(\sigma)}\,. \tag{L.38}$$

In this equation the only reference to the configuration σ is when computing the new observable values $x_+(\eta\vee\sigma)$ and $x_-(\eta\vee\sigma)$. The iteration of the function that gives these values in terms of $x_+(\sigma)$ and $x_-(\sigma)$ is the dynamical system that will reproduce the properties of the spin system. For the simple exponentially decaying potential this is given by two maps, F_+ and F_-. The map F_+ takes $\{x_+(\sigma),x_+(\sigma)\}$ into $\{x_+(+\vee\sigma),x_-(+\vee\sigma)\}$ which is $\{a(1+x_+), ax_-\}$ and the map F_- takes $\{x_+, x_-\}$ into $\{ax_+, a(1+x_-)\}$. In a more general case we have maps F_η that take $x(\sigma)$ to $x(\eta \vee \sigma)$.

We can now define a new operator $\mathcal{L}$

$$\mathcal{L}G\,(x) \stackrel{\text{def}}{=} \mathcal{T}G(x(\sigma)) = \sum_{\eta \in \{+,-\}} G\,(F_\eta(x))\, e^{-\beta x_\eta}\,, \tag{L.39}$$

where all dependencies on σ have disappeared — if we know the value of the state observables x, the action of $\mathcal{L}$ on G can be computed.

A dynamical system is formed out of the maps F_η. They are chosen so that one of the state variables is the interaction energy. One can consider the two maps F_+ and F_- as the inverse branches of a hyperbolic map f, that is, $f^{-1}(x) = \{F_+(x), F_-(x)\}$. Studying the thermodynamics of the interaction ϕ is equivalent to studying the long term behavior of the orbits of the map f, achieving the transformation of the spin system into a dynamical system.

Unlike the original transfer operator, the $\mathcal{L}$ operator — acting in the space of observables that depend only on the state variables — is of trace-class (its trace is finite). The finite trace gives us a chance to relate the trace of $\mathcal{L}^n$ to the partition function of a system of size n. We can do better. As most properties of interest (thermodynamics, fall-off of correlations) are determined directly from its spectrum, we can study instead the zeros of the Fredholm determinant $\det(1 - z\mathcal{L})$ by the technique of cycle expansions developed for dynamical systems [15.2]. A cycle expansion consists of finding a power series expansion for the determinant by writing $\det(1 - z\mathcal{L}) = \exp(\mathrm{tr}\,\ln(1 - z\mathcal{L}))$. The logarithm is expanded into a power series and one is left with terms of the form $\mathrm{tr}\,\mathcal{L}^n$ to evaluate. For evaluating the trace, the $\mathcal{L}$ operator is equivalent to

$$\mathcal{L}G(x) = \int_{\mathbf{R}^D} dy\, \delta(y - f(x)) e^{-\beta y} G(y) \tag{L.40}$$

from which the trace can be computed:

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{x = f^{(\circ n)}(x)} \frac{e^{-\beta H(x)}}{|\det\left(1 - \partial_x f^{(\circ n)}(x)\right)|} \tag{L.41}$$

with the sum running over all the fixed points of $f^{(\circ n)}$ (all spin configurations of a given length). Here $f^{(\circ n)}$ is f composed with itself n times, and $H(x)$ is the energy of the configuration associated with the point x.

In practice the map f is never constructed and the energies are obtained directly from the spin configurations.

To compute the value of $\operatorname{tr}\mathcal{L}^n$ we must compute the value of $\partial_x f^{(\circ n)}$; this involves a functional derivative. To any degree of accuracy a number x in the range of possible interaction energies can be represented by a finite string of spins ϵ, such as $x = \phi(+, \epsilon_0, \epsilon_1, \ldots, -, -, \ldots)$. By choosing the sequence ϵ to have a large sequence of spins $-$, the number x can be made as small as needed, so in particular we can represent a small variation by $\phi(\eta)$. As $x_+(\epsilon) = \phi(+ \vee \epsilon)$, from the definition of a derivative we have:

$$\partial_x f(x) = \lim_{m\to\infty} \frac{\phi(\epsilon \vee \eta^{(m)}) - \phi(\epsilon)}{\phi(\eta^{(m)})}, \tag{L.42}$$

where $\eta^{(m)}$ is a sequence of spin strings that make $\phi(\eta^{(m)})$ smaller and smaller. By substituting the definition of ϕ in terms of its pairwise interaction $J(n) = n^s a^{n^\gamma}$ and taking the limit for the sequences $\eta^{(m)} = \{+, -, -, \ldots, \eta_{m+1}, \eta_{m+2}, \ldots\}$ one computes that the limit is a if $\gamma = 1$, 1 if $\gamma < 1$, and 0 if $\gamma > 1$. It does not depend on the positive value of s. When $\gamma < 1$ the resulting dynamical system is not hyperbolic and the construction for the operator $\mathcal{L}$ fails, so one cannot apply it to potentials such as $(1/2)^{\sqrt{n}}$. One may solve this problem by investigating the behavior of the formal dynamical system as $\gamma \to 0$.

The manipulations have up to now assumed that the map f is smooth. If the dimension D of the embedding space is too small, f may not be smooth. Determining under which conditions the embedding is smooth is a complicated question [L.15]. But in the case of spin systems with pairwise interactions it is possible to give a simple rule. If the interaction is of the form

$$\phi(\sigma) = \sum_{n\geq 1} \delta_{\sigma_0,\sigma_n} \sum_k p_k(n) a_k^{n^\gamma} \tag{L.43}$$

where p_k are polynomials and $|a_k| < 1$, then the state observables to use are $x_{s,k}(\sigma) = \sum \delta_{+,\sigma_n} n^s a_k^n$. For each k one uses $x_{0,k}, x_{1,k}, \ldots$ up to the largest power in the polynomial p_k. An example is the interaction with $J_1(n) = n^2(3/10)^n$. It leads to a 3-dimensional system with variables $x_{0,0}$, $x_{1,0}$, and $x_{2,0}$. The action of the map F_+ for this interaction is illustrated fig. L.3. Plotted are the pairs $\{\phi(+\vee\sigma), \phi(+\vee+\vee\sigma)\}$. This can be seen as the strange attractor of a chaotic system for which the variables $x_{0,0}$, $x_{1,0}$, and $x_{2,0}$ provide a good (analytic) embedding.

The added smoothness and trace-class of the $\mathcal{L}$ operator translates into faster convergence towards the thermodynamic limit. As the reconstructed

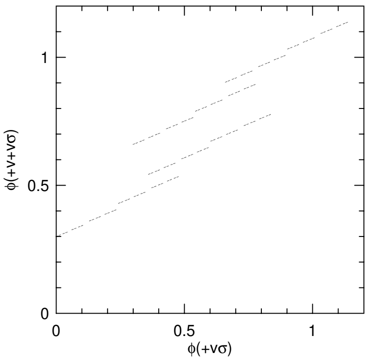


Figure L.3: The spin adding map F_+ for the potential $J(n) = \sum n^2 a^{\alpha n}$. The action of the map takes the value of the interaction energy between + and the semi-infinite configuration $\{\sigma_1, \sigma_2, \sigma_3, \ldots\}$ and returns the interaction energy between + and the configuration $\{+, \sigma_1, \sigma_2, \sigma_3, \ldots\}$.

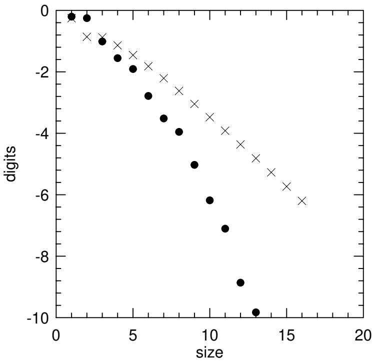


Figure L.4: Number of digits for the Fredholm method (•) and the transfer function method (×). The size refers to the largest cycle considered in the Fredholm expansions, and the truncation length in the case of the transfer matrix.

dynamics is analytic, the convergence towards the thermodynamic limit is faster than exponential [2.12, L.16]. We will illustrate this with the polynomial-exponential interactions (L.43) with $\gamma = 1$, as the convergence is certainly faster than exponential if $\gamma > 1$, and the case of a^n has been studied in terms of another Fredholm determinant by Gutzwiller [L.17]. The convergence is illustrated in fig. L.4 for the interaction $n^2(3/10)^n$. Plotted in the graph, to illustrate the transfer matrix convergence, are the number of decimal digits that remain unchanged as the range of the interaction is increased. Also in the graph are the number of decimal digits that remain unchanged as the largest power of $\mathrm{tr}\,\mathcal{L}^n$ considered. The plot is effectively a logarithmic plot and straight lines indicate exponentially fast convergence. The curvature indicates that the convergence is faster than exponential. By fitting, one can verify that the free energy is converging to its limiting value as $\exp(-n^{(4/3)})$. Cvitanović [2.12] has estimated that the Fredholm determinant of a map on a D dimensional space should converge as $\exp(-n^{(1+1/D)})$, which is confirmed by these numerical simulations.

Commentary

Remark L.1 Presentation functions. The best place to read about Feigenbaum's work is in his review article published in *Los Alamos Science* (reproduced in various reprint collections and conference proceedings, such as ref. [15.5]). Feigenbaum's *Journal of Statistical Physics* article [21.13] is the easiest place to learn about presentation functions.

Remark L.2 Interactions are smooth In most computational schemes for thermodynamic quantities the translation invariance and the smoothness of the basic interaction are never used. In Monte Carlo schemes, aside from the periodic boundary conditions, the interaction can be arbitrary. In principle for each configuration it could be possible to have a different energy. Schemes such as the Sweneson-Wang cluster flipping algorithm use the fact that interaction is local and are able to obtain dramatic speed-ups in the equilibration time for the dynamical Monte Carlo simulation. In the geometrization program for spin systems, the interactions are assumed translation invariant and smooth. The smoothness means that any interaction can be decomposed into a series of terms that depend only on the spin arrangement and the distance between spins:

$$\phi(\sigma_0, \sigma_1, \sigma_2, \ldots) = J_0\sigma_0 + \sum \delta(\sigma_0, \sigma_n) J_1(n) + \sum \delta(\sigma_0, \sigma_{n_1}, \sigma_{n_2}) J_2(n_1, n_2) + \cdots$$

where the J_k are symmetric functions of their arguments and the δ are arbitrary discrete functions. This includes external constant fields (J_0), but it excludes site dependent fields such as a random external magnetic field.

Résumé

The geometrization of spin systems strengthens the connection between statistical mechanics and dynamical systems. It also further establishes the value of the Fredholm determinant of the $\mathcal{L}$ operator as a practical computational tool with applications to chaotic dynamics, spin systems, and semiclassical mechanics. The example above emphasizes the high accuracy that can be obtained: by computing the shortest 14 periodic orbits of period 5 or less it is possible to obtain three digit accuracy for the free energy. For the same accuracy with a transfer matrix one has to consider a 256×256 matrix. This make the method of cycle expansions practical for analytic calculations.

References

[L.1] Ya. Sinai. Gibbs measures in ergodic theory. *Russ. Math. Surveys*, 166:21–69, 1972.

[L.2] R. Bowen. Periodic points and measure for axiom-A diffeomorphisms. *Transactions Amer. Math. Soc.*, 154:377–397, 1971.

[L.3] D. Ruelle. Statistical mechanics on a compound set with Z^ν action satisfying expansiveness and specification. *Transactions Amer. Math. Soc.*, 185:237–251, 1973.

[L.4] E. B. Vul, Ya. G. Sinai, and K. M. Khanin. Feigenbaum universality and the thermodynamic formalism. *Uspekhi Mat. Nauk.*, 39:3–37, 1984.

[L.5] M.J. Feigenbaum, M.H. Jensen, and I. Procaccia. Time ordering and the thermodynamics of strange sets: Theory and experimental tests. *Physical Review Letters*, 57:1503–1506, 1986.

[L.6] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Physical Review Letters*, 45:712 – 716, 1980.

[L.7] F. Takens, Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics 898*, pages 366–381. Springer, Berlin, 1981.

[L.8] R. Mainieri. Thermodynamic zeta functions for Ising models with long range interactions. *Physical Review A*, 45:3580, 1992.

[L.9] R. Mainieri. Zeta function for the Lyapunov exponent of a product of random matrices. *Physical Review Letters*, 68:1965–1968, March 1992.

[L.10] D. Wintgen. Connection between long-range correlations in quantum spectra and classical periodic orbits. *Physical Review Letters*, 58(16):1589–1592, 1987.

[L.11] G. S. Ezra, K. Richter, G. Tanner, and D. Wintgen. Semiclassical cycle expansion for the Helium atom. *Journal of Physics B*, 24(17):L413–L420, 1991.

[L.12] H. A. Kramers and G. H. Wannier. Statistics of the two-dimensional ferromagnet. Part I. *Physical Review*, 60:252–262, 1941.

[L.13] D. Ruelle. Statistical mechanics of a one-dimensional lattice gas. *Communications of Mathematical Physics*, 9:267–278, 1968.

[L.14] David Ruelle. *Thermodynamic Formalism.* Addison-Wesley, Reading, 1978.

[L.15] P. Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Text in Mathematics.* Springer-Verlag, New York, 1982.

[L.16] H.H. Rugh. *Time evolution and correlations in chaotic dynamical systems.* PhD thesis, Niels Bohr Institute, 1992.

[L.17] M.C. Gutzwiller. The quantization of a classically ergodic system. *Physica D*, 5:183–207, 1982.

[L.18] M. Feigenbaum. The universal metric properties of non-linear transformations. *Journal of Statistical Physics*, 19:669, 1979.

[L.19] G.A. Baker. One-dimensional order-disorder model which approaches a second order phase transition. *Phys. Rev.*, 122:1477–1484, 1961.

[L.20] M. E. Fisher. The theory of condensation and the critical point. *Physics*, 3:255–283, 1967.

[L.21] G. Gallavotti. Funzioni zeta ed insiemi basilari. *Accad. Lincei. Rend. Sc. fis. mat. e nat.*, 61:309–317, 1976.

[L.22] R. Artuso. Logarithmic strange sets. *J. Phys. A.*, 21:L923–L927, 1988.

[L.23] Dieter H. Mayer. *The Ruelle-Araki transfer operator in classical statistical mechanics.* Springer-Verlag, Berlin, 1980.

[L.24] Hutchinson

Exercises

Exercise L.1 **Not all Banach spaces are also Hilbert** *If we are given a norm $\|\cdot\|$ of a Banach space B, it may be possible to find an inner product $\langle\cdot\,,\cdot\rangle$ (so that B is also a Hilbert space H) such that for all vectors $f \in B$, we have*

$$\|f\| = \langle f, f\rangle^{1/2}\,.$$

This is the norm induced by the scalar product. If we cannot find the inner product how do we know that we just are not being clever enough? By checking the parallelogram law for the norm. A Banach space can be made into a Hilbert space if and only if the norm satisfies the parallelogram law. The parallelogram law says that for any two vectors f and g the equality

$$\|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2\,,$$

must hold.

Consider the space of bounded observables with the norm given by $\|a\| = \sup_{\sigma\in\Omega^{\mathbb{N}}} |a(\sigma)|$. Show that ther is no scalar product that will induce this norm.

Exercise L.2 **Automaton for a droplet** *Find the Markov graph and the weights on the edges so that the energies of configurations for the dropolet model are correctly generated. For any string starting in zero and ending in zero your diagram should yield a configuration the weight $e^{H(\sigma)}$, with H computed along the lines of (L.13) and (L.18).*

Hint: the Markov graph is infinite.

Exercise L.3 **Spectral determinant for a^n interactions** *Compute the spectral determinant for one-dimensional Ising model with the interaction*

$$\phi(\sigma) = \sum_{k>0} a^k \delta(\sigma_0, \sigma_k)\,.$$

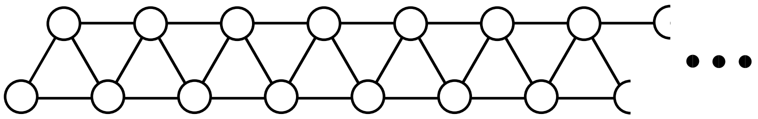
Take a as a number smaller than $1/2$.

(a) *What is the dynamical system this generates? That is, find F_+ and F_- as used in (L.39).*

(b) *Show that*

$$\frac{d}{dx}F_{\{+\ \mathrm{or}-\}} = \left[\begin{array}{cc} a & 0 \\ 0 & a \end{array}\right]$$

Exercise L.4 **Ising model on a thin strip** *Compute the transfer matrix for the Ising model defined on the graph*



Assume that whenever there is a bond connecting two sites, there is a contribution $J\delta(\sigma_i, \sigma_j)$ to the energy.

Exercise L.5 **Infinite symbolic dynamics** *Let σ be a function that returns zeo or one for every infinite binary string: $\sigma : \{0,1\}^{\mathbb{N}} \to \{0,1\}$. Its value is represented by $\sigma(\epsilon_1, \epsilon_2, \ldots)$ where the ϵ_i are either 0 or 1. We will now define an operator $\mathcal{T}$ that acts on observables on the space of binary strings. A function a is an observable if it has bounded variation, that is, if*

$$\|a\| = \sup_{\{\epsilon_i\}} |a(\epsilon_1, \epsilon_2, \ldots)| < \infty\,.$$

For these functions

$$\mathcal{T}a(\epsilon_1, \epsilon_2, \ldots) = a(0, \epsilon_1, \epsilon_2, \ldots)\sigma(0, \epsilon_1, \epsilon_2, \ldots) + a(1, \epsilon_1, \epsilon_2, \ldots)\sigma(1, \epsilon_1, \epsilon_2, \ldots)\,.$$

The function σ is assumed such that any of $\mathcal{T}$'s "matrix representations" in (a) have the Markov property (the matrix, if read as an adjacency graph, corresponds to a graph where one can go from any node to any other node).

(a) *(easy) Consider a finite version T_n of the operator $\mathcal{T}$:*

$$\begin{aligned} &T_n a(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = \\ &\qquad a(0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\sigma(0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1}) + \\ &\qquad a(1, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\sigma(1, \epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1})\,. \end{aligned}$$

Show that T_n is a $2^n \times 2^n$ matrix. Show that its trace is bounded by a number independent of n.

(b) *(medium) With the operator norm induced by the function norm, show that $\mathcal{T}$ is a bounded operator.*

(c) *(hard) Show that $\mathcal{T}$ is not trace-class. (Hint: check if $\mathcal{T}$ is compact).*

Classes of operators are nested; trace-class $\leq$ compact $\leq$ bounded.

Appendix M

Noise/quantum corrections

(G. Vattay)

The Gutzwiller trace formula is only a good approximation to the quantum mechanics when $\hbar$ is small. Can we improve the trace formula by adding quantum corrections to the semiclassical terms? A similar question can be posed when the classical deterministic dynamics is disturbed by some way gaussian white noise with strength D. The deterministic dynamics then can be considered as the weak noise limit $D \to 0$. The effect of the noise can be taken into account by adding noise corrections to the classical trace formula. A formal analogy exists between the noise and the quantum problem. This analogy allows us to treat the noise and quantum corrections together.

M.1 Periodic orbits as integrable systems

From now on, we use the language of quantum mechanics, since it is more convenient to visualize the results there. Where it is necessary we will discuss the difference between noise and quantum cases.

First we would like to introduce periodic orbits from an unusual point of view, which can convince you, that chaotic and integrable systems are in fact not as different from each other, than we might think. If we start orbits in the neighborhood of a periodic orbit and look at the picture on the Poincaré section we can see a regular picture. For stable periodic orbits the points form small ellipses around the center and for unstable orbits they form hyperbolas (See Fig. M.1).

The motion close to a periodic orbits is regular in both cases. This is

Figure M.1: Poincaré section close to a stable and an unstable periodic orbit

due to the fact, that we can linearize the Hamiltonian close to an orbit, and linear systems are always integrable. The linearized Hamilton's equations close to the periodic orbit $(q_p(t) + q, p_p(t) + p)$ look like

$$\begin{aligned} \dot{q} &= +\partial^2_{pq} H(q_p(t), p_p(t))q + \partial^2_{pp} H(q_p(t), p_p(t))p, & \text{(M.1)} \\ \dot{p} &= -\partial^2_{qq} H(q_p(t), p_p(t))q - \partial^2_{qp} H(q_p(t), p_p(t))p, & \text{(M.2)} \end{aligned}$$

where the new coordinates q and p are relative to a periodic orbit. This linearized equation can be regarded as a d dimensional oscillator with time periodic frequencies. These equations are representing the equation of motion in a redundant way since more than one combination of q, p and t determines the same point of the phase space. This can be cured by an extra restriction on the variables, a constraint the variables should fulfill. This constraint can be derived from the time independence or stacionarity of the full Hamiltonian

$$\partial_t H(q_p(t) + q, p_p(t) + p) = 0. \tag{M.3}$$

Using the linearized form of this constraint we can eliminate one of the linearized equations. It is very useful, although technically difficult, to do one more transformation and to introduce a coordinate, which is parallel with the Hamiltonian flow $(x_{\|})$ and others which are orthogonal. In the orthogonal directions we again get linear equations. These equations with $x_{\|}$ dependent rescaling can be transformed into normal coordinates, so that we get tiny oscillators in the new coordinates with constant frequencies. This result has first been derived by Poincaré for equilibrium points and later it was extended for periodic orbits by V.I. Arnol'd and co-workers. In the new coordinates, the Hamiltonian reads as

$$H_0(x_{\|}, p_{\|}, x_n, p_n) = \frac{1}{2} p_{\|}^2 + U(x_{\|}) + \sum_{n=1}^{d-1} \frac{1}{2}(p_n^2 \pm \omega_n^2 x_n^2), \tag{M.4}$$

which is the general form of the Hamiltonian in the neighborhood of a periodic orbit. The $\pm$ sign denotes, that for stable modes the oscillator potential is positive while for an unstable mode it is negative. For the unstable modes, ω is the Lyapunov exponent of the orbit

$$\omega_n = \ln \Lambda_{p,n} / T_p, \tag{M.5}$$

where $\Lambda_{p,n}$ is the expanding eigenvalue of the Jacobi matrix. For the stable directions the eigenvalues of the Jacobi matrix are connected with ω as

$$\Lambda_{p,n} = e^{-i\omega_n T_p}. \tag{M.6}$$

The Hamiltonian close to the periodic orbit is integrable and can be quantized by the Bohr-Sommerfeld rules. The result of the Bohr-Sommerfeld quantization for the oscillators gives the energy spectra

$$\begin{aligned} E_n &= \hbar\omega_n\left(j_n + \frac{1}{2}\right) \text{ for stable modes,} \\ E_n &= -i\hbar\omega_n\left(j_n + \frac{1}{2}\right) \text{ for unstable modes,} \end{aligned} \tag{M.7}$$

where $j_n = 0, 1, ...$ It is convenient to introduce the index $s_n = 1$ for stable and $s_n = -i$ for unstable directions. The parallel mode can be quantized implicitly trough the classical action function of the mode:

$$\frac{1}{2\pi}\oint p_{\|}dx_{\|} = \frac{1}{2\pi}S_{\|}(E_m) = \hbar\left(m + \frac{m_p\pi}{2}\right), \tag{M.8}$$

where m_p is the topological index of the motion in the parallel direction. This latter condition can be rewritten by a very useful trick into the equivalent form

$$(1 - e^{iS_{\|}(E_m)/\hbar - im_p\pi/2}) = 0. \tag{M.9}$$

The eigenenergies of a semiclassically quantized periodic orbit are all the possible energies

$$E = E_m + \sum_{n=1}^{d-1} E_n. \tag{M.10}$$

This relation allows us to change in (M.9) E_m with the full energy minus the oscillator energies $E_m = E - \sum_n E_n$. All the possible eigenenergies of the periodic orbit then are the zeroes of the expression

$$\Delta_p(E) = \prod_{j_1,\ldots,j_{d-1}} (1 - e^{iS_{\|}(E - \sum_n \hbar s_n\omega_n(j_n+1/2))/\hbar - im_p\pi/2}). \tag{M.11}$$

If we Taylor expand the action around E to first order

$$S_{\|}(E+\epsilon) \approx S_{\|}(E) + T(E)\epsilon, \tag{M.12}$$

where $T(E)$ is the period of the orbit, and use the relations of ω and the eigenvalues of the Jacobi matrix, we get the expression of the Selberg product

$$\Delta_p(E) = \prod_{j_1,\cdots,j_{d-1}} \left(1 - \frac{e^{iS_p(E)/\hbar - im_p\pi/2}}{\prod_n \Lambda_{p,n}^{(1/2+j_n)}}\right). \tag{M.13}$$

If we use the right convention for the square root we get exactly the d dimensional expression of the Selberg product formula we derived from the Gutzwiller trace formula in ? . Just here we derived it in a different way! The function $\Delta_p(E)$ is the semiclassical zeta function for one prime orbit.

Now, if we have many prime orbits and we would like to construct a function which is zero, whenever the energy coincides with the BS quantized energy of one of the periodic orbits, we have to take the product of these determinants:

$$\Delta(E) = \prod_p \Delta_p(E). \tag{M.14}$$

The miracle of the semiclassical zeta function is, that if we take infinitely many periodic orbits, the infinite product will have zeroes not at these energies, but close to the eigenenergies of the whole system !

So we learned, that both stable and unstable orbits are integrable systems and can be individually quantized semiclassically by the old Bohr-Sommerfeld rules. So we almost completed the program of Sommerfeld to quantize general systems with the method of Bohr. *Let us have a remark here. In addition to the Bohr-Sommerfeld rules, we used the unjustified approximation (M.12). Sommerfeld would never do this ! At that point we loose some important precision compared to the BS rules and we get somewhat worse results than a semiclassical formula is able to do. We will come back to this point later when we discuss the quantum corrections.* To complete the program of full scale Bohr-Sommerfeld quantization of chaotic systems we have to go beyond the linear approximation around the periodic orbit.

The Hamiltonian close to a periodic orbit in the parallel and normal coordinates can be written as the 'harmonic' plus 'anharmonic' perturbation

$$H(x_{\|}, p_{\|}, x_n, p_n) = H_0(x_{\|}, p_{\|}, x_n, p_n) + H_A(x_{\|}, x_n, p_n), \tag{M.15}$$

where the anharmonic part can be written as a sum of homogeneous polynomials of x_n and p_n with $x_{||}$ dependent coefficients:

$$H_A(x_{||}, x_n, p_n) = \sum_{k=3} H^k(x_{||}, x_n, p_n) \tag{M.16}$$

$$H^k(x_{||}, x_n, p_n) = \sum_{\sum l_n + m_n = k} H^k_{l_n, m_n}(x_{||}) x_n^{l_n} p_n^{m_n} \tag{M.17}$$

This classical Hamiltonian is hopeless from Sommerfeld's point of view, since it is non integrable. However, Birkhoff in 1927[3] introduced the concept of normal form, which helps us out from this problem by giving successive integrable approximation to a non-integrable problem. Let's learn a bit more about it!

M.2 The Birkhoff normal form

Birkhoff studied the canonical perturbation theory close to an equilibrium point of a Hamiltonian. Equilibrium point is where the potential has a minimum $\nabla U = 0$ and small perturbations lead to oscillatory motion. We can linearize the problem and by introducing normal coordinates x_n and conjugate momentums p_n the quadratic part of the Hamiltonian will be a set of oscillators

$$H_0(x_n, p_n) = \sum_{n=1}^{d} \frac{1}{2}(p_n^2 + \omega_n^2 x_n^2). \tag{M.18}$$

The full Hamiltonian can be rewritten with the new coordinates

$$H(x_n, p_n) = H_0(x_n, p_n) + H_A(x_n, p_n), \tag{M.19}$$

where H_A is the anharmonic part of the potential in the new coordinates. The anharmonic part can be written as a series of homogeneous polynomials

$$H_A(x_n, p_n) = \sum_{j=3}^{\infty} H^j(x_n, p_n), \tag{M.20}$$

$$H^j(x_n, p_n) = \sum_{|l|+|m|=j} h^j_{lm} x^l p^m, \tag{M.21}$$

[3]It is really a pity, that in 1926 Schrödinger introduced the wave mechanics and blocked the development of Sommerfeld's concept.

where h^j_{lm} are real constants and we used the multi-indices $l := (l_1, ..., l_d)$ with definitions

$$|l| = \sum l_n, x^l := x_1^{l_1} x_2^{l_2} ... x_d^{l_d}.$$

Birkhoff showed, that that by successive canonical transformations one can introduce new momentums and coordinates such, that in the new coordinates the anharmonic part of the Hamiltonian up to any given n polynomial will depend only on the variable combination

$$\tau_n = \frac{1}{2}(p_n^2 + \omega_n^2 x_n^2), \tag{M.22}$$

where x_n and p_n are the new coordinates and momentums, but ω_n is the original frequency. This is called the Birkhoff normal form of degree N:

$$H(x_n, p_n) = \sum_{j=2}^{N} H^j(\tau_1, ..., \tau_d), \tag{M.23}$$

where H^j are homogeneous degree j polynomials of τ-s. This is an integrable Hamiltonian, the non-integrability is pushed into the remainder, which consists of polynomials of degree higher than N. We run into trouble only when the oscillator frequencies are commensurate e.g. it is possible to find a set of integers m_n such that the linear combination

$$\sum_{n=1}^{d} \omega_n m_n,$$

vanishes. This extra problem has been solved by Gustavson in 1966 and we call the the object Birkhoff-Gustavson normal form. The procedure of the successive canonical transformations can be computerized and can be carried out up to high orders (~ 20).

Of course, we pay a price for forcing the system to be integrable up to degree N. For a non-integrable system the high order terms behave quete widely and the series is not convergent. Therefore we have to use this tool carefully. Now, we learned how to approximate a non-integrable system with a sequence of integrable systems and we can go back and carry out the BS quantization.

M.3 Bohr-Sommerfeld quantization of periodic orbits

There is some difference between equilibrium points and periodic orbits. The Hamiltonian (M.4) is not a sum of oscillators. One can transform the parallel part, describing circulation along the orbit, into an oscillator Hamiltonian, but this would make the problem extremelly difficult. Therefore, we carry out the canonical transformations dictated by the Birkhoff procedure only in the orthogonal directions. The $x_{\|}$ coordinate plays the role of a parameter. After the tasformation up to order N the Hamiltonian (M.17) is

$$H(x_{\|}, p_{\|}, \tau_1, ...\tau_{d-1}) = H_0(x_{\|}, p_{\|}, \tau_1, ..., \tau_{d-1}) + \sum_{j=2}^{N} U^j(x_{\|}, \tau_1, ..., \tau_{d-1}), \quad \text{(M.24)}$$

where U^j is a jth order homogeneous polynomial of τ-s with $x_{\|}$ dependent coefficients. The orthogonal part can be BS quantized by quantizing the individual oscillators, replacing τ-s as we did in (M.8). This leads to a one dimensional effective potential indexed by $j_1, ..., j_{d-1}$

$$H(x_{\|}, p_{\|}, j_1, ..., j_{d-1}) = \frac{1}{2}p_{\|}^2 + U(x_{\|}) + \sum_{n=1}^{d-1} \hbar s_n \omega_n (j_n + 1/2) + \quad \text{(M.25)}$$
$$+ \sum_{k=2}^{N} U^k(x_{\|}, \hbar s_1 \omega_1(j_1 + 1/2), \hbar s_2 \omega_2(j_2 + 1/2), ..., \hbar s_{d-1} \omega_{d-1}(j_{d-1} + 1/2)),$$

where j_n can be any non-negative integer. The term with index k is proportional with $\hbar^k$ due to the homogeneity of the polynomials.

The parallel mode now can be BS quantized for any given set of j-s

$$S_p(E, j_1, ..., j_{d-1}) = \oint p_{\|} dx_{\|} = \quad \text{(M.26)}$$
$$= \oint dx_{\|} \sqrt{E - \sum_{n=1}^{d-1} \hbar s_n \omega_n (j_n + 1/2) - U(x_{\|}, j_1, ..., j_{d-1})} = 2\pi\hbar(m + m_p/2),$$

where U contains all the $x_{\|}$ dependent terms of the Hamiltonian. The spectral determinant becomes

$$\Delta_p(E) = \prod_{j_1, ..., j_{d-1}} (1 - e^{iS_p(E, j_1, ..., j_{d-1})/\hbar - m_p \pi/2}). \quad \text{(M.27)}$$

This expression completes the Sommerfeld method and tells us how to quantize chaotic or general Hamiltonian systems. Unfortunately, quantum mechanics postponed this nice formula until our book.

This formula has been derived with the help of the semiclassical Bohr-Sommerfeld quantization rule and the classical normal form theory. Indeed, if we expand S_p in the exponent in the powers of $\hbar$

$$S_p = \sum_{k=0}^{N} \hbar^k S_k,$$

we get more than just a constant and a linear term. This formula already gives us corrections to the semiclassical zeta function in all powers of $\hbar$. There is a very attracting feature of this semiclassical expansion. $\hbar$ in S_p shows up only in the combination $\hbar s_n \omega_n (j_n + 1/2)$. A term proportional with $\hbar^k$ can only be a homogeneous expression of the oscillator energies $s_n \omega_n (j_n + 1/2)$. For example in two dimensions there is only one possibility of the functional form of the order k term

$$S_k = c_k(E) \cdot \omega_n^k (j + 1/2)^k,$$

where $c_k(E)$ is the only function to be determined.

The corrections derived sofar are *doubly* semiclassical, since they give semiclassical corrections to the semiclassical approximation. What can quantum mechanics add to this ? As we have stressed in the previous section, the exact quantum mechanics is not invariant under canonical transformations. In other context, this phenomenon is called the operator ordering problem. Since the operators $\hat{x}$ and $\hat{p}$ do not commute, we run into problems, when we would like to write down operators for classical quantities like $x^2 p^2$. On the classical level the four possible orderings $xpxp$, $ppxx$, $pxpx$ and $xxpp$ are equivalent, but they are different in the quantum case. The expression for the energy (M.26) is not exact. We have to go back to the level of the Schrödinger equation if we would like to get the exact expression.

M.4 Quantum calculation of $\hbar$ corrections

The Gutzwiller trace formula has originally been derived from the saddle point approximation of the Feynman path integral form of the propagator. The exact trace is a pathsum for all closed paths of the system

$$\mathrm{Tr} G(x, x', t) = \int dx G(x, x, t) = \int \mathcal{D}x e^{iS(x,t)/\hbar}, \tag{M.28}$$

where $\int \mathcal{D}x$ denotes the discretization and summation for all paths of time length t in the limit of the infinite refination and $S(x,t)$ is the classical action calculated along the path. The trace in the saddle point calculation is a sum for classical periodic orbits and zero length orbits, since these are the extrema of the action $\delta S(x,t) = 0$ for closed paths:

$$\mathrm{Tr}G(x,x',t) = g_0(t) + \sum_{p \in PO} \int \mathcal{D}\xi_p e^{iS(\xi_p + x_p(t),t)/\hbar}, \tag{M.29}$$

where $g_0(t)$ is the zero length orbit contribution. We introduced the new coordinate ξ_p with respect to the periodic orbit $x_p(t)$, $x = \xi_p + x_p(t)$. Now, each path sum $\int \mathcal{D}\xi_p$ is computed in the vicinity of periodic orbits. Since the saddle points are taken in the configuration space, only spatially distinct periodic orbits, the so called prime periodic orbits, appear in the summation. Sofar nothing new has been invented. If we continue the standard textbook calculation scheme, we have to Taylor expand the action in ξ_p and keep the quadratic term in the exponent while treating the higher order terms as corrections. Then we can compute the path integrals with the help of gaussian integrals. The key point here is that we don't compute the path sum directly. We use the correspondence between path integrals and partial differential equations. This idea comes from Maslov [M.6] and a good summary is in ref. [M.7]. We search for that Schrödinger equation, which leads to the path sum

$$\int \mathcal{D}\xi_p e^{iS(\xi_p + x_p(t),t)/\hbar}, \tag{M.30}$$

where the action around the periodic orbit is in a multi dimensional Taylor expanded form:

$$S(x,t) = \sum_{\mathbf{n}}^{\infty} s_{\mathbf{n}}(t)(x - x_p(t))^{\mathbf{n}}/\mathbf{n}!. \tag{M.31}$$

The symbol $\mathbf{n} = (n_1, n_2, ..., n_d)$ denotes the multi index in d dimensions, $\mathbf{n}! = \prod_{i=1}^{d} n_i!$ the multi factorial and $(x - x_p(t))^{\mathbf{n}} = \prod_{i=1}^{d}(x_i - x_{p,i}(t))^{n_i}$, respectively. The expansion coefficients of the action can be determined from the Hamilton-Jacobi equation

$$\partial_t S + \frac{1}{2}(\nabla S)^2 + U = 0, \tag{M.32}$$

in which the potential is expanded in a multidimensional Taylor series around the orbit

$$U(x) = \sum_{\mathbf{n}} u_{\mathbf{n}}(t)(x - x_p(t))^{\mathbf{n}}/\mathbf{n}!. \tag{M.33}$$

The Schrödinger equation

$$i\hbar\partial_t\psi = \hat{H}\psi = -\frac{\hbar^2}{2}\Delta\psi + U\psi, \tag{M.34}$$

with this potential also can be expanded around the periodic orbit. Using the WKB ansatz

$$\psi = \varphi e^{iS/\hbar}, \tag{M.35}$$

we can construct a Schrödinger equation corresponding to a given order of the Taylor expansion of the classical action. The Schrödinger equation induces the Hamilton-Jacobi equation (M.32) for the phase and the transport equation of Maslov and Fjedoriuk [M.8] for the amplitude:

$$\partial_t\varphi + \nabla\varphi\nabla S + \frac{1}{2}\varphi\Delta S - \frac{i\hbar}{2}\Delta\varphi = 0. \tag{M.36}$$

This is the partial differential equation, solved in the neighborhood of a periodic orbit with the expanded action (M.31), which belongs to the local pathsum (M.30).

If we know the Green's function $G_p(\xi,\xi',t)$ corresponding to the local equation (M.36), then the local path sum can be converted back into a trace:

$$\int \mathcal{D}\xi_p e^{i/\hbar\sum_{\mathbf{n}} S_{\mathbf{n}}(x_p(t),t)\xi_p^{\mathbf{n}}/\mathbf{n}!} = \mathrm{Tr}G_p(\xi,\xi',t). \tag{M.37}$$

The saddle point expansion of the trace in terms of local traces then becomes

$$\mathrm{Tr}G(x,x',t) = \mathrm{Tr}G_W(x,x',t) + \sum_p \mathrm{Tr}G_p(\xi,\xi',t), \tag{M.38}$$

where $G_W(x,x',t)$ denotes formally the Green's function expanded around zero length (non moving) periodic orbits, known as the Weyl term [M.9]. Each Green's function can be Fourier-Laplace transformed independently and by definition we get in the energy domain:

$$\mathrm{Tr}G(x,x',E) = g_0(E) + \sum_p \mathrm{Tr}G_p(\xi,\xi',E). \tag{M.39}$$

Notice, that we do not need here to take further saddle points in time, since we are dealing with exact time and energy domain Green's functions. indexGreen's function!energy dependent

The spectral determinant is a function which has zeroes at the eigenenergies E_n of the Hamilton operator $\hat{H}$. Formally it is

$$\Delta(E) = \det(E - \hat{H}) = \prod_n (E - E_n).$$

The logarithmic derivative of the spectral determinant is the trace of the energy domain Green's function:

$$\mathrm{Tr}G(x, x', E) = \sum_n \frac{1}{E - E_n} = \frac{d}{dE}\log\Delta(E). \tag{M.40}$$

We can define the spectral determinant $\Delta_p(E)$ also for the local operators and we can write

$$\mathrm{Tr}G_p(\xi, \xi', E) = \frac{d}{dE}\log\Delta_p(E). \tag{M.41}$$

Using (M.39) we can express the full spectral determinant as a product for the sub-determinants

$$\Delta(E) = e^{W(E)}\prod_p \Delta_p(E),$$

where $W(E) = \int^E g_0(E')dE'$ is the term coming from the Weyl expansion.

The construction of the local spectral determinants can be done easily. We have to consider the stationary eigenvalue problem of the local Schrödinger problem and keep in mind, that we are in a coordinate system moving together with the periodic orbit. If the classical energy of the periodic orbit coincides with an eigenenergy E of the local Schrödinger equation around the periodic orbit, then the corresponding stationary eigenfunction fulfills

$$\psi_p(\xi, t+T_p) = \int d\xi' G_p(\xi, \xi', t+T_p)\psi_p(\xi', t) = e^{-iET_p/\hbar}\psi_p(\xi, t), \tag{M.42}$$

where T_p is the period of the prime orbit p. If the classical energy of the periodic orbit is not an eigenenergy of the local Schrödinger equation, the non-stationary eigenfunctions fulfill

$$\psi_p^1(\xi, t+T_p) = \int d\xi' G_p(\xi, \xi', t+T_p)\psi_p(\xi', t) = e^{-iET_p/\hbar}\lambda_p^1(E)\psi_p^1(t), \tag{M.43}$$

where $\mathbf{l} = (l_1, l_2, ...)$ is a multi-index of the possible quantum numbers of the local Schrödinger equation. If the eigenvalues $\lambda_p^{\mathbf{l}}(E)$ are known the local functional determinant can be written as

$$\Delta_p(E) = \prod_{\mathbf{l}} (1 - \lambda_p^{\mathbf{l}}(E)), \tag{M.44}$$

since $\Delta_p(E)$ is zero at the eigenenergies of the local Schrödinger problem. We can insert the ansatz (M.35) and reformulate (M.43) as

$$e^{\frac{i}{\hbar}S(t+T_p)}\varphi_p^{\mathbf{l}}(t+T_p) = e^{-iET_p/\hbar}\lambda_p^{\mathbf{l}}(E)e^{\frac{i}{\hbar}S(t)}\varphi_p^{\mathbf{l}}(t). \tag{M.45}$$

The phase change is given by the action integral for one period $S(t+T_p) - S(t) = \int_0^{T_p} L(t)dt$. Using this and the identity for the action $S_p(E)$ of the periodic orbit

$$S_p(E) = \oint pdq = \int_0^{T_p} L(t)dt + ET_p, \tag{M.46}$$

we get

$$e^{\frac{i}{\hbar}S_p(E)}\varphi_p^{\mathbf{l}}(t+T_p) = \lambda_p^{\mathbf{l}}(E)\varphi_p^{\mathbf{l}}(t). \tag{M.47}$$

Introducing the eigenequation for the amplitude

$$\varphi_p^{\mathbf{l}}(t+T_p) = R_{\mathbf{l},p}(E)\varphi_p^{\mathbf{l}}(t), \tag{M.48}$$

the local spectral determinant can be expressed as a product for the quantum numbers of the local problem:

$$\Delta_p(E) = \prod_{\mathbf{l}} (1 - R_{\mathbf{l},p}(E)e^{\frac{i}{\hbar}S_p(E)}). \tag{M.49}$$

Since $\hbar$ is a small parameter we can develop a perturbation series for the amplitudes $\varphi_p^{\mathbf{l}}(t) = \sum_{m=0}^{\infty} \left(\frac{i\hbar}{2}\right)^m \varphi_p^{\mathbf{l}(m)}(t)$ which can be inserted into the equation (M.36) and we get an iterative scheme starting with the semiclassical solution $\varphi^{\mathbf{l}(0)}$:

$$\begin{aligned} \partial_t\varphi^{\mathbf{l}(0)} + \nabla\varphi^{\mathbf{l}(0)}\nabla S + \frac{1}{2}\varphi^{\mathbf{l}(0)}\Delta S &= 0, \\ \partial_t\varphi^{\mathbf{l}(m+1)} + \nabla\varphi^{\mathbf{l}(m+1)}\nabla S + \frac{1}{2}\varphi^{\mathbf{l}(m+1)}\Delta S &= \Delta\varphi^{\mathbf{l}(m)}. \end{aligned} \tag{M.50}$$

The eigenvalue can also be expanded in powers of $i\hbar/2$:

$$
\begin{aligned}
R_{\mathbf{l},p}(E) &= \exp\left\{\sum_{m=0}^{\infty}\left(\frac{i\hbar}{2}\right)^m C_{\mathbf{l},p}^{(m)}\right\} && \text{(M.51)} \\
= \exp(C_{\mathbf{l},p}^{(0)})\,\{1 &+ \frac{i\hbar}{2}C_{\mathbf{l},p}^{(1)} + \left(\frac{i\hbar}{2}\right)^2\left(\frac{1}{2}(C_{\mathbf{l},p}^{(1)})^2 + C_{\mathbf{l},p}^{(2)}\right) + \ldots && \text{(M.52)}
\end{aligned}
$$

The eigenvalue equation (M.48) in $\hbar$ expanded form reads as

$$
\begin{aligned}
\varphi_p^{\mathbf{l}(0)}(t+T_p) &= \exp(C_{\mathbf{l},p}^{(0)})\varphi_p^{\mathbf{l}(0)}(t), \\
\varphi_p^{\mathbf{l}(1)}(t+T_p) &= \exp(C_{\mathbf{l},p}^{(0)})[\varphi_p^{\mathbf{l}(1)}(t) + C_{\mathbf{l},p}^{(1)}\varphi_p^{\mathbf{l}(0)}(t)], \\
\varphi_p^{\mathbf{l}(2)}(t+T_p) &= \exp(C_{\mathbf{l},p}^{(0)})[\varphi_p^{\mathbf{l}(2)}(t) + C_{\mathbf{l},p}^{(1)}\varphi_p^{\mathbf{l}(1)}(t) + (C_{\mathbf{l},p}^{(2)} + \frac{1}{2}(C_{\mathbf{l},p}^{(1)})^2)\varphi_p^{\mathbf{l}(0)}(t)], && \text{(M.53)}
\end{aligned}
$$

and so on. These equations are the conditions selecting the eigenvectors and eigenvalues and they hold for all t.

It is very convenient to expand the functions $\varphi_p^{\mathbf{l}(m)}(x,t)$ in Taylor series around the periodic orbit and to solve the equations (M.51) in this basis [M.11], since only a couple of coefficients should be computed to derive the first corrections. This technical part we are going to publish elsewhere [M.10]. One can derive in general the zero order term $C_{\mathbf{l}}^{(0)} = i\pi\nu_p + \sum_{i=1}^{d-1}\left(l_i + \frac{1}{2}\right)u_{p,i}$, where $u_{p,i} = \log\Lambda_{p,i}$ are the logarithms of the eigenvalues of the monodromy matrix $\mathbf{J}_p$ and ν_p is the topological index of the periodic orbit. The first correction is given by the integral

$$
C_{\mathbf{l},p}^{(1)} = \int_0^{T_p} dt\,\frac{\Delta\varphi_p^{\mathbf{l}(0)}(t)}{\varphi_p^{\mathbf{l}(0)}(t)}.
$$

When the theory is applied for billiard systems, the wave function should fulfill the Dirichlet boundary condition on hard walls, e.g. it should vanish on the wall. The wave function determined from (M.36) behaves discontinuously when the trajectory $x_p(t)$ hits the wall. For the simplicity we consider a two dimensional billiard system here. The wave function on the wall before the bounce (t_{-0}) is given by

$$
\psi_{in}(x,y(x),t) = \varphi(x,y(x),t_{-0})e^{iS(x,y(x),t_{-0})/\hbar}, \tag{M.54}
$$

where $y(x) = Y_2x^2/2! + Y_3x^3/3! + Y_4x^4/4! + \ldots$ is the parametrization of the wall around the point of reflection (see Fig 1.). The wave function on the wall after the bounce (t_{+0}) is

$$
\psi_{out}(x,y(x),t) = \varphi(x,y(x),t_{+0})e^{iS(x,y(x),t_{+0})/\hbar}. \tag{M.55}
$$

The sum of these wave functions should vanish on the hard wall. This implies that the incoming and the outgoing amplitudes and the phases are related as

$$S(x, y(x), t_{-0}) = S(x, y(x), t_{+0}), \tag{M.56}$$

and

$$\varphi(x, y(x), t_{-0}) = -\varphi(x, y(x), t_{+0}). \tag{M.57}$$

The minus sign can be interpreted as the topological phase coming from the hard wall.

Now we can reexpress the spectral determinant with the local eigenvalues:

$$\Delta(E) = e^{W(E)} \prod_p \prod_{\mathbf{l}} (1 - R_{\mathbf{l},p}(E) e^{\frac{i}{\hbar} S_p(E)}). \tag{M.58}$$

This expression is the quantum generalization of the semiclassical Selberg-product formula [M.12]. A similar decomposition has been found for quantum Baker maps in ref. [M.13]. The functions

$$\zeta_{\mathbf{l}}^{-1}(E) = \prod_p (1 - R_{\mathbf{l},p}(E) e^{\frac{i}{\hbar} S_p(E)}) \tag{M.59}$$

are the generalizations of the Ruelle type [27.37] zeta functions. The trace formula can be recovered from (M.40):

$$\mathrm{Tr}G(E) = g_0(E) + \frac{1}{i\hbar} \sum_{p,\mathbf{l}} (T_p(E) - i\hbar \frac{d \log R_{\mathbf{l},p}(E)}{dE}) \frac{R_{\mathbf{l},p}(E) e^{\frac{i}{\hbar} S_p(E)}}{1 - R_{\mathbf{l},p}(E) e^{\frac{i}{\hbar} S_p(E)}}. \tag{M.60}$$

We can rewrite the denominator as a sum of a geometric series and we get

$$\mathrm{Tr}G(E) = g_0(E) + \frac{1}{i\hbar} \sum_{p,r,\mathbf{l}} (T_p(E) - i\hbar \frac{d \log R_{\mathbf{l},p}(E)}{dE}) (R_{\mathbf{l},p}(E))^r e^{\frac{i}{\hbar} r S_p(E)}. \tag{M.61}$$

The new index r can be interpreted as the repetition number of the prime orbit p. This expression is the generalization of the semiclassical trace formula for the exact quantum mechanics. We would like to stress here, that the perturbation calculus introduced above is just one way to compute the eigenvalues of the local Schrödinger problems. Non-perturbative methods can be used to calculate the local eigenvalues for stable, unstable and marginal orbits. Therefore, our trace formula is not limited to integrable or hyperbolic systems, it can describe the most general case of systems with mixed phase space.

The semiclassical trace formula can be recovered by dropping the subleading term $-i\hbar d \log R_{\mathbf{l},p}(E)/dE$ and using the semiclassical eigenvalue $R^{(0)}_{\mathbf{l},p}(E) = e^{C_p^{\mathbf{l}(0)}} = e^{-i\nu_p\pi}e^{-\sum_i(l_i+1/2)u_{p,i}}$. Summation for the indexes l_i yields the celebrated semiclassical amplitude

$$\sum_{\mathbf{l}}(R^{(0)}_{\mathbf{l},p}(E))^r = \frac{e^{-ir\nu_p\pi}}{\mid\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\mid^{1/2}}. \tag{M.62}$$

To have an impression about the improvement caused by the quantum corrections we have developed a numerical code [M.14] which calculates the first correction $C^{(1)}_{p,l}$ for general two dimensional billiard systems . The first correction depends only on some basic data of the periodic orbit such as the lengths of the free flights between bounces, the angles of incidence and the first three Taylor expansion coefficients Y_2, Y_3, Y_4 of the wall in the point of incidence. To check that our new local method gives the same result as the direct calculation of the Feynman integral, we computed the first $\hbar$ correction $C^{(1)}_{p,0}$ for the periodic orbits of the 3-disk scattering system [M.15] where the quantum corrections have been We have found agreement up to the fifth decimal digit, while our method generates these numbers with any desired precision. Unfortunately, the $l \neq 0$ coefficients cannot be compared to ref. [M.16], since the l dependence was not realized there due to the lack of general formulas (M.58) and (M.59). However, the l dependence can be checked on the 2 disk scattering system [M.17]. On the standard example [M.15, M.16, M.17, M.19], when the distance of the centers (R) is 6 times the disk radius (a), we got

$$C^{(1)}_l = \frac{1}{\sqrt{2E}}(-0.625l^3 - 0.3125l^2 + 1.4375l + 0.625).$$

For $l = 0$ and 1 this has been confirmed by A. Wirzba [M.18], who was able to compute $C^{(1)}_0$ from his exact quantum calculation. Our method makes it possible to utilize the symmetry reduction of Cvitanović and Eckhardt and to repeat the fundamental domain cycle expansion calculation of ref. [M.19] with the first quantum correction. We computed the correction to the leading 226 prime periodic orbits with 10 or less bounces in the fundamental domain. Table I. shows the numerical values of the

Figure M.2: A typical bounce on a billiard wall. The wall can be characterized by the local expansion $y(x) = Y_2x^2/2! + Y_3x^3/3! + Y_4x^4/4! + ...$

Table M.1: Real part of the resonances (Re k) of the 3-disk scattering system at disk separation 6:1. Semiclassical and first corrected cycle expansion versus exact quantum calculation and the error of the semiclassical δ_{SC} divided by the error of the first correction δ_{Corr}. The magnitude of the error in the imaginary part of the resonances remains unchanged.

Quantum	Semiclassical	First correction	$\delta_{SC}/\delta_{Corr}$
0.697995	0.758313	0.585150	0.53
2.239601	2.274278	2.222930	2.08
3.762686	3.787876	3.756594	4.13
5.275666	5.296067	5.272627	6.71
6.776066	6.793636	6.774061	8.76
...	...	...	...
30.24130	30.24555	30.24125	92.3
31.72739	31.73148	31.72734	83.8
32.30110	32.30391	32.30095	20.0
33.21053	33.21446	33.21048	79.4
33.85222	33.85493	33.85211	25.2
34.69157	34.69534	34.69152	77.0

exact quantum calculation [M.17], the semiclassical cycle expansion [M.11] and our corrected calculation. One can see, that the error of the corrected calculation vs. the error of the semiclassical calculation decreases with the wave number. Besides the improved results, a fast convergence up to six decimal digits can be observed, which is just three decimal digits in the full domain calculation [M.16].

References

[M.1] M. C. Gutzwiller, J. Math. Phys. **12**, 343 (1971); *Chaos in Classical and Quantum Mechanics* (Springer-Verlag, New York, 1990)

[M.2] A. Selberg, J. Indian Math. Soc. **20**, 47 (1956)

[M.3] See examples in : CHAOS **2** (1) Thematic Issue; E. Bogomolny and C. Schmit, Nonlinearity **6**, 523 (1993)

[M.4] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948)

[M.5] L. S. Schulman, *Techniques and Applications of Path Integration*, Wiley-Interscience Publication (1981)

[M.6] We thank E. Bogomolny for bringing this reference to our attention.

[M.7] V. M. Babić and V. S. Buldyrev, *Short Wavelength Diffraction Theory*, Springer Series on Wave Phenomena, Springer-Verlag (1990)

[M.8] V. P. Maslov and M. V. Fjedoriuk, *Semiclassical Approximation in Quantum Mechanics*, Dordrecht-Reidel (1981)

[M.9] R. B. Balian and C. Bloch, Ann. Phys. (New York) **60**, 81 (1970);ibid. **63**, 592 (1971); M.V. Berry, M.V., C.J. Howls, C.J. Proceedings of the Royal Society of London. **447**, 1931 (1994)

[M.10] P. E. Rosenqvist and G. Vattay, in progress.

[M.11] P. Cvitanović, P. E. Rosenqvist, G. Vattay and H. H. Rugh, CHAOS **3** (4), 619 (1993)

[M.12] A. Voros, J. Phys. **A21**, 685 (1988)

[M.13] A. Voros, Prog. Theor. Phys. Suppl. **116**,17 (1994); M. Saraceno and A. Voros, to appear in Physica D.

[M.14] The FORTRAN code is available upon e-mail request to G. Vattay.

[M.15] P. Gaspard and S. A. Rice, J. Chem. Phys. **90** 2225, 2242, 2255 (1989) **91** E3279 (1989)

[M.16] D. Alonso and P. Gaspard, Chaos **3**, 601 (1993); P. Gaspard and D. Alonso, Phys. Rev. **A47**, R3468 (1993)

[M.17] A. Wirzba, CHAOS **2**, 77 (1992); Nucl. Phys. **A560**, 136 (1993)

[M.18] A. Wirzba, private communication.

[M.19] P. Cvitanović and B. Eckhardt, Phys. Rev. Lett. **63**, 823 (1989)

Appendix N

What reviewers say

N.1 Niels Bohr

"The most important work since that Schrödinger killed the cat."

N.2 Richard P. Feynman

"Great doorstop!"

N.3 Divakar Viswanath

I have read the first six chapters. The introduction chapter is fabulous. The computation of the escape rate for the pin-ball game in the first chapter is absolutely riveting. There can't be a better way to begin. I find the book to be very stimulating.

N.4 Benny Lautrup

I am now reading your book as meticulously. I have lots of little comments, and one big one, which I can tell you immediately. I think that your opening chapter with its many literary references, jokes, and finesses, loses out in the end, because the main message gets obscured. Don't make section quotes unless they are really pertinent (although I love the first one about standing behind giants), but why cite Kierkegaard in Danish (or at all for that matter)?

N.5 Professor Gatto Nero

This book, which I have received unsolicited from the Szczsyrk Oblast Agricultural and Mazuth Office Press appears to be a collage of LaTeX clips from random papers authored by the motley collection of co-authors whose number one usually associates with an experimental high energy *Phys. Rev. Letter*, rather than a book that aspires to be the Landau-Lifshitz of chaos.

Entire rain forests went down so this not inconsiderable tome can be printed and bound. Why these ravings were not left on the Web where they more properly belong is not explained anywhere in the text. If it is an antiBourbaki, as one has in the antimatter world, then why not do the real thing? A Landau-Lifshitz for nonlinear Science, written as it should be done. The nonScience book to end all nonScience books.

Appendix O

Solutions

Chapter 1

Solution 1.1: 3-disk symbolic dynamics. *As explained in sect. 1.4, each orbit segment can be characterized by either of the two symbols* 0 *and* 1*, differentiating topologically bouncing back or going onto the third disk. So, there are* 2^n *topologically different orbits starting from each disk. Altogether, the 3-disk pinball has* $3 \cdot 2^n$ *itineraries of length* n*.*
Periodic orbits (prime cycles in fundamental domain)

- *Of length* 2*: 12,13,32; or (0).*
- *Of length* 3*: 123,321; or (1).*
- *Of length* 4*: 1213,2321,3231; or (01).*
- *Of length* 5*: 12123,13132,23231,21213,32321,31312; or (00111).*

Some of the cycles are listed in table 9.1 and drawn in fig. 19.3.

(Yuheng Lan)

Solution 1.2: Sensitivity to initial conditions. *To estimate the pinball sensitivity we consider a narrow beam of point particles bouncing between two disks, fig. O.1(a). Or if you find this easier to visualize, think of a narrow ray of light. We assume that the ray of light is focused along the axis between the two points. This is where the least unstable periodic orbit lies, so its stability should give us an upper bound on the number of bounces we can expect to achieve. To estimate the stability we assume that the ray of light has a width* $w(t)$ *and a "dispersion angle"* $\theta(t)$ *(we assume both are small), fig. O.1(b). Between bounces the dispersion angle stays constant while the width increases as*

$$w(t) \approx w(t') + (t - t')\theta$$

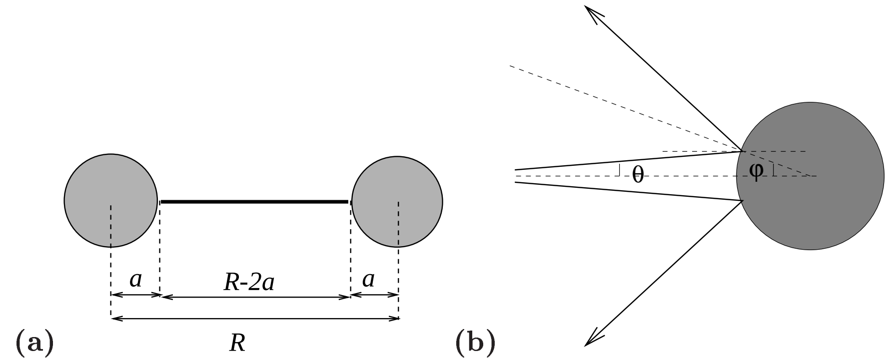


Figure O.1: The 2-disk pinball (a) geometry, (b) defocusing of scattered rays.

At each bounce the width stays constant while the angle increases by

$$\theta_{n+1} = \theta_n + 2\phi \approx \theta_n + w(t)/a.$$

where θ_n denotes the angle after bounce n. Denoting the width of the ray at the nth bounce by w_n then we obtain the pair of coupled equations

$$w_{n+1} = w_n + \left(R - 2a\right)\theta_n \tag{O.1}$$

$$\theta_n = \theta_{n-1} + \frac{w_n}{a} \tag{O.2}$$

where we ignore corrections of order w_n^2 and θ_n^2. Solving for θ_n we find

$$\theta_n = \theta_0 + \frac{1}{a}\sum_{j=1}^{n} w_n.$$

Assuming $\theta_0 = 0$ then

$$w_{n+1} = w_n + \frac{R-2a}{a}\sum_{j=1}^{n} w_n$$

Plugging in the values in the question we find the width at each bounce in Ångstrøms grows as 1, 5, 29, 169, 985, etc. To find the asymptotic behavior for a large number of bounces we try an solution of the form $w_n = ax^n$. Substituting this into the equation above and ignoring terms that do not grow exponentially we find solutions

$$w_n \approx a w_n^{asym} = a(3 \pm 2\sqrt{2})^n$$

The solution with the positive sign will clearly dominate. The constant a we cannot determine by this local analysis although it is clearly proportional to w_0. However, the *asymptotic solution is a good approximation even for quite a small number of bounces. To find an estimate of a we see that w_n/w_n^{asym} very rapidly converges to 0.146447, thus*

$$w_n \approx 0.146447 w_0 (3 + 2\sqrt{2})^n \approx 0.1 \times w_0 \times 5.83^n$$

The outside edges of the ray of light will miss the disk when the width of the ray exceeds 2 cm; this occurs after 11 bounces.

(Adam Prügel-Bennett)

Solution 1.2: Sensitivity to initial conditions, another try. *Adam's estimate is not very good - do you have a better one? The first problem with it is that the instability is very underestimated. As we shall check in exercise 5.5, the exact formula for the 2-cycle stability is $\Lambda = R - 1 + R\sqrt{1 - 2/R}$. For $R = 6$, $a = 1$ this yields $w_n/w_0 \approx (5 + 2\sqrt{6})^n = 9.898979^n$, so if that were the whole story, the pinball would be not likely to make it much beyond 8 bounces.*

The second problem is that local instability overestimates the escape rate from an enclosure; trajectories are reinjected by scatterers. In the 3-disk pinball the particle leaving a disk can be reinjected by hitting either of other 2 disks, hence $w_n/w_0 \approx (9.9/2)^n$. This interplay between local instability and global reinjection will be cast into the exact formula involving "Lyapunov exponent" and "Kolmogorov entropy". In order to relate this estimate to our best continuous time escape rate estimate $\gamma = 0.4103\ldots$ (see table 15.2), we will have to also compute the mean free flight time (15.20). As a crude estimate, we take the shortest disk-to-disk distance, $\langle T \rangle = R - 2 = 4$. The continuous time escape rate result implies that $w_n/w_0 \approx e^{(R-2)\gamma n} = (5.16)^n$, in the same ballpark as the above expansion-reinjection estimate.

(Predrag Cvitanović)

Chapter 2

Solution 2.1: Trajectories do not intersect. *Suppose that two trajectories C_x and C_y intersect at some point z. We claim that any points $\tilde{x}$ on C_x is also a point on C_y and vice versa. We only need to prove the first part of the statement. According to the definition of C_x, there exist $t_x, t_y, t_1 \in \mathbb{R}$ such that $f^{t_x}(x) = z, f^{t_y}(y) = z, f^{t_1}(x) = \tilde{x}$. It is easy to check that $f^{t_y - t_x + t_1}(y) = \tilde{x}$. So, $\tilde{x} \in C_y$. Therefore, if two trajectories intersect, then they are the same trajectory.*

(Yueheng Lan)

Solution 2.2: Evolution as a group. *Let's check the basic defining properties of a group. The members of the set are $f^t, t \in \mathbb{R}$ and the "product law" is given by '$\circ$'.*

- *As $f^{t+s} = f^t \circ f^s$, the set is closed,* i.e., *the product of any two members generates another member of the set.*

- *It is associative, as* $(f^t \circ f^s) \circ f^r = f^{t+s+r} = f^t \circ (f^s \circ f^r)$.
- $I = f^0$ *is the identity, as* $f^t \circ f^0 = f^t$.
- f^{-t} *is the inverse of* f^t, *as* $f^{-t} \circ f^t = I$.

So, $\{f^t, \circ\}_{t \in \mathbb{R}}$ *forms a group. As* $f^t \circ f^s = f^{t+s} = f^s \circ f^t$, *it is a commutative (Abelian) group.*
Any Abelian group can replace the continuous time. For example, $\mathbb{R}$ *can be replaced by* $\mathbb{Z}_6$. *To mess things up try a non-commutative group.*

(Yueheng Lan)

Solution 2.3: Almost ODE's. *What is an ODE on* $\mathbb{R}$ *? An ODE is an equality which reveals explicitly the relation between function* $x(t)$ *and its time derivatives* $\dot{x}, \ddot{x}, \cdots$, i.e., $F(t, x, \dot{x}, \ddot{x}, \cdots) = 0$ *for some given function* F. *Let's check the equations given in the exercise.*
(a) $\dot{x} = \exp(\dot{x})$ *is an ODE.*
(b) $\dot{x} = x(x(t))$ *is not an ODE, as* $x(x(t))$ *is not a known function acting on* $x(t)$.
(c) $\dot{x} = x(t+1)$ *is not an ODE, as* $x(t+1)$ *is not a value at current time. Actually, it is a difference-differential equation.*

(Yueheng Lan)

Solution 2.4: All equilibrium points are fixed points. *Given a vector field* $v(x)$, *the phase space dynamics is defined by*

$$\frac{d}{dt}x(t) = v(x(t))\,. \tag{O.3}$$

An equilibrium point a *of* v *is defined by* $v(a) = 0$, *so* $x(t) = a$ *is a constant solution of (O.3). For the flow* f^t *defined by (O.3), this solution satisfies* $f^t(a) = a\,, t \in \mathbb{R}$. *So, it is a fixed point of the dynamics* f^t.

(Yueheng Lan)

Solution 2.5: Gradient systems. *(a) The directional derivative*

$$\frac{d}{dn}\phi = n \cdot \nabla\phi$$

produces the increasing rate along the unit vector n. *So, along the gradient direction* $\nabla\phi/|\nabla\phi|$, ϕ *has the largest increasing rate. The velocity of the particle has the opposite direction to the gradient, so* ϕ *deceases most rapidly in the velocity direction.*

(b) An extremum a *of* ϕ *satisfies* $\nabla\phi(a) = 0$. *According to Exercise 2.4,* a *is a fixed point of the flow.*

(c) Two arguments lead to the same conclusion here.
First, near an equilibrium point, the equation is always linearizable. For gradient system, after orthogonal transformation it is even possible to write the linearized equation in diagonal form so that we need only to consider one eigendirection. The corresponding scalar equation is $\dot{x} = \lambda x$. *Notice that we moved the origin to the equilibrium point. The solution of this equation is* $x(t) = x(0)\exp(\lambda t),$ *for* $\lambda \neq 0$. *if* $x(0) \neq 0$, *it will take infinite amount of time (positive or negative) for* $x(t) \to 0$. *For* $\lambda = 0$, *the approach to zero is even slower as then only higher orders of* x *take effect.*
The second argument seems easier. We know that the solution curve through an equilibrium point is the point itself. According to Excercise 2.1, no other solution curve will intersect it, which means that if not starting from the equilibrium point itself, other point can never reach it.

(d) On a periodic orbit, the velocity is bounded away from zero. So ϕ *is always decreasing on a periodic orbit, but in view of the periodicity, we know that this can not happen (at each point, there is only one value of* ϕ*.). So, there is no periodic orbit in a gradient system.*

(Yueheng Lan)

Solution 2.9: Classical collinear helium dynamics. *An example of a solution are A. Prügel-Bennett's programs, available at* `www.nbi.dk/ChaosBook/extras`.

Appendix P

Projects

You are urged to try to work through the essential steps in a project that combines the techniques learned in the course with some application of interest to you for other reasons. It is OK to share computer programs and such, but otherwise each project should be distinct, not a group project. The essential steps are:

- **Dynamics**
  1. construct a symbolic dynamics
  2. count prime cycles
  3. prune inadmissible itineraries, construct Markov graphs if appropriate
  4. implement a numerical simulator for your problem
  5. compute a set of the shortest periodic orbits
  6. compute cycle stabilities
- **Averaging, numerical**
  1. estimate by numerical simulation some observable quantity, like the escape rate,
  2. or check the flow conservation, compute something like the Lyapunov exponent
- **Averaging, periodic orbits**
  1. implement the appropriate cycle expansions
  2. check flow conservation as function of cycle length truncation, if the system is closed

3. implement desymmetrization, factorization of zeta functions, if dynamics possesses a discrete symmetry
4. compute a quantity like the escape rate as a leading zero of a spectral determinant or a dynamical zeta function.
5. or evaluate a sequence of truncated cycle expansions for averages, such as the Lyapunov exponent or/and diffusion coefficients
6. compute a physically intersting quantity, such as the conductance
7. compute some number of the classical and/or quantum eigenvalues, if appropriate

P.1 Deterministic diffusion, zig-zag map

To illustrate the main idea of chapter 20, tracking of a globally diffusing orbit by the associated confined orbit restricted to the fundamental cell, we consider a class of simple 1-d dynamical systems, chains of piecewise linear maps, where all transport coefficients can be evaluated analytically. The translational symmetry (20.10) relates the unbounded dynamics on the real line to the dynamics restricted to a "fundamental cell" - in the present example the unit interval curled up into a circle. An example of such map is the sawtooth map

$$\hat{f}(x) = \begin{cases} \Lambda x & x \in [0, 1/4 + 1/4\Lambda] \\ -\Lambda x + (\Lambda+1)/2 & x \in [1/4 + 1/4\Lambda, 3/4 - 1/4\Lambda] \\ \Lambda x + (1-\Lambda) & x \in [3/4 - 1/4\Lambda, 1] \end{cases} . \tag{P.1}$$

The corresponding circle map $f(x)$ is obtained by modulo the integer part. The elementary cell map $f(x)$ is sketched in fig. P.1. The map has the symmetry property

$$\hat{f}(\hat{x}) \;=\; -\hat{f}(-\hat{x})\,, \tag{P.2}$$

so that the dynamics has no drift, and all odd derivatives of the generating function (20.3) with respect to β evaluated at $\beta = 0$ vanish.

The cycle weights are given by

$$t_p = z^{n_p} \frac{e^{\beta \hat{n}_p}}{|\Lambda_p|}\,. \tag{P.3}$$

The diffusion constant formula for 1-d maps is

$$D = \frac{1}{2} \frac{\langle \hat{n}^2 \rangle_\zeta}{\langle n \rangle_\zeta} \tag{P.4}$$

where the "mean cycle time" is given by

$$\langle n \rangle_\zeta \;=\; z \frac{\partial}{\partial z} \frac{1}{\zeta(0,z)}\bigg|_{z=1} = -\sum{}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{P.5}$$

the mean cycle displacement squared by

$$\langle \hat{n}^2 \rangle_\zeta \;=\; \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta,1)}\bigg|_{\beta=0} = -\sum{}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{P.6}$$

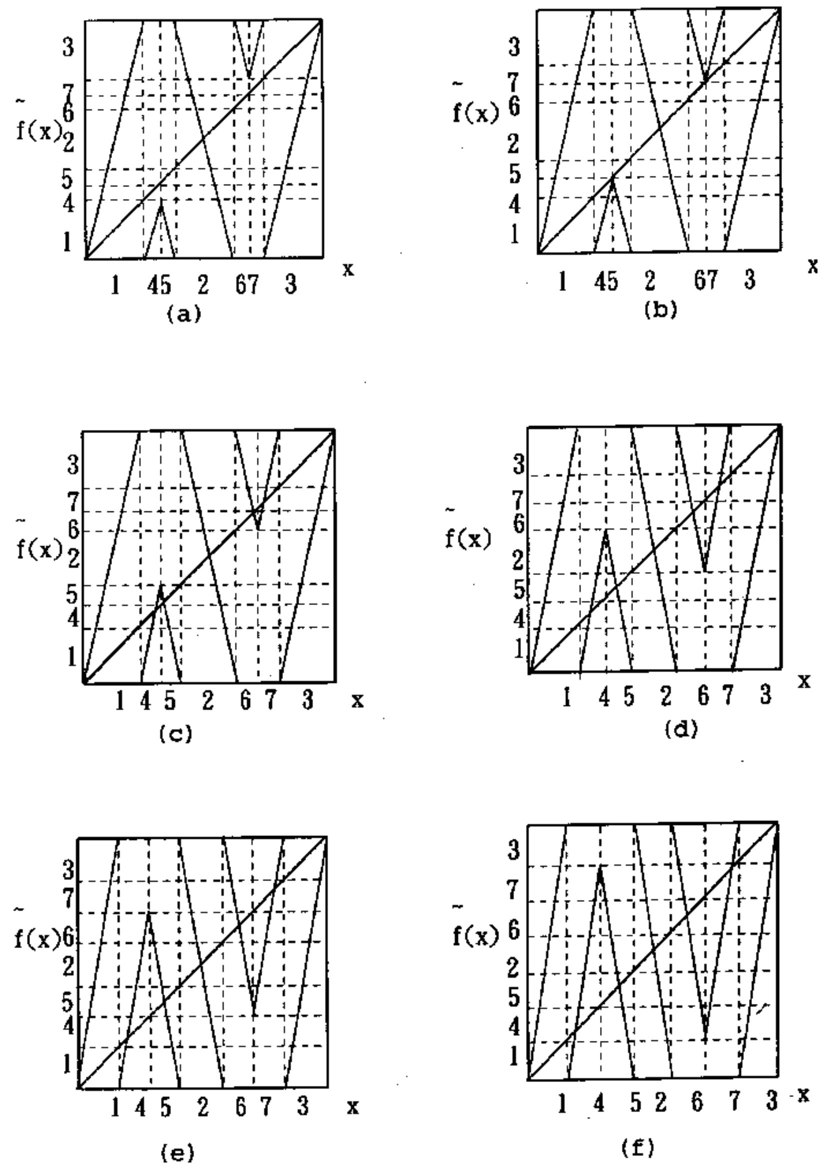


Figure P.1: (a)-(f) The sawtooth map (P.1) for the 6 values of parameter a for which the folding point of the map aligns with the endpoint of one of the 7 intervals and yields a finite Markov partition (from ref. [P.1]). The corresponding Markov graphs are given in fig. P.2.

and the sum is over all distinct non-repeating combinations of prime cycles. Most of results expected in this projects require no more than pencil and paper computations.

Implementing the symmetry factorization (20.35) is convenient, but not essential for this project, so if you find sect. 19.1.2 too long a read, skip the symmetrization.

P.1.1 The full shift

Take the map (P.1) and extend it to the real line. As in example of fig. 20.3, denote by a the critical value of the map (the maximum height in the unit cell)

$$a = \hat{f}(\frac{1}{4} + \frac{1}{4\Lambda}) = \frac{\Lambda + 1}{4} \,. \tag{P.7}$$

Describe the symbolic dynamics that you obtain when a is an integer, and derive the formula for the diffusion constant:

$$D = \frac{(\Lambda^2 - 1)(\Lambda - 3)}{96\Lambda} \qquad \text{for } \Lambda = 4a - 1,\ a \in \mathbb{Z}\,. \tag{P.8}$$

If you are going strong, derive also the fromula for the half-integer $a = (2k+1)/2$, $\Lambda = 4a+1$ case and email it to DasBuch@nbi.dk. You will need to partition $\mathcal{M}_2$ into the left and right half, $\mathcal{M}_2 = \mathcal{M}_8 \cup \mathcal{M}_9$, as in the derivation of (20.21).

20.1
page 413

P.1.2 Subshifts of finite type

We now work out an example when the partition is Markov, although the slope is not an integer number. The key step is that of having a partition where intervals are mapped *onto* unions of intervals. Consider for example the case in which $\Lambda = 4a - 1$, where $1 \leq a \leq 2$. A first partition is constructed from seven intervals, which we label $\{\mathcal{M}_1, \mathcal{M}_4, \mathcal{M}_5, \mathcal{M}_2, \mathcal{M}_6, \mathcal{M}_7, \mathcal{M}_3\}$, with the alphabet ordered as the intervals are laid out along the unit interval. In general the critical value a will not correspond to an interval border, but now we choose a such that the critical point is mapped onto the right border of $\mathcal{M}_1$, as in fig. P.1(a). The critical value of $f()$ is $f(\frac{\Lambda+1}{4\Lambda}) = a - 1 = (\Lambda - 3)/4$. Equating this with the right border of $\mathcal{M}_1$, $x = 1/\Lambda$, we obtain a quadratic equation with the expanding solution

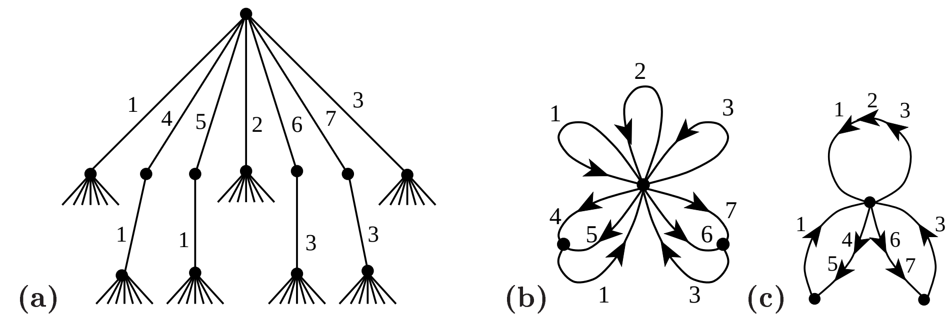


Figure P.2: (a) The sawtooth map (P.1) partition tree for fig. P.1(a); while intervals $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ map onto the whole unit interval, $f(\mathcal{M}_1) = f(\mathcal{M}_2) = f(\mathcal{M}_3) = \mathcal{M}$, intervals $\mathcal{M}_4, \mathcal{M}_5$ map onto $\mathcal{M}_1$ only, $f(\mathcal{M}_4) = f(\mathcal{M}_5) = \mathcal{M}_1$, and similarly for intervals $\mathcal{M}_6, \mathcal{M}_7$. An initial point starting out in the interval $\mathcal{M}_1$, $\mathcal{M}_2$ or $\mathcal{M}_3$ can land anywhere on the unit interval, so the subtrees originating from the corresponding nodes on the partition three are similar to the whole tree and can be identified (as, for example, in fig. 9.12), yielding (b) the Markov graph for the Markov partition of fig. P.1(a). (c) the Markov graph in the compact notation of (20.26).

$\Lambda = 4$. We have that $f(\mathcal{M}_4) = f(\mathcal{M}_5) = \mathcal{M}_1$, so the transition matrix (9.2) is given by

$$\phi' = T\phi \; = \; \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_4 \\ \phi_5 \\ \phi_2 \\ \phi_6 \\ \phi_7 \\ \phi_3 \end{pmatrix} \tag{P.9}$$

and the dynamics is unrestricted in the alphabet

$$\{1, \underline{41}, \underline{51}, 2, \underline{63}, \underline{73}, 3, \} \, .$$

One could diagonalize (P.9) on the computer, but, as we saw in sect. 9.6, the Markov graph fig. P.2(b) corresponding to fig. P.1(a) offers more insight into the dynamics. The dynamical zeta function

$$\begin{aligned} 1/\zeta &= 1 - (t_1 + t_2 + t_3) - 2(t_{14} + t_{37}) \\ 1/\zeta &= 1 - 3\frac{z}{\Lambda} - 4\cosh\beta\frac{z^2}{\Lambda^2} \, . \end{aligned} \tag{P.10}$$

follows from the loop expansion (10.13) of sect. 10.3.

The material flow conservation sect. 16.2 and the symmetry factorization (20.35) yield

$$0 = \frac{1}{\zeta(0,1)} = \left(1 + \frac{1}{\Lambda}\right)\left(1 - \frac{4}{\Lambda}\right)$$

which indeed is satisfied by the given value of Λ. Conversely, we can use the desired Markov partition topology to write down the corresponding dynamical zeta function, and use the $1/\zeta(0,1) = 0$ condition to fix Λ. For more complicated transition matrices the factorization (20.35) is very helpful in reducing the order of the polynomial condition that fixes Λ.

The diffusion constant follows from (20.36) and (P.4)

$$\langle n \rangle_\zeta = -\left(1 + \frac{1}{\Lambda}\right)\left(-\frac{4}{\Lambda}\right), \quad \langle \hat{n}^2 \rangle_\zeta = \frac{4}{\Lambda^2}$$

$$D = \frac{1}{2}\frac{1}{\Lambda + 1} = \frac{1}{10}$$

Think up other non-integer values of the parameter for which the symbolic dynamics is given in terms of Markov partitions: in particular consider the cases illustrated in fig. P.1 and determine for what value of the parameter a each of them is realized. Work out the Markov graph, symmetrization factorization and the diffusion constant, and check the material flow conservation for each case. Derive the diffusion constants listed in table P.1. It is not clear why the final answers tend to be so simple. Numerically, the case of fig. P.1(c) appears to yield the maximal diffusion constant. Does it? Is there an argument that it should be so?

The seven cases considered here (see table P.1, fig. P.1 and (P.8)) are the 7 simplest complete Markov partitions, the criterion being that the critical points map onto partition boundary points. This is, for example, what happens for unimodal tent map; if the critical point is preperiodic to an unstable cycle, the grammar is complete. The simplest example is the case in which the tent map critical point is preperiodic to a unimodal map 3-cycle, in which case the grammar is of golden mean type, with $\_00\_$ substring prohibited (see fig. 9.12). In case at hand, the "critical" point is the junction of branches 4 and 5 (symmetry automatically takes care of the other critical point, at the junction of branches 6 and 7), and for the cases considered the critical point maps into the endpoint of each of the seven branches.

One can fill out parameter a axis arbitrarily densely with such points - each of the 7 primary intervals can be subdivided into 7 intervals obtained

fig. P.1	Λ	D
	3	0
(a)	4	$\frac{1}{10}$
(b)	$\sqrt{5}+2$	$\frac{1}{2\sqrt{5}}$
(c)	$\frac{1}{2}(\sqrt{17}+5)$	$\frac{2}{\sqrt{17}}$
(c')	5	$\frac{2}{5}$
(d)	$\frac{1}{2}(\sqrt{33}+5)$	$\frac{1}{8}+\frac{5}{88}\sqrt{33}$
(e)	$2\sqrt{2}+3$	$\frac{1}{2\sqrt{2}}$
(f)	$\frac{1}{2}(\sqrt{33}+7)$	$\frac{1}{4}+\frac{1}{4\sqrt{33}}$
	7	$\frac{2}{7}$

Table P.1: The diffusion constant as function of the slope Λ for the $a = 1, 2$ values of (P.8) and the 6 Markov partitions of fig. P.1

by 2-nd iterate of the map, and for the critical point mapping into any of those in 2 steps the grammar (and the corresponding cycle expansion) is finite, and so on.

P.1.3 Diffusion coefficient, numerically

(optional:)
Attempt a numerical evaluation of

$$D = \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} \left\langle \hat{x}_n^2 \right\rangle . \tag{P.11}$$

Study the convergence by comparing your numerical results to the exact answers derived above. Is it better to use few initial $\hat{x}$ and average for long times, or to use many initial $\hat{x}$ for shorter times? Or should one fit the distribution of $\hat{x}^2$ with a gaussian and get the D this way? Try to plot dependence of D on Λ; perhaps blow up a small region to show that the dependance of D on the parameter Λ is fractal. Compare with fig. 20.5 and figures in refs. [P.1, P.2, 20.7, 20.8].

P.1.4 D is a nonuniform function of the parameters

(optional:)
The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. An interpretation of this lack of monotonicity would be interesting.

You can also try applying periodic orbit theory to the sawtooth map (P.1) for a random "generic" value of the parameter Λ, for example $\Lambda = 6$. The idea is to bracket this value of Λ by the nearby ones, for which higher and higher iterates of the critical value $a = (\Lambda + 1)/4$ fall onto the partition boundaries, compute the exact diffusion constant for each such approximate Markov partition, and study their convergence toward the value of D for $\Lambda = 6$. Judging how difficult such problem is already for a tent map (see sect. 10.6 and appendix E.1), this is too ambitious for a week-long exam.

References

[P.1] H.-C. Tseng, H.-J. Chen, P.-C. Li, W.-Y. Lai, C.-H. Chou and H.-W. Chen, "Some exact results for the diffusion coefficients of maps with pruned cycles", *Phys. Lett.* **A** 195, 74 (1994).

[P.2] C.-C. Chen, "Diffusion Coefficient of Piecewise Linear Maps", *Phys. Rev.* **E51**, 2815 (1995).

[P.3] H.-C. Tseng and H.-J. Chen, "Analytic results for the diffusion coefficient of a piecewise linear map", *Int. J. Mod. Phys.***B 10**, 1913 (1996).

P.2 Deterministic diffusion, sawtooth map

To illustrate the main idea of chapter 20, tracking of a globally diffusing orbit by the associated confined orbit restricted to the fundamental cell, we consider in more detail the class of simple 1-d dynamical systems, chains of piecewise linear maps (20.9). The translational symmetry (20.10) relates the unbounded dynamics on the real line to the dynamics restricted to a "fundamental cell" - in the present example the unit interval curled up into a circle. The corresponding circle map $f(x)$ is obtained by modulo the integer part. The elementary cell map $f(x)$ is sketched in fig. 20.3. The map has the symmetry property

$$\hat{f}(\hat{x}) \,=\, -\hat{f}(-\hat{x})\,, \tag{P.12}$$

so that the dynamics has no drift, and all odd derivatives of the generating function (20.3) with respect to β evaluated at $\beta = 0$ vanish.

The cycle weights are given by

$$t_p = z^{n_p} \frac{e^{\beta \hat{n}_p}}{|\Lambda_p|}\,. \tag{P.13}$$

The diffusion constant formula for 1-d maps is

$$D = \frac{1}{2} \frac{\left\langle \hat{n}^2 \right\rangle_\zeta}{\left\langle n \right\rangle_\zeta} \tag{P.14}$$

where the "mean cycle time" is given by

$$\left\langle n \right\rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0,z)} \bigg|_{z=1} = -\sum{}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{P.15}$$

the mean cycle displacement squared by

$$\left\langle \hat{n}^2 \right\rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta,1)} \bigg|_{\beta=0} = -\sum{}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|}\,, \tag{P.16}$$

and the sum is over all distinct non-repeating combinations of prime cycles. Most of results expected in this projects require no more than pencil and paper computations.

fig. 20.4	Λ	D
	4	$\frac{1}{4}$
(a)	$2+\sqrt{6}$	$1-\frac{3}{4}\sqrt{6}$
(b)	$2\sqrt{2}+2$	$\frac{15+2\sqrt{2}}{16+4\sqrt{2}}$
(c)	5	1
(d)	$3+\sqrt{5}$	$\frac{5}{2}\frac{\Lambda-1}{3\Lambda-4}$
(e)	$3+\sqrt{7}$	$\frac{5\Lambda-4}{3\Lambda-2}$
	6	$\frac{5}{6}$

Table P.2: The diffusion constant as function of the slope Λ for the $\Lambda = 4, 6$ values of (20.20) and the 5 Markov partitions like the one indicated in fig. 20.4.

P.2.1 The full shift

Reproduce the formulas of sect. 20.2.1 for the diffusion constant D for Λ both even and odd integer.

P.2.2 Subshifts of finite type

We now work out examples when the partition is Markov, although the slope is not an integer number. The key step is that of having a partition where intervals are mapped *onto* unions of intervals.

Start by reproducing the formula (20.28) of sect. 20.2.3 for the diffusion constant D for the Markov partition, the case where the critical point is mapped onto the right border of I_{1_+}.

Think up other non-integer values of the parameter Λ for which the symbolic dynamics is given in terms of Markov partitions: in particular consider the remaing four cases for which the critical point is mapped onto a border of a partion in one iteration. Work out the Markov graph symmetrization factorization and the diffusion constant, and check the material flow conservation for each case. Fill in the diffusion constants missing in table P.2. It is not clear why the final answers tend to be so simple. What value of Λ appears to yield the maximal diffusion constant?

The 7 cases considered here (see table P.2 and fig. 20.4) are the 7 simplest complete Markov partitions in the $4 \le \Lambda \le 6$ interval, the criterion being that the critical points map onto partition boundary points. In case at hand, the "critical" point is the highest point of the left branch of the map (symmetry automatically takes care of the other critical point, the lowest point of the left branch), and for the cases considered the critical point maps into the endpoint of each of the seven branches.

One can fill out parameter a axis arbitrarily densely with such points - each of the 6 primary intervals can be subdivided into 6 intervals obtained by 2-nd iterate of the map, and for the critical point mapping into any of those in 2 steps the grammar (and the corresponding cycle expansion) is finite, and so on.

P.2.3 Diffusion coefficient, numerically

(optional:)
Attempt a numerical evaluation of

$$D = \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} \left\langle \hat{x}_n^2 \right\rangle . \tag{P.17}$$

Study the convergence by comparing your numerical results to the exact answers derived above. Is it better to use few initial $\hat{x}$ and average for long times, or to use many initial $\hat{x}$ for shorter times? Or should one fit the distribution of $\hat{x}^2$ with a gaussian and get the D this way? Try to plot dependence of D on Λ; perhaps blow up a small region to show that the dependance of D on the parameter Λ is fractal. Compare with fig. 20.5 and figures in refs. [P.1, P.2, 20.7, 20.8].

P.2.4 D is a nonuniform function of the parameters

(optional:)
The dependence of D on the map parameter Λ is rather unexpected - even though for larger Λ more points are mapped outside the unit cell in one iteration, the diffusion constant does not necessarily grow. Fig. 20.5 taken from ref. [20.7] illustrates the fractal dependence of diffusion constant on the map parameter. An interpretation of this lack of monotonicity would be interesting.

You can also try applying periodic orbit theory to the sawtooth map (20.9) for a random "generic" value of the parameter Λ, for example $\Lambda = 4.5$. The idea is to bracket this value of Λ by the nearby ones, for which higher and higher iterates of the critical value $a = \Lambda/2$ fall onto the partition boundaries, compute the exact diffusion constant for each such approximate Markov partition, and study their convergence toward the value of D for $\Lambda = 4.5$. Judging how difficult such problem is already for a tent map (see sect. 10.6 and appendix E.1), this is too ambitious for a week-long exam.