Table of Lectures

Rudiments of dynamical systems

What should we learn?

'Nonlinear systems' usually discusses nonlinear dynamics and, these days, the so-called complex systems, because complex systems are purportedly closely related to chaos, fractal, etc.[1] In my perspective course I confine myself to the discussion of time-evolution of nonlinear systems and to that of (sometimes) complicated outcomes of seemingly rather simple systems. That is, my scope will cover a subset of the topics discussed in the standard theory of dynamical systems[2] and some needed rudiments of the theory of computation.

Since the theory of dynamical systems is a highly developed branch of mathematics, its precise understanding demands rather sophisticated mathematics. However, the 'perspective' will try to make theories as intuitively accessible as possible (with illustrations), although I will not avoid precise mathematical statements; physicists must learn how to read math intuitively. Thus, I will try to go back and forth between rigorous math and intuitive arguments; if intuitive understanding strongly suggests a certain statements very natural, I will not give a rigorous argument.

A dynamical system is a deterministic system (i.e., its future is uniquely determined by the information about its past), whose significance was first clearly recognized by Newton. In many cases such systems are described by (ordinary) differential equations[3] or by time-evolution maps (that give the next time slice from the current one).[4] Such systems are called dynamical systems. We are especially interested in the global (often this means 'long-time') qualitative features of dynamical systems.[5] Most dynamical systems are nonlinear, so we are interested in the long-time quali-

[1] As Prigogine totally screwed up: Nicolis and Prigogine's *Exploration of Complexity* (W. H. Freeman and Co., New York, 1989) wrote on p8: Since the 1960s a revolution in both mathematical and physical sciences has imposed a new attitude in the description of nature. Parallel development in the thermodynamic theory of irreversible phenomena, in the theory of dynamical systems, and in classical mechanics have converged to show in a compelling way that the gap between 'simple' and 'complex,' between 'disorder' and 'order,' is much narrower than previously thought." Immediately after this is a mention of chaos as an example of complex behavior. As we will learn we cannot be this naive. We must reflect on what 'complexity' really mean.

[2] Cf. *Dynamical Systems* I-X of *Encyclopaedia of Mathematical Sciences* (Springer).

[3] We will discuss vector fields and flows on manifolds, unique existence of the solution to the initial value problem/rectification theorem for Lipschitz continuous ODE.

[4] We will discuss diffeomorphisms, the relation between time continuous and discrete dynamical systems (i.e., Poincaré map and suspension).

[5] We will see the classification of low dimensional systems, ω-limit sets, limit cycles, the Poincaré-Bendixson theorem, etc.

tative behavior of nonlinear systems (general ODEs): what is the fate of the system after a long time? Will there be eventually no time dependence, or periodic oscillations, etc.?[6]

Since we are interested in naturally occurring dynamical systems, their features must be reproducibly observable. This requires the stability argument of the systems themselves and that of their behaviors.[7],[8]

Incidentally, one might ask what such studies are for when we can numerically solve differential equations efficiently these days. I must point out that still long-time study of nonlinear dynamics is numerically hard.[9] Furthermore, it is hard to conclude something general for a class of systems numerically. We must not forget a dictum: What we can show only with computers is not general; what is general should be demonstrable without computers.

The study of qualitative behaviors makes it clear that there are many stable but complicated behaviors called chaos often supported by strange attractors. To characterize such behaviors we must have a clear understanding of randomness. Therefore, before going to such topics requiring some tools that may not be familiar to physicists, we go to a special branch of dynamical systems: classical mechanical systems or the Hamiltonian systems which were the original target of the qualitative study: is a planetary motion predictable or a star system bounded forever?[10] I will outline a history of celestial mechanics culminating in Poincaré's topological study.[11] The study of integrability of mechanics leads to the study of completely integrable systems and eventually to the theory of solitons.[12] I will discuss solitons only very briefly, and then will go to the perturbation of integrable systems. The original question was: suppose we have a completely integrable system, then what happens if we perturb it? Poincaré realized that the outcome is complicated 'beyond imagination.' Then, Kolmogorov realized that still most periodic motions stay periodic (the KAM

[6] Weak nonlinear perturbations (the so-called singular perturbations) will be discussed as an application of RG. Chiba's theorem will also be discussed.

[7] Andronov-Pontryagin's structural stability, hyperbolicity, etc., will be discussed.

[8] The stability of solutions will be discussed. Thus, the Grobman-Hartman theorem will be discussed to justify linear stability analysis.

[9] A general discussion of numerical analysis may be given. When the size of the time increment is not sufficiently small, the numerical results can be qualitatively different from the true behavior of the ODE.

[10] Three-body problems, Bruns' theorem. etc., will be outlined historically.

[11] This will review the canonical transformation and Hamilton-Jacobi theory, classical solvability condition, etc. This culminates in Arnold's theorem about integrability of Hamiltonian systems.

[12] The Lax pair and complete integrability will be discussed. Kortweg-de Vries equation will be looked at briefly.

theorem). The destroyed periodic orbits, however, behave, as Poincaré expected, chaotically, leading to almost random motion of planets (Arnold diffusion). These topics are also related to the almost periodic quantum systems.

Thus, we have encountered deterministic but very complicate motion of dynamical systems. How can we characterize complicate behavior? How can we quantify the extent of complicatedness? I believe this is most unambiguously answered with the theory of computation and computational complexity.

After introducing classic examples of complicate behaviors (e.g., Lorenz system and interval endomorphisms)[13] I will embark on characterization of chaotic behaviors, starting with the discussion: what is computation? The universal Turing machine and Church's thesis is introduced and then Kolmogorov complexity will be discussed. With this preparation we can unambiguously discuss what chaos should be.[14]

Incidentally, there is a deep question: even if an ODE has a unique solution, if we cannot compute it, what is its significance in physics? Such questions will also be discussed briefly.[15]

Now, we have all the machineries to discuss general dynamical systems. The general question is: what is the behavior of typical dynamical systems? I wish to discuss the Palis conjecture trying to answer the question once and for all. I will outline how the theory of dynamical systems had culminated in this conjecture, following the theory of hyperbolic dynamical systems[16]

[13] This portion will not only introduce popular chaotic systems, but also introduce mathematical concepts and tools such as the Kolmogorov-Sinai entropy, Lyapunov exponents, etc. and related topics.

[14] The argument will introduce the symbolic dynamical system (Markov partition, ε-tracing property), endomorphism of an interval ("Period $\neq 2^n$ implies chaos", Li-Yorke's theorem, Sharkovski's theorem), Kolmogorov-Sinai entropy and its relation to Kolmogorov complexity (Brudno's theorem), Artin-Mazur ζ-function and thermodynamic formalism. We need the Shannon-McMillan-Breiman theorem.

[15] This will cover the theory of computable numbers and functions. There are functions that are differentiable but their derivatives cannot be evaluated, because they are not computable. See M. B. Pour-El and J. I. Richards, *Computability in analysis and physics* (Springer,1989).

[16] This will cover open-density, Baire properties (residual or generic features). Then, Peixoto's theorem, Kupka-Smale and Morse-Smale systems will be discussed as standard topics, but these will come after all the rudiments such as the Sinai-Ruelle-Bowen measure is introduced.

1 Introductory overview

1.1 What is a dynamical system?

If the states of a system[17] change in time[18] according to a definite rule, the system is called a dynamical system.

'Dynamics' here follows a definite rule; that is, the state at present (or more generally, up to present) determines the future states.[19] Actually, we discuss only two types of dynamical systems:

A discrete-time dynamical system is defined by a map f from a set M into itself (= endomorphism). For an 'initial condition $x_0 \in M$[20] the state x_n at time step $n \in$ (say) $\mathbb{N}$[21] is defined by $f^n x_0 \equiv f(f(\cdots f(x_0)\cdots))$ (there are n f's). $M = [0,1]$, $f = 4x(1-x)$ (the *logistic map*) is a typical example (Fig. 1.1 Left).

A continuous dynamical system is defined by a vector field X on M, and the rule is given by a differential equation[22] (Example Fig. 1.1 Right)

$$\dot{x}(t) = X(x(t)). \tag{1.1}$$

1.2 Topological and measure-theoretical dynamical systems

[17] ⟪**System**⟫ A 'system' is a part of the universe (or nature) which is more or less with a sort of integrity (for example, an electronic circuit, a dog, the Earth, the solar system, etc.). The word 'system' consists of 'syn' (together) and 'histanai' (stand) in Greek. Thus, its original meaning is "a set of objects supporting each other to make an entity." Therefore, ideally,

(1) we should be able to distinguish things belonging to the system and those not, and

(2) things belonging to the system interact (have relations) with each other through system-specific interactions; their interactions with the things outside the system may be treated separately.

[18] ⟪**Time**⟫ What is 'time'? Although I wish you to think this philosophico-physics question at least on and off, in these lectures, we understand 'time' as the one we understand in common sense: it passes homogeneously and has a direction. Carlo Rovelli, *The order of time* (Riverhead Books 2018) is perhaps the best book so far written by a physicist on the topic, although I do not share some key opinions with the author.

[19] The rule usually gives a deterministic result (unique future), but whether a set of rules allows determinism (= unique future from the data at and/or before present) or not must be checked carefully and may depend on contexts.

[20] x_0 is usually a multidimensional vector, but I will not use the vector notation such as $\boldsymbol{x}_0$.

[21] $\mathbb{N} = \{0, 1, 2, \cdots\}$, nonnegative integers. Other standard number set notations will be used freely: $\mathbb{R}$: real numbers; $\mathbb{Q}$: rational numbers; $\mathbb{Z}$: integers; $\mathbb{C}$: complex numbers.

[22] We will discuss whether this can consistently define a time evolution rule later. This is the fundamental problem of ODE.

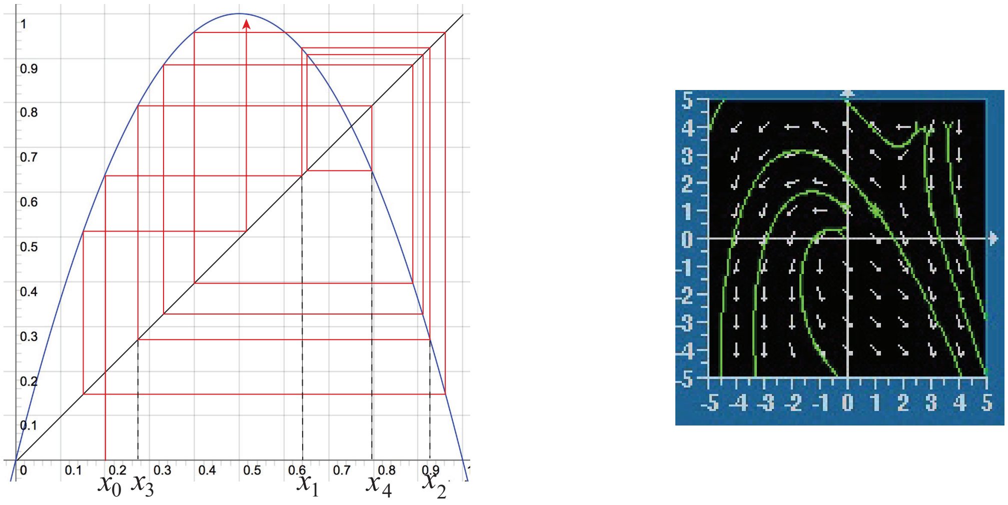


Figure 1.1: Left: The discrete dynamics defined by the logistic map, and how to follow the trajectory $\{x_i\}$. Right: Vector field and its flows for $\dot{x} = x - y$, $\dot{y} = y - x^2$.

Dynamical systems are often classified by their main features we pay attention to, geometrical features of their trajectories (topological dynamical systems) or probabilisitic features (measure-theoretical dynamical systems). When we will discuss the theory of dynamical systems in earnest, we discuss these aspects separately.

1.3 Linear vs nonlinear dynamical systems

The title of the course is 'nonlinear ...,' so we must discuss what 'nonlinear' means.

A system is called a linear system if its observable y depends on the 'control' variable x linearly: the map $y = f(x)$ is linear. f is linear if

$$f(ax_1 + bx_2) = af(x_1) + bf(x_2). \tag{1.2}$$

That is, f is linear, if it satisfies linear scaling $f(ax) = af(x)$ and additivity: $f(x_1 + x_2) = f(x_1) + f(x_2)$. (1.2) is called the superposition principle. If f is not linear, we say the system is nonlinear.[23]

The reader must have heard that nothing strange occurs for linear dynamical systems, but, in contrast, we may encounter strange complicate time evolutions for nonlinear systems. This is largely true, but we will learn that 'strange behaviors (called 'chaos') are caused, in essence, due to (local) 'expansion' in the 'bounded'

[23] A more careful definition may be found in Chapter 1 of YO *The nonlinear world* (Springer, 2012) [henceforth abbreviated as 'TNW'.

phase space. Nonlinearity is required to bound the domain where the states live, if the domain is not naturally bounded (say, $\mathbb{R}^n$). If the domain is intrinsically bounded like an n-torus T^n, chaotic behavior does not require nonlinearity.

The main reason why we are interested in nonlinear dynamical systems is that their behaviors are often unexpected and complicate (e.g., chaos = deterministic unpredictability). However, as we learn, complication is due to local expansion (small errors are magnified) with 'confinement'; if not spatially confined, we would find explosion, which is not usually complicate.

1.4 Nonlinearity means scale interference[24]

Still it is true that nonlinearity is often a key to nontrivial dynamical behavior, because many systems are not geometrically confined. A nonlinear system is a system that is not a linear system as discussed in **1.3**. We know most phenomena in the world are nonlinear,[25] and to characterize this important 'nonlinearity' by the negation of 'linearity' is quite unsatisfactory. We should characterize 'nonlinear systems' more positively rather than through negation of something else.

Thanks to the superposition principle, even if we superpose high-frequency perturbation to a linear system, its behavior (esp., time-coarse-grained behavior) is not affected. However, for a nonlinear system such a guarantee does not exist, because different scales (length and/or time scales) interfere. Typical nonlinear phenomena are often due to such interference. For example, critical phenomena in phase transition are due to the nonlinear coupling of small scale fluctuations producing mesoscale or even macroscale fluctuations. Thus, we may characterize nonlinear systems as systems having scale interferences. We will see that chaos clearly teaches us that the very microscopic scale of the Universe affects our scale sometimes significantly.

1.5 Unknowable affects our lives due to nonlinearity

Since what happens on extreme small scales are never knowable/observable, even if we totally ignore quantum mechanics, our world even on the scale we can directly observe is inevitably riddled with the absolute unknowable.

Perhaps the geometrical study of dynamical systems try to understand what sort of mechansims cause this, and the probabilistic or statistical study of dynamical systems try to understand what is certain despite the unknowable. Thus, statistical and

[24] see Chapter 1 of 'TNW'.

[25] The word 'nonlinear system' is a shorthand for the fact that the aspect of the system we are interested in does not have a description that admits the superposition principle.

topological qualitative studies of dynamical systems are the main goal of the theory of dynamical systems.

1.6 What features should we pay attention to?

As noted in **1.1** we pay attention to topological and probabilistic features of dynamical systems. Although we will discuss some details of concrete examples, we are much more interested in generic (universal) or typical qualitative features of dynamical systems.

Being 'typical' may mean what we can sample from a class of dynamical systems 'casually' without any premeditation. How can we 'formalize' this? This is not so simple, so we will discuss this in a separate section with some mathematical rudiments.

1.7 Where do we encounter dynamical systems?

This may be a rather stupid question, since differential equations appear almost everywhere in physics. Thus, we physicists are very familiar with dynamical systems, or are we? Many elementary examples are linear and analytically simply solvable. However, the study of dynamical systems was initiated by Poincaré to understand unsolvable equations of motion (the three body problem). Lurking in benign-looking dynamical systems are bewilderingly complicate motions. Thus it is not surprising that simply looking systems with complicate unpredictable behaviors appear in, e.g., stellar systems, particle accelerators, plasma confinement, chemical reaction kinetics, population dynamics. We usually do not pay particular attention to such complications in physical systems.

Apart from such relevance to physics, however, it should be simply interesting to know what is possible in apparently not so complicate systems. Thus, application of theory of dynamical systems is not the main concern of the course.[26]

Some old examples are in Figs. 1.2 and 1.4.

[26] "Unlearn this "for," you creators; your virtue itself wants that you do nothing "for" and "in order" and "because." You should plug your ears against these false little words." (F. Nietzsche, *Thus spoke Zarathustra*, On the higher man 11) [Cambridge Texts in the History of Philosophy, edited by A Del Caro and R Pippin].

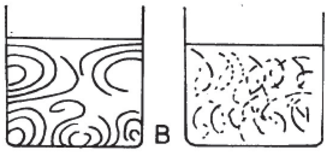


Fig. 5. Sketches of representative patterns. a: ordered phase, b: turbulent phase. In the turbulent phase much more diffuse patterns were often observed.

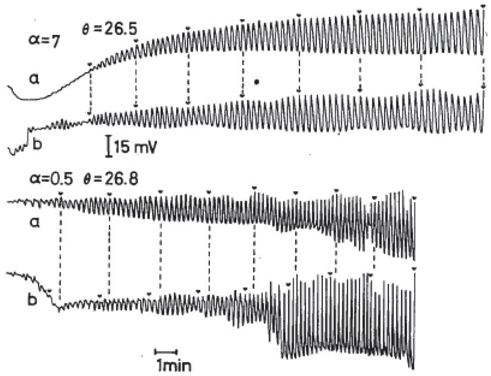


Fig. 6. Comparison of two emf signals observed by two small electrodes separated by a 0.2 mm gap.

Figure 1.2: 'Chemical turbulence' [Fig. 5,6 of Yamazaki et al. JPSJ 46 722 (1978)]

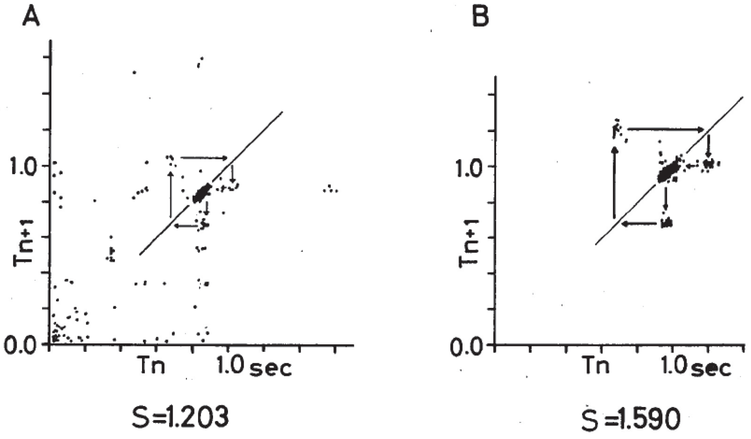


Fig. 8. The interval Lorenz plot for the pulses of a patient suffering from arrythmia. The pattern

Figure 1.3: Arhythmia Lorenz plot [Fig. 5,6 of YO et al JPSJ 48 733 (1980)]

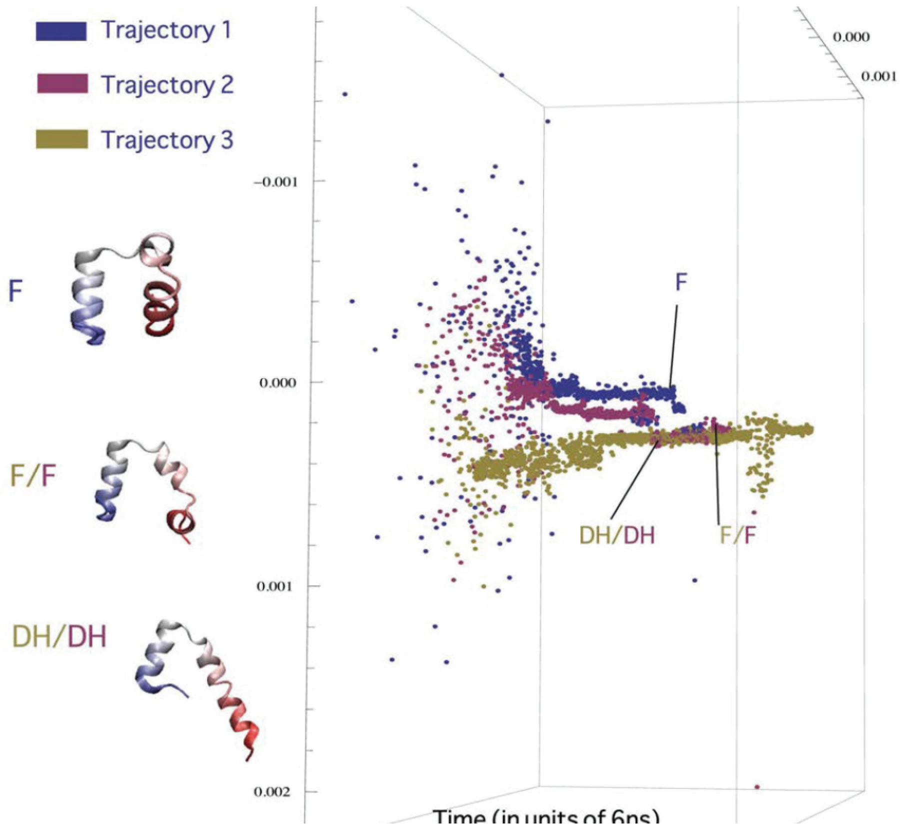


Figure 1.4: Villin head piece full MD[A Rajan Thesis UIUC (2009)]

1.8 Outline of the course

An outline of the course is given. Here, no explanation of technical terms (in italic) will be given (for example, manifold, compact, C^r-vector field, etc.). I know physicists (in the US) hate math,[27] but I wish you to have intuitive grasp of core math concepts.

Basically, after warmup with or without Hamiltonians, we will understand that 'chaos' is algorithmically random behaviors of dynamical systems, whose statistical behaviors we can understand 'statistical mechanically.'

Warmup with ODE

We will begin with a geometrical study of ODE

$$\dot{x} = X(x), \tag{1.3}$$

where $x \in M$, $X \in \mathcal{X}^r(M)$. Here, M is a *n-manifold* (usually *compact*), and $\mathcal{X}^r(M)$ is a *C^r-vector field* on M. Typical discussion topics are:

$*$ When can we say $x(t)$ is unique, or determinism legitimate? X must be *Lipshitz continuous*: $\|X(x) - X(x')\| \leq L\|x - x'\|$. Under this condition if $X(x) \neq 0$, in an *neighborhood* of x, X is transformed to a constant field (say, $X = e_1$) [*Rectification theorem*].

$*$ What happens in the $t \to \infty$ limit? We will discuss *attractors*, *invariant manifolds*, etc.

$*$ If $X(x_f) = 0$, x_f is a fixed point (= *critical point* or *singular point*; some examples of flow around it are in Fig. 1.5). Is $x(t) = x_f$ stable against small perturbations? Can we study its stability from the linearized equation around x_f? We will discuss *hyperbolicity* and the fundamental theorem: *Hartman's theorem* that justifies the linear stability analysis.[28]

$*$ A qualitative change of the solution (say, from a stationary to oscillatory behavior) is called a bifurcation. What sort of bifurcations do we generically encounter? When a bifurcation occurs, can we compute the new qualitatively different solution by per-

[27] Boltzmann said, "What the brain is to man, mathematics is to science" (quoted in K. Sigmund, *Exact thinking in demented times: the Vienna Circle and the epic quest for the foundation of science* (Basic Books, 2017; original in 2015).

[28] Quiz: is the following argument legitimate?
Suppose $X(0) = 0$. Since we are interested in the situation very close to $x = 0$, let us linearize X around $x = 0$ as $\dot{x} = Ax$ (i.e., we compute $A = DX/Dx|_{x=0}$). It happens that all the eigenvalues of A has negative real part. Thus, $x = 0$ is a stable fixed point.

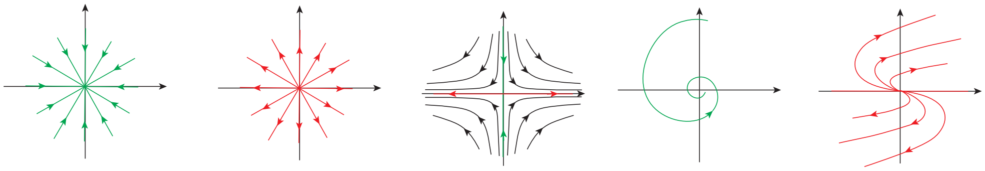


Figure 1.5: Some examples of flows around singular points on $\mathbb{R}^2$.

turbation? We will discuss a renormalization group theory for *singular perturbations* (= perturbations that qualitatively alter the nature of the system).
$*$ What kind of singular points can a system have on M? Topological constraints matter (the *Poincaré-Hopf theorem*); *degree theory* can tell you something.

Classical mechanics
Next, we discuss a special class of ODE: the Hamiltonian systems
$*$ After reviewing classical mechanics (extremely briefly) *canonical transformation* will be discussed (with related topics as *Lagrange* and *Poisson brackets*, canonical invariants).
$*$ Integrable cases will be discussed generally; *Liouville-Arnold's theorem* and *action-angle variables* will be discusssed.
$*$ Then we go to the cradle of the theory of dynamical systems: celestial mechanics.
$*$ (Bruns and) Poincaré realized perturbation has problems.
$*$ However Kolmogorov realized that still for many energies, perturbation series converges [*KAM theorem*]. A prototypical theorem due to Siegel will be proved following Kolmogorov's idea. When the series do not converge, 'chaos' show up as Poincaré realized. An illustration of what happens is in Fig. 1.6

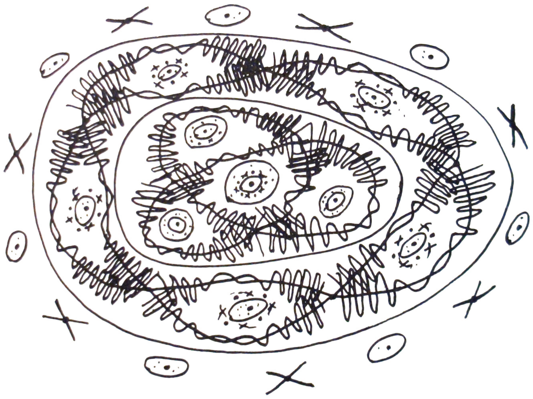


Figure 1.6: Many tori due to integrability are destroyed except for the KAM tori; the destroyed tori exhibit chaotic behavior such as the *Arnold diffusion.* [Fig. 6.7 of A E Jackson, *Perspective of nonlinear dynamics* Vol. 2 (Cambridge, 1990) p56]

$\ast$ This chaotic behavior will be discussed with the aid of the standard map (related to accelerator dynamics).

Up to this point is introductory; although mathematically delicate issues (or at least its delicate nature) should be appreciated (for Hartman, Siegel and Poincaré theorems), technical details will not be emphasized (details are already posted as supplements). However, the flow of the logic should be understood.

An exposition of the 'modern' theory of dynamical systems begins here. Physicists should realize that most of the main ingredients had been finished before 1980 when physicists, esp., in the US started to pay serious attention to the topic. Computers were in large part irrelevant (contrary to the claim of Gleick, whose book I cannot recommend) and many key ingredients were done in USSR.[29]

We must not forget a dictum: What we can show only with computers is not general; what is general should be demonstrable without computers (if we understand it). Needless to say, however, for exploratory work computers are great especially in the hand of those who understand mathematics.

Chaos gallery:
We begin with typical and historical examples to become familiar with the real issues of dynamical systems.
$\ast$ Weather forecasting and *Lorenz system*. The iconic figure must be familiar to most of you:
https://upload.wikimedia.org/wikipedia/commons/1/13/A_Trajectory_Through_Phase_Space_in_a_Lorenz_Attractor.gif
$\ast$ *Strange attractors* and how to observe them (*Takens' reconstruction theorem*).
$\ast$ Interval maps and Theorem: "Period $\neq 2^n$ implies chaos." I will follow the path I took more than 40 years ago.[30]
$\ast$ But what is chaos? We use coding of dynamical behaviors into symbol sequence (*shift dynamical systems*), and discuss its complexity with typical examples: *baker's*

[29] Y. G. Sinai, "Chaos Theory Yesterday, Today and Tomorrow," J. Stat. Phys. **138**, 2 (2010). "My personal experience shows that people in the West consider the development of Chaos Theory differently from their Russian colleagues, mathematicians and physicists." [This seems a great euphemism or sarcasm.] "Many people share the point of view that the beginning of chaos theory dates back to 1959 when the Kolmogorov's paper ["New metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces," Dokl. Acad. Nauk SSSR **119**, 861 (1958)] on the entropy of dynamical system appeared."

Since my collaborator in this field (Y. Takahashi) was a postdoc of Sinai, naturally I am close to the Russian school.

[30] before 1980 in the US only the people at UCSC [Rob Shaw, Doyne Farmer, Normal Packard and Jim Crutchfield] were active in the field, and we were in correspondence.

transformation and *Bernoulli systems*.
∗ We must know how to measure complexity or the extent of being chaotic. This requires *Kolmogorov complexity* which requires *Turing machines*. Thus we will discuss the ABC of *algorithmic complexity* and *computability* of physical processes: I wish you to think of the scientific significance of a theoretical result that is not computable.
∗ Eventually *Brudno's theorem* tells us chaos = algorithmically complex trajectories.

After these elementary examples we go into measure theoretical and topological dynamical systems.

Measure-theoretical dynamical systems
We will prove the Poincaré recurrence theorem and the Birkhoff ergodic theorem. The latter was long misunderstood as the key to statistical mechanics (even long after Boltzmann himself realized this). The *Kolmogorov-Sinai entropy* is a (the?) measure of the extent of chaos, which is related to statistical mechanical entropy. We will discuss the *thermodynamic formalism* for dynamical systems. This allows us to characterize observable invariant measures (i.e., observable or numerical-experimentally detectable stationary states) with a variational principle (mathematically isomorphic to Gibbs variational principle in statistical mechanics).

General theory of dynamical systems
Hopefully, the course will conclude with the typical behaviors of dynamical systems (e.g., *Axiom A systems*), which are *structurally stable* (i.e., stable against small modification of the system). Palis proposed the conjecture (*Palis conjecture*): typical systems have a finite number of (strange) attractors which support the observable invariant measures (called the *Sinai-Ruelle-Bowen measure*). Or in the physicist-friendly words: Any typical and structurally stable finite dimensional dynamical system with its phase space being compact may be understood by using a statistical mechanical device. The demonstration of this conjecture will close a very active phase that started with *Smale's horseshoe*.

1.9 Some 'practical' topics we will discuss

When we observe various time-dependent phenomena in hydrodynamics, cell biology, population dynamics, etc., often we do not have any explicit mathematical description at the beginning. Can we visualize 'strange attractors' directly from the observed time data? This is broadly answered affirmatively by Takens' theorem and related topics. The 'UCSC gang of four' proposed a similar method (see 1.7).[31] Even if clean

[31] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Geometry from a Time Series PRL 45 712 (1979).

geometrical fatures may not be obtained, still something suggestive may be gleaned (for example about our brains). Even some discrete plots could suggest something, although no simple dynamics behind the data can be guessed.

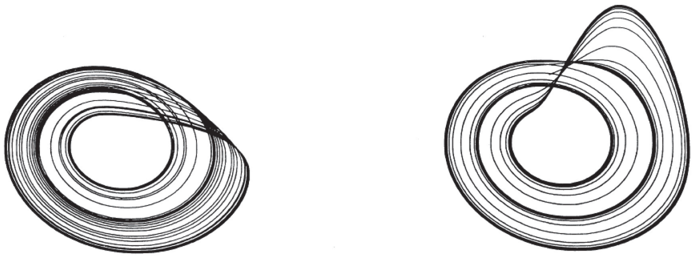


FIG. 1. (x,y) projection of Rossler (Ref. 7).

FIG. 2. $(x,\dot{x})$ reconstruction from the time series.

Figure 1.7: Reconstruction from a time series [Fig. 1,2 of Packard et al., PRL 45 712 (1979)]

Most natural phenomena are high-dimensional phenomena described by partial differential equations and by field theories. However, their asymptotic behaviors may not be very high-dimensional, and at least qualitatively we can reduce them to a low dimensional systems via various attractor theories and analytic methods like the Galerkin method. The famous Lorenz system is obtained by Salzmann in such an attempt.[32]

Such dimensional reduction methods may be of some interest in summarizing large scale MD simulation of biomolecules.

1.10 Video: *Chaos*

The following 9 (artistic) videos from http://www.chaos-math.org/en may be close to the spirit of the course. Most of you feel the pace is too slow and the explanation too elementary, but the movie actually explains sophisticated topics (very quietly esp beyond Chapter 6); the movie exhibits nice European taste.

Chapter 1: Motion and determinism—$\pi\alpha\nu\tau\alpha\ \rho\varepsilon\iota$.
https://www.youtube.com/watch?v=c0gDLEHbYCk&t=4s&frags=pl%2Cwn

Chapter 2: The vector fields—The lego race
https://www.youtube.com/watch?v=_Y68GX2UpQ0&frags=wn

Chapter 3: Mechanics—The apple and the Moon (Newton, universal law of gravity, etc.)

[32] For a systematic approach to the Couette flow is worked out by H. Yahata, Temporal development of the Taylor vortices in a rotating fluid, Prog. Theor. Phys. Suppl. 64 176 (1978).

https://www.youtube.com/watch?v=ZwTGAW0b_bo&frags=wn

Chapter 4: Oscillations—the swing (including Lotka-Volterra; Poincare-Bendixson theorem including the idea to prove it)
https://www.youtube.com/watch?v=uEfB5DG9x9M&frags=wn

Chapter 5: Billiards—Duhem's bull (geodesics on negative curvature surface included, symbolic dynamics a bit)
https://www.youtube.com/watch?v=3u2SJKxJhh8&frags=wn

Chapter 6: Chaos and the horseshoe—Smale in Copacabana (Poincaré map, Smale's horseshoe, symbolic dynamics, structural stability)
https://www.youtube.com/watch?v=ItZLb5xI_1U

Chapter 7: Strange attractors—the butterfly effect (Lorenz system, Lorenz template, symbolic dynamics
https://www.youtube.com/watch?v=aAJkLh76QnM&frags=wn

Chapter 8: Statistics—Lorenz' mill (measure-theoretical aspect, physical model of Lorenz system, sensitivity to initial conditions, SRB measure)
https://www.youtube.com/watch?v=SlwEt5QhAGY&frags=wn

Chapter 9: Chaotic or not—research today (bifurcation diagram, heteroclinic connection, non-generic case, Palis conjecture)
https://www.youtube.com/watch?v=_xfi0NwoqX8

1.11 Classic books still fresh

Recommended books for very serious students are listed here. They are now classic but still almost fresh. Readable? Yes, if you know basic math well (after one semester of my course you know the outline of many key portions of these classics).

∗ V. I. Arnold and A. Avez: *Ergodic Problems of Classical Mechanics* (Advanced Book Classics; Addison-Wesley; Reprint edition 1989; original in French 1967).

∗ J. Moser: *Stable and random motions in dynamical systems* (Annals of Mathematical Studies 77, Princeton UP 1973).

∗ J. Palis, Jr. and W. de Melo: *Geometric theory of dynamical systems* (Springer,1982).

∗ J. Palis and F. Takens: *Hyperbolicity & sensitive chaotic dynamics at homoclinic bifurcations* (Cambridge studies in advanced mathematics 35, Cambridge UP, 1993).

∗ R. E. Bowen: *Equilibrium states and the ergodic theory of Anosov diffeomorphisms* (Springer Lecture Notes in Mathematics 470; Second Ed 2008).

1.12 Do not misunderstand complexity

It is often said that the study of chaos is an important part of complexity study. I totally disagree with this popular view. If you admit that biological systems are complex systems, which is not merely complicated, you must recognize the distinction between the truly complex systems and pseudo complex systems that have been studied under the name of complexity study. If something is easy to (re)produce without any special preparation, that something must not be complex. Chaos is a typical example. As you learn it is easy to produce; perhaps the surprise was that benign-looking simple systems readily exhibit bewilderingly complicated behaviors (as illustrated in Fig. 1.6). In contradistinction, you cannot readily produce life from scratch; we do not even understand how life began at all. Complex systems are systems requiring a lot of prerequisite that we cannot (at leat readily) construct; you have your parents, because you are complex systems. Pasteur realized that complex systems are produced only by complex systems.

Thus, chaos has nothing to do with complex systems. The so-called complex-systems study in physics[33] studied only pseudo-complex systems that can self-organize almost from scratch.

[33] Even in our department this is the case, unfortunately.

2 Lecture 2: Setting the stage

2.1 Two main purposes of this lecture

There are two main topics today. Dynamical systems are defined as maps or flows defined by vector fields on manifolds. Therefore, first, we discuss manifolds and vector fields on them.

For practical physicists, basically we have only to understand what happens in the ordinary n-dimension Euclidean space $\mathbb{E}^n$, because manifolds are patchworks of Euclidean spaces. Thus, the purpose of the first part of this lecture is to introduce concepts mathematically properly, but to tell you how to 'ignore' their technicality. To discuss a dynamical system we need a stage: a manifold.

The second part reflects on what 'common' or 'general' means, because we are interested in 'general pictures' of dynamical systems. We review very basic concepts, openness, denseness, etc. Cantor sets are introduced and then we will discuss what 'dimension' is.

Although you must be able to find (and to understand) real math definitions, as an active physicist it is very important to grasp math concepts and theorems intuitively (and even emotionally).

2.2 Discrete-time dynamical systems

A discrete time dynamical system is a map $f : M \to M$, where M is a compact[34] (often C^r-)manifold (see justbelow).

I hope you know what map is.

The stage (or the totality of the states of the system; the phase space in classical statistical mechanics is an example) of the dynamical system may not be a simple space like $\mathbb{R}^n$ (n-dimensional Euclidean space; $\mathbb{E}^n$ may be a better notation). It could be a 2-torus (T^2), so it cannot be mapped continuously to $\mathbb{R}^2$. Usually, we choose the stage to be a *manifold*.

An n-manifold is a geometrical object that can be constructed by 'smoothly' pasting a n-Cartesian coordinate patches that are 'flexibly deformed' (Fig. 2.1):

A formal definition is in **2.3**. To understand this we need a concept 'diffeomorphism': Let A and B be sets (actually, topological spaces). A map $f : A \to B$ is a C^r-diffeomorphism, if it is continuously r-times differentiable (C^r-map), and its

[34] In a finite dimensional space, 'compact' means that the object is covered with a finite number of open sets. See **3.10**

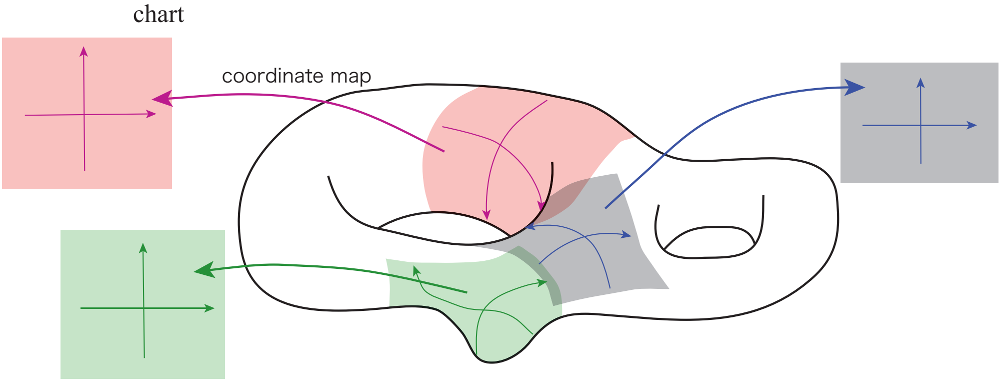


Figure 2.1: Intuitive idea of a manifold; here only some coordinate patches are illustrated; you must fill the whole object with such a cover.

inverse f^{-1} is also a C^r-map. If f and f^{-1} are continuous maps, we say f is a homeomorphism.

Practically, whenever we work on a manifold, we take a chart and then locally work in a Cartesian system. We must seamlessly connect the local results, but that part is usually only technical, so for physicists intuitively the core discussions are over on the chart.

2.3 Manifold

A manifold $M = (\mathcal{M}, \mathcal{U})$ consists of two elements:
(1) Basic set $\mathcal{M}$, which is usually a Hausdorff space.[35]
(2) Local coordinate system $\mathcal{U} = \{U_\alpha, \phi_\alpha\}_{\alpha \in A}$, where A is a subscript set. Here
(i) $U_\alpha \subset \mathcal{M}$ is open and $\cup_\alpha U_\alpha = \mathcal{M}$ (i.e., $\{U_\alpha\}$ is an open cover of $\mathcal{M}$.
(ii) ϕ_α is a homeomorphism (or diffeomorphism)[36] from U_α to an open set in $\mathbb{R}^n$.
(iii) If $U_\alpha \cap U_\beta \neq \emptyset$, then $\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$ is a homeo or diffeomorphism.

(U_α, ϕ_α) is called a chart, and $\mathcal{U}$ is called an atlas.[37] Needless to say, charts in an atlas of a manifold must be consistent as (iii).

[35] ⟪**Hausdorff space**⟫ A topological space X is a Hausdorff space if all distinct points in X are pairwise neighborhood-separable. That is, if $x, y \in X$ and $x \neq y$, there is a neighborhood U of x and that V for y such that $U \cap V = \emptyset$.

[36] This choice depends on how smoothly you wish to set up your manifold.

[37] The largest atlas including all the atlases containing a given chart is called the maximal atlas (or the differentiable structure). It may not be unique even diffeomorphically. S^7 has 28 different differentiable structures.

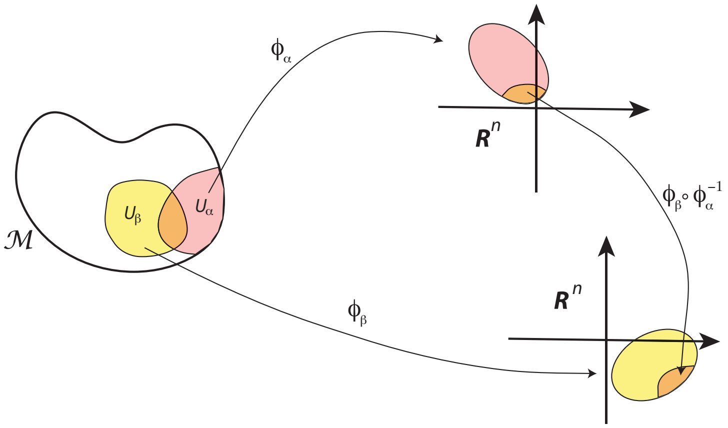


Figure 2.2: Manifold: local coordinate system/charts

2.4 'Topological continuation'

When we discuss mathematics on a manifold, it is almost always the case that we assume some convenient charts and on each chart we discuss what we are interested in. Then, we show that we can continue the conclusion to different charts using the 'smoothness' of the relations between the overlapping charts. The latter steps may be tedious, but intuitively without any trouble. Thus, in these lecture notes, we almost always discuss mathematics on a chart and will not discuss the 'second step,' simply saying 'according to the topological continuation' we claim the statements on the entire manifold. As you realize, this means that we may discuss things on an appropriate Euclidean space.

2.5 Vector field on manifold

We wish to consider an ordinary differential equation (ODE) defined on a manifold at each point $x \in M$. We take a chart (local coordinates) $(x_1, x_2, \cdots, x_n)$. With respect to this coordinates a vector field $X(x)$ at x may be expressed as

$$X(x) = (X_1, X_2, \cdots). \tag{2.1}$$

Here, X_i is the x_i-component of X. See Fig. 2.3.

Mathematicians express X as follows:

$$X(x) = \sum_i X_i \frac{\partial}{\partial x_i}. \tag{2.2}$$

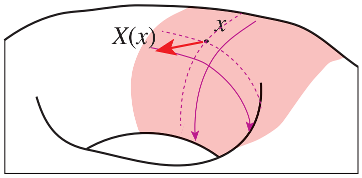


Figure 2.3: Vector field on manifold. The dotted lines correspond to the tangential directions $\frac{\partial}{\partial x_1}$ and $\frac{\partial}{\partial x_2}$.

Here $\frac{\partial}{\partial x_i}|_x$ points the tangent direction along the local coordinate x_i as illustrated in Fig. 2.3. The vector space spanned by $\{\frac{\partial}{\partial x_i}|_x\}$ is called the tangent space of M at x, and is denoted by the symbol T_xM.

The notation (2.2) is very reasonable as you can see from the directional derivative of a function f on M at x along the curve $\xi(t) \in M$. Let the tangent vector for ξ at x be X:

$$\frac{d}{dt}\xi(t) = X = (X_1, X_2, \cdots). \tag{2.3}$$

Then,

$$\frac{d}{dt}f(\xi(t)) = \sum_i X_i \frac{\partial}{\partial x_i} f(x) = Xf(x) = \sum_i X_i \frac{\partial f}{\partial x_i}. \tag{2.4}$$

Besides, the notation (2.2) automatically tells us how to rewrite the vector components when we change a chart from (U, x) to another overlapping chart (V, y):

$$X = \sum_j Y_j \frac{\partial}{\partial y_j} = \sum_j Y_j \sum_i \frac{\partial x_i}{\partial y_j} \frac{\partial}{\partial x_i}. \tag{2.5}$$

Thus, we have

$$X_i = \sum_j Y_j \frac{\partial x_i}{\partial y_j}. \tag{2.6}$$

A more mathematically respectable explanation can be found in 'indented' **2.7-2.9**.

2.6 Continuous-time dynamical systems

An ODE on M may be defined as

$$\frac{d}{dt}x = X. \tag{2.7}$$

This implies

$$\sum_i \dot{x}_i \frac{\partial}{\partial x_i} = \sum_i X_i(x) \frac{\partial}{\partial x_i}. \tag{2.8}$$

that is, as we know in the elementary calculus, $\dot{x}_i = X_i(x)$.

If (3.1) is well-posed (i.e., there exists a unique solution to the Cauchy problem = initial value problem), it can define a continuous-time dynamical system $\phi_t : M \to M$, where ϕ_t is the time evolution operator ($\{\phi_t\}$ is a group if $t \in \mathbb{R}$ or a monoid if $t \in [0, +\infty)$):
(i) $\phi_0 = 1$,[38]
(ii) $\phi_t \circ \phi_s = \phi_{t+s}$.

The totality of the C^r-vector fields on M is written as $\mathcal{X}^r(M)$. As we will know, $X \in \mathcal{X}^r(M)$ defines a continuous-time dynamical system.

Remark We discuss only the autonomous systems for which X never depends on t explicitly. If you wish to discuss a time-dependent vector field $X(t)$, you could introduce a new component z satisfying $\dot{z} = 1$.

2.7 Tangent vector

Let $\xi(t)$ be a (differentiable) curve in a manifold M in the chart (U, ϕ). The coordinate system is denoted by $(x_1, \cdots, x_n)$. Let $f : M \to \mathbb{R}$ be a differentiable function. If you wish to differentiate f along the tangential direction of ξ we can compute

$$\frac{d}{dt} f(\xi(t)) = \sum_i \frac{d\xi_i}{dt} \frac{\partial}{\partial x_i} f \equiv X_\xi f \tag{2.9}$$

where

$$\frac{\partial}{\partial x_i} f = D_i f(\phi^{-1}). \tag{2.10}$$

Here, D_i is the differentiation with respect to the coordinate x_i on the chart.

We use the following notational convention:

$$X_\xi = \frac{d\xi}{dt} = \sum_i \frac{d\xi_i}{dt} \frac{\partial}{\partial x_i}. \tag{2.11}$$

X_ξ denotes a vector in the tangential space of M. If $p = \xi(0)$, then the totality of X_ξ at $t = 0$ spans the tangential space of M at p denoted by T_pM:

$$T_pM = \left\langle \frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n} \right\rangle. \tag{2.12}$$

[38] '1' means 'multiplying 1' or the identity operator: $1x = x$.

2.8 Vector bundle

A vector bundle (V, M, π) on a manifold M is a triple of a manifold M, a vector space V and a map $\pi : V \to M$.

$\pi^{-1}(x)$ for $x \in M$ is called the fiber at x.

$s : M \to V$ such that $\pi s(x) = x$.

If s is C^r, the vector field X is said to be a C^r-vector field. The totality of the C^r-vector field on M is written as $\mathcal{X}^r(M)$.

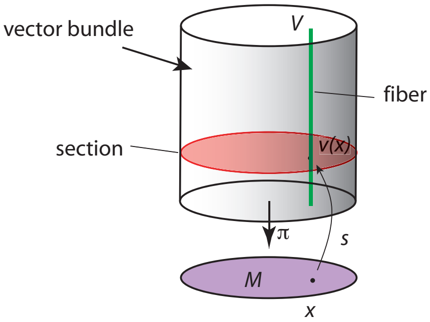


Figure 2.4: Vector bundle

If M is C^∞, then $\mathcal{X}^r(M)$ is a separable Banach space.[39]

2.9 Tangent vector bundle

The vector field on a manifold M whose fiber at $x \in M$ is T_xM is called the tangent vector bundle of M and is denoted as TM. Thus, $\mathcal{X}^r(M)$ consists of smooth sections of TM.

2.10 Pursuit of general pictures

Although particular examples, if representative enough in some sense, are often important to give us deep insights, our main goal is to have an overall general picture of dynamical systems. We study discrete dynamical systems $C^r(M)$ or continuous dynamical systems defined by the corresponding vector fields $\mathcal{X}^r(M)$. We wish to know the common features of dynamical systems, or wish to characterize typical examples in these collections of dynamical systems.

What do we wish to mean by ‘general’, ‘common’ or ‘typical’?

The most natural or desirable idea is that a general property is a property shared by all the instances in $C^r(M)$, for example. However, in most cases exceptions exist for any apparently very general properties, so ‘all’ should be relaxed to ‘almost all’

[39] A Banach space is a linear space with a complete norm (‘complete’ = any Cauchy sequence converges). ‘Separable’ means that the space contains an everywhere dense countable set.

or 'most.' Therefore, how to relax 'all' and still to secure 'many' in a precise sense[40] is our key issue.

In the following we first discuss denseness **2.12** and openness **2.13**. If systems with property A are dense in the totality of dynamical systems (say, $\mathcal{X}^r(M)$), we can always find (need not be easy) a system with property A as close as you wish to a system you pick up. If systems with property A make an open subset of the totality of dynamical systems, then a system with property A is surrounded by systems with property A, so property A is stable against perturbations. However, this does not guarantee 'many,' although it tells us that 'all' (really all, not almost all) around the example.

Then, 'open-denseness' = openness + denseness sounds desirable. Not quite as we will see below.

2.11 Topology[41]

If we wish to say somethings are common, we must be able to classify the objects we are interested in, or at least, we must be able to tell which objects are more similar or less so than other objects. Thus, we need a concept to judge 'closeness' or similarity. You might immediately have in your mind some sort of distances, but a more basic concept is some sort of 'nearness' or being in the same neighborhood. A topology furnishes this concept.

Let τ be a family of subsets of a set X. τ is a topology defined for a set X, if

(i) $\tau \ni \emptyset$ and $\ni X$.
(ii) If $\{G_a \in \tau\}$, then $\cup_a G_a \in \tau$.
(iii) If $\{G_a \in \tau\}$ is a finite collection, then $\cap_a G_a \in \tau$.

(X, τ) is called a topological space, and the elements in τ are called open sets.

Any open set containing $x \in X$ is called a neighborhood of x (in this topological space).

We can define continuous functions between two topological spaces X and Y: $f : X \to Y$ is continuous around $x \in X$, if for any nbh U of $f(x) \in Y$ is a nbh V of $x \in X$ such that if $y \in V \Rightarrow f(y) \in U$.

Thus, geometrical properties that are preserved by continuity are called topo-

[40] Here, 'precise' means, ultimately, 'can be axiomatized'.

[41] For such basic concepts the following books are strongly recommended:
I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry* (Springer Undergraduate Texts in Mathematics 1976)
A. N. Kolmogorov and S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis* (Martino Fine Books, 2012) \$9.5 !

logical properties.

2.12 Denseness

Let Ω be a set. A subset $U \subset \Omega$ is dense if every neighborhood of any element in Ω contains an element in U.

As $\mathbb{Q} \subset \mathbb{R}$ shows, denseness can still mean shear minority. If we evenly and randomly sample a point from $[0, 1]$ almost surely you are in $\mathbb{Q}^c$. It is a countable set, so its total length is zero (= measure zero).

Thus, a bigger subset than mere dense subset is desirable.

A countable set Q is measure zero (volume is zero; more precisely its Lebesgue measure is zero). Since Q is countable, it is one-to-one correspondent to $\mathbb{N}^+$. Thus we can write $Q = \{a_k\}_{k\in\mathbb{N}^+}$. For example, let us assume $Q \in \mathbb{R}^n$. Then, we choose a ball B_k of volume $\varepsilon/2^{k+1}$ ($\varepsilon > 0$, arbitrary) centered at a_k. Obviously $Q \subset \cup_k B_k$. Therefore, the total volume of Q (denoted as $|Q|$) must be less than the total volume of these balls $B_1, B_2, \cdots$:

$$|Q| \le \sum_k |B_k| = \sum_k \frac{\varepsilon}{2^{k+1}} = \varepsilon. \tag{2.13}$$

Since $\varepsilon > 0$ is arbitrary, $|Q|$ must be smaller than any positive number: 0.

2.13 Openness, structural stability

A subset U is open in Ω, if any element of U has a neighborhood contained in U.

Thus, intuitively speaking, when $x \in U$ is perturbed to $x + \delta x$ with any sufficiently small δx $x + \delta x \in U$. This is a stability of the property. If a dynamical system itself (in contrast to perturbing its initial condition) is perturbed and if still the system 'looks (behaves) similar' to the original system, we say the system is structurally stable.

In a certain sense 'openness' can be too strong or too weak, depending on the situation. It can be too strong, because there can be special direction to destroy the property we are interested in, although in all other directions it is stable against small perturbations. A 'too weak' case is in **2.14**. Also if we recall that irrational numbers are with full measure but not open in $\mathbb{R}$, demanding openness may not be wise to characterize 'general.'

2.14 Open denseness

If a set U is open, then it is measure positive for 'natural sampling measure' as an open set in $[0, 1]$ illustrates. Therefore, 'open denseness' could be a candidate of 'naturalness,' and certainly it is used to assert certain generality. However, the compliment of an open dense set could be as close to the full (measure) set as possible as the fat Cantor set illustrates.[42]

However, unfortunately, it is often the case that most properties for dynamical systems are vulnerable to a particular perturbation in a certain 'direction', so 'openness' does not usually hold. Thus, we wish to relax 'open denseness'.

Can we find an open dense set easily? Not so. See **2.19**.

2.15 Cantor set

Cantor discussed a perfect nowhere dense set C (Cantor set in the general sense). 'Perfect' means no point is isolated. That is, for any neighborhood of $x \in C$ is a point in C. 'Nowhere dense' means $[C]^\circ = \emptyset$.

The most famous example of a Cantor set is the one invented by H. J. S. Smith, the set built by removing the open middle thirds of a line segment (see 2.7). Cantor mentioned this as an example, in passing, but it is now usually recognized as 'the Cantor set.'

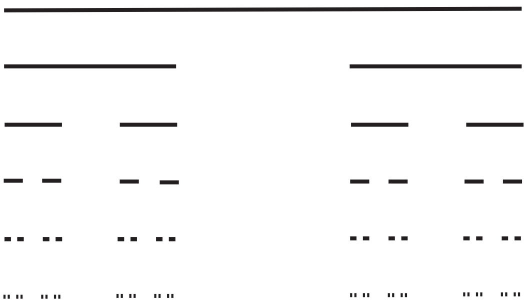


Figure 2.5: 'The' Cantor set due to Smith (1874) constructed by removing the middle third open set

This Cantor set may be analytically expressed as

$$C = \left\{ x = \sum_{n=1}^{\infty} \frac{a_n}{3^n} \,\middle|\, a_n \in \{0, 2\} \right\}. \tag{2.14}$$

It is a measure zero uncountable and self-similar set. Notice that the number of gaps is a countable infinity. $[0,1] \setminus C$ is an open dense set and with full measure.

[42] B. R. Gelbaum and J. M. H. Olmsted, *Counterexamples in Analysis* (Holden-Day, Inc. San Francisco, 1964) is a great source book for delicate issues.

2.16 Why Cantor sets are relevant to dynamical systems

Perhaps, you might think Cantor sets are rather contrived artificial sets. It is actually not. As noted before, complicated behavior due to nonlinearity is due to cofinement of the phase space in a finite domain by 'folding.' If the invariant set[43] is not identical with M, then it is very natural to exhibit a self-similar structure (see Fig. 2.8). Therefore, U itself is a Cantor set or a direct product of Cantor set and 'an ordinary set' (interval, circle, etc.).

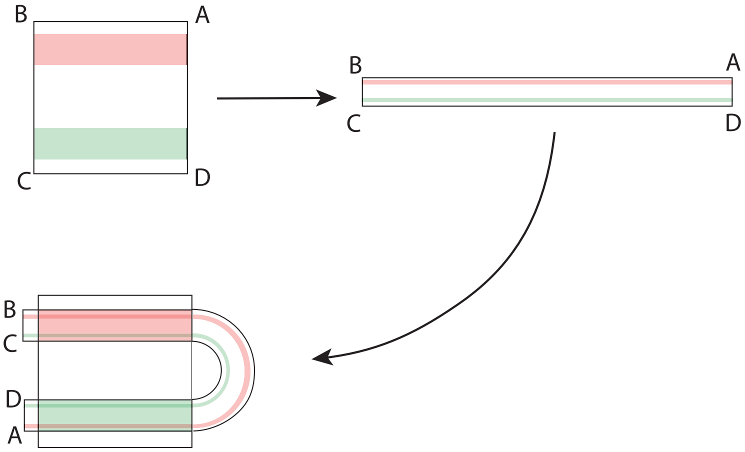


Figure 2.6: Folding can produce self similar Cantor-like invariant set. This is a part of the famous Smale's horseshoe.

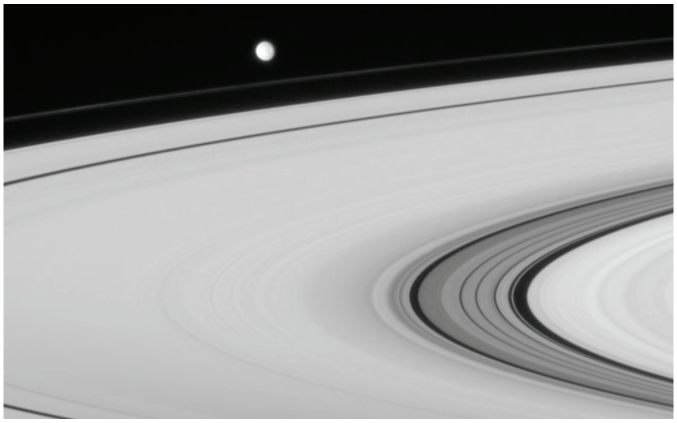


Figure 2.7: Is the ring of Saturn cantor?

[43] We will discuss this in detail, but here you may intuitively understand that a set $U \subset M$ such that $f(U)$ or $\phi_t(U) = U$.

2.17 Cantor set due to 'excessively tall' tent maps

Consider a tent map $f : \mathbb{R} \to \mathbb{R}$

$$f(x) = \begin{cases} kx \text{ for } x \le 1/2, \\ k(1-x) \text{ for } x > 1/2. \end{cases} \tag{2.15}$$

The points remaining in $[0,1]$ make a Cantor set. For $k = 3$ we get the 'standard Cantor set' explained in Fig. 2.7.

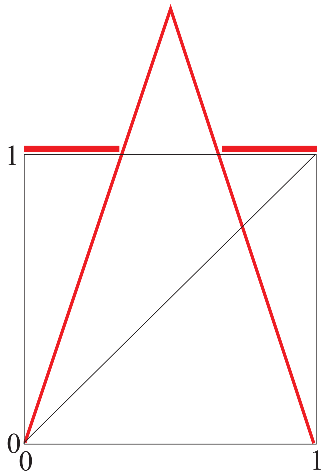


Figure 2.8: Adjusting the slopes of a tent map, we can make a self-similar middle removed Cantor set as an invariant set

2.18 Fat Cantor set

In the standard Cantor set, we remove 1/3 of the interval in a self-similar fashion. What if we remove much smaller interval? If the construction is self similar, the remaining (the resultant) Cantor set is measure zero. What if, then we remove center pieces that are shrinking faster?

Let $\alpha \in (0,1)$.
(i) Remove $(1/2 - \alpha/4, 1/2 + \alpha/4)$ from $[0,1]$.
(ii) Then, remove the middle $\alpha/24$ fractions from the remaining two intervals.
(iii) Repeat this procedure ad infinitum (see Fig. 2.9).

The remaining set is a closed set and has no interior and perfect. That is, a Cantor set = nowhere dense perfect set.

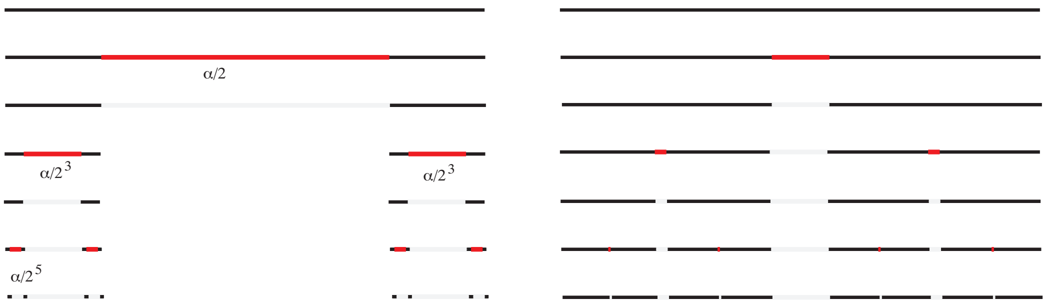


Figure 2.9: Construction of fat cantor set. Remove successively the scaled copies of red chunks. Left the usual measure zero cantor set; Right: a fat cantor set with a positive length (only the values of α are different).

The total length of the resultant cantor set is

$$1-\frac{\alpha}{2}-2\frac{\alpha}{2^3}-2^2\frac{\alpha}{2^5}-\cdots=1-\frac{\alpha}{2}\left(1+\frac{1}{2}+\frac{1}{2^2}+\cdots\right)=1-\frac{\alpha}{2}\frac{1}{1-1/2}=1-\alpha. \tag{2.16}$$

If $\alpha < 1$, then the resultant Cantor set is no where dense but with positive measure, so it is called a *fat Cantor set*. We will actually encounter such a set in chaos.

2.19 Very small open-dense set

Now, it should be clear how to make an open and dense set which is far from the major part of the total set.

Let us construct a fat Cantor set of measure $1-\varepsilon$ for small $\varepsilon > 0$. Then, make its complement in $[0,1]$. It is open and dense, but with measure ε (> 0), which can be as small as you wish. Thus, we have constructed an open dense set that is very hard to sample.

2.20 What is dimension?[44]

We usually say our space is 3D because we need three independent coordinates to specify a point in the space uniquely. However, the space may not be metric (i.e., distances may not be clearly defined; what is space?). Perhaps the simplest idea is that our space is modeled by a 3-manifold. 3-manifold is defined by 3D charts. The latter is defined as a 3D vector space, which we can define mathematically (will not be discussed).

However, the concept of dimension should be 'more primitive'; we should

[44] Detailed reference: Y. B. Pesin: *Dimension theory in dynamical systems* (Chicago Lectures in Mathematics, Chicago UP 1997).

not be required to have a metric, for example.

2.21 Inductive topological dimension[45]

A motivation comes from the fact that a bounded geometric object B its dimension is the dimension of $\partial B + 1$. However, if B is an open set $\partial B = \emptyset$, even though $\partial[B]$ is not empty. Thus, the induction suggested above must be formulated with some care.
The inductive dimension $d_{\rm I}$ is defined as follows:
(i) $d_{\rm I}(\emptyset) = -1$.
(ii) $d_{\rm I}(B) \le n$ if for all $x \in B$, there is a neighborhood U such that an open set $V \subset U$ such that $x \in V$, $[V] \subset U$ with $d_I(\partial[V]) \le n-1$.
(iii) $d_{\rm I}(B) = m$, if m is the smallest number satisfying (ii).

Any totally disconnected sets[46] have dimension zero. Therefore, the union of two dimension zero sets can have a positive dimension. This does not happen for the Lebesgue cover dimension (see **2.22**): $\dim(X \cup Y) = \max\{\dim(X), \dim(Y)\}$.

2.22 Lebesgue covering dimension

Consider an open cover $\mathcal{A}$ of a topological space X. Suppose for any $x \in X$ we can refine $\mathcal{A}$ so that the covering order (= the min number of the open sets in $\mathcal{A}$ covering a point is its covering order) is less than $n+1$. Then the minimum value of n is called the topological dimension of X.[47]

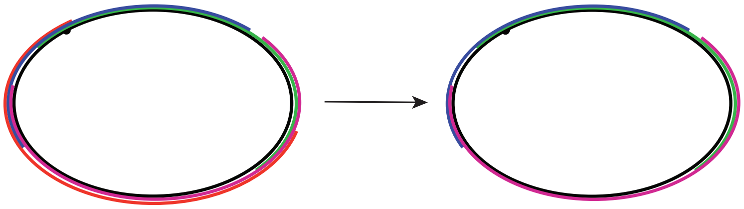


Figure 2.10: How to obtain the Lebesgue covering dimension of S^1

If X and Y are homeomorphic, both have the same topological dimension. Cantor set has dimension zero, because we can alway find a disjoint subcover.

Every compact n-topological space can be embedded in $\mathcal{R}^{2m+1}$ (Whitney's theorem https://en.wikipedia.org/wiki/Whitney_embedding_theorem), which

[45] There are 'small' and 'large' inductive dimensions, but here only the former is mentioned.

[46] A topological space X is totally disconnected if the connected components in X are the one-point sets.

[47] The illustrations in Wikipedia are wrong (or at best misleading).

will be discussed later.

A problem of this definition is that finding refinements could be daunting.

2.23 Box counting or Minkowski dimension

Let $N_\delta(A)$ be the smallest number of diameter δ sets covering A. If the following limit exists, it is called the box counting (or Minkowski) dimension of A:

$$\dim_M(A) = \lim \frac{\log N_\delta(A)}{-\log \delta}. \tag{2.17}$$

Notice A and $[A]$ have the same $\dim_M$. This is not a very desirable feature, but since it is easy to compute and in many cases it agrees with the following Hausdorff dimension **2.25**. Also an unpleasant example is $\dim_M(\{1/n\}_{n\in\mathbb{N}}) = 1/2$.

2.24 Hausdorff measure

The s-dimensional Hausdorff measure of a set $A \subset \mathbb{R}^n$ is defined by

$$H^s(A) = \lim_{\delta\to 0} H^s_\delta(A), \tag{2.18}$$

where

$$H^s_\delta(A) = \inf_{\mathcal{U}} \left\{ \sum_i \mathrm{diam}(U_i)^s \,\middle|\, U_i \in \mathcal{U}(A) \right\} \tag{2.19}$$

with $\mathcal{U}(A)$ being the open cover of A.

H^s is a decreasing function of s.

2.25 Hausdorff dimension[48]

The Hausdorff dimension of a set A is defined as

$$\dim_H(A) = \sup_s \{s \mid H^s(A) = \infty\} = \inf_s \{s \mid H^s(A) = 0\}. \tag{2.20}$$

The H-dimension of a totally disconnected set is less than 1.

If A is self similar, then $\dim_M(A) = \dim_H(A)$.

[48] B. Simon*Real Analysis* (A comprehensive course in analysis, Part 1.) (AMS 2015) 8.2 is much more detailed.

2.26 Hausdorff dimension of Cantor set

The Hausdorff dimension of the middle third Cantor set is $\log 2/\log 3$. This can be confirmed due to the self-similarity of the Cantor set.

Let us consider the middle one ninth Cantor set. The first step removed 1/9. The next step removes 1/9 of the remaining segments. Thus, $(8/9)^n \to 0$ that is, the resultant set is not fat. As is clear, self-similar Cantor set is measure zero. However, intuition tells us that the middle 1/9th Cantor set should be 'larger' than that of the middle third Cantor set. Thus, the Hausdorff dimension is useful.[49]

2.27 Fractals

A geometric object whose topological dimension is different from its Hausdorff dimension is called a fractal object. A typical examples is the von Koch curve. Its Hausdorff dimensiion is $\log 4/\log 3$.

[49] ⟪**Hausdorff**⟫ Felix Hausdorff (1868-1942) did his work on Hausdorff dimension in 1919. As a World War I veteran, he could avoid the laws against Jews, but by the late 1930s he was dismissed. He became concerned about being shipped to the camps. On the night of Jan 25, 1942, having learned they were to be picked up the next day to be sent to the Endenich camp, Hausdorf, his wife and his wife's sister committed suicide by overdose of barbital [B Simon p49 + Wikipedia].

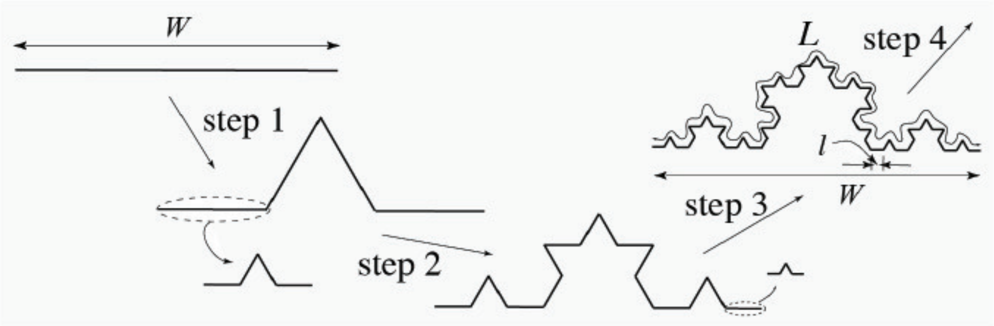


Fig. 3.4 How to construct the von Koch curve. First, take a line segment of length W and make a segment of length $1/3$ of the former. Prepare four such segments and make a piecewise linear figure with a triangular mound at the center. This is Step 1. Next, each linear segment of this figure (a representative is encircled with a broken ellipse) is replaced by a $1/3$ shrunk copy of the whole figure. This is Step 2. The resultant figure at the bottom of the figure is made of unit segments (monomer units) of length $\ell = W/9$. These length $W/9$ monomer units are all replaced by the small copy of the figure constructed in Step 1 as shown just below the Step 3 arrow. After Step 3 the length of the 'monomer unit' is $W/3^3$. Now, we repeat this procedure *ad infinitum* beyond Step 4, and we will obtain the von Koch curve. In this section, we regard the length of the 'monomer unit' to be finite, so we repeat these steps only finite times. However, since the number of repetition n is large, the length of the monomer unit $\ell = W/3^n$ is invisibly small.

Figure 2.11:

2.28 Residual property

A nowhere dense set E is defined by $\overline{E}^\circ = \emptyset$. A meager set is a set described as a countable union of nowhere dense sets. A set A is *almost open* (or Baire set) if there is an open set such that $A\Delta U$ is meager.

A property that holds on Ω except for a meager set is called a residual property. Needless to say, a residual set that is $\Omega\setminus$ a meager set (i.e., the complement of a meager set) is 'smaller' than an open dense set. A residual set can be measure zero (but an open set is always with positive measure).

2.29 Ultimate conjectures on discrete dynamical systems due to Palis [no explanation given here.]

Palis conjectured (See Lect 44):

> Every diffeomorphism in $\mathrm{Diff}^1(M)$ can be approximated by an Axiom A diffeomorphism or else by one exhibiting a homoclinic bifurcation involving a homoclinic tangency or a cycle of hyperbolic periodic saddles with different indices.

For physicists, the following Milnor-Palis conjecture may be more interesting:[50]

> For a typical smooth dynamical system $f : M \to M$, the global attractor A_f is decomposed into finitely many minimal attractors A_i. Moreover, for almost every point $x \in M$, the ω-limit set $\omega(x)$ is equal to one of the A_i. Typically each minimal attractor supports a unique SRB measure μ that governs behavior of Lebesgue almost all points $x \in M$. The latter means that as $n \to \infty$
>
> $$\frac{1}{n}\sum_{i=0}^{n-1}\phi(f^k x) \to \int \phi(x) d\mu(x) \tag{2.21}$$
>
> for any continuous function $\phi \in C(M)$.

Here 'typical'[51] means:

A certain property is considered to be typical if it is satisfied for almost all pa-

[50] in *Abel Prize 2008-2012* by M Lyubich.

[51] The appropriate probabilistic notion (in infinitely dimensional space of systems) goes back to

rameters in a generic one-parameter family of systems.

For one-dimensional unimodal analytical maps the conjecture was proved.[52]

For example, it is still conceivable that from probabilistic point of view, the Newhouse phenomenon is negligible.

[52] Lyubich, M.: Dynamics of quadratic polynomials, Ill. Parapuzzle and SBR measure. 11 Géométric complexe et systémes dynamiques. Asterisque Volume in Honor of Douady's 60th Birthday, vol. 261, pp. 173- 200 (2000); Lyubich, M.: Almost every real quadratic map is either regular or stochastic. Ann. Math. 156, 1-78 (2002); Avila, A., Lyubich, M., de Melo, W.: Regular or stochastic dynamics in real analytic families of unimodal maps, Inv Math 154 451 (2003).

3 Lecture 3: ODE review

3.1 Ordinary differential equation on manifold

As we have already discussed in **2.6** an ordinary differential equation on a manifold M is

$$\dot{x} = X(x), \tag{3.1}$$

where X is a section of TM. You may understand this as an ordinary n-vector ODE in a flat space $\mathbb{R}^n$ (or its subset).

More generally, an ODE is a functional relation among a function and its derivatives. Thus, (3.1) is not the most general form (see **3.2**), but is the most natural object to study what can happen for a time-evolving system whose phase space is M.

3.2 General ODE

Let y be a n-times differentiable function of $t \in \boldsymbol{R}$. A functional relation

$$f(t, y(t), y'(t), \cdots, y^{(n)}(t)) = 0 \tag{3.2}$$

among t, $y(t)$, $y'(t)$, $\cdots$, $y^{(n)}(t)$ is called an *ordinary differential equation* (ODE) for $y(t)$, and n is called its *order*, where the domain of f is assumed to be appropriate. Such $y(t)$ that satisfies $f = 0$ is called a *solution* to the ODE.

If the highest order derivative of y is explicitly solved as

$$y^{(n)}(t) = F(t, y, y', \cdots, y^{(n-1)}) \tag{3.3}$$

from $f = 0$, we say the ODE is in the *normal form*.[53]

3.3 Normal form ODE is essentially first order.

Let $y_j \equiv y^{(j-1)}$ $(j = 1, \cdots, n)$. Then (3.3) can be rewritten as

$$\begin{aligned} \frac{dy_1}{dt} &= y_2, \\ &\cdots \\ \frac{dy_{n-1}}{dt} &= y_n, \\ \frac{dy_n}{dt} &= F(y, y_1, y_2, \cdots, y_n). \end{aligned} \tag{3.4}$$

[53] Notice that not normal ODE's may have many pathological phenomena, but we will not pay any attention to the non-normal form cases henceforth.

That is, (3.3) has been converted into a first order ODE for a vector $y = (y_1, y_2, \cdots, y_n)^T$.[54] Any normal form n-th order scalar ODE can be converted into the n-vector first order ODE of the form

$$\frac{dy}{dt} = X(t, y). \tag{3.5}$$

Any solution $y(t)$ can be understood as an orbit (or trajectory) parametrized with 'time' t in the n-space (= phase space) in which y lives.

3.4 Autonomous vs nonautonomous

For (3.5) the vector field X explicitly depends on time. This means that there is a certain 'external' agent modifying the vector field. Thus, generally, we are not interested in such a system that is 'not self-contained.'[55] Such a system is called a non-autonomous system. We are interested in autonomous systems as described by (3.1).

3.5 When does ODE define dynamical system?

If (3.1) has a unique solution for any intial condition $x \in M$, we may define a continuous-time dynamical system. We know the following:
(1) Peano's theorem: If X is continuous, (3.1) has a solution.
(2) Cauchy-Lipshitz uniqueness theorem: If X satisfies a Lipshitz condition $\|X(x) - X(x')\| < L\|x - x'\|$ (see **3.6**), (3.1) has a unique solution.
(3) If X is not Lipshitz, then the uniqueness of the solution is not guaranteed (counterexamples exist).

Thus, we confine our attention to differentiable vector fields that are automatically Lipshitz.[56] However, we should understand why (1)-(3). This is the purpose of the rest of the lecture. To prove (1) an approximate solution sequence is constructed (via the Euler approximation), and then we prove the existence of a limit. This requires a knowledge of functional analysis, but at least the gist of the demonstration or its delicate point should be recognized. (2) can be understood almost geometrically (the rectifiability theorem **3.19**).

3.6 Lipschitz condition.

[54] You may prefer $\boldsymbol{y}$ for y, I will maximally avoid explicit vector notation throughout the lecture notes.

[55] Except perhaps the perturbation is periodic.

[56] cf. the mean-value theorem

Let X be a continuous vector function whose domain is a region $D \subset \mathbb{R}^n$. For any compact[57] set $K \subset D$, if for any y_1 and y_2 both in K there is a positive constant L_K (which is usually dependent on K) such that

$$|X(y_1) - X(y_2)| \leq L_K |y_1 - y_2|, \tag{3.6}$$

then X is said to satisfy a *Lipschitz condition* on D.

A C^1 function is Lipschitz continuous due to the mean value theorem. If a vector field is C^1, then it is Lipshitz.

3.7 Peano's existence theorem

Suppose X in (3.1) is continuous in a bounded closed region $G \subset M$, then for any $x \in G$ there is at least one integral curve passing through it in G.

The proof may be obtained with the aid of Arzela's theorem **3.11** that can show the existence of the convergence of the Euler approximation sequence.[58]

3.8 Euler approximation

Consider

$$\dot{x} = X(x(t)) \tag{3.7}$$

for the time span $[0.T]$. if we approximate the derivative with a finite difference $\dot{x} \simeq [x(t + \Delta t) - x(t)]/\Delta t$ we may write the ODE as

$$x(t + \Delta t) = x(t) + \Delta t X(x(t)). \tag{3.8}$$

Therefore, we can make an approximate function φ_i by making a piecewise connection of adjacent time points $\{x(n\Delta t_i)\}$, where Δt_i is the time increment. We make an approximation sequence $\{\varphi_i\}$ for $\Delta t_i > \Delta t_{i+1} \to 0$.

Does this sequence converges to a solution (have an accumulation point corresponding to a solution)?

3.9 Strategy to prove Peano's theorem

First, we must show that $\{\varphi_i\}$ for $\Delta t_i > \Delta t_{i+1} \to 0$ has an accumulation point.

[57] 'Compact' means in a finite dimensional space 'closed and bounded'.

[58] Kolmogorov-Fomin p102

Since the totality of continuous functions $[0,T] \to M$ is not compact (nor relative compact), to show the existence of an accumulation point is not trivial. However, Arzela's theorem tells us that $\{\varphi_i\}$ is compact. Thus, accumulation points exist (this is why we cannot prove the uniqueness).

Then, we show that the limit indeed satisfies the original ODE.

As you see, we must understand the concept of compactness in a functional space (or infinite dimensional space).

3.10 Review of compactness[59]

If any open covering of a set S has a finite subcover,[60] S is called a compact set.

If the closure of S is compact, we say S is relative compact.

If a space is finite-dimensional, then bounded closed set is automatically compact. It is thanks to the Bolzano-Weierstrass theorem (= bounded sequences must have an accumulation point; a finite dimensional bounded closed set is countably compact). However, if the dimension is not finite, this is not true: think of $\{e_n\}$, where $e_n = (0, \cdots, 0, \overset{n}{1}, 0, \cdots)$.

We can make a function space $C_{[0,T]}$ as a metric space by introducing a sup metric $\rho(f,g) = \sup_{t\in[0,T]} |f(t) - g(t)|$, but the space is obviously not finite dimensional. We use

Theorem A necessary and sufficient condition for a metric space to be compact is: (i) totally bounded and (ii) complete.

To understand this theorem we must understand:

$*$ 'totally bounded': A metric space M is totally bounded if for any $\varepsilon > 0$ there is an ε net A consisting of finitely many points. That is $\forall y \in M\ \exists x \in A$ such that $\rho(x,y) < \varepsilon$.

$*$ complete: any Cauchy sequence converges.[61]

[Demo]

If a metric space A is compact, then it is totally bounded: If not, then there is $e_0 > 0$ such that there is no finite ε_0-net for A. Thus, we can have an infinite point set $\{a_i \in A\}$ such that $\rho(a_i, a_j) > \varepsilon_0$. Thus, $\{a_i\}$ is an infinite point set without an accumulation point. Thus, A is not compact. The necessity of completeness is obvious.

To show (i)+(ii) implying compactness, we have only to show that any bounded infinite set has an accumulation point. Consider an infinite sequence

[59] https://www.dropbox.com/home/ApplMath?preview=AMII-ElementaryCheckList.pdf may be useful as your analysis rudiment checklist.

[60] a cover consisting of a subset of the original cover.

[61] If the distance is L^2,then C is not complete. KF3.1

$\{x_i\}$. Take a 1-net. Then in a ball B within distance one from at least one of the net point are infinitely many points in this sequence. B is totally bounded, we can repeat the argument with distance $1/2$. Repeating this, we can construct a Cauchy sequence. Thanks to the completeness, there must be a limit point.

Thanks to this theorem and the fact that a closed subset of a complete space is complete, we can conclude that in a complete metric space the total boundedness is enough to guarantee the relative compactness os any subset M.

3.11 Arzela's theorem[62]

Theorem [Arzela] A set of function $\Phi \subset C_{[0,T]}$ is relative compact iff Φ is uniformly bounded and equicontinuous.

To understand this theorem we must understand:
* Uniformly bounded: For any $f \in \Phi$ and for any $t \in [0,T]$, $|f(x)| \leq K$ for some positive K.
* Equicontinuous: For any $\varepsilon > 0$ there is $\delta > 0$ such that for any $f \in \Phi$ $|f(t) - f(t')| < \varepsilon$ if $|t - t'| < \delta$.
[Demo]
Here we prove the sufficiency: $C_{[0,T]}$ is a complete metric space, since the convergence in sup norm means the uniform convergence on $[0,T]$. Therefore, we have only to check the total boundedness of $C_{[0,T]}$. Since Φ is uniformly bounded and equicontinuous, we can choose $\delta > 0$ appropriately so that for all $\varphi \in \Phi$

$$|\varphi| < K, \ |\varphi(x) - \varphi(x')| < \varepsilon \text{ if } |x - x'| < \delta. \tag{3.9}$$

We construct an ε-net consisting of piecewise linear functions. The idea must be intuitively grasped from the figure 3.1. Make the totality of piecewise linear functions connecting NE, E or SE arrows on the lattice. This is a finite ε-net. Therefore, Φ is compact.

3.12 Proof of Peano's theorem

We have constructed the piecewise linear continuous approximation sequence $\{\varphi_i(t)\} \subset C_{[0,T]}$. It is uniformly bounded and equicontinuous. Therefore, Arzela's theorem tells us that there is a uniformly convergent subsequence in $\{\varphi^{(i)}(t)\}$, converging to $\varphi(t)$. The remaining task is to show that for any $\varepsilon > 0$ we can choose Δt small enough to

[62] A readable proof is given in Kolmogorov-Fomin.

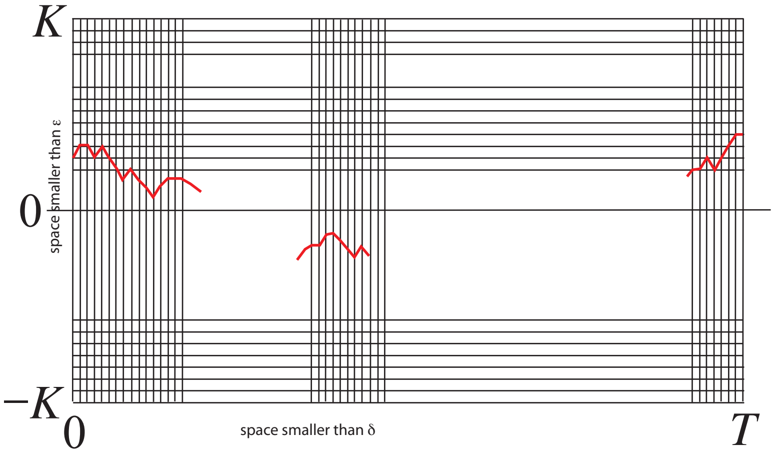


Figure 3.1: A representative piecewise linear approximate function (red) making an ε-net for bounded functions in $C_{[0.T]}$.

make

$$\left|\frac{\varphi(t+\Delta t)-\varphi(t)}{\Delta t}-X(\varphi(t))\right|<\varepsilon. \tag{3.10}$$

To this end we have only to show for sufficiently large k

$$\left|\frac{\varphi^{(k)}(t+\Delta t)-\varphi^{(k)}(t)}{\Delta t}-X(\varphi^{(k)}(t))\right|<\varepsilon. \tag{3.11}$$

We can formally demonstrate this,[63] but intuitively the closeness of $\varphi^{(k)}$ to φ and continuity of X implies all the terms are close to the limits.

3.13 Nonuniqueness cases

Peano noted that for $X(x) = 3x^{2/3}$ $x = 0$ and $x = t^3$ are solutions satisfying $(0, 0)$ as the starting point.[64] That is, continuity of X is not enough for the determinacy. However, differentiability is not needed.

If X is Hölder continuous with the exponent less than 1 at a point, the uniqueness is lost at the point.

[Hölder continuity].

[63] e.g., see Kolmogorov-Fomin

[64] More generally, $X = x^{1-1/n}$ $(n \in \mathbb{N}^+)$.

If a function f satisfies

$$|f(x) - f(y)| \le L|x-y|^\alpha \tag{3.12}$$

on its domain for constants L and $\alpha \in (0,1)$, f is said to be *Hölder continuous* of order α.

The physical reason for this nonuniqueness is 'forgetting the initial condition' due to super-ballistic acceleration.

3.14 Relevance to physics of nonuniqueness cases?[65]

In a fully developed turbulence the velocity field becomes not Lipshitz but only Hölder with $\alpha \simeq 1/3$ (if we believe in the Kolmogorov spectrum). Thus, the motion of a particle advected by the flow becomes non-deterministic.

If we consider a particle in a potential for $\alpha \in (0, `)$

$$V(x) = -\frac{1}{1+\alpha}|x|^{1+\alpha} \tag{3.13}$$

the classical mechanics gives the following equation of motion:

$$\dot{x} = v, \;\; \dot{v} = |x|^\alpha \mathrm{sign}(x). \tag{3.14}$$

Note the superballistic nature of $\dot{v}$ around $x = 0$. However, 'a physically realizable potential will however exhibit a power-law scaling as in (3.13) only over a limited range of x-values, with an inner or short-distance cutoff ℓ and an outer or large-distance cut-off L. The latter may not be very serious, but the former is serious.

Eyink and Drivas 'mollified' the potential with a smooth short-range cutoff, and then considered (1D) quantum mechanical potential (although not trapping) problem. The wave packet is not max at the origin, and even in the classical limit 'stochasticity' remains.

3.15 Phase flow

Let a n-dynamical system

$$\dot{x} = X(x) \tag{3.15}$$

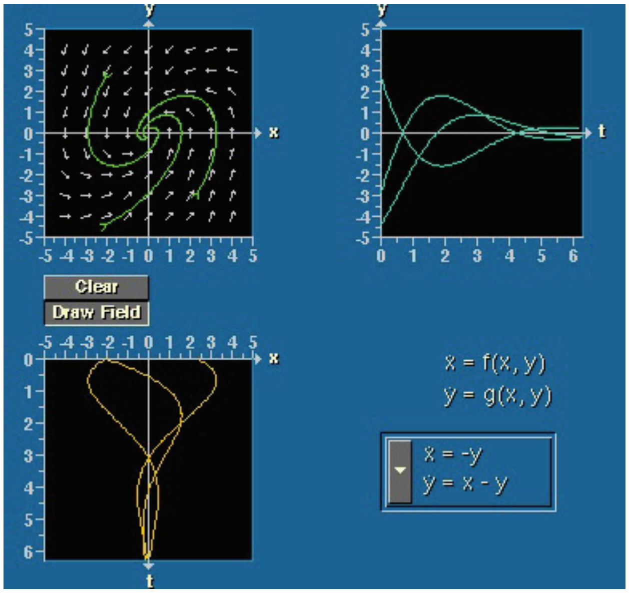
be defined on a domain U. We can imagine a flow field on U described by the vector field X. It is called the *phase flow* because it flows the phase space = the state space; in our case the state is specified by $x(t)$ at time t, so the space in which x lives is

[65] G L Eyink and T D Drivas, Quantum spontaneous stochasticity arXiv: 1509:04941 (2015).

our phase space.

We can imagine a trajectory of a point passively flowing with this flow field. It is called the phase flow.

> You can use the following software to see examples of 2-vector fields and corresponding phase flows for various initial conditions: https://media.pearsoncmg.com/aw/ide/idefiles/media/JavaTools/twoddfeq.html



$\dot{x} = f(x, y)$
$\dot{y} = g(x, y)$

$\dot{x} = -y$
$\dot{y} = x - y$

Figure 3.2: Vector-Flow demo

3.16 Direction field and solution curve

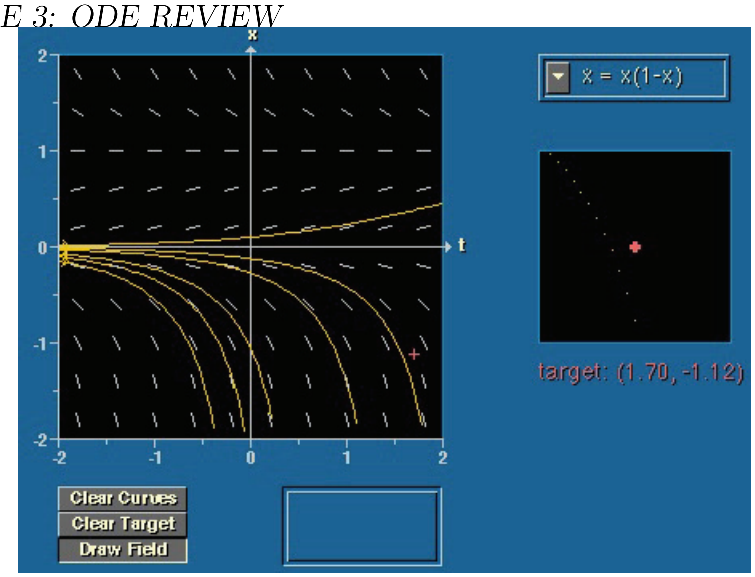
For (3.15) its *direction field* is a vector field $(1, X(x))$ on each (t, x) in $T \times U$, where T is the set of time under consideration. Again we can imagine a trajectory of a point passively flowing with this flow field. It is called the graphs of the solutions of (3.15).

> You can use the following software to see examples of 1-dynamical systems (may be non-autonomous) and corresponding solution graphs for various initial conditions:https://media.pearsoncmg.com/aw/ide/idefiles/media/JavaTools/exunqtrg.html

3.17 Singular points

A singular point of a vector field X is point where the vector field vanishes.

The essence of the general theory of ODE is that as long as the vector field is



$\dot{x} = x(1-x)$

target: (1.70, -1.12)

Figure 3.3: Direction field demo. You can play a shooting game.

nonsingular, unique existence of the solutions and their 'maximally' nice properties (**3.21**, **3.24**) are guaranteed. We proceed as geometrically and intuitively as possible.

3.18 Cauchy-Lipschitz uniqueness theorem.[66]

For (3.15), if X satisfies a Lipschitz condition on D, and if there is a solution passing through $x_0 \in D$, it is unique.

Remark: Even if the Lipshitz condition is not satisfied, If the variables are separable as

$$\frac{dy}{dx} = \frac{Y(y)}{X(x)}, \tag{3.16}$$

and X and Y are continuous and not zero near (x_0, y_0), then the solution near this point is unique. However, the condition is important as we see in the next.

3.19 Rectifiability theorem for vector field

In a sufficiently small neighborhood of any nonsingular point a differentiable vector field is diffeomorphic to the constant field $e_1 = (1, 0, \cdots, 0)^T$.

The graph of the solution never crosses at a non-singular point.

If the original field is C^r, the diffeo can be C^r.

All the basic theorems are more or less straightforward corollaries of the funda-

[66] AMM 116 61 Does Lipschitz with Respect to x Imply Uniqueness for the Differential Equation $y = f(x, y)$? Author(s): José Ángel Cid and Rodrigo López Pouso

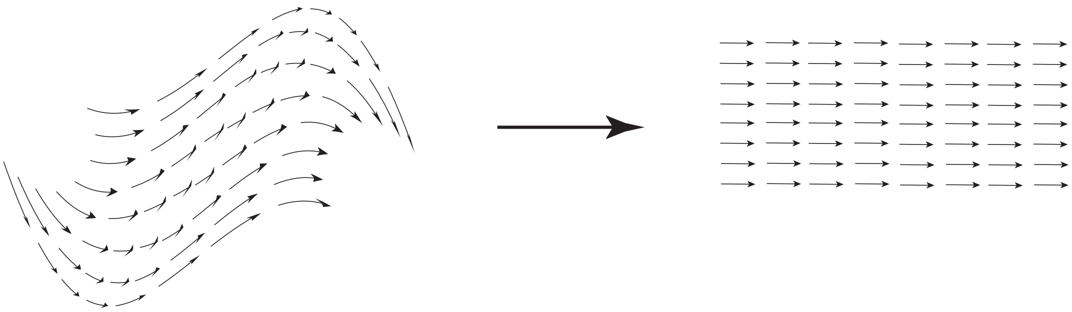


Figure 3.4: Rectification of a vector field

mental theorem.

3.20 Extension theorem[67]

An extension of the solution φ is a solution which coincides with φ on the (time) interval on which φ is defined and which is defined on a greater (time) interval.

Theorem [The extension theorem] Let K be a compact subset of the domain of the ODE (3.15). Then, every solution of this equation with an initial condition in K can be extended to the the boundary of K of infinitely in time (to $\pm\infty$).

3.21 Continuous dependence on initial conditions.

If the vector field is Lipschitz continuous (**3.6**), then the solution at time t depends on the initial condition continuously.

Although this should be intuitively clear, since we need a more quantitative statement later, let us estimate the bounds. We need an important inequality:

3.22 Gronwall's inequality

Let $u, v : [a, b] \to \mathbb{R}$ be continuous nonnegative functions satisfying

$$u(t) \le \alpha + \int_a^t u(s)v(s)ds \tag{3.17}$$

for some α (≥ 0) and for $\forall t \in [a, b]$. Then,

$$u(t) \le \alpha \exp\left(\int_a^b v(s)ds\right). \tag{3.18}$$

[67] Arnold I 2.5 p17

[Demo]
If $\alpha = 0$, then $u(t) = 0$, so we assume $\alpha > 0$. Let us define $\omega(t)$ as

$$\omega(t) = \alpha + \int_a^t u(s)v(s)ds. \tag{3.19}$$

Obviously, $u(t) \leq \omega(t)$.

$\omega(a) = \alpha$ and $\omega(t) \geq \alpha > 0$. As $\omega'(t) = u(t)v(t) \leq v(t)\omega(t)$, we have

$$\omega'(t)/\omega(t) \leq v(t). \tag{3.20}$$

Integrating this, we get the inequality.

3.23 Initial condition dependence

We assume $X \in \mathcal{X}(M)$ is Lipschitz (usually C^r) with the Lipschitz constant L and M is compact. Make two solutions starting from $x_0, y_0 \in M$:

$$x(t) = x_0 + \int_0^t X(x(s))ds, \;\; y(t) = y_0 + \int_0^t X(y(s))ds. \tag{3.21}$$

Then,

$$x(t) - y(t) = x_0 - y_0 + \int_0^t [X(x(s)) - X(y(s))]ds. \tag{3.22}$$

This means

$$\|x(t)-y(t)\| \leq \|x_0-y_0\| + \left\| \int_0^t [X(x(s)) - X(y(s))]ds \right\| \leq \|x_0-y_0\| + \int_0^t \|X(x(s))-X(y(s))\|ds. \tag{3.23}$$

Using the Lipschitz constant we have

$$\|x(t) - y(t)\| \;\leq \|x_0 - y_0\| + L\int_0^t \|x(s) - y(s)\|ds. \tag{3.24}$$

Now, we can apply Gronwall's inequality to obtain

$$\|x(t) - y(t)\| \leq e^{LT}\|x_0 - y_0\|. \tag{3.25}$$

That is, the solution must also be Lipschitz continuous. This explicitly proves **3.21**.

3.24 Smooth dependence on parameter.

If the vector field is smooth, then the solution at finite time is as smooth as the vector field. If the vector field is holomorphic, then the solution is also holomorphic. Then, we can use perturbation theory to obtain the solution in powers of the parameter. This was the idea of Poincaré.

3.25 Picard's successive approximation method[68]

$$\begin{aligned} x_0(t) &= x(0), && (3.26)\\ x_{k+1}(t) &= x(0) + \int_0^t v(s, x_k(s))ds. && (3.27) \end{aligned}$$

[Demo] (may never be given explicitly)
First, we get formally

$$x(t) = x(0) + \int_0^t v(s, x(s))ds. \tag{3.28}$$

If the limit $k \to \infty$ of $x_k(t)$ exists, then obviously (3.27) gives (3.28). Therefore, we need a uniform convergence of the sequence. See Arzela **3.11**.

3.26 History[69]

'Differential equations' began with Leibniz, the Bernoulli brothers and others from the 1680s, not long after Newton's ' fluxional equations' in the 1670s. Applications were made largely to geometry and mechanics; isoperimetrical problems were exercises in optimisation."
According to:
The role of the concept of construction in the transition from inverse tangent problems to differential equations. Henk J. M. Bos p2733
Tangent problems—given a curve, to find its tangents at given points—are as old as classical Greek mathematics. 'Inverse tangent problems' was the name coined in the seventeenth century for problems of the type: given a property of tangents, find a curve whose tangents have that property. It seems that the first such problem was proposed by Florimod De Beaune in 1639. Much of the activities in the early

[68] Arnold 2.4c p16

[69] The History of Differential Equations, 1670-1950 Organised by Thomas Archibald (Wolfville) Craig Fraser (Toronto) Ivor Grattan-Guinness (Middlesex)

infinitesimal calculus (second half of the seventeenth century) were motivated by inverse tangent problems, many of them suggested by the new mechanical theory.

The transition to differential equations occurred around 1700. This transition was much more than a simple translation from figure to formula, from geometry to analytical formalism.

In the seventeenth century to solve this problem is to construct the curve required in the problem. Descartes had restricted geometry to algebraic curves. But inverse tangent problems often had non-algebraic curves as solution. Consequently mathematicians went outside the Cartesian demarcation of geometry and consequently lost a clear and shared conception of what it meant to solve a differential equation; indeed, the status of differential equations became fuzzy: were they problems? were they objects? When were their solutions satisfactory? Many puzzling developments in early analysis, and especially delays in developments expected with hindsight, can be explained by the tenacity of the older ideas on problem solving.

4 Lecture 4: Singularity

4.1 General study of flow of vector field

We have studied what happens in the domain where there is no singularity of X. We can rectify local patches and then connect them maximally to obtain a unique flow. Now, we are left with singularities.

Since X vanishes there, the flow should be slow in the neighborhood, so we have only to study it in some small neighborhood of the singularities. Thus, often linearization of the original X around singularities is effective.

Especially when the singularity is hyperbolic (i.e., linearized vector field may be expressed in terms of linear operator whose spectrum is not on the imaginary axis), the linearized system and the original system are homeomorphic (Hartman's theorem), so we may study the stability of singularities by linearization.

4.2 Simple singularity[70]

$p \in M$ is a simple singularity of $X \in \mathcal{X}^r(M)$ if $DX_p : T_pM \to T_pM$ does not have zero as an eigenvalue.[71]; in other words, the linearized vector field at p may be written as Ax ($A = DX_p$) and A is non-singular.

4.3 Simple singularities are isolated

M as in **4.2**. All the C^r-vector fields on M sufficiently close to X has a simple singularity near p. The position of the zero depends continuously on the vector field.[72] This should be intuitively clear, since the solution to $X = 0$ is locally unique, because A is non-singular, and the solution should depend continuously on X.

Since the derivative of X 'does not vanish', X and the vector field everywhere 0 (= the zero section of TM) should cross transversally.[73] Thus, the isolation of simple singularities should be obvious.

[70] Palis-de Melo p55-

[71] hyperbolicity ⇒ simplicity, but not vice versa.

[72] An official expression is: There exist a neighborhood $\mathcal{N}(X)$, a neighborhood U_p of p and a continuous function $\rho : \mathcal{N}(M) \to U_p$ such that all $Y \in \mathcal{N}(M)$ has a unique zero $\rho(Y) \in U_p$. Palis-de Melo Proposition 3.1.

[73] Proposition 3.2 of Palis-de Melo.

4.4 Vector fields with only simple singularities are structurally stable

More precisely, vector fields $\mathcal{G}_0$ with only simple singularities are open dense in $\mathcal{X}^r(M)$. This follows from **4.3**. Hyperbolic vector fields are also open dense.

The concept for diffeo f corresponding to simplicity is being elementary: if 1 (i.e., identity) is not an eigenvalue of D_pf, f is elementary.

4.5 Linearization of ODE around singularity

For $\dot{(x)} = X(x)$ with a smooth X, near its singularity X should be small, so if we take the singularity at the origin, x is small and $\dot{x}$ is small. Thus, it is a natural idea that the linearized equation

$$\dot{x} = Ax \tag{4.1}$$

with

$$A = \frac{dX}{dx}_{x=0} \tag{4.2}$$

can tell us the local behavior of the system near the singularity.

The justification of the idea is not always possible, but if A has no pure imaginary eigenvalues (the so-called hyperbolic case),[74] the idea goes through. Therefore, let us study linear systems (4.1) fairly in detail first.

You must sense the logical error in the following argument physicists always use: "Let us assume $\|x\|$ is small near the singularity p. Then, we may use linearization (4.1) around p. Since all the eigenvalues have negative real parts, we may conclude p is a stable fixed point," but we have assumed from the start that δx doe not grow (and small to linearize the system!).

4.6 Exponential function of linear operators

If a linear operator is bounded (that is, $\sup_{\|x\|=1} \|Ax\| = \|A\| < \infty$)

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} \tag{4.3}$$

is well defined ($\|e^A\| \leq e^{\|A\|}$).

if $[A, B] = 0$, then

$$e^A e^B = e^{A+B} = e^B e^A. \tag{4.4}$$

[74] Such matrices are open dense. Cf,˙ **5.7**.

Using this, the general solution to (4.1) reads

$$x(t) = e^{At}x_0. \tag{4.5}$$

To compute the matrix representation of e^A it is convenient to 'diagonalize' A. However, unless A is normal ($AA^* = A^*A$) (and also allow the use of complex eigenvalues), this is not possible. If A is defined on a complex vector field, we can still use the Jordan normal form (see appendix). Therefore, we use a trick called 'complexification.'

4.7 Complexification[75]

Since eigenvalues of real matrices are generally complex, if we wish to use linear operator theory fully, it is convenient to consider A as an operator on $\mathbb{C}^n$ instead of its original domain $\mathbb{R}^n$. To realize this, we introduce 'complexification' $\mathcal{C}$.

Since any vector in $\mathbb{C}^n$ may be written as $a + ib$ ($a, b \in \mathbb{R}^n$), we can complexify $A \to \mathcal{C}(A) = \tilde{A}$ according to

$$\mathcal{C}(A)(a + ib) = Aa + iAb. \tag{4.6}$$

Since $\mathcal{C}$ is linear by definition, to preserve the algebraic (i.e., the ring) structure of matrices, we should show the following:

$$\mathcal{C}(AB) = \mathcal{C}(A)\mathcal{C}(B). \tag{4.7}$$

From this we know

$$\mathcal{C}(e^A) = e^{\mathcal{C}(A)}. \tag{4.8}$$

We can also show (recall the definition of the norm)

$$\|\mathcal{C}(A)\| = \|A\|. \tag{4.9}$$

4.8 Complex and real diagonalization of real matrix

We know normal matrices (satisfying $AA^* = A^*A$) may be unitary diagonalized on

[75] A very kind explanation is found in M. W. Hirsch and S. Smale, *Differential equations, dynamical systems and linear algebra* (Academic Press, 1974), Chapter 4.

the complex vector space.[76] However, this is not generally possible on the real vector space. The eigenvectors corresponding to λ and $\overline{\lambda}$ e_1 and e_2 (respectively):

$$\tilde{A}e_1 = \lambda e_1, \quad \tilde{A}e_2 = \overline{\lambda} e_2 \tag{4.10}$$

cannot be chosen in $\mathbb{R}^n$. Notice that we may choose $e_2 = \overline{e_1}$, so we choose two real vectors $a = e_1 + e_2$ and $b = (e_1 - e_2)/i$ among the basis.[77] Diagonalization of A is not possible, but still 'almost diagonalized' real matrix may be obtained.

To make the situation crisp clear, consider a 2×2 matrix that can be diagonalized on $\mathbb{C}^2$ as

$$\begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix} \tag{4.11}$$

with the basis $\{e_1, e_2\}$. $\lambda = \alpha + i\beta$. If we use the basis $\{a, b\}$ the above diagonalized matrix reads

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \tag{4.12}$$

Since the 'real translation' is straightforward, we consider everything on $\mathbb{C}^n$.

4.9 Complexification of linear ODE

(4.1) on $\mathbb{R}^n$ can be complexified as

$$\dot{z} = \mathcal{C}(A)z, \tag{4.13}$$

where $z = x + iy$. This consists of two real equations $\dot{x} = Ax$ and $\dot{y} = Ay$.

How to recover the real solution from the full complexified computation is explained in **4.8**.[78]

4.10 Singularity of 2-real vector field

The singularity of 2-dimensional system may be classified according to the Jordan normal form of A.

(1) 0 is a sink: stable fixed point.

$$\text{(a)} \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} (\lambda, \mu < 0), \quad \text{(b)} \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, (a < 0, b \neq 0), \quad \text{(c)} \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} (\lambda < 0). \tag{4.14}$$

[76] Normality is a necessary and sufficient condition for A to be diagonalized by a unitary transformation.

[77] If you wish to normalize them, divide them with $\sqrt{2}$.

[78] Detailed examples can be seen in Hirsch+Smale, so I will not dwell on examples.

For case (b) the origin is called a focus. (c) is a bicritical node.[79]

(2) 0 is a source: in any direction it is unstable.

$$\text{(d)}\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} (\lambda, \mu > 0), \quad \text{(e)}\begin{pmatrix} a & b \\ -b & a \end{pmatrix}, (a > 0, b \neq 0), \quad \text{(f)}\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} (\lambda > 0). \tag{4.15}$$

For case (e) the origin is called a focus. (f) is a bicritical node.

(3) 0 is a saddle: there is one stable direction and one unstable direction.

$$\text{(g)}\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} (\lambda > 0 > \mu). \tag{4.16}$$

These are hyperbolic cases. The phase portraits look like Fig. 4.1.

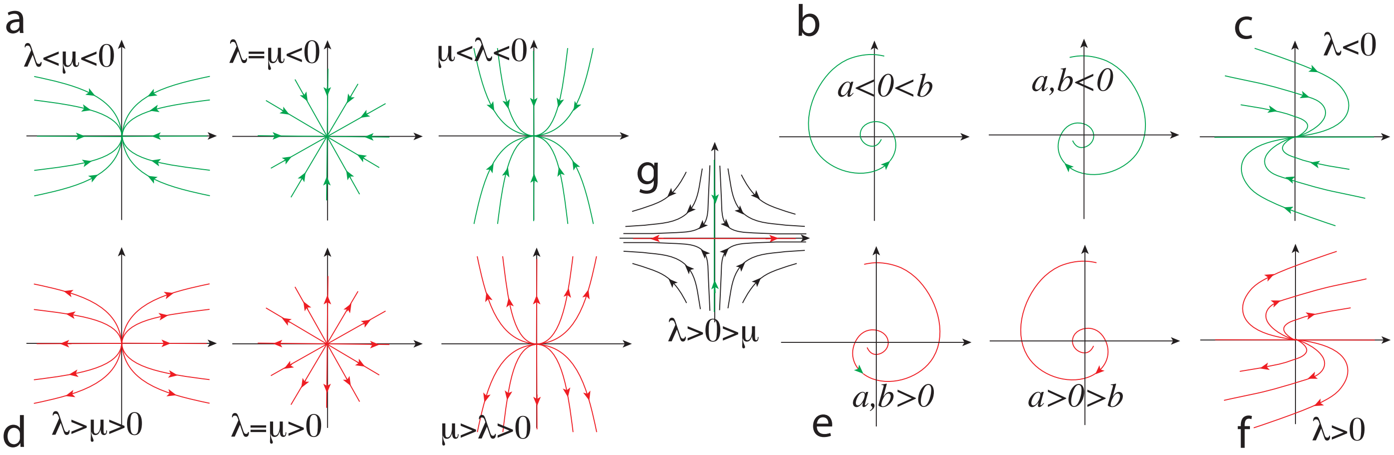


Figure 4.1: Linear flows on the plane

The non-hyperbolic cases cannot be classified without referring to the higher order terms. That is, linearization does not preserve qualitative nature of the singularity.

(4) Non-hyperbolic case.

$$\text{(h)}\begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix} (b \neq 0), \quad \text{(i)}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{4.17}$$

[79] In this case the flow looks, e.g., like $x = e^{t\lambda}x_0 + te^{t\lambda}y_0$, $y = e^{t\lambda}y_0$.

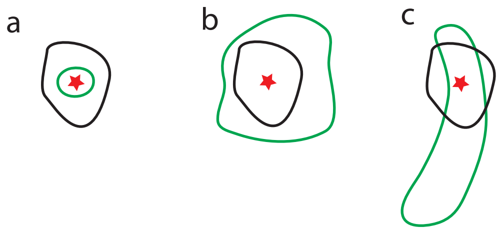
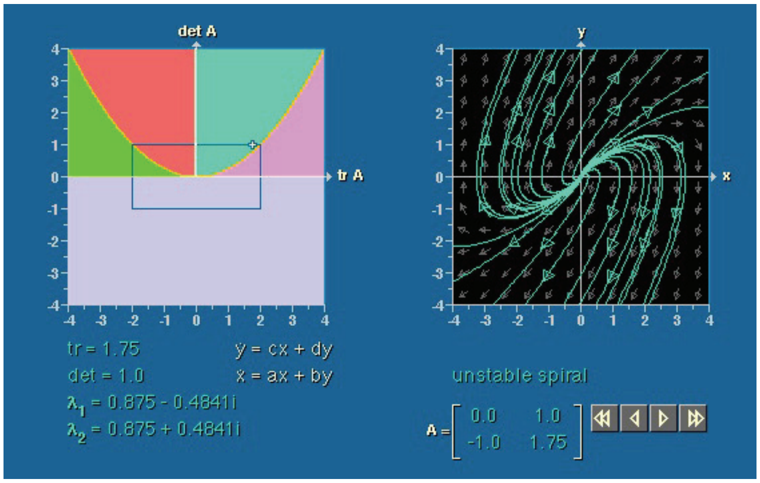


Figure 4.2: Another illustration of (hyperbolic) singularities Black 'before', Green, 'after' The start is the critical point = fixed points. a: sink, b: source, c: saddle

Classification (see Fig. 4.3 for an example): `https://media.pearsoncmg.com/aw/ide/idefiles/media/JavaTools/lnclppan.html`. This demo inevitably includes non-simple zero cases. Watch what happens.



tr = 1.75

det = 1.0

$\lambda_1 = 0.875 - 0.4841i$

$\lambda_2 = 0.875 + 0.4841i$

y = cx + dy

x = ax + by

unstable spiral

$$A = \begin{bmatrix} 0.0 & 1.0 \\ -1.0 & 1.75 \end{bmatrix}$$

Figure 4.3: Classification

Appendix 1. Jordan normal form

4.11 Generalized eigenspace[80]

The eigenvalues are the roots of the characteristic equation

$$\det(A - xI) = \prod_k (\lambda_k - x)^{n_k} = 0. \tag{4.18}$$

[80] A very kind explanation of the Jordan normal form is found in M. W. Hirsch and S. Smale, *Differential equations, dynamical systems and linear algebra* (Academic Press, 1974), Chapter 6.

Here n_k is the multiplicity of λ_k. $\mathrm{Ker}(A-\lambda_k I) = V_k$ is called the eigenspace of λ_k.[81]

$\mathrm{Ker}(A-\lambda_k I)^{n_k} = V_k$ is called the generalized eigenspace of A belonging to λ_k.

4.12 The fundamental decomposition theorem

Let A be a linear operator on $V = \mathbb{C}^n$. Then, the dimension of the generalized eigenspace V_k is n_k and $\oplus_k V_k = V$.

This is proved in **4.13** and **4.14**.

4.13 Basic decomposition lemma

The key part of **4.12** is: for any linear $A : V \to V$

$$V = M \oplus N, \tag{4.19}$$

where $M = \cap_{j\in\mathbb{N}^+} A^j V$ and $N = \cup_{j\in\mathbb{N}^+}\mathrm{Ker}(A^j)$.[82]
[Demo] Let $M_j = A^j V$ and $N_j = \mathrm{Ker}(A^j)$. Then,

$$\begin{aligned} 0 &= N_0 \subset N_1 \subset \cdots \subset N_j \subset N_{j+1} \subset \cdots \subset V, & (4.20)\\ V &= M_0 \supset M_1 \supset \cdots \supset M_j \supset M_{j+1} \supset \cdots \supset 0. & (4.21)\end{aligned}$$

Since V is finite dimensional, there must be $n, m \in \mathbb{N}^+$ such that $j \geq m \Rightarrow M_j = M_m$ and $j \geq n \Rightarrow N_j = N_n$. Let $M = M_m$ and $N = N_n$. Since $A^n M = M$, if $x \in M$ is not zero, then $A^n x \neq 0$. On the other hand, $A^n N = 0$, so for any $y \in N$ $A^n y = 0$. Therefore, $M \cap N = \{0\}$.

For any $y \in V$ let $x \in M$ such that $A^m x = A^m y$; such x exists, because $A^m y \in M$ and A^m is reversible on M. Then, $z = y - x$ is in $N_m \subset N$, so $V = M \oplus N$.

4.14 Concluding demo of fundamental decomposition theorem

Instead of A, let us consider $A_k = A - \lambda_k I$. Let us define $M_k = \cap_{j\in\mathbb{N}^+} A_k^j V$ and $N_k = \cup_{j\in\mathbb{N}^+}\mathrm{Ker}(A_k^j)$. Applying the fundamental decomposition theorem to A_1, we get $V = N_1 \oplus M_1$.

Now, restricting A_2 on M_1 we can repeat the argument above. $M_1 = N_2 \oplus M_2'$, where M_2' is the orthogonal complement of N_2 in M_1. Thus, we would arrive at

$$V = \oplus_k N_k. \tag{4.22}$$

If we prove $N_k = V_k$, we are done. Obviously $V_k \subset N_k$. This follows from the following Lemma. We need one definition:

N is a nilpotent operator if for some $m \in \mathbb{N}^+$ $N^m = 0$ (in the domain). Notice that $A_k = A - \lambda_k I$ is a nilpotent operator, if restricted to N_k.

Lemma. Let n be the smallest positive integer such that $N^n x = 0$ for $x \in V$. Then, $\{x, Nx, \cdots, N^{n-1}x\}$ is a basis of subspace $Z = \{x, Nx, \ldots, N^j x, \cdots\}$ (called a cyclic subspace) of V.

[81] If $\oplus_k V_k = \mathbb{C}^n$, we say A is unitary diagonalizable. Notice that, generally speaking, even if you collect all the eigenvectors, the result can only span a genuine subset of the original vector space.

[82] Here, the demonstration follows Hirsch-Smale, but if you know the basic theorem of linear algebra $\mathrm{Im}(A) + \mathrm{Ker}(A) = V$ for any linear operator A, it is obvious.

[Demo] Since $N^n x = 0$, obviously $Q = \{x, Nx, \cdots, N^{n-1}x\}$ can generate Z. Thus, we have only to show that Q is linearly independent. If not,

$$\sum_{k=1}^{n-1} a_k N^k x = 0 \tag{4.23}$$

has a nontrivial solution with the first nonzero coefficient a_j. Operating N^{n-j-1} we have

$$a_j N^{n-1}x + N^{n-j-1} \sum_{k=j+1}^{n-1} a_k N^k x = a_j N^{n-1} x = 0. \tag{4.24}$$

This contradicts the definition of n.

4.15 A on V_k

Since V is decomposed into the direct sum of V_k, we have handle each subspace separately. Let us consider A restricted on V_k. Let $S = \lambda_k I$ (here I is n_k-dimensional identity) and $N = A - \lambda_k I$. N is nilpotent, $A = N + S$ and $SN = NS$. This decomposition is unique.

This implies that V can be decomposed into the diagonal S and nilpotent N uniquely as $A = N + S$, $NS = SN$ on V.

The lemma in 4.14 tells us that we can choose $U = \{x, Nx, \cdots N^{n_k-1}x\}$ as the basis of V_k. The subspace spanned by $U_{-1} = \{x, Nx, \cdots N^{n_k-2}x\}$ is one dimension smaller than V_k, so we can choose a vector y in V_k but orthogonal to this smaller space U_{-1}. Now $Ny \in U_{-1}$ and is orthogonal to y, Notice that Ny is not in $U_{-2} = \{x, Nx, \cdots N^{n_k-3}x\}$ but may not be orthogonal to U_{-2}. Thus we make y' such that $Ny = Ny' + z$, where Ny' is orthogonal to U_{-2} and $z \in U_{-2}$. Repeating this argument, we can make $N^j q$ orthogonal to $N^k q$ ($j \neq k$). That is, we can choose q such that $\{q, Nq, \cdots, N^{n_k-1}q\}$ makes an orthogonal basis. With this basis, N is expressed as an $n_k \times n_k$ matrix:

$$N = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ & & & \cdots & & & \\ & & & \cdots & & & \\ & & & \cdots & & & \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \tag{4.25}$$

This means A on V_k has the following $n_k \times n_k$ matrix:

$$A = \begin{pmatrix} \lambda_k & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & \lambda_k & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \lambda_k & \cdots & 0 & 0 & 0 \\ & & & \cdots & & & \\ & & & \cdots & & & \\ & & & \cdots & & & \\ 0 & 0 & 0 & \cdots & 1 & \lambda_k & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_k \end{pmatrix} \tag{4.26}$$

This is called a Jordan cell.

4.16 Jordan normal form

A we have been considering can be represented as the direct sum of the Jordan cells. The number of cells with the same eigenvalue λ is given by dim Ker$(A - \lambda I)$.

4.17 Detailed example[83]

Let us solve the following linear ODE $\dot{x} = Tx$ that reads wrt a coordinate system as :

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{pmatrix} -1 & 1 & -2 \\ 0 & -1 & 4 \\ 0 & 0 & 1 \end{pmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{4.27}$$

Let us write

$$T_0 = \begin{pmatrix} -1 & 1 & -2 \\ 0 & -1 & 4 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.28}$$

The solution is $e^{T_0 t}\boldsymbol{x}_0$, where $\boldsymbol{x}_0$ is the initial condition. We wish to calculate this explicitly. Notice that

$$\begin{pmatrix} -1 & 1 & -2 \\ 0 & -1 & 4 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ -2 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 6 & -9 & -2 \\ -9 & 17 & 4 \\ -2 & 4 & 1 \end{pmatrix} \tag{4.29}$$

which is not equal to

$$\begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ -2 & 4 & 1 \end{pmatrix}\begin{pmatrix} -1 & 1 & -2 \\ 0 & -1 & 4 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 2 & -6 \\ 2 & -8 & 21 \end{pmatrix} \tag{4.30}$$

so we cannot diagonalize T_0 with an orthogonal transformation. The eigenvalues are obviously -1, -1 and 1. The rank of

$$T_0 + 1 = \begin{pmatrix} 0 & 1 & -2 \\ 0 & 0 & 4 \\ 0 & 0 & 2 \end{pmatrix} \tag{4.31}$$

[83] Hirsh-Smale Ex2 Chapter 6, but all the details are filled so that you do not need any pencil nor paper for follow all the details.

is 2, so its kernel is 1 dimension spanned by $(1,0,0)^T$.[84] Thus the eigenvalue -1 has algebraic multiplicity 2 and geometrical multiplicity 1. Thus we need the generalized eigenspace of -1 defined by

$$0=(T_0+1)^2\boldsymbol{x}=\begin{pmatrix}0&1&-2\\0&0&4\\0&0&2\end{pmatrix}\begin{pmatrix}0&1&-2\\0&0&4\\0&0&2\end{pmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{pmatrix}0&0&0\\0&0&8\\0&0&4\end{pmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}\tag{4.32}$$

Therefore, the generalized eigenspace is spanned by $(1,0,0)^T$ and $(0,1,0)^T$. The (generalized) eigenspace of 1 is given by

$$0=(T_0-1)\boldsymbol{x}=\begin{pmatrix}-2&1&-2\\0&-2&4\\0&0&0\end{pmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}\tag{4.33}$$

This implies that z is arbitrary, say 1, and $-y+2z=0$ implies $y=2$ and $x=0$: Thus, $\{|a\rangle,|b\rangle,|c\rangle\}=\{(1,0,0)^T,(0,1,0)^T,(0,2,1)^T\}$ is a basis (not orthogonal!) wrt which $T=S+N$, where S is diagonal $-1\oplus-1\oplus1$ and N is nilpotent according to the general theorem.

Let us rewrite T wrt to this basis (which is denoted as T_1). $PT_1=T_0P$ or $T_1=P^{-1}T_0P$. We need[85]

$$\langle a|T|b\rangle=\sum_{x,y}\langle a|x\rangle\langle x|T|y\rangle\langle y|b\rangle\tag{4.34}$$

Notice that

$$P=\langle y|b\rangle=\begin{pmatrix}1&0&0\\0&1&2\\0&0&1\end{pmatrix}\tag{4.35}$$

Therefore, P^{-1} is given by (Note that $\det P=1$)

$$P^{-1}=\begin{pmatrix}1&0&0\\0&1&-2\\0&0&1\end{pmatrix}\tag{4.36}$$

[84] Here we have used the basic theorem: for a linear map T from a vector space V to another vector space W, dim V = dim(Im T) + dim(Ker T).

[85] We use the bra-ket notation for the ‘mnemonic sake’; since our kets are not orthonormal, the ‘transposition’ is actually to compute the transposition of the inverse matrix. The mnemonics works perfectly. Look at

$$\sum_x\langle a|x\rangle\langle x|b\rangle=\delta_{ab}.$$

$(P)_{xb}=\langle x|b\rangle$, $(P^{-1})_{ax}=\langle a|x\rangle$.

Thus, $T_1 = P^{-1}T_0P$:

$$T_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} T_0 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & -2 \\ 0 & -1 & 4 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.37}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.38}$$

Thus

$$S_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad N_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.39}$$

Obviously, $[S_1, N_1] = 0$. This implies

$$N_0 = PN_1P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.40}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.41}$$

This tells us that

$$T_0 = \begin{pmatrix} -1 & 1 & -2 \\ 0 & -1 & 4 \\ 0 & 0 & 1 \end{pmatrix} = S_0 + N_0 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 4 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.42}$$

Notice that, as the general theory tells us, $[S_0, N_0] = 0$. Therefore,

$$e^{tT_0} = e^{tS_0}e^{tN_0}. \tag{4.43}$$

Since N_0 is nilpotent (actually $N_0^2 = 0$:

$$\begin{pmatrix} 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.44}$$

we have

$$e^{tN_0} = 1 + tN_0 = \begin{pmatrix} 1 & t & -2t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.45}$$

To compute e^{tS_0}, we should use

$$e^{tS_0} = Pe^{tS_1}P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{t} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.46}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-t} & -2e^{-t} \\ 0 & 0 & e^{t} \end{pmatrix} = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-t} & -2e^{-t}+2e^{t} \\ 0 & 0 & e^{t} \end{pmatrix} \tag{4.47}$$

Thus we have arrived at the final answer:

$$e^{tT_0} = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-t} & -2e^{-t}+2e^{t} \\ 0 & 0 & e^{t} \end{pmatrix} \begin{pmatrix} 1 & t & -2t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{-t} & te^{-t} & -2te^{-t} \\ 0 & e^{-t} & -2e^{-t}+2e^{t} \\ 0 & 0 & e^{t} \end{pmatrix} \tag{4.48}$$

Appendix 2. Degree theory

The aim of this appendix is very restricted: to understand the degree of the vector field with only simple singularities. Illustrations are only in 2-space.

4.18 Index of singularity
Take a ball B containing only one singularity of a vector field $X : B \to \mathbb{R}^n$. Thus, X has no singularity on $\partial B = S^{d-1}$ if the vector field is d-dimensional. deg(X, B) is defined as the degree of the map $f : \partial B \to \mathbb{R}^d$, which is defined as follows:

$$\deg(X, B) = p - q, \tag{4.49}$$

where p (resp., q) is the number of points in $f^{-1}(x)$ $(x \in \partial B)$ with det df being positive (resp., negative).

4.19 Examples of simple singularities
For 2-vector fields,
The indices of a sink, a source or a center is $+1$.
The index of a hyperbolic saddle is -1.
See Fig. 4.4.

The map $f : D \to \mathbb{R}^2$, where D is a disk around the origin of the base space, corresponds to the relation between the red and the black arrows. Thus, to obtain the degrees for 2-space examples, we have only to see how many times the black arrows rotate when the base vector rotate once (i.e., when we ground ∂D once).

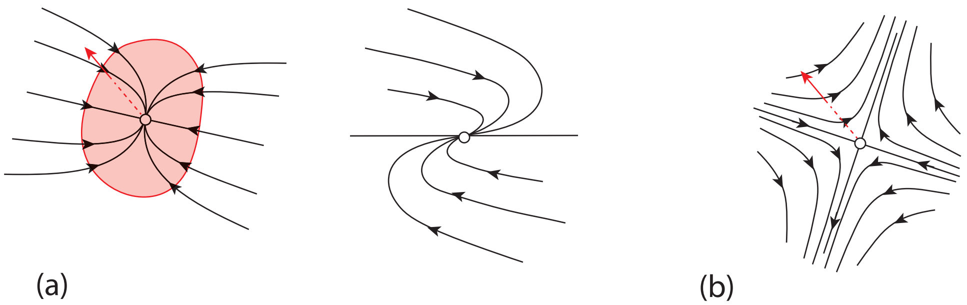


Figure 4.4: (a) with index +1, (b) −1. Red arrows show the base space vectors.

4.20 Total sum of indices = degree

The degree of a vector field X with simple singularities is defined as

$$\deg(X, D) = \sum_{s \in D} \text{index} s, \tag{4.50}$$

where s is a singularity.

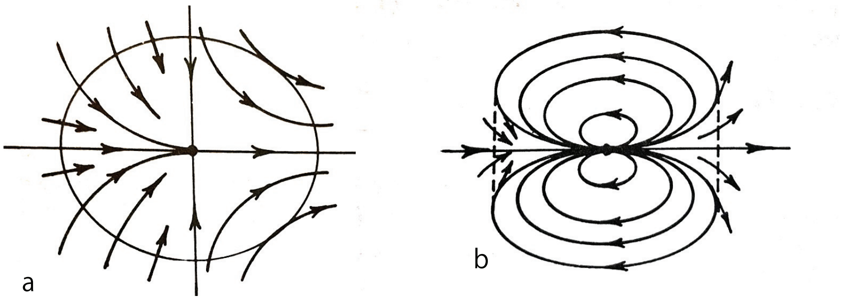


Figure 4.5: (a) with degree 0 (index zero), (b) index +2.

In Fig. 4.5 (a) we may interpret the zero due to merging of a sink and a saddle. Thus, $+1 + (-1) = 0$ is the degree of the disk in the figure. We get the same answer even if we follow the original definition. For (b) we could follow the rotation of the vector along a small circle around the singular point (or from the merging of a sink and a source).

4.21 Poincaré-Hopf's theorem on degrees

Let M be a compact orientable manifold, and X be a differentiable vector field on M with finitely many singular points. Then,

$$\sum_{z \in M} \text{index}(z) = \chi(M), \tag{4.51}$$

where the sum is over all the singular points on M, and $\chi(M)$ is the Euler index[86] That is, the degree of $\mathcal{X}^r(M)$ is identical to the Euler index of M.
[Demo]
If we accept that the LHS of (4.51) does not depend on X on M, then we can compute it with a convenient X. For a polyhedron assign a sink to a vertex, a source to a surface, and saddle to an edge (Fig. 4.6).

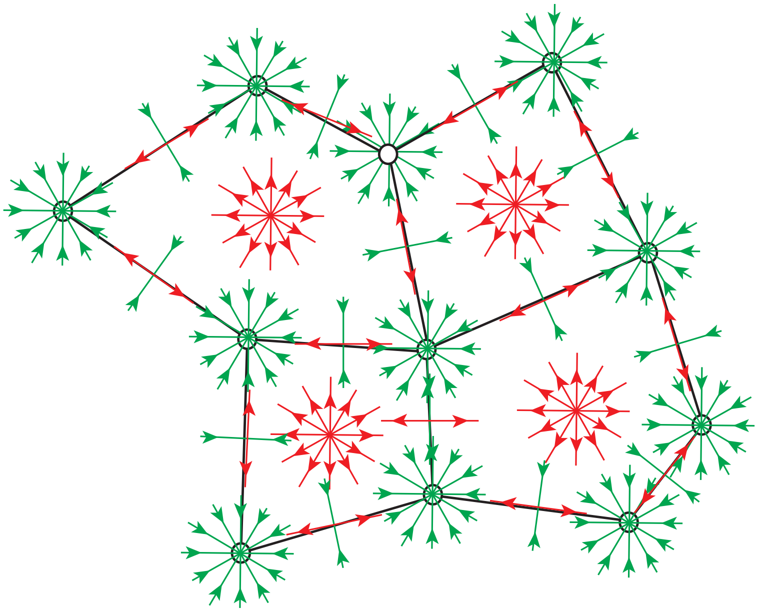


Figure 4.6: Demo of Poincare-Hopf theorem

We can construct X. For this obviously (4.51) holds.

[86] or Euler characteristics defined by

$$\chi(M)=\sum_{i=0}^{\infty}(-1)^i b_i,$$

where b_i is the ith Betti number. For a CW complex,

$$\chi(M)=\sum_{i=0}^{\infty}(-1)^i n_i,$$

where n_i is the number of i-simplexes in M. In particular, we get $\chi(M)=V-E+F+\cdots+(-1)^n X_n+\cdots=1$, where V: number of vertices, E: number of edges, F: number of faces, B: number of 3-simplexes. In 3-space of a polyhedron (surface) this is always 2. More generally, $\chi-2-2g$, where g is the genus of the manifold (the number of handles attached to the sphere or the number of holes; Solid torus has $g=1$.

4.22 Degree-theoretical constraints on singularities

Poincare-Hopf's theorem imposes a strong constraints on the existence of various singularities of the vector field on a manifold. For example, if the vector field is on S^2, it is impossible to have a single source or saddle. If there is a saddle, there must be at least three other sources/sinks. However, on T^2 a single source-saddle pair can live happily.

In Fig. 4.5 (a) can be on T^2 without any other singularity, and (b) on S^2. See Fig. 5.2.

5 Lecture 5: Hyperbolicity

5.1 Hyperbolic fixed points

We have already introduced a singular point x of a vector field $X \in \mathcal{X}^r(M)$, where $r \geq 1$ and M is (usually) a compact n-manifold: $X(x) = 0$. Its derivative at x is a linear operator $D_xX = A$. A is a $n \times n$ matrix[87].

If A has eigenvalues whose real parts are nonzero (i..e, not neutral), we say x is a hyperbolic fixed point.

5.2 Invariant set

For a flow ϕ_t for $\mathcal{X}^r(M)$ (or map $f \in C^r(M)$), a subset $S \subset M$ is an invariant set, if for any $x \in S$, $\phi_t(x) \in S$ for all $t \in \mathbb{R}$ ($f^n(x) \in S$ for any $n \in \mathbb{Z}$).

Needless to say, singular points (= fixed points) are in the invariant set of the dynamical system.

5.3 Non-wandering set[88]

A nonwandering point of a dynamical system $\mathcal{X}^r(M)$ (or map $f \in C^r(M)$) is a point such that for its any nbh U there is t such that $\phi_t(U) \cap U \neq \emptyset$ ($n \in \mathbb{Z}$ such that $f^n(U) \cap U \neq \emptyset$). The totality of non-wandering points is the non-wandering set of the dynamical system.

5.4 ω and α limit sets

Long time behaviors of a dynamical system is studied by limit sets describing the $t \to \pm\infty$ behaviors of trajectories.

ω-limit set of x: the set of accumulation points of $\phi_t(x)$ for $t > 0$. That is,
$\omega(x) = \{y \mid \lim_i \phi_{t_i}(x) = y, t_i \to \infty\}$.

α-limit set of x: the set of accumulation points of $\phi_t(x)$ for $t < 0$. That is,

[87] Of course, we choose an appropriate chart, but as we have discussed at length in Section 2, we may assume our world is a (bounded) subset of $\mathbb{R}^n$.

[88] There are several different definitions. For example,
(i) x is a non-wandering point if there is a nbh U and τ such that $\phi_t(U) \cap U \neq \emptyset$ for $t > \tau$.
(ii) x is a non-wandering point if for any nbh U there is $t > 0$ such that $\phi_t(U) \cap U \neq \emptyset$.
Notice that $\phi_{-t}(\phi_t(U) \cap U) = U \cap \phi_{-t}(U)$, so the direction of time does not matter.

$\alpha(x) = \{y \mid \lim_i \phi_{t_i}(x) = y, t_i \to -\infty\}$.

5.5 Some properties of limit sets

The following statement should be intuitively clear.

Proposition. ω and α limit sets of a point x are non-wandering closed sets. [Wandering sets are open.]

Proposition. If N is a invariant set, then ∂N, N°, $\overline{N}$ and N^c are invariant. [This is due to the continuity of ϕ_t.]

Proposition. The totality of the non-wandering sets is an invariant set.

Any point in periodic orbits and fixed points is non-wandering but there are more subtle nonwandering points. If an orbit densely fill a domain, then any point in(side) the domain is non-wandering, although it need not be periodic nor fixed point.

Proposition. For any point in the non-empty compact invariant set, its α and ω limit sets are non-empty. [If x_0 is in a compact invariant set, then $\{\phi_t(x_0)\}$ is in a compact set, so its accumulation points are in the same compact set.]

5.6 Attracting set

A closed invariant set $A \subset M$ is an attracting set, if there is a nbh of A such that any $x \in U$ stays in U (i.e., $\phi_t(x) \in U$ for all $t > 0$) and $\phi_t(x) \to A$.

The domain of attraction of A (basin of A) is $\cup_{T \leq 0} \phi_t(U)$.

By reversing time $t \to -t$ we can analogously define repelling sets.

Even in 1D an attracting set can be complicated as the following example by Ruelle shows:

$$\dot{x} = -x^4 \sin\frac{\pi}{x}. \tag{5.1}$$

5.7 Hyperbolic linear vector field

If the spectrum of $L \in \mathcal{L}(\mathbb{R}^n)$ is disjoint from the imaginary axis, L is called hyperbolic. The number of eigenvalues with negative real part is called the index of L. Hyperbolic linear fields are open dense in $\mathcal{L}(\mathbb{R}^n)$.

If L is hyperbolic, then there are invariant subspaces E^s and E^u such that $\mathbb{R}^n = E^s \oplus E^u$. $L|_{E^u}$ has positive real eigenvalues and $L|_{E^s}$ has negative real eigenvalues.

Two hyperbolic linear vector fields are topologically conjugate iff their indices are identical.

5.8 Hyperbolicity

For a hyperbolic fixed point p we can linearize the dynamics, and then study the linearized system as a dynamics on $\mathbb{R}^n$. Its eigenspace for stable eigenvalues (on the left half space for vectors and inside the unit circle for maps) may be understood as a tangent subspace E^s of T_pM. We can collect trajectories tangent to the vectors that is a locally invariant submanifold of M. This is the local stable mfd for p denoted as $W_U^{s\,\mathrm{loc}}(p)$, where U is an appropriate nbh of p (see **5.9**).

Reversing the time we can define a local unstable manifold as well.

We have seen in **4.4** that $\mathcal{G}_0$ the vector fields with only simple singularities) is open-dense in $\mathcal{X}^r(M)$.[89] We can show that

Theorem. The vector fields $\mathcal{G}_\tau$ whose singularities are all hyperbolic is open-dense in $\mathcal{X}^r(M)$.[90]
[Demo']
We have only to show that $\mathcal{G}_\tau$ is open-dense in $\mathcal{G}_0$, but at least intuitively this should not be surprising.

5.9 Local stable manifold

Let p be a fixed point of $f \in C^r(M, M)$ and U be a neighborhood. The local stable manifold $W_U^{\mathrm{loc}}(p)$ in U is given by

$$W_U^{s\,\mathrm{loc}}(p) = \{q \in U \mid f^n(q) \in U \text{ for } \forall n \in \mathbb{N}\}. \tag{5.2}$$

5.10 Stable manifold

The stable manifold $W^s(p)$ of p is defined as

$$W^s(p) = \cup_{n=0}^{\infty} f^{-n}\left(W_U^{s\,\mathrm{loc}}(p)\right) \tag{5.3}$$

That is, the totality of the points eventually mapped to p is its stable manifold.

The stable manifold of f^{-1} is the unstable manifold.

[89] PdM p56

[90] PdM p58

Remark: $W^s(p)$ need not a a submanifold of M.

5.11 Stable manifold theorem

Let p be a hyperbolic fixed point of $f \in C^r(M,M)$. Thus, $T_pM = V^s \oplus V^u$, where V^s (resp. V^u) is the vector space on which T_pf is contracting (resp., expanding). Then, there is a contraction $g : W^s(p) \to W^s(p)$ and embedding $J : W^s(p) \to M$ such that $Jg = fJ$. Furthermore, $T_pJ : T_p(W^s(p)) \to V^s$ is an isomorphism.

5.12 Example: Renormalization group flow in the Hamiltonian space

We may interpret the renormalization group transformation as a map from a (generalized) canonical distribution μ to another (generalized) canonical distribution $\mu' = \mathcal{R}\mu$. We can imagine effective Hamiltonians H and H' (it is customary that β is absorbed in H's) according to

$$\mu = \frac{1}{Z}e^{-H}, \ \mu' = \frac{1}{Z'}e^{-H'}. \tag{5.4}$$

We may write $H' = \mathcal{R}H$. Therefore, we can imagine that successive applications of $\mathcal{R}$ defines a flow (RG flow) in the space of Hamiltonians (or models or systems). This idea is illustrated in Fig. 5.1.

In Fig. 5.1 H^* is a fixed point with an infinite correlation length of the RG flow. Its stable manifold is called the *critical surface*. The Hamiltonian of the actual material, say, magnet A, changes (do not forget that β is included in the definition of the Hamiltonian in (5.4)) as the temperature changes along the trajectory denoted by the curve with 'magnet A.' It crosses the critical surface at its critical temperature. The renormalization transformation uses the actual microscopic Hamiltonian of magnet A at various temperatures as its initial conditions. Three representative RG flows for magnet A are depicted. 'a' is slightly above the critical temperature, 'b' exactly at T_c of magnet A ('b′' is the corresponding RG trajectory for magnet B, a different material; both b and b′' are on the critical surface), 'c' slightly below the critical temperature. Do not confuse the trajectory (black curve) of the actual microscopic system as temperature changes and the trajectories (successive arrows; RG flow) produced by the RG transformation.

If we understand H^*, we understand all the universal features of the critical behaviors of all the magnets crossing its critical surface.

http://www.youtube.com/watch?v=MxRddFrEnPc a video by Douglas Ashton.

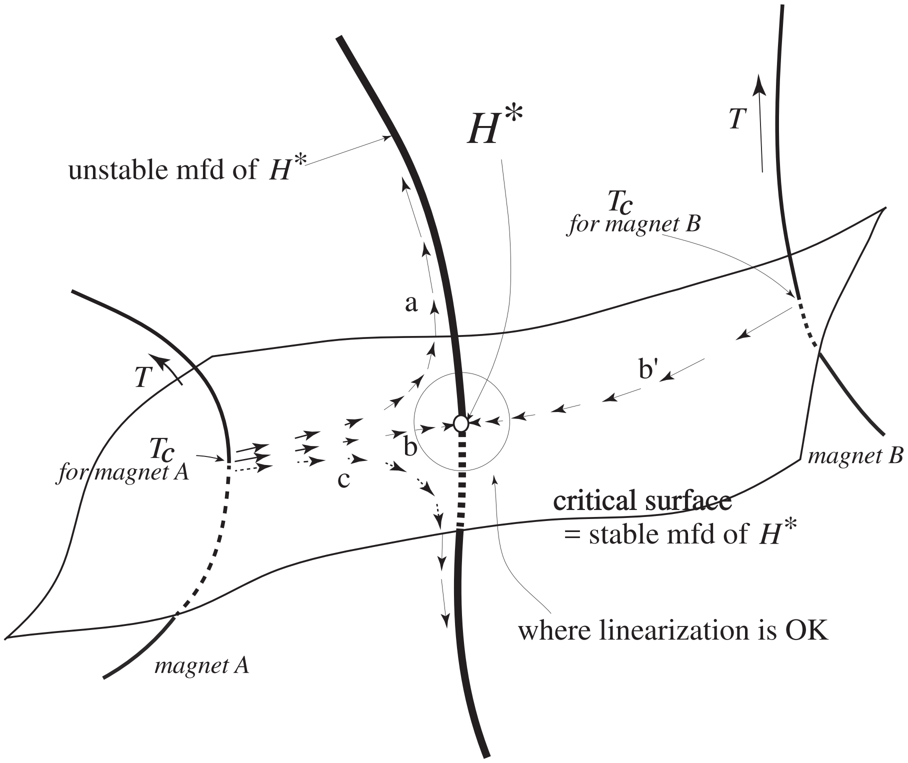


Figure 5.1: A global picture of renormalization group flow in the Hamiltonian space $\mathcal{H}$. The explanation is in the text. 'mfd' = manifold. The thick curves emanating from H^* denote the direction that the Hamiltonians are driven away from the fixed point by renormalization.

5.13 Hartman's theorem

Let $X \in \mathcal{X}^r(M)$ and $p \in M$ be a hyperbolic singularity of X. Then, X is locally equivalent to its linearization.

Here equivalence means topological conjugacy. That is, for $A = DX_p$, there is a continuous map h such that $hx(t) = e^{At}hx_0$.

Notice that h is a homeo, not a diffeo. If we wish to have a diffeomorphic conjugation, then there is a strong relation between the two flow velocities, but such relations are already fixed by the two vector fields we are comparing. Thus, such h may not be chosen.[91]

To prove this theorem is to construct h.

5.14 Strategy to show Hartman's theorem

The linearized system has $L_t = e^{At}$ as the evolution operator. The evolution operator for the original system may be written as a sum of L_t and the deviation from it ϕ_t (i.e., $\varphi_t = L_t + \phi_t$).
(i) There is a homeo h such that $h(L_\tau + \phi_\tau) = L_\tau h$ for some (perhaps small) τ.
(ii) The following H

$$H = \int_0^\tau e^{-As} h\varphi_s ds \tag{5.5}$$

is actually $H = h$ and $H\varphi_s = L_s H$ for $\forall s \in \mathbb{R}$.

Let us show the last statement holds, assuming we have constructed h. Its construction is in **5.15**-.

$$L_{-s}H\varphi_s = L_{-s}\int_0^\tau L_{-t}h\varphi_t dt \varphi_s = \int_0^\tau L_{-t-s}h\varphi_{t+s}dt. \tag{5.6}$$

Let us introduce $u = t + s - \tau$. Then,

$$\begin{aligned}\int_0^\tau L_{-t-s}h\varphi_{t+s}dt &= \int_{s-\tau}^s du\, L_{-u-\tau}h\varphi_{u+\tau} && (5.7)\\ &= \int_{s-\tau}^0 du\, L_{-u-\tau}h\varphi_{u+\tau} + \int_0^s du\, L_{-u-\tau}h\varphi_{u+\tau} && (5.8)\\ &= \int_{s-\tau}^0 du\, L_{-u-\tau}h\varphi_{u+\tau} + \int_0^s du\, L_{-u}L_{-\tau}h\varphi_\tau\varphi_u && (5.9)\end{aligned}$$

[91] See a comment on p33-4 of Palis-de Melo.

$$= \int_{s-\tau}^{0} du\, L_{-u-\tau} h\varphi_{u+\tau} + \int_{0}^{s} du\, L_{-u} h\varphi_u. \tag{5.10}$$

We have used $h(L_\tau + \phi_\tau) = h\varphi_\tau = L_\tau h$. Introduce $v = u + \tau$. Thus, we have shown

$$L_{-s} H\varphi_s = \int_{s}^{\tau} dv\, L_{-v} h\varphi_v + \int_{0}^{s} du\, L_{-v} h\varphi_v = H. \tag{5.11}$$

This equality is true for $s = \tau$, so $H = h$.

5.15 Formal construction of h

Let $h = 1 + u$

$$(1+u)(L_\tau + \varphi_\tau) = L_\tau(1+u) \tag{5.12}$$

implies

$$\mathcal{L}u \equiv L_\tau u - u(L_\tau + \phi_\tau) = \phi_\tau. \tag{5.13}$$

We write

$$\mathcal{L} = L_\tau \mathcal{L}^* \text{ with } \mathcal{L}^* u = u - L_\tau^{-1} u(L_\tau + \phi_\tau). \tag{5.14}$$

Since L_τ is invertible and $L_\tau + \varphi_\tau$ is a homeomorphism, $\mathcal{L}^*$ is invertible. Thus, u exists. We have shown that the formal solution is actually real. However, we have not guaranteed that h is homeomorphism. that is, invertible.

5.16 h is invertible

We must show that h is a homeo: can we invert $1 + u$? u must be small: we must estimate

$$\|u\| \le \|L_\tau^{-1}\| \|(\mathcal{L}^*)^{-1}\| \|\phi_\tau\|. \tag{5.15}$$

$\|L_\tau^{-1}\|$ is bounded. Therefore, we must show $\|(\mathcal{L}^*)^{-1}\|$ is bounded and $\|\phi_\tau\|$ must be small.

5.17 $\|(\mathcal{L}^*)^{-1}\|$ is bounded

Let us write $\mathcal{L}^* - 1 = \mathcal{K}$. Actually $\mathcal{K}u = -L_\tau^{-1} u(L_\tau + \phi_\tau)$. $\mathcal{K}$ is invertible: Notice that formally

$$\mathcal{K}^{-1} u = L_\tau u(L_\tau + \phi_\tau)^{-1}. \tag{5.16}$$

Since $L_\tau + \phi_\tau$ is the time evolution operator for the system, it is at least locally homeomorphic. Therefore, its inverse is well defined. Notice that $\mathcal{K}$ is a linear map keeping the stable (E^s) and unstable (E^u) subspaces at 0 intact.[92] $\mathcal{K}$.

Thanks to the hyperbolicity $\|\mathcal{K}\| \le a < 1$ on the stable subspace thanks to the 'shrinking nature' of L_τ. Also $\|\mathcal{K}^{-1}\| \le a < 1$ on the unstable subspace thanks to the 'expanding nature' of L_τ. Then,
(a) $1+\mathcal{K}$ is isomorphic on the stable subspace and $\|(1+\mathcal{K})^{-1}\| \le 1/(1-a)$.
(b) $1+\mathcal{K}$ is isomorphic on the unstable subspace and $\|(1+\mathcal{K})^{-1}\| \le a/(1-a)$.
Thus, $1+\mathcal{K} = \mathcal{L}^*$ is isomorphic and $\|(\mathcal{L}^*)^{-1}\| \le \max\{1/(1-a), a/(1-a)\} = 1/(1-a)$.[93]In that case We must show (a) and (b).

(a): To compute the norm consider $(1+\mathcal{K})^-1y = x$ for y in the tangential space of the stable manifold with $\|y\| = 1$. Then $y = x + \mathcal{K}x$ implies $1 \ge \|x\| - \|\mathcal{K}\|\|x\|$ or $\|x\| \le 1/(1-\|\mathcal{K}\|)$. Thus, $\|(1+\mathcal{K})^{-1}\| \le 1/(1-a)$.
(b): Analogously, $(1+\mathcal{K})^-1y = \mathcal{K}^{-1}(1+\mathcal{K}^{-1})^-1y = x$ implies $y = (1+\mathcal{K}^{-1})\mathcal{K}x$. Therefore, $1 \ge (1-\|\mathcal{K}^{-1}\|)\|x\|/\|\mathcal{K}^{-1}\| = (1-a)\|x\|/a$. Thus, $\|(1+\mathcal{K})^{-1}\| \le a/(1-a)$.

5.18 $\|\phi_\tau\|$ is small for small τ

We make $\varphi_\tau = L_\tau + \varphi_\tau$. We assume X is Lipschitz with some constant K. (3.25) in **3.23** tells us

$$\|\varphi_t(x) - \varphi_t(y)\| \le e^{Kt}\|x-y\|. \tag{5.17}$$

We must evaluate $\phi_t = \varphi_t - L_t$. Solving the equations formally, we get

$$\varphi_t(x) = x + \int_0^t A\varphi_s(x)ds + \int_0^t \psi(\varphi_s(x))ds, \tag{5.18}$$

$$L_t(x) = x + \int_0^t AL_s(x)ds, \tag{5.19}$$

where we write $X = Ax + \psi$. Therefore,

$$\phi_t(x) = \int_0^t A[\varphi_s(x) - L_s(x)]ds + \int_0^t \psi(\varphi_s(x))ds = \int_0^t A\phi_s(x)ds + \int_0^t \psi(\varphi_s(x))ds. \tag{5.20}$$

Hence, we obtain

$$\|\phi_t(x) - \phi_t(y)\| \le \int_0^t \|A(\phi_s(x) - \phi_s(y))\|ds + \int_0^t \|\psi(\varphi_s(x)) - \psi(\varphi_s(y))\|ds. \tag{5.21}$$

[92] $L_\tau + \phi_\tau$ is a homeomorphism at least locally, so it keeps the stable and unstable manifolds intact. Thus, their tangent spaces at the fixed point are intact. These tangent spaces are just eigensubspaces of L_τ.

[93] Here, a convenient norm satisfying $\|x+y\| \le \max\{\|x\|, \|y\|\}$ is chosen for $x \in E^s$ and $y \in E^u$ (as in most books), but the usual one is OK if you allow a constant multiple in front of $1/(1-a)$.

Now we use (5.17) and the Lipschitz property of ψ with sufficiently small constant δ to get

$$\|\psi(\varphi_s(x)) - \psi(\varphi_s(y))\| \leq \delta\|\varphi_s(x) - \varphi_s(y)\| \leq \delta e^{Kt}\|x - y\| \tag{5.22}$$

Therefore, (5.21) has the Gronwall form:

$$\|\phi_t(x) - \phi_t(y)\| \leq \int_0^t \delta e^{Ks}\|x - y\|ds + \|A\| \int_0^t \|\phi_s(x) - \phi_s(y)\|ds. \tag{5.23}$$

Or if δ (and τ) is small enough, we can choose a small positive number ε and have

$$\|\phi_\tau(x) - \phi_\tau(y)\| \leq \varepsilon\|x - y\| + \|A\| \int_0^\tau \|\phi_s(x) - \phi_s(y)\|ds. \tag{5.24}$$

We use Gronwall's inequality **3.22**

$$\|\phi_\tau(x) - \phi_\tau(y)\| \leq \varepsilon e^{\|A\|\tau}\|x - y\| \tag{5.25}$$

or

$$\|\phi_\tau(x)\| \leq \varepsilon e^{\tau\|A\|}\|x\|. \tag{5.26}$$

Thus, we have shown that ϕ_τ is Lipschitz and sufficiently small.

5.19 What if singularity is not hyperbolic? Center manifold[94]

If the fixed point (at 0) we consider is not hyperbolic, then the linearization gives us a matrix with vanishing real parts. Let us consider still the stable case. That is, the fixed point is not stable linearly but thanks to higher order terms the point is a ω-limit point of itself.

In such a case the ODE looks like

$$\begin{aligned} \dot{x} &= Ax + f(x, y), & (5.27)\\ \dot{y} &= By + g(x, y), & (5.28) \end{aligned}$$

where x and y correspond to the neutral and stable subspaces: All the eigenvalues of A have no real part, and the eigenvalues of B have negative real parts. f and g are higher order terms and vanish at the origin. If $f = g = 0$ then, $x = 0$ is a stable

[94] A good introduction is J. Carr *Applications of canter manifold theory* (Springer1981).

manifold, and $y = 0$ is called a center manifold. More generally, if $y = h(x)$ is an invariant manifold, it is called a center manifold (which is not unique, generally). At least locally, a center manifold exists which is C^2. Let us proceed formally.

We solve

$$\dot{y} = h'(x)\dot{x} \;\Rightarrow, \; Bh(x) + g(x,h) = h'(x)(Ax + f(x,h)). \tag{5.29}$$

to determine h. $h(0) = h'(0) = 0$ is the auxiliary condition.

On $y = h(x)$ the flow u is governed by

$$\dot{u} = Au + f(u, h(u)). \tag{5.30}$$

Thus, the dynamics is reduced to the one on a lower dimensional space.

How can we obtain h? Solving (5.29) is equivalent to solving the original system. However, if we can solve (5.29) approximately, we can get a reasonable approximation to h. Set

$$M(h) = Bh(x) + g(x,h) - h'(x)(Ax + f(x,h). \tag{5.31}$$

If $M(\phi) = O[|x|^q]$ $(q > 1)$, then $|h - \phi| = O[|x|^q]$.

5.20 Lyapunov stability

A time-independent solution (stationary solution, fixed point p) of an autonomous differential equation is said to be Lyapunov stable, if all the trajectories starting from a neighborhood of p is defined for all $t > 0$ and converges uniformly in time to p.

Thus, $t \to \infty$ behavior need not be a convergence to p. More formally, for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$\|p - x(0)\| < \delta \;\Rightarrow\; \|p - x(t)\| < \varepsilon \text{ for } \forall t > 0, \tag{5.32}$$

we say the fixed point p is Lyapunov stable.

5.21 Asymptotic stability

A fixed point p is asymptotically stable, if it is Lyapunov stable and $\lim_{t\to\infty} x(t) = p$ for some nbh of p.

Look at counterexamples in Fig. 5.2.

As can be seen from these examples, convergence to an equilibrium point in the $t \to \infty$ limit of all the solutions starting at near p is not a sufficient condition for its

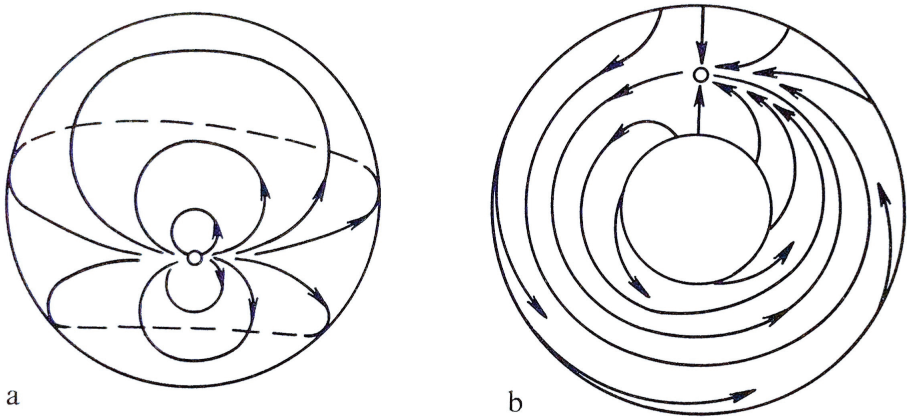


Figure 5.2: Unstable singular points to which all the trajectories which start at nearby points converge. [Fig. 3 of Arnold DS I]

asymptotic stability.

Note that a center is Lyapunov stable, but not asymptotically stable. In such a case the stability is called marginal.

5.22 Stability by linearization

If all the eigenvalues of the linearized equation at a singular point have negative real parts then the singular point is asymptotically stable.

This is a special case of what Hartman's theorem implies.

5.23 Lyapunov function[95]

A differentiable function f is called a Lyapunov function for a singular point x_0 of a vector field X if $Xf \le 0$[96] and x_0 is its strict local minimum in a neighborhood of x_0.

5.24 Lyapunov's stability theorem A singularity of a differentiable vector field for which a Lyapunov function exists is stable.

[95] A DS I p24; For other attractors we can define an analogous concept. See p202-3 of DS1 by Anosov.

[96] Note that $(d/dt)f(x) = Xf(x)$; We use the 'standard notation' $X = \sum X_i \frac{\partial}{\partial x_i}$.

6 Lecture 6: Periodic orbit and limit cycle

6.1 Three types of trajectories

For real ODE $\dot{x} = v(x)$ with a smooth vector field, trajectories are diffeomorphic to a point, a circle (S^1) or a line.

Thus a phase curve of an equation always has a simple intrinsic geometry.[97]

6.2 Periodic orbit (cycle)

A trajectory diffeomorphic to S^1 is called a periodic orbit or a cycle.

6.3 First return map = Poincare map = monodromy transformation

For a smooth vector field, if there is a periodic orbit, we can take a transversal hypersurface (= codimension one surface perpendicular to the orbit; often called a Poincare surface) crossing the orbit at a point p. The orbits starting sufficiently close to p on this surface will return to a neighborhood of p and are again transversal to the surface (Fig. 6.1). Thus, we can locally define a map from the surface into itself. This map[98] is called the first return map, Poincare map or monodromy transformation.

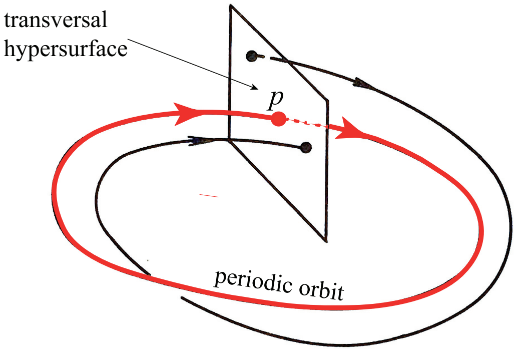


Figure 6.1: First return map

The periodic orbit corresponds to a fixed point of this map.

The first return map does not depend on the choice of the Poincare surface (all

[97] However, complicated knots can be formed as we will see in the Lorenz system.

[98] More precisely, its germ

diffeomorphic locally).

6.4 Linearization around periodic orbit

In a tubular nbh of a periodic orbit γ of period T we can linearize the flow as

$$\dot{x} = DX(\gamma(t))x = A(t)x, \tag{6.1}$$

where $A(t) = DX(\gamma(t))$ is a linear operator with period T: $A(t+T) = A(t)$. Its solution may be written as

$$x(t) = F(t)x_0, \tag{6.2}$$

where $F(t)$ is a linear operator that may be formally written in terms of time ordered exponential of the integral of $DX(\gamma(t))$. F is called a fundamental solution matrix.

$F(T)$ must be the linearization of the Poincare map. It is called the monodromy matrix.

6.5 Floquet's theorem about fundamental matrix for periodic system[99]

Theorem. The fundamental matrix $F(t)$ in **6.4** maybe written as

$$F(t) = B(t)e^{2\pi\Lambda t}, \tag{6.3}$$

where $B(t+T) = B(t)$ and Λ a constant matrix.
[Demo]
Notice that $\Phi(t) = F(t+T)$ is also a fundamental matrix:

$$\frac{d}{dt}\Phi(t) = A(t+T)\Phi(t) = A(t)\Phi(t). \tag{6.4}$$

Since $F(t)$ is invertible for all t, so is $\Phi(t)$. Therefore, Φ is also a fundamental solution matrix. Therefore, there is an invertible matrix C such that $F(t+T) = F(t)C$ for all t (for example $C = F(T)$ is a possible choice). Its log is well defined, so we can introduce Λ as

$$2\pi T\Lambda = \log C. \tag{6.5}$$

Let us write

$$B(t) = F(t)e^{-2\pi\Lambda t}. \tag{6.6}$$

[99] Its 3D version is Bloch's theorem for solid state physics.
'Time crystal' is somewhat related, which is a quantum many-body phenomenon. See Zhang et al., Observation of a discrete time crystal, Nature **543**, 217 (2017) and papers quoted in its introduction; I do not recommend the recent Phys Today article.

Then,

$$B(t+T) = F(t+T)e^{-2\pi\Lambda(t+T)} = F(t)e^{2\pi\Lambda T}e^{-2\pi\Lambda(t+T)} = F(t)e^{-2\pi\Lambda t} = B(t). \tag{6.7}$$

Thus, B is invertible and periodic.

The eigenvalues of $e^{2\pi T\Lambda}$ are called the Floquet multipliers governing the behavior of the Poincaré map, and the eigenvalues ($\times\pi$) of Λ is called the characteristic exponents.[100]

We can restate the theorem as

Theorem There is an invertible periodic linear transformation $B(t)$ such that $x = B(t)y$ transforms the original equation to

$$\dot{y} = \Lambda y. \tag{6.8}$$

Thus, the real part of the eigenvalues of Λ is called a Lyapunov exponent. In terms of Lyapunov exponents, we can discuss asymptotic stability of the periodic orbit.

6.6 Limit cycle

A limit cycle is an isolated phase curve diffeomorphic to a circle. In other words, a closed curve is called a limit cycle if it corresponds to an isolated fixed point of the first return map.

The multiplicity of a limit cycle is the multiplicity of the corresponding fixed point of its return map.

6.7 Stability of limit cycle

A periodic orbit is orbitally stable (Lyapunov stable), if any orbit staring in a certain tubular neighborhood of the orbit stays init.

A periodic orbit is orbitally asymptotically stable. if it is orbitally stable and in the $t \to \infty$ limit any orbit stating from a point in its tubular neighborhood converges to the orbit (its fixed point of the return map is asymptotically stable).[101]

[100] In some books, any number μ such that e^{μ} becomes a Floquet multiplier is called a characteristic exponent. In this case its imaginary part is not unique.

[101] ⟪**Stability of biological clock**⟫ Michele Monti, David K. Lubensky, and Pieter Rein ten Wolde, Robustness of Clocks to Input Noise, PRL 121 078101 (2018) "Here, using models of the Kai system of cyanobacteria, we compare a limit-cycle oscillator with two hourglass models, one

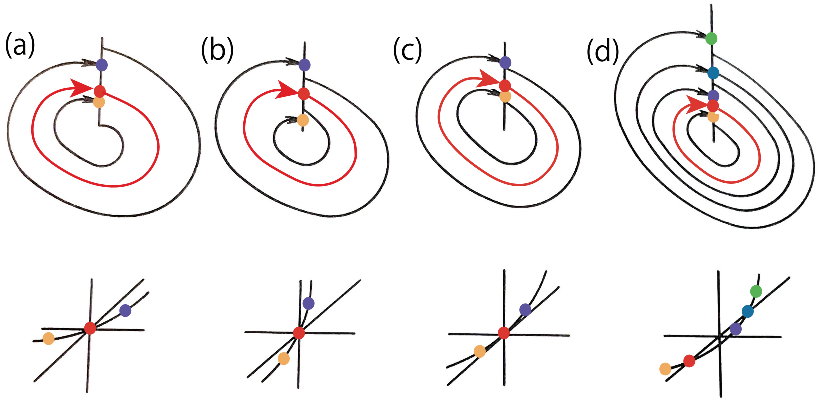


Figure 6.2: Limit cycle on 2-space and its return maps: $I \to I$, where I is an appropriate interval. (a) Stable limit cycle, (b) Unstable limit cycle, (c) semistable limit cycle which is not structurally stable, (d) an outcome of perturbation of (c).

As noted already, we can use Lyapunov exponents for the periodic orbit to study its stability (**6.5**).

6.8 Suspension

Let $f : M \to M$ be a map (say, diffeo). Then, the suspension manifold M_f is obtained from $M \times [0, 1]$ by identifying $(x, 1)$ and $f(x), 0)$ for $x \in M$. The suspension flow is a constant vertical flow (see 6.3).

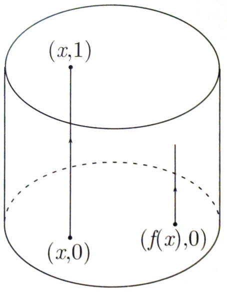


Figure 6.3: Suspension flow [Fig. 0.3.1 of Katok and Hasselblatt p8]

If a dynamical system has a cross section that is transversal to 'all the trajectories,' then the suspension of the Poincaré map is diffeomorphic to the original flow

that without driving relaxes exponentially and one that does so in an oscillatory fashion. In the limit of low input noise, all three systems are equally informative on time, yet in the regime of high input-noise the limit-cycle oscillator is far superior."

(Smale).[102]

6.9 Poincare-Bendixson's theorem

Let D be a subset of S^2.[103] For a flow on D if α and ω-limit sets of $p \in D$ do not contain a singular point, it is a periodic orbit.

A proof is given in **6.10-6.11**. To begin with we need to know that the limit sets of a point is a connected set.
https://www.youtube.com/watch?v=uEfB5DG9x9M&frags=pl%2Cwn illustrates the idea of a proof of the theorem after ca 10 min.

6.10 Limit sets are connected

We consider a flow on a compact manifold M or its subset.

Take $\omega(p)$ and suppose it is not connected. Limit sets are closed sets, so $\omega(p)$ must consist of at least two closed sets ω_1 and ω_2. Since they are closed, they have unoverlapping neighborhoods U_1 and U_2,[104] Since they are ω-limit sets, there must be a sequence of points on the orbit such that $x_i \to x \in \omega_1$ and $y_i \to y \in \omega_2$. We can choose these sequences as $x_1 < y_1 < x_2 < y_2 < \cdots$ along the orbit. For sufficiently large n $x_n \in U_1$ and $y_n \in U_2$, so we can choose z_n on the orbit between x_n and y_n but outside $U_1 \cup U_2$. Since M is compact, we can choose a converging subsequence from $\{z_k\}$, but it converges somewhere other than ω_1 nor ω_2, so this sequence misses $\omega(p)$, a contradiction.

6.11 Demonstration of Poincare-Bendixson's theorem

If p is on a periodic orbit, there is nothing to show. Let us assume p is not on a periodic orbit and $\omega(p)$ does not contain any singular point.
(i) Let $p' \in \omega(p)$. Then, the orbit $C(p')$ going through p' is a periodic orbit.
(ii) $C(p') \subset \omega(p)$. If they do not agree, since both must be closed, $\omega(p) \setminus C(p')$ is not a closed set. Since ω must be connected, there must be a point $a \in \omega(p)$ such that $a \in C(p') \cap \overline{\omega(p) \setminus C(p')}$.
(iii) Take a small neighborhood U of a. Since $a \in C(p') \cap \overline{\omega(p) \setminus C(p')}$, there must be $b \in U \cap (\omega(p) \setminus C(p'))$. Since a is not a singular point, we can make a transversal

[102] Shiraiwa p179-80.
[103] or its open subset or P^2.
[104] We assume Hausdorff.

line ℓ through a (see Fig. 6.4). Since $b \in U$, the vector through it must be close to that at a, so the orbit through b crosses ℓ at $c \in U$.
(iv) This $c \notin C(p')$, since if in $C(p')$, so is $b \in C(p')$, contradicting $b \in \omega(p) \setminus C(p')$. Thus, c is recurrent and not periodic.
(v) Since $a \in \omega(p)$, there are $\{a_k\}$ converging to a and on ℓ in this order along the orbit. Take a_1 and a_2. Then the orbit extended beyond a_2 must be in the green shaded region in Fig. 6.4. Thus, a_3 must be between a_2 and a, etc. That is, $\lim a_k = a$, implying $\omega(p) \cap \ell = \{a\}$, unique. This contradicts the existence of c. Therefore, $\omega(p) = C(p') = C(p)$, a periodic orbit.

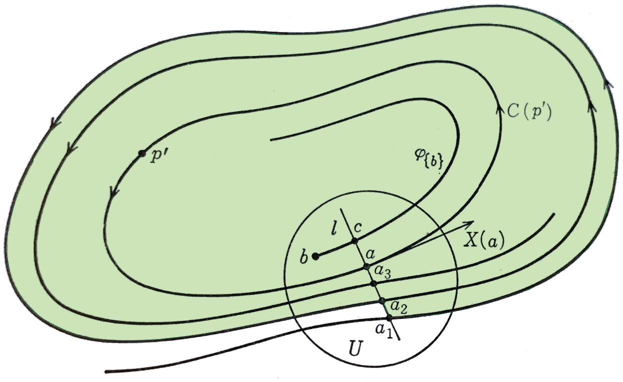


Figure 6.4: [Fig. 3.2 of Tamura, color added]

6.12 Poincare-Bendixson's theorem: a more general version

Let $X \in \mathcal{X}^r(S^2)$ be a vector field with a finite number of singularities. For any $p \in S^2$ one of the followings holds:
(1) $\omega(p)$ is a singularity = fixed point.
(2) $\omega(p)$ is a periodic orbit.
(3) $\omega(p)$ consists of singularities $\{p_i\}$ and regular orbits γ such that if $\gamma \subset \omega(p)$, then $\alpha(p_i)$ and $\omega(p_j)$.

6.13 Bendixson's criterion for no closed orbit

On $\mathbb{R}^2$, consider $\dot{x} = X(x)$. On a simply connected region $D \subset \mathbb{R}^2$, div$X$ is positive or negative semidefinite, then there is no closed orbit lying entirely in D.
[Demo]

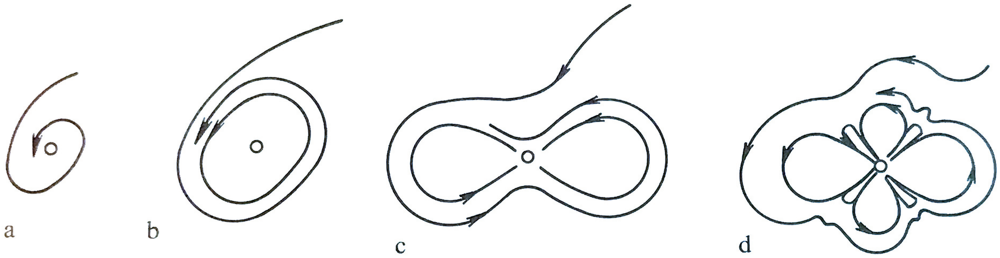


Figure 6.5: Possible $\omega(p)$ [Fig. 7 of DS I]

For any closed curve γ Green's theorem tells us

$$\int_\gamma X \times (dx, dy) \neq 0. \tag{6.9}$$

However, if γ is a solution curve, then X must be parallel to (dx, dy), so the integral must vanish.

6.14 How many limit cycles are there?

There are a set of theorems on various kinds of vector fields called the finiteness theorems. For example,

A polynomial vector field on the real plane has only a finite number of limit cycles.[105]

This follows from a much more general theorem:

Limit cycles of an analytic vector field on a 2-surface cannot accumulate to a compound cycle of the field.[106].

Hilbert's 16th problem is to prove:

The number of limit cycles of a polynomial vector field (of order n) in the real plane is bounded by a number N depending only on n.

This is still open even for $n = 2$.

6.15 Structurally unstable examples

Although we will pay our main attention to generic cases, we must know simple structurally unstable behaviors as illustrated in Fig. 6.6

[105] Theorem 1 on p106 of DS I.

[106] Theorem e of DS I p 106.

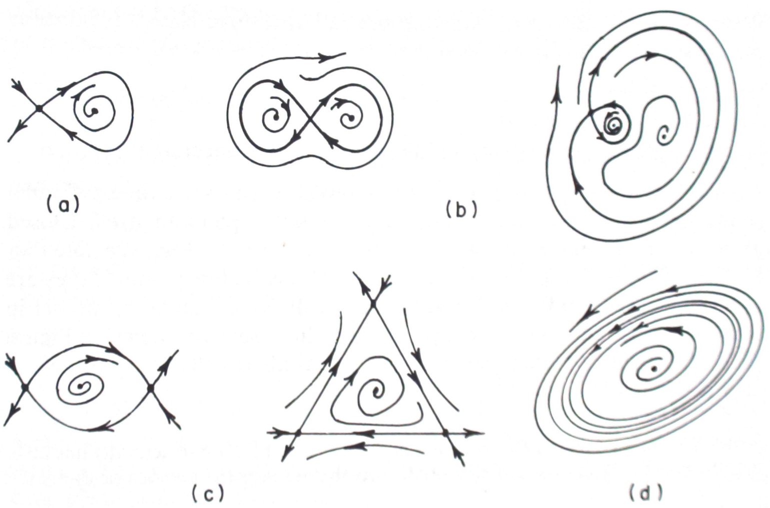


Figure 6.6: Some structurally unstable limit sets (a) homoclinic orbits (saddle loops); (b) Double saddle loop; (c) Homoclinic cycles; (d) periodic orbit band [Fig. 1.81 of GH]

6.16 Chemical oscillation

Belousov discovered the so-called BZ reaction (Belousov-Zhabotinsky reaction). This reaction is basically the oxidation of malonic acid by bromic acid $HBrO_3$. If ferroin is used as a catalyst the oscillation may be observed as a color oscillation of the solution between blue and red.[107]
Stirred: https://www.youtube.com/watch?v=eSWyxMWXw00&frags=pl%2Cwn.
Not-stirred: https://www.youtube.com/watch?v=IBa4kgXI4Cg&frags=wn

The reaction involves numerous (likely to be > 50) chemical species, but a simplified model based on actual observations was proposed as the Oregonator model:[108]

Fig. 6.7 TWO SETS OF REACTIONS can account for the oscillation of ferroin from red to blue and back to red. The concentration of Br^- determines which of the two sets of reactions will dominate. In the first set (left) the bromide and bromate both brominate malonate to form bromomalonate. During this process ferroin (II) is red. If the concentration of the bromide drops below a threshold level, then the second set of reactions (right) starts to dominate. The last vestige of bromide is consumed and the $BrO_3{}^-$ takes over the bromination of the malonate. Simultaneously it oxidizes

[107] Belousov quit science because he could not publish his fundamental work on this reaction in any established journals. Even after this reaction became famous, he never showed up in any meeting on the reaction. I am sympathetic to Belousov who quit science for the reason that science (or the science community) was not scientific enough. S. E. Shnoll was interested in the reaction, to whom Belousov gave the prescription and promised to publish his report in the annual report of the institute he belonged to. Shnoll told his student Zhabotinski to study the mechanism.

[108] A = $BrO_3{}^-$, B = all oxidized organic species, X = $HBrO_3$, Y = Br^-, Z = ferroin(III).

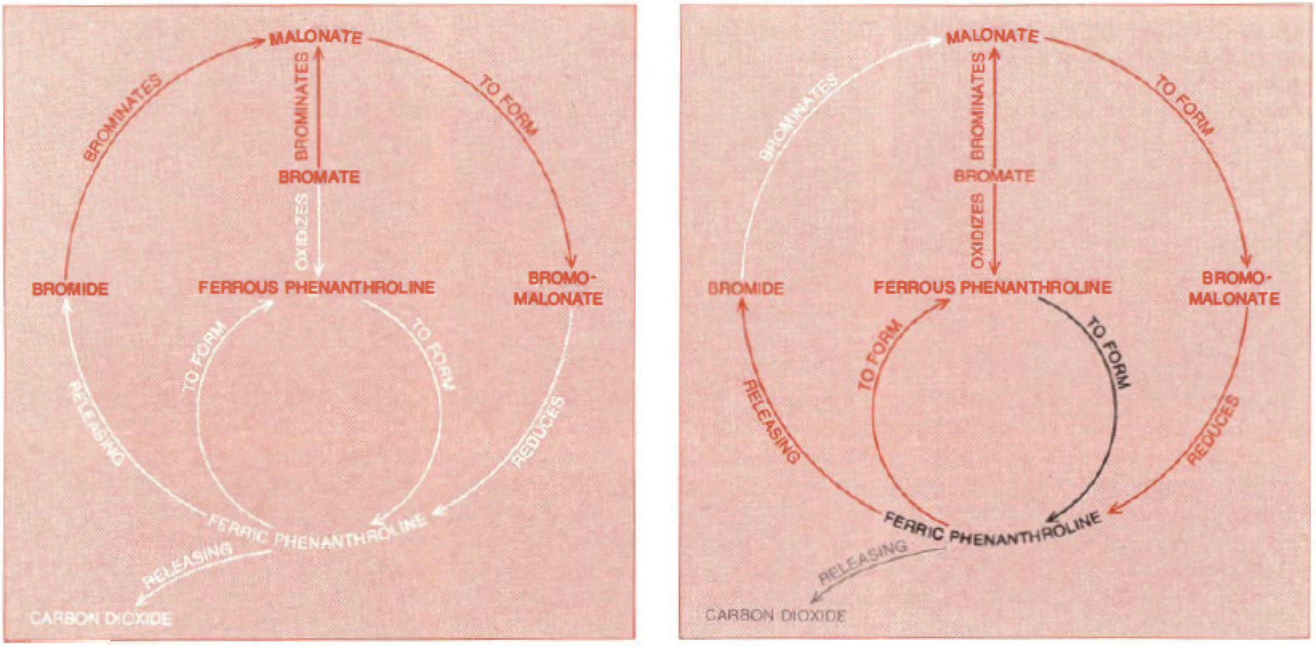


Figure 6.7: Outline of th BZ reaction [Fig. of Winfree Sci Am 1974 p82]

ferroin changing it from red to blue. Accumulated bromomalonate now reduces ferroin(III) back to its red form ferrous, releasing Br^- and carbon dioxide. High concentration of bromide shuts off this reaction sequence and restarts the red stage.

$$
\begin{aligned}
&\mathrm{A} + \mathrm{Y} \longrightarrow \mathrm{X} + \mathrm{P}\\
&\mathrm{X} + \mathrm{Y} \longrightarrow 2\mathrm{P}\\
&\mathrm{A} + \mathrm{X} \longrightarrow 2\mathrm{X} + 2\mathrm{Z}\\
&2\mathrm{X} \longrightarrow \mathrm{A} + \mathrm{P}\\
&\mathrm{B} + \mathrm{Z} \longrightarrow (f/2)\mathrm{Y}
\end{aligned}
$$

A further simplification is possible, because Y is slaved to other concentrations. Eventually, we get

$$\varepsilon\dot{x} = x(1-x) + f(q-x)z/(q+x) = g(x,z), \quad \dot{z} = x - z = h(x,z). \tag{6.10}$$

Here $\varepsilon \simeq 10^{-2}$ and $q \sim 10^{-3}$. Let us simplify these further to

$$\varepsilon\dot{x} = x(1-x) - fz = g(x,z), \quad \dot{z} = x - z = h(x,z). \tag{6.11}$$

However, this is an oversimplification, because $x < 0$ must not happen. For very small x ($x < q$) the sign of the coefficient of z must be negative. Thus, a simplified model must be

$$\varepsilon \dot{x} = x(1-x) - f(x)z = g(x,z), \quad \dot{z} = x - z = h(x,z). \tag{6.12}$$

where $f(x) = f - a\delta(x)$ $(a < 0)$ for $x \geq 0$ to prevent x falling to the negative world.

The fixed points are $x = z$, $x^2 - (1-f)x = 0$. Therefore, $(x,z) = (0,0)$ and $(1-f, 1-f)$ are fixed points. The former is a saddle:

$$\frac{d}{dt}\begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} 1/\varepsilon & -f/\varepsilon \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ z \end{pmatrix}. \tag{6.13}$$

Around the other fixed point, we have

$$\frac{d}{dt}\begin{pmatrix} \delta x \\ \delta z \end{pmatrix} = \begin{pmatrix} (2f-1)/\varepsilon & -f/\varepsilon \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \delta x \\ \delta z \end{pmatrix}. \tag{6.14}$$

Notice that the characteristic equation is $\lambda^2 - \lambda \mathrm{Tr}\, A + \mathrm{det} A = 0$. In our case $\mathrm{det} A = (1-f)/\varepsilon$ and $(1/\varepsilon)\mathrm{Tr}\, A = 2f - 1 - \varepsilon$. Assume $f < 1$. Then, the eigenvalues are complex. Therefore, (ignoring the small ε) for $f = 1/2$ the fixed point is a center. If $f \in (1/2, 1)$ $\mathrm{Tr}\, A > 0$, so orbits near the fixed point spiral out. This bifurcation is called a Hopf bifurcation as we will discuss later.

6.17 Nullcline approach

Do we have a limit cycle? In this case the flow is certainly confined: x and z mut be positive, and cannot be too large. There is no attracting fixed point anywhere. To see the situation closer, a good way is to draw the nullclines $g = 0$ and $h = 0$ (Fig. 6.8).

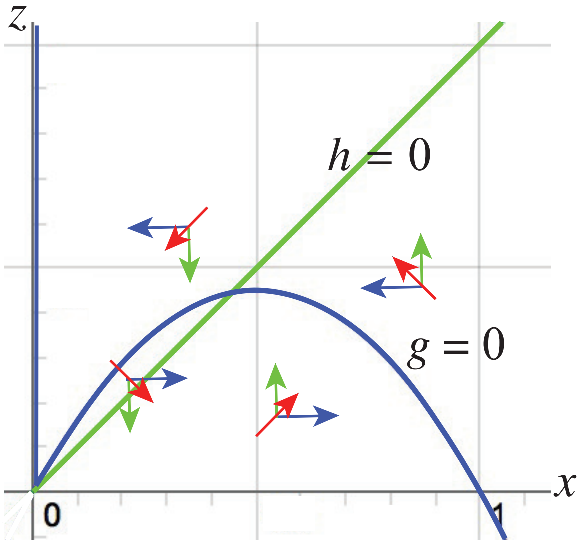


Figure 6.8: Nullclines and the vector field for a simplified Oregonator.

7 Lecture 7: Bifurcation of vector fields

7.1 Family of dynamical systems and bifurcation

Let $\{X_\mu \in \mathcal{X}^r(M)\}_{\mu \in B}$ be a family of C^r vector field on M, where B is a set of parameter values (can be a set of vectors). We say this family exhibits a bifurcation at $\mu^* \in B$, if every neighborhood of μ^* contains μ such that X_{μ^*} and X_μ are topologically distinct (i.e., there is no (local) homeomorphism between them; intuitively speaking, qualitatively distinct). We may say X_{μ^*} is not structurally stable.

We have already encountered the Hopf bifurcation.

7.2 What are the key questions about bifurcations?

Suppose we have a vector field $X \in \mathcal{X}^r(M)$ which is not structurally stable. The most interesting question must be: what happens if we modify X a bit in $\mathcal{X}^r(M)$? Needless to say, we are not interested in reparametrization (or chart change) of the vector field, so we wish to classify the fields near X modulo homeomorphism of the fields and reparametrization of the 'bifurcation parameters' that describe deviation of the fields from the original X.

7.3 Versal unfolding

Thus, we wish to make a family $\{X_\mu\}$ with $X_0 = X$ that is most general and the 'simplest.' Such a family is called the versal unfolding of X.

Here, any X_μ such that $X_0 = X$ is called an unfolding of X.

An unfolding is the versal unfolding, if any other unfolding is equivalent to it. Here 'equivalence' means the same modulo homeo and reparametrization of the parameters.

To construct the versal unfolding, we use two steps: reduction of vector fields to the normal form, and a much more subtle mathematics suh as Malgrange's preparation theorem.

7.4 Malgrange's preparation theorem

Let $F(\mu, x) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a smooth function defined on a neighborhood of the origin of $\mathbb{R}^n \times \mathbb{R}$. Here we discuss the simplest cas where the dynamics is in 1 space (but the parameter is n-vector).

Theorem.[109] Suppose $F(0,x) = x^k g(x)$, where g is smooth in a neighborhood of $x = 0$ and $g(0) \neq 0$. Then, there is a smooth function $q(\mu, x)$ in a neighborhood of $(\mu, x) = (0, 0)$, and functions $s_i(\mu)$ ($i \in \{0, 1, \cdots, k-1\}$) smooth in a neighborhood of $\mu = 0$ such that

$$q(\mu, x)F(\mu, x) = x^k + \sum_{i=0}^{k-1} s_i(\mu)x^i. \tag{7.1}$$

7.5 Versal unfolding of $-x^2$

Let us consider $X = -x^2$. Certainly, this is structurally unstable. The above theorem tells us that

$$X_{a,b} = -x^2 + 2ax + b \tag{7.2}$$

is a versal unfolding. However, this can be rewritten as $-(x-a)^2 + b + a^2$, so a homeo can change this to $-x^2 + \mu$ form. Therefore, we can conclude that the versal unfolding is $X_\mu = -x^2 + \mu$. We will use this result later.

A versal unfolding of $\dot{x} = -x^3$ is $\dot{x} = \nu_1 x + \nu_2 x^2 - x^3$.

7.6 Normal form

Nere $x = 0$ formally a vector field has the following form

$$X = Ax + \sum_{r \geq 2} X^r(x), \tag{7.3}$$

where A is the derivative at $x = 0$ and where X^r is an order r polynomial vector field Since we are interested in a bifurcation at the bifurcation point usually A has a special feature, say, some 0 eigenvalues. If we look at Malgrange type theorems **7.4**, we see that the versal family is determined by the lowest nontrivial order X^r at least locally. Thus, it is very convenient to eliminate lower order polynomial terms as much as possible by a coordinate transformation $x \to y$ so that

$$X = Ay + \sum_{r \geq s} X^r(y), \tag{7.4}$$

as much large s as possible. This form is called the normal form of X.

The conversion of the original X to its normal form is done order by order.

[109] Needless to say there is a version for higher-dimensional spaces. This is a key theorem for catastrophe theory, but not easy to prove.

7.7 Formal elimination of degree r term

Consider

$$\dot{x} = X(x) = Ax + X^r(x) + O[|x|^{r+1}], \tag{7.5}$$

where A is $dA/dx|_{x=0}$, and X^r is a sum of x^k with $|k| = r$ (degree r polynomial term).

Introduce

$$x = y + h^r(y), \tag{7.6}$$

where h^r is a degree r polynomial. Using this, we try to eliminate X^r from (7.5).

(7.6) may be inverted as

$$y = x - h^r(x) + O[|x|^{r+1}]. \tag{7.7}$$

Thus, $(d/dx = D)$

$$\begin{aligned} \dot{y} &= \dot{x} - Dh^r(x)\dot{x} + O[|x|^r]\dot{x} && (7.8)\\ &= Ax + X^r(x) - Dh^r(x)Ax + O[|x|^r]\dot{x} && (7.9)\\ &= A[y + h^r(y)] + X^r(y) - Dh^r(y)Ay + O[|y|^{r+1}] && (7.10)\\ &= Ay - [Dh^r(y)Ay - Ah^r(y)] + X^r(y) + O[|y|^{r+1}] && (7.11) \end{aligned}$$

Notice that (with the summation convention)

$$(Dh^r(y)Ay)_k = D_i h^r_k(y) A_{ij} y_j = \frac{\partial h^r_k}{\partial y_i} A_{ij} y_j = \left(A_{ij} y_j \frac{\partial}{\partial y_i} \right) h^r_k \tag{7.12}$$

and

$$(Ah^r(y))_k = A_{kj} h^r_j \tag{7.13}$$

Introduce a linear operator L_A (called the Lie bracket) as[110]

$$Dh^r(y)Ay - Ah^r(y) = \left(A_{ij} y_j \frac{\partial}{\partial y_i} \right) h^r_k - A_{kj} h^r_j \equiv L_A h^r. \tag{7.14}$$

[110] L_A is the Lie derivative operator. In a more standard language (7.14) reads

$$\left[A_{ij} y_j \frac{\partial}{\partial y_i}, h^r_k \frac{\partial}{\partial y_k} \right] = A_{ij} y_j \frac{\partial h^r_k}{\partial y_i} \frac{\partial}{\partial y_k} - h^r_j A_{kj} \frac{\partial}{\partial y_k} = \left\{ A_{ij} y_j \frac{\partial h^r_k}{\partial y_i} - h^r_j A_{kj} \right\} \frac{\partial}{\partial y_k}.$$

Then, (7.11) reads

$$\dot{y} = Ay - L_A h^r(y) + X^r(y) + O[|y|^{r+1}]. \tag{7.15}$$

If we can solve

$$L_A h^r(y) = X^r(y) \tag{7.16}$$

we are done. This is the non-resonance condition.

7.8 Lie bracket with resonance

When the 'non-resonance condition' is not satisfied, the procedure in **7.7** cannot remove X^r totally. The elements in the cokernel ($= [\mathrm{Im}(L_A)]^c$) of L_A survive.

7.9 Normal form theorem

$X \in \mathcal{X}^r$ with $X(0) = 0$, the differential equation with $DX(0) = 0$

$$\dot{x} = Ax + X(x) \tag{7.17}$$

may be transformed by a polynomial transformation $y = x + h(x)$, where h is a polynomial of second or higher degree, to

$$\dot{y} = Ay + \sum_{r=2}^{N} Y^r + O[|y|^{N+1}]., \tag{7.18}$$

where Y^r is a polynomial of degree r in the cokernel of L_A.

Let us consider the following example:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \sum_{r\geq 2} X^r(x). \tag{7.19}$$

The nonresonance condition (7.12) reads as follows.

$$L_A = \left[2y_1\frac{\partial}{\partial y_1} + y_2\frac{\partial}{\partial y_2}, \quad\right] \tag{7.20}$$

Let h^r be

$$\sum_{m+n=r} a^1_{mn} y_1^m y_2^n \frac{\partial}{\partial y_1} + \sum_{m+n=r} a^2_{mn} y_1^m y_2^n \frac{\partial}{\partial y_2}. \tag{7.21}$$

Therefore,

$$\begin{aligned}[L_A, h^r] &= 2y_1\frac{\partial}{\partial y_1}\sum_{m+n=r}\left\{a^1_{mn}y_1^m y_2^n\frac{\partial}{\partial y_1} + a^2_{mn}y_1^m y_2^n\frac{\partial}{\partial y_2}\right\} - \sum_{m+n=r}\left\{a^1_{mn}y_1^m y_2^n\frac{\partial}{\partial y_1} + a^2_{mn}y_1^m y_2^n\frac{\partial}{\partial y_2}\right\}2y_1\frac{\partial}{\partial y_1} \\ &\quad + y_2\frac{\partial}{\partial y_2}\sum_{m+n=r}\left\{a^1_{mn}y_1^m y_2^n\frac{\partial}{\partial y_1} + a^2_{mn}y_1^m y_2^n\frac{\partial}{\partial y_2}\right\} - \sum_{m+n=r}\left\{a^1_{mn}y_1^m y_2^n\frac{\partial}{\partial y_1} + a^2_{mn}y_1^m y_2^n\frac{\partial}{\partial y_2}\right\}y_2\frac{\partial}{\partial y_2}\end{aligned} \tag{7.22}$$

From this the y_1 component reads

$$\sum_{m+n=r}(2ma^1_{mn}y_1^m y_2^n + na^1_{mn}y_1^m y_2^n - 2a^1_{mn}y_1^m y_2^n)\frac{\partial}{\partial y_1} = \sum_{m+n=r}a^1_{mn}(2m+n-2)y_1^m y_2^n\frac{\partial}{\partial y_1}, \tag{7.23}$$

and similarly the y_2 component reads

$$\sum_{m+n=r}a^2_{mn}(2m+n-1)y_1^m y_2^n\frac{\partial}{\partial y_2}. \tag{7.24}$$

The terms with vanishing coefficients are resonant terms and we cannot eliminate them. For $r = 2$ for y_1 we must solve $m + n = 2$ and $2m + n = 2$. That is $m = 0$, $n = 2$. For y_2 we must solve $m + n = 2$ and $2m + n = 1$. There is no solution. For larger r $m = 1 - r$ means not solution at all. That is, all the terms can be removed. Thus the lowest order normal form reads

$$\dot{y}_1 = 2y_1 + Ky_2^2, \tag{7.25}$$

$$\dot{y}_2 = y_2. \tag{7.26}$$

7.10 Saddle-node bifurcation

Ler us study the case illustrated in Fig. 4.5(a). In this case for A one eigenvalue maintains its sign, and the second one changes its sign:

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & 0 \end{pmatrix} \tag{7.27}$$

at the bifurcation point. Let us convert the original equation into a normal form. The resonance condition is (a good exercise):

For the first component: $m + n = r$, $m\lambda = \lambda$ (i.e., $m = 1$). $m = n = 1$ for $r = 2$, for

$r \geq 3$ $m = 1$ $n = r - 1$. Thus, $x + Axy + B_r xy^{r-1}$.
For the second component: $m + n = r$, $m = 0$. That is, $\sum_{r \geq 2} a_r y^r$.
In summary, the normal form around the saddle-node bifurcation point is

$$\dot{x} = \lambda x + \sum a_r xy^{r-1}, \tag{7.28}$$

$$\dot{y} = \sum b_r y^r. \tag{7.29}$$

To determine a convenient versal unfolding, we truncate this at $r = 2$

$$\dot{x} = \lambda x + Axy, \tag{7.30}$$

$$\dot{y} = By^2. \tag{7.31}$$

As Malgrange's theorem tells us, adding lower order polynomials to the above normal form gives a versal unfolding. However, as we noted in **7.5** we can eliminate some terms by coordinate transformation

$$\dot{x} = \lambda x + Axy, \tag{7.32}$$

$$\dot{y} = \nu + By^2. \tag{7.33}$$

For the $\lambda > 0$ $B < 0$ case, the bifurcation diagram looks like Fi.g 7.1.

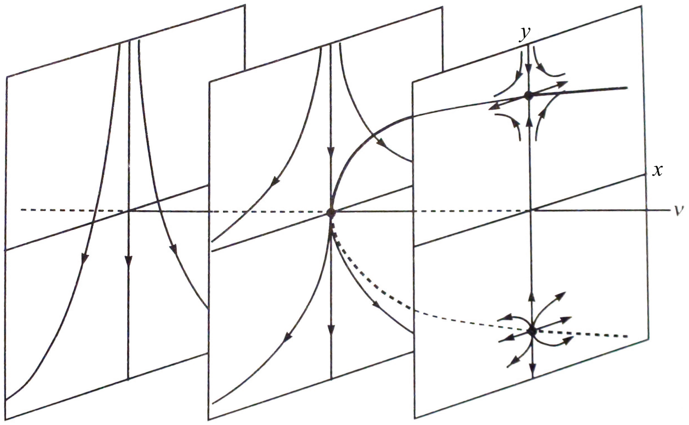


Figure 7.1: Supercritical saddle-node bifurcation [Fig. 4.5 of Arrowsmith and Place p201]

7.11 Hopf singularity

In this case for A conjugate complex eigenvalues change the sign of their real part.

At the bifurcation point

$$A = \begin{pmatrix} 0 & -\beta \\ \beta & 0 \end{pmatrix}. \tag{7.34}$$

In this case after complexification we can still diagonalize A as $A = [i\beta, 0i\beta]$.

7.12 Normal form around Hopf singularity

Let us convert the original equation into a normal form. The resonance condition is For the first component z: $m + n = r$, $im\beta - in\beta = i\beta$ (i.e., $m - n = 1$). Thus, $i\beta z + z \sum B_r(z\overline{z})^r$.
For the second component $\overline{z}$: $m+n = r$, $im\beta - in\beta = -i\beta$ (i.e., $m-n = -1$). Thus, $-i\beta\overline{z} + \overline{z}\sum C_r(z\overline{z})^r$.
Therefore, the normal form reads

$$\dot{z} = i\beta z + z \sum B_r(z\overline{z})^r, \tag{7.35}$$

and its conjugate. In terms of the original variables

$$\begin{aligned} \dot{x} &= -\beta y + \sum(Re(B_r)x - Im(B_r)y)(x^2 + y^2)^r, & (7.36) \\ \dot{y} &= \beta x + \sum(Im(B_r)x + Re(B_r)y)(x^2 + y^2)^r. & (7.37) \end{aligned}$$

7.13 Versal unfolding of Normal form for Hopf bifurcation

Locally we may truncate (7.37) as

$$\begin{aligned} \dot{x} &= -\beta y + (ax - by)(x^2 + y^2) + O[|x|^5], & (7.38) \\ \dot{y} &= \beta x + (bx + ay)(x^2 + y^2) + O[|x|^5] & (7.39) \end{aligned}$$

We wish to keep the circular symmetry and mirror symmetry, Then there should not be any even order terms and the unfolding looks like

$$\begin{aligned} \dot{x} &= \nu x - \beta y + (ax - by)(x^2 + y^2) + O[|x|^5], & (7.40) \\ \dot{y} &= \beta x + \nu y + (bx + ay)(x^2 + y^2) + O[|x|^5] & (7.41) \end{aligned}$$

7.14 Hopf bifurcation diagram

We can plot the local family described by (7.41). If $a > 0$ it is called the subcritical bifurcation; if $a < 0$ it is called the supercritical bifurcation (Fig. 7.2).

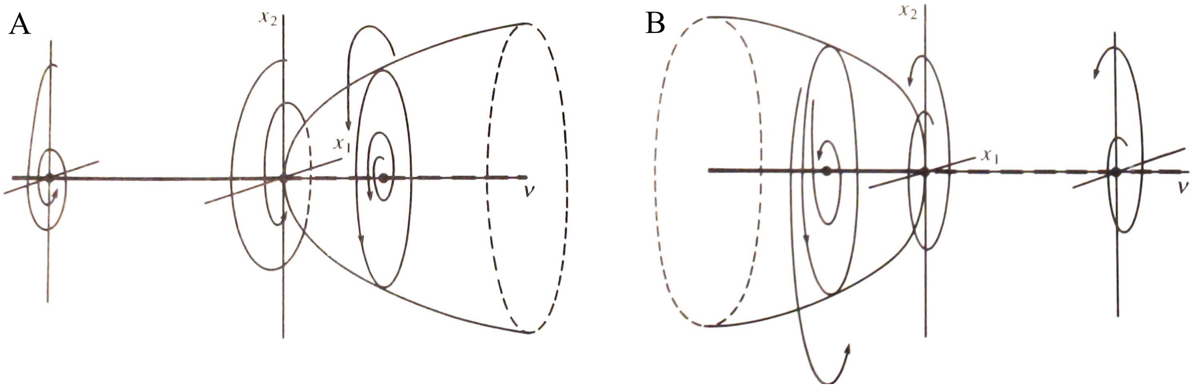


Figure 7.2: A: supercritical Hopf bifurcation, B: subcritical (or inverted) [Fig. 4.7 of AP]

8 Lecture 8: Bifurcation 2, maps

8.1 Singularities of maps

Consider $f \in C^r(M)$. $x \in M$ such that $f(x) = x$ is called a fixed point of f, which corresponds to a singularity in a vector field: no change in time. We can Taylor expand f around it (often we choose the coordinate system so that $x = 0$) as

$$f(x) = Ax + f_2 + \cdots + f_r + \cdots, \tag{8.1}$$

where f_r is a homogeneous polynomial of degree r. If A has no eigenvalue on the unit circle, the fixed point is called a hyperbolic fixed point. It is structurally stable.

We can introduce all the analogues of normal forms and unfoldings.

8.2 Representative hyperbolic fixed points

According to the eigenvalues of A, we can classify the fixed points. The 2-dimensional case is in Fig. 8.1.

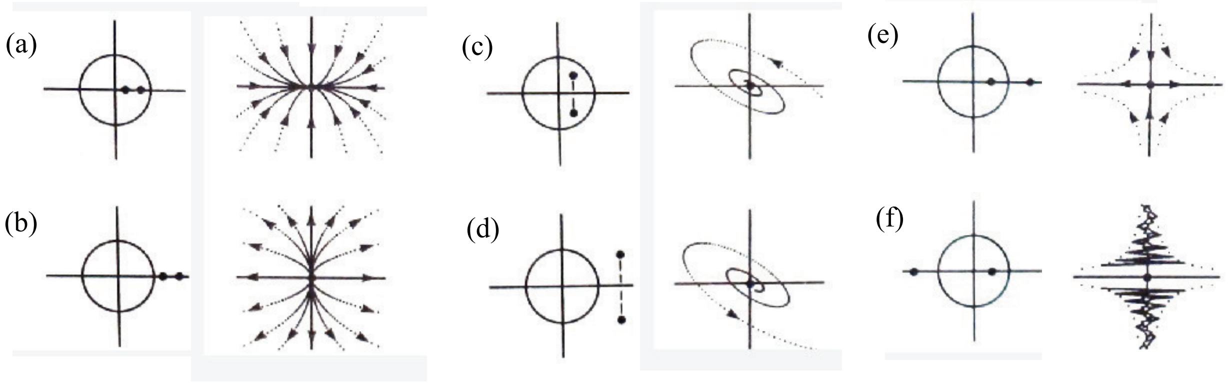


Figure 8.1: Representative hyperbolic fixed points [Fig. 2.1 of AP]

In the rest of this lecture, we discuss an interval map $f : [0,1] \to [0,1]$. That is an endomorphism of $[0,1]$.

8.3 Nonhyperbolic interval endomorphism

At a fixed point if $|f'| = 1$ f is non-hyperbolic, and structurally unstable.

8.4 Versal unfolding of critical map with slope 1

We must distinguish two cases: the fixed point for cases with near the critical parameter value (1) depends on the parameter or (2) not.
Case (1): The normal form with the linear term x generally reads x plus a higher degree polynomial. Thus, $x + Ax^2$ is locally (i.e., around the origin) enough. Then, unfolding should be

$$f(\mu, x) = \nu + (1 + \mu)x + Ax^2. \tag{8.2}$$

However, we can scale x and reparameterize this as

$$f(\mu, x) = \nu + x \pm x^2. \tag{8.3}$$

Here, ± may be replaced by + with an appropriate sign change of ν so we have only to consider $f(y) = \nu + y + y^2$. The bifurcation diagram looks like Fig. 8.2.

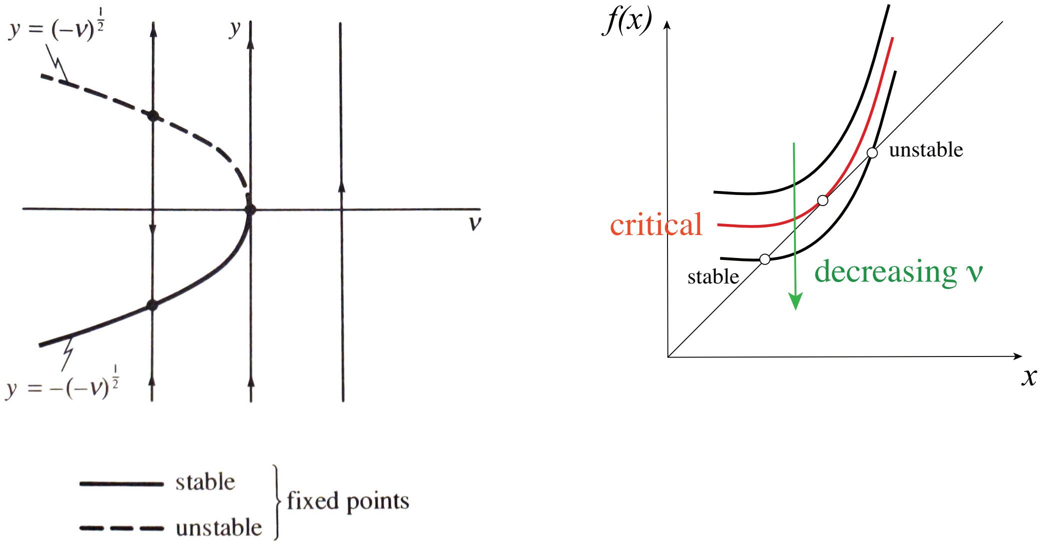


Figure 8.2: Fold: $\nu + y + y^2$; the right is a interval map example indicating a fold bifurcation [Left: Fig. 4.16b of AP]

For case (2): Since we cannot add a constant term, the versal unfolding reads

$$f(\mu, x) = (1 + \mu)x \pm x^2. \tag{8.4}$$

In this case the lowest order term that changes the local topology (a new fixed point) is $\pm x^2$.

The bifurcation diagram reads as Fig 8.3:

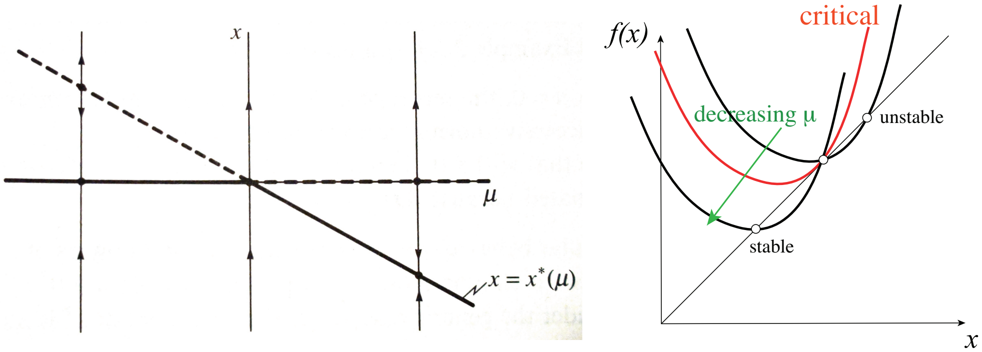


Figure 8.3: Bifurcation diagram for $(1+\mu)x + x^2$; Right. An interval map example.

8.5 Versal unfolding of critical map with slope −1

The case with the slope -1 is a bit complicated, because locally it seems that f does not change much, but f^2 can change qualitatively. Notice that for $f(x) = -x$ $f^2(x) = x$, so it has a fixed line.

Adding $\pm x^2$ (and all other even powers) does not change the topology: suppose $f = -x + x^2$. Then

$$f^2(x) = -(-x + x^2) + (-x + x^2)^2 = x - 2x^3 + x^4 = x(1 + 2x^2 + x^3) \tag{8.5}$$

so no new local fixed point appears.

Consider $f = -x \pm x^3$. Then,

$$f^2(x) = -(-x \pm x^3) \pm (-x \pm x^3)^3 = x - \pm 2x^3 + \cdots, \tag{8.6}$$

so with perturbation new fixed points show up (see Fig. 8.4).

Adding other terms does not alter the qualitative picture. Therefore

$$f(\nu, x) = (-1 + \nu)x \pm x^3 \tag{8.7}$$

is a versal unfolding. This describes the so-called pitch-fork bifurcation.

The bifurcation diagram is Fig. 8.5.

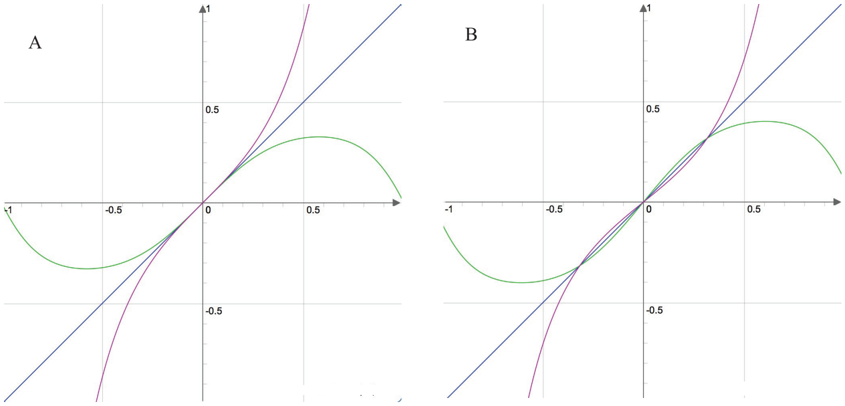


Figure 8.4: f^2 near the slope -1 fixed point. A: at the bifurcation point , B: unfolding. A: Purple: $-x+x^3$, green: $-x-x^3$. B: purple: $-0.9x+x^3$, green: $-1.1x-x^3$

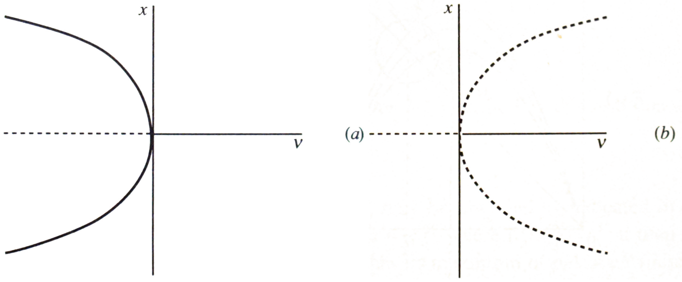


Figure 8.5: pitchfork bifurcation [Fig. 4.18 of AP]

8.6 Pitchfork bifurcations in logistic maps

The logistic map is a map defined on $[0,1]$ as

$$f(x) = ax(1-x). \tag{8.8}$$

Here the parameter $a \in [0,4]$. As a is increased, the map exhibits a series of pitchfork bifurcations as illustrated in the following bifurcation diagram (Fig. 8.6):

These bifurcations correspond to period-doubling bifurcations.

8.7 Feigenbaum critical phenomenon

Initially, Feigenbaum found numerically that
(1) The successive bifurcation parameter value behaves as $a_n = a_\infty - A\delta^{-n}$, where $\delta = 4.66\cdots$.
(2) The pattern size at the n-th bifurcation is compressed as $(-a)^{-n}$, where $a =$

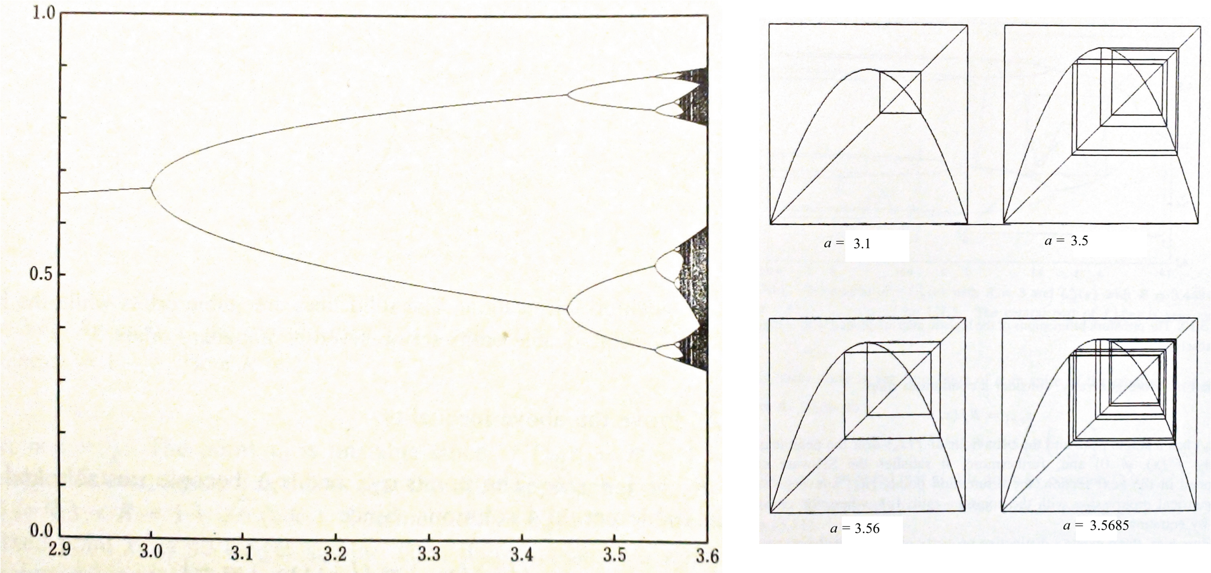


Figure 8.6: Pitchfork accumulating to chaos ($a = 3.5699456\cdots$) in the logistic map [Fig. 3.8. 3.9 of Nagashima & Baba]

$2.5029\cdots$, where $-$ sign implies that the pattern is flipped as n increases by one. He found the universality as well: as long as the map is smooth (C^1), these results do not depend on the map.

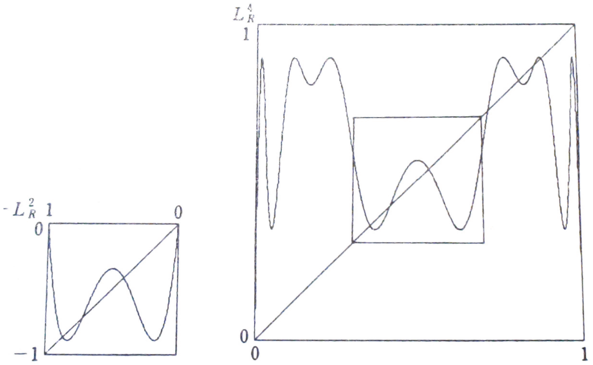


Figure 8.7: f^4 for $a = 3$ is similar to f^4 for $a = 4.44949$ [Fig. 3.11 of Nagashima & Baba]

Let us define

$$g^{[n+1]} = g^{[n]} \circ g^{[n]} \tag{8.9}$$

with $g^{[0]} = g = f^2$.

Notice that we study $g = f \circ f$ or its iterates around $x = 1/2$ (see Fig. 8.7). From the figure we can guess that between the n-th and the $n+1$-pitchfork bifurcations (near $x = 1/2$)

$$g^{[n]}(a, x) = (-\alpha)^{-n} h^{[n]}(\varepsilon_n, y_n) + \frac{1}{2}, \tag{8.10}$$

where

$$y_n = (x - 1/2)(-\alpha)^n, \ \varepsilon_n = A(a - a_\infty)\delta^n. \tag{8.11}$$

Feigenbaum's conjecture is that the following limit is well-defined:

$$h^{[n]}(\varepsilon_n, y_n) \to h(\varepsilon, y) \tag{8.12}$$

for any 'smooth' g.

8.8 Renormalization around a_∞

(8.9) in terms of $h^{[n]}$ can be obtained as follows (note that $y_n = h^{[n]}(\varepsilon_n, y_n)$):

$$(-\alpha)^{-n-1} h^{[n+1]}(\varepsilon_{n+1}, y_{n+1}) + \frac{1}{2} = (-\alpha)^{-n} h^{[n]}(\varepsilon_n, h^{[n]}(\varepsilon_n, y_n)) + \frac{1}{2}. \tag{8.13}$$

Therefore,

$$(-\alpha)^{-(n+1)} h^{[n+1]}(\varepsilon_{n+1}, y_{n+1}) = (-\alpha)^{-n} h^{[n]}(\varepsilon_n, h^{[n]}(\varepsilon_n, y_n)) \tag{8.14}$$

or

$$(-\alpha)^{-1} h^{[n+1]}(\delta\varepsilon_n, -\alpha y_n) = h^{[n]}(\varepsilon_n, h^{[n]}(\varepsilon_n, y_n)). \tag{8.15}$$

Therefore, h satisfies

$$-\alpha^{-1} h(\delta\varepsilon, -\alpha y) = h(\varepsilon, h(\varepsilon, y)). \tag{8.16}$$

This is the RG fixed point equation.

8.9 Approximate solution of the RG equation

(8.16) is not easy to solve. Therefore, we use the following Ansatz: $\varepsilon = 0$ is the critical point, so

$$h(0, y) = 1 - cy^2. \tag{8.17}$$

For this to satisfy

$$-\alpha^{-1} h(0, -\alpha y) = h(0, h(0, y)) \tag{8.18}$$

we have

$$-\alpha^{-1}(1 - c(\alpha y)^2) = 1 - c(1 - cy^2)^2 = 1 - c + 2c^2y^2 - c^3y^4 \tag{8.19}$$

Truncating this at $O[y^2]$, we require

$$-\alpha^{-1} = 1 - c \tag{8.20}$$
$$c\alpha = 2c^2. \tag{8.21}$$

We can determine $\alpha = 1 + \sqrt{3} = 2.732$. The empirical value is 2.5029. Not too bad.

8.10 Determination of δ

To determine δ we mus point function to study the $\varepsilon \neq 0$ case, so we study the deviation ψ_n from the fixed point.

$$h^{[n]}(\varepsilon_n, y_n) = h(0, y_n) + \varepsilon_n \psi_n(y_n). \tag{8.22}$$

We introduce this into (8.15) and linearize as

$$(-\alpha)^{-1}[h(0,-\alpha y_n)+\varepsilon_{n+1}\psi_n(-\alpha y_n)] = h(0,h(0,y_n))+\varepsilon_n\partial_y h(0,h(0,y_n))\psi_n(y_n)+\varepsilon_n\psi_n(h(0,y_n)). \tag{8.23}$$

Define $\mathcal{G}$ as

$$\mathcal{G}\phi(y) = -\alpha[\partial_y h(0, h(0,-y/\alpha))\phi(-y/\alpha) + \phi(h(0,-y/\alpha))] \tag{8.24}$$

Since $(-\alpha)^{-1}h(0,-\alpha y_n) = h(0,h(0,y_n))$, we get the following equation ($y_{n+1} = -\alpha y_n$)

$$\varepsilon_{n+1}\psi_n(y) = \varepsilon_n \mathcal{G}\psi_n(y), \tag{8.25}$$

That is,

$$\delta\psi_n(y) = \mathcal{G}\psi_n(y). \tag{8.26}$$

The eigenvalue of $\mathcal{G}$ gives δ. We know $h(0,y) = 1 - (1+\alpha^{-1})y^2$ so we can compute $\mathcal{G}$. However, I could not get a good approximate value for $\delta = 4.66\cdots$. Try.

9 Lecture 9: Singular perturbation and renormalization

9.1 Regular and singular perturbations[111]

Let us consider an ODE whose vector field is $X_0 \in \mathcal{X}^r(M)$. When we add a 'small' term εX_1 to X_0, where ε (> 0) is a small number, and $X_1 \in \mathcal{X}^r(M)$, the resultant systems is said to be a perturbed system

$$\dot{x} = X \equiv X_0 + \varepsilon X_1. \tag{9.1}$$

Perturbations may be classified broadly into two classes, regular and singular perturbations. Pragmatically speaking, if the result as a power series in ε is convergent (for some fixed time range independent of ε), the perturbation is regular; if not, singular. If the perturbation is regular, there is no qualitative change of the dynamical system due to perturbation. If singular, generally, the perturbed system exhibit new features, e.g., new appearance of limit cycles.

Singular perturbations may be classified into two major classes; X_0 is actually defined on a submanifold of M (in other words X_1 requires new vector components; this is probably the singular perturbation in the original sense) and X_0 and X have no such submanifold structure (e.g., the resonance problems).

Since the perturbation series formally obtained by a straightforward perturbation calculation for singular perturbation problems give (at best) asymptotic series just as interacting field theories, it may not be so surprising that renormalization-group

[111] The best reference book (practical book) of singular perturbation is C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, 1978). Numerous examples and numerical confirmation of the results make this book unrivaled. It is a book to be kept at one's side whenever singular perturbation problems are studied. However, the book is wonderfully devoid of any mathematical theory. To have an overview of various singular perturbation methods within a short time, E. J. Hinch, *Perturbation Methods* (Cambridge UP., Cambridge, 1991) is recommended. J. Kevorkian and J. D. Cole *Perturbation methods in applied mathematics* (Springer, 1981) was a standard reference.

ideas can be useful.[112],[113]

9.2 Simple example of singular perturbation

Consider

$$\epsilon \frac{d^2y}{dt^2} + \frac{dy}{dt} + y = 0, \tag{9.2}$$

where $\epsilon > 0$ is a small constant. If this is zero, the solution decays exponentially; Ae^{-t} is its general solution, where A is a numerical constant. If $\epsilon > 0$, for a sufficiently small time, the second order derivative term is important (notice that however small the mass may be, the inertial effect is crucial for a short time at the beginning of the motion). However, after a long time, the system behavior should be similar to the $\epsilon = 0$ case. Then, why don't we take this term into account through perturbation? This problem is easily solved by hand exactly, but let us pretend that we cannot do so, and perform a perturbative calculation.

9.3 Naive perturbation and its difficulty

Let us expand the solution to (9.2) formally as

$$y = y_0 + \epsilon y_1 + \cdots \tag{9.3}$$

and then introduce this into the equation (9.2). Equating the terms with the same power of ε, we obtain

$$\frac{dy_0}{dt} + y_0 \quad = \quad 0, \tag{9.4}$$

[112] Key ideas are in N. D. Goldenfeld, O. Martin and Y. Oono, "Intermediate asymptotics and renormalization group theory," J. Scientific Comp. **4**, 355 (1989); L.-Y. Chen, N. Goldenfeld, Y. Oono, and G. Paquette, "Selection, stability and renormalization," Physica A **204**, 111 (1993). The latter clearly recognized the relation between the reductive perturbation theory (fully exploited by Y. Kuramoto) and RG. Thus, the crux of many singular perturbation methods is to construct the equation of motion that governs the asymptotic solutions of perturbed systems. A paper with many examples is the 'CGO': Lin-Yuan Chen, Nigel Goldenfeld, and Y. Oono, Renormalization group and singular perturbations: Multiple scales, boundary layers, and reductive perturbation theory, Phys Rev E **54**, 376 (1996). Some misunderstandings (even by my collaborators) were corrected in Y. Oono and Y. Shiwa, Reductive renormalization of the phase-field crystal equation, Phys. Rev. E **86**, 061138 (2012).

[113] In these days, applied mathematicians have organized the results of the renormalization group method as a method of nonlinear variable transformation that requires no idea of renormalization. I have no interest in this direction.

$$\frac{dy_1}{dt} + y_1 \quad = \quad -\frac{d^2 y_0}{dt^2}, \tag{9.5}$$

etc. Let us write the solution to the first equation as $y_0 = A_0 e^{-t}$, where A_0 is an integration constant. Then, the general solution to the second equation reads

$$y_1 = A_1 e^{-t} - A_0 t e^{-t}, \tag{9.6}$$

where A_1 is also an integration constant. Combining these two results, we obtain to order ϵ

$$y = A_0 e^{-t} - \epsilon A_0 t e^{-t} + O(\epsilon^2). \tag{9.7}$$

Here, $A_0 + \epsilon A_1$ is redefined as A_0 (we ignore $\varepsilon^2 A_1$ in the second term; as can be seen from this, in the perturbative expansion we have only to find special solutions).

In this way, we can compute any higher order terms, but this is usually regarded as a bad solution, because ϵ appears with t which increases indefinitely. Thus, the perturbation effect that should be small becomes not small, and the perturbation method breaks down. In other words, the perturbation result may be used only for the time span much shorter than ϵ^{-1}. Mathematically speaking, the convergence is not uniform in time. The term with multiplicative t is traditionally called a *secular term.*

9.4 How we extract long time behaviors

The analogy with the standard RG problem (or critical phenomena) is as follows: We are interested in the long term behavior $t \to \infty$ that is insensitive to initial details (= the long-term qualitative behaviors). In this limit secular terms diverge. If we could remove such divergences, then naive perturbation series as obtained above could make sense.

For the present problem to watch the behavior just in front of us ('at present time') corresponds to macroscopic observations and the behaviors long ago correspond to microscopic scales.[114] That is, we are interested in the global behavior that does not change very much even if the initial condition is modified. In the ordinary renormalization problem, the microscopic-detail-sensitive responses (that diverge in the $L/\ell \to \infty$ limit, where L is our scale and ℓ the atomic scale) are separated and renormalized into materials constants.

[114] For ordinary deterministic systems it is hard to obtain the initial condition from the observation result at t (it is asymptotically impossible in the $t \to \infty$ limit). For chaotic systems, however, this is not only always possible, but the estimate of the initial condition becomes more accurate if we observe longer time asymptotic behaviors (however, we assume there is strictly no noise). That is, in a certain sense, chaos is the antipode of renormalizability.

Therefore, in the present problem what must be renormalized is the sensitively dependent behavior on the initial condition, and the place it should be pushed into must be the integration constants (the quantities connecting what we observe now and the initial condition); it is a natural observation, because the integration constants are determined by the initial condition.

Notice that what we can renormalize is the relation between what we can observe and what we cannot. Since there is no arbitrariness in the relationships among observable quantities, there is no room for renormalization constants for observable relationships. It is crucial to distinguish what we can observe and what we cannot.[115]

9.5 Renormalization along time axis

Separate the secular divergence as $(t-\tau)+\tau$, and then absorb τ by modifying unobservable A_0 (since we do not know the precise initial condition) as $A(\tau)$, which is understood as an adjustable parameter to be determined so that the solution agrees with the behavior we directly observe at present, i.e., around time t.[116] After this renormalization, the perturbation series (9.7) reads

$$y = A(\tau)e^{-t} - \epsilon(t-\tau)A(\tau)e^{-t} + O(\epsilon^2). \tag{9.8}$$

(As can be seen from this, τ is actually $\tau -$ initial time, i.e., the time lapse from the initial time). Such a series is called a renormalized perturbation series.

This equation makes sense only when $\epsilon(t-\tau)$ is small, but it is distinct from the original perturbation series (9.7) we started with (which is often called a 'bare' perturbation series), because we can choose τ to be large enough.

9.6 Renormalization-group equation

Since τ is a parameter not existing in the problem itself, $\partial y/\partial\tau = 0$. This is the renormalization group equation for the current problem:

$$\frac{\partial y}{\partial \tau} = \frac{dA}{d\tau}e^{-t} + \epsilon(t-\tau)\frac{dA}{d\tau} + \epsilon A e^{-t} + \cdots = 0. \tag{9.9}$$

The equation tells us that $dA/d\tau$ must be of order ϵ (the terms proportional to e^{-t} must cancel each other), so we may discard the second term of (9.9) as a higher order

[115] Thus, the following ancient teaching becomes a crucial renormalization instruction: "When you know a thing, recognize that you know it, and when you do not, recognize that you do not." *Analects* Book 2, 17 [A. Waley, *The Analects of Confucius* (Vintage, 1989)].

[116] If you know the ordinary field-theoretic RG, you must have recognized that log(length scale) corresponds to time; $t-\tau \leftrightarrow \log(L/\ell)$.

term. Therefore, the renormalization group equation reads, to order ε,

$$\frac{dA}{d\tau} = -\epsilon A. \tag{9.10}$$

The renormalized perturbation series (9.8) is simplified if we set $\tau = t$ (those who question this procedure should see the note below):

$$y = A(t)e^{-t}. \tag{9.11}$$

(9.10) implies that $A(t)$ obeys the following amplitude equation:

$$\frac{dA(t)}{dt} = -\epsilon A(t). \tag{9.12}$$

The equation indicates that A changes significantly only in the long time scale of order $t \sim 1/\epsilon$; only when ϵt has a visible magnitude can A change significantly. Finally, solving (9.12), we get the following asymptotic behavior,

$$y = Be^{-(1+\epsilon)t} + O(\epsilon^2), \tag{9.13}$$

where B is an adjustable parameter. This is our conclusion about the asymptotic behavior.[117] Here, notice that the form of $A(t)$ as a function of t is universal in the sense that it does not directly depend on the initial condition ((9.12) is determined by the original differential equation itself).

As can be seen from the above example, the core of the renormalization group method for the problems with secular terms is to derive equations that govern the slow systematic motions such as (9.12). This may be interpreted as coarse-graining or reduction of the system behavior.

9.7 More systematic approach

To perform calculation more systematically, we introduce the renormalization constant Z as $A = ZA_0$ or more conveniently as (notice that the following Z is the reciprocal of the Z in $A = ZA_0$)

$$A = ZA_R, \tag{9.14}$$

where A is the 'bare' microscopic quantity and A_R the renormalized counterpart (in the current context it is what we observe long time later). We are performing a perturbation calculation, so we expand $Z = 1 + \epsilon Z_1 + \cdots$, and the coefficient are determined order by order to remove divergences. In the lowest order calculation as we have done, there is no danger of making any mistake, so a simple calculation as explained above is admissible, but, as we will see in Note 3.7.1, a formal expansion helps systematic studies.

[117] In terms of the two roots $\lambda_\pm = (-1 \pm \sqrt{1-4\epsilon})/2\epsilon$ of the characteristic equation $\epsilon s^2 + s + 1 = 0$ the analytic solution for (9.2) is $y(t) = Ae^{\lambda_+ t} + Be^{\lambda_- t}$, where for small ϵ $\lambda_+ = -1 - \epsilon + O(\epsilon^2)$, $\lambda_- = -1/\epsilon + 1 + \epsilon + O(\epsilon^2)$, so (9.13) is a uniformly correct order ε solution up to time $\epsilon t \sim 1$.

9.8 Why we can set $t = \tau$

In the above calculation putting $\tau = t$ makes everything simple, but there are people who feel that it is a bit too convenient and *ad hoc* a procedure, so let us avoid this procedure. The result of renormalized perturbation series has the following structure:

$$y(t) = f(t; \epsilon\tau) + \epsilon(t - \tau)g(t) + O(\epsilon^2). \tag{9.15}$$

Since f is differentiable with respect to the second variable, with the aid of Taylor's formula we may rewrite it as

$$y(t) = f(t; \epsilon t) + \epsilon(\tau - t)\partial_2 f(t, \epsilon t) + \epsilon(t - \tau)g(t) + O(\epsilon^2), \tag{9.16}$$

where ∂_2 denotes the differentiation with respect to the second variable. The second and the third terms must cancel each other, since the original problem does not depend on τ. That is, the procedure to remove the secular term by setting $\tau = t$ is always correct.

9.9 RG equation as envelope equations

To construct an envelop is a renormalization procedure.[118] Suppose a family of curves $\{x = F(t, \alpha)\}$ parameterized with α is given. Its envelop is given by

$$x = F(t, \alpha), \quad \frac{\partial}{\partial \alpha} F(t, \alpha) = 0. \tag{9.17}$$

The second equation can be interpreted as a renormalization group equation. The envelop curve is such a set of points among the points $\{x, t\}$ satisfying $x = F(t, \alpha)$ that stay invariant under change of α to $\alpha + \delta\alpha$. Therefore, it must satisfy the second equation describing the condition that the (x, t) relation does not change under perturbation of α. This is exactly the same idea as searching features that stay invariant even if microscopic details are perturbed.

However, renormalization group theory must not be misunderstood[119] as a mere theory of special envelop curves. The theory of (or the procedure to make) envelop curves is meaningful only after a one-parameter family of curves is supplied. Thus, from the envelop point of view, the most crucial point is that renormalization provides a principle to construct the one parameter family to which the theory of envelop may be applied. Needless to say, the key to singular perturbation is this principle and not the envelop interpretation, which may not always be useful.

9.10 What the simple example suggests

The above simple example suggests the following:

[118] This was pointed out by T. Kunihiro. T. Wall seems to be the first to construct a result that can be obtained naturally by a renormalization group method as an envelop of approximate solutions: F. T. Wall, "Theory of random walks with limited order of non-self-intersections used to simulate macromolecules," J. Chem. Phys. **63**, 3713 (1975); F. T. Wall and W. A. Seitz, "The excluded volume effect for self-avoiding random walks," J. Chem. Phys. **70**, 1860 (1979). Later, the same method (called the coherent anomaly method) was systematically and extensively used by M. Suzuki to study critical phenomena.

[119] As Kunihiro did

(1) The secular term is a divergence, and renormalization procedure removes this divergence to give the same result singular perturbation methods give. Singular perturbation methods are 'renormalized ordinary perturbations.'[120]
(2) The renormalization group equation is an equation governing slow phenomena. The core of singular perturbation theories is to extract such a slow motion equation, which can be obtained in a unified fashion with the aid of renormalization.[121]

9.11 Resonance due to perturbation

The second class of singular perturbation is due to divergence caused by resonance. If a harmonic oscillator is perturbed by an external perturbation with the frequency identical to the oscillator itself, its amplitude increases indefinitely. For a globally stable nonlinear system, this divergence is checked sooner or later by some nonlinear effect, so there is no genuine divergence. However, if the nonlinear term is treated as perturbation, this effect disappears from the perturbation equations, so singularity due to resonance shows up. Even if the average external force is zero, resonance has a 'secular effect.' That is, there is an effect that accumulates with time. The etymology of 'secular term' lies here.

9.12 Weakly nonlinear oscillators

A typical example illustrating that an ordinary perturbation series is plagued by resonance is the following weak nonlinear oscillator:

$$\frac{d^2y}{dt^2} + y = \epsilon(1-y^2)\frac{dy}{dt}, \tag{9.18}$$

where ϵ is a small positive constant (so the nonlinearity is weak). This equation is a famous equation called the van der Pol (1889-1959) equation. Introducing the following expansion

$$y = y_0 + \epsilon y_1 + O(\epsilon^2) \tag{9.19}$$

into (9.18), and equating terms with the same power in ε, we obtain

$$\frac{d^2y_0}{dt^2} + y_0 \quad = \quad 0, \tag{9.20}$$

[120] As long as the lecturer has experienced, many (almost all?) problems solved by named singular perturbation methods can be solved by renormalization method in a unified fashion without any particular prior knowledge.

[121] Many (all?) famous equations governing phenomenological behaviors (e.g., the nonlinear Schrödinger equation, the Burgers equation, the Boltzmann equation, etc.) can be derived as renormalization group equations.

$$\frac{d^2y_1}{dt^2} + y_1 \quad = \quad (1 - y_0^2)\frac{dy_0}{dt}, \tag{9.21}$$

etc.

9.13 Appearance of secular terms

The general solution to the first equation (9.20) may be written as

$$y_0(t) = Ae^{it} + \text{ c.c.}, \tag{9.22}$$

where A is a complex constant and c.c. implies complex conjugate. Using this in the second equation (9.21), we get

$$\frac{d^2y_1}{dt^2} + y_1 = iA(1 - |A|^2)e^{it} - iA^3e^{3it} + \text{ c.c.} \tag{9.23}$$

We have only to obtain its special solution. To this end it is the easiest to use Lagrange's method of varying coefficients.[122] Thus, we obtain

$$y_1 = \frac{1}{2}A(1 - |A|^2)te^{it} + \frac{i}{8}A^3e^{3it} + \text{ c.c.} \tag{9.24}$$

We have obtained the naive perturbation series as

$$y(t) = Ae^{it} + \epsilon\left[\frac{1}{2}A(1 - |A|^2)te^{it} + \frac{i}{8}A^3e^{3it}\right] + \text{ c.c.} + O(\varepsilon^2). \tag{9.25}$$

Clearly, there is a secular term. The reason for it is that the right-hand side of the equation for y_1 contains the term proportional to e^{it} that has the same frequency as the harmonic oscillator expressed by the left-hand side.

[122] For example, a special solution u to the second order ordinary differential equation

$$\frac{d^2y}{dx^2} + a\frac{dy}{dx} + by = f$$

may be constructed as follows in terms of the fundamental solutions of the corresponding homogeneous equation ϕ_1 and ϕ_2:

$$u = C_1\phi_1 + C_2\phi_2,$$

where the coefficients (functions) C_1 and C_2 are obtained by solving the following equations:

$$\frac{dC_1}{dx} = -\frac{f\phi_2}{W}, \quad \frac{dC_2}{dx} = \frac{f\phi_1}{W}.$$

Here, W is the Wronskian $W = \phi_1\phi_2' - \phi_2\phi_1'$.

9.14 Renormalization of resonance

Again, numerous singular perturbation methods have been developed to cure secular terms, but our procedure is exactly the same as in the simple example above **9.5**. We separate t as $(t-\tau)+\tau$, and then absorb τ into the constant A that depends on the initial condition. The renormalized result is

$$y(t) = A(\tau)e^{it} + \epsilon\left[\frac{1}{2}A(\tau)(1-|A(\tau)|^2)(t-\tau)e^{it} + \frac{i}{8}A(\tau)^3e^{3it}\right] + \text{ c.c.} + O(\epsilon^2). \quad (9.26)$$

Since y cannot depend on τ, the renormalization group equation $\partial y/\partial\tau = 0$ becomes

$$\frac{dA}{dt} = \epsilon\frac{1}{2}A(1-|A|^2) + O(\epsilon^2), \quad (9.27)$$

where τ is already replaced with t. This is an equation governing the long time behavior of the amplitude. Setting $t = \tau$ in (9.26), we get

$$y(t) = A(t)e^{it} + \epsilon\frac{i}{8}A(t)^3e^{3it} + \text{ c.c.} + O(\epsilon^2). \quad (9.28)$$

The key result is the amplitude equation (9.27).

In this example, the case $\epsilon = 0$ and the case $\epsilon > 0$ are qualitatively different. For a harmonic oscillator, any amplitude is allowed. In contrast, too large or too small amplitudes are not stable for (9.18) as can be seen from its right-hand side term: if $(1-y^2) < 0$, then it is an acceleration term that reduces the amplitude; otherwise, it is an acceleration term injecting energy to the oscillator. Indeed, according to the above approximate calculation, the first order solution slowly converges to a limit cycle expressed by $|A| = 1$. Pragmatically, much simpler calculational method called proto-RG approach exists. See **9.15**

Notice that the amplitude equation (9.27) contains ϵ. As expected from the result of the preceding section, the renormalization group equation describes a slow change of the amplitude (the actual motion is a busy rotation of period about 2π). As already suggested in the preceding section, the renormalization group approach supplies a new point of view for singular perturbation: to extract such a slow motion equation is the key point of the singular perturbation problems.

9.15 Proto RG equation

Let us consider an autonomous equation

$$Ly = \varepsilon N(y), \quad (9.29)$$

where L is a linear operator and N is a nonlinear operator. We make a formal expansion as

$$y = y_0 + \varepsilon y_1 + \varepsilon^2 y_2 + \cdots. \tag{9.30}$$

We know these terms generally contain secular terms. Let us renormalize y_0: according to the spectrum of L, we may write

$$y_0 = \sum_i A_i e_i(t). \tag{9.31}$$

Each y_k has its own secular term Y_k: let us write $y_k = \eta_k + Y_k$. Let us dissect Y_k as

$$Y_k = t \sum_i P_k^{(i)}(A) e_i(t) + Q_k(t, A). \tag{9.32}$$

Thus, we can write y generally as follows:

$$y(t) = \sum_i A_i e_i(t) + t \sum_i P_i(A) e_i(t) + Q(t, A) + R(t, A), \tag{9.33}$$

where $Q(t, A) = \sum_k \varepsilon^k Q_k(t, A)$ is the secular term containing higher powers of t, and R is the rest.

After renormalization (9.33) reads

$$y(t, \tau) = \sum_i A_{Ri}(\tau) e_i(t) + R(t, A_R(\tau)). \tag{9.34}$$

Here, note that $y(t, t) = y(t)$. We have

$$A_{Ri}(\tau) = A_I + \tau P_i(A). \tag{9.35}$$

Let us introduce $\tilde{L}_i$ by

$$L(f(t)e_i(t)) = (\tilde{L}_i f(t)) e_i. \tag{9.36}$$

Then, (9.35) reads

$$\tilde{L}_i A_{Ri} = \tilde{L}_i P_i(A_R). \tag{9.37}$$

This is the proto-RG equation. Notice that the RHS is just $(-)$ the perturbation term containing e_i, so we can read it off from the perturbation equation without solving it.

9.16 Proto RG applied to van der Pol

As an illustration, let us go back to the van der Pol equation (9.18). We have e^{it} and its conjugate as $e_i(t)$. Thus, (9.21) tells us

$$-\tilde{L}_i P_i(A) = iA(1-|A|^2). \tag{9.38}$$

From (9.36) we get $\mathcal{L}_i$ for this case is

$$\left[\frac{d^2}{dt^2}+1\right] f(t)e^{it} = -f(t)e^{it} + 2ie^{it}\frac{d}{dt}f(t) + e^{it}\frac{d^2}{dt}f(t) + f(t)e^{it}, \tag{9.39}$$

so we have

$$\mathcal{L}_i = \frac{d^2}{dt^2} + 2i\frac{d}{dt}. \tag{9.40}$$

Therefore, the protoRG equation reads

$$\left[\frac{d^2}{dt^2} + 2i\frac{d}{dt}\right] A_R(t) = \varepsilon i A_R(1-|A_R|^2), \tag{9.41}$$

but differentiation wrt t gives $O[\varepsilon]$ quantity, so we may keep only the first derivative. Thus, we have obtained the lowest order amplitude equation (9.27) almost for free.

9.17 How reliable is the renormalization group method?

It is easy to prove that (9.28) stays with the true solution within the error of $O[\varepsilon]$ for the time scale $1/\varepsilon$ by a standard argument with the aid of the Grönwall inequality **3.22**. However, such a result never tells us anything definite about the long-term behavior of the system. For example, even the existence of a limit cycle cannot be demonstrated.

A recent work by H. Chiba[123] considerably clarified this problem. Apart from some technicality, his conclusion is: "the long-time behavior of the system with small ε can be qualitatively inferred from its renormalization group equation." That is, roughly speaking, the invariant manifold of the renormalization group equation is diffeomorphic to that of the original equation. In the resonance example above, it is trivial that the renormalization group equation has a hyperbolic limit cycle, so we can conclude that the original equation also has a hyperbolic limit cycle. At least an intuitive explanation of the qualitative reliability of the RG results is attempted in the following.

[123] H. Chiba, "C^1 approximation of vector fields based on the renormalization group method," SIAM J. Appl. Dyn. Syst. **7**, 895 (2008).

9.18 Chiba's logic[124]

(1) The invariant manifold of the renormalization group equation and that of the equation governing the (truncated) renormalized perturbation series are (crudely put; see **9.19**) diffeomorphic.
(2) The differential equation governing the (truncated) renormalized perturbation series is at least C^1-close to the original differential equation.
With (2) and Fenichel's theorem (see **9.20**),
(3) The invariant manifold of the equation governing the (truncated) renormalized perturbation series is diffeomorphic to the invariant manifold of the original equation.

We may expect that the invariant manifolds of C^1-close vector fields are 'close' in some sense. This is, however, a bit delicate question even under hyperbolicity if we demand C^1-closeness of the manifolds.

At least in the case of diffeomorphisms it is known that normal hyperbolicity (see below) is a necessary and sufficient condition for an invariant manifold to persist.[125] Thus, hyperbolicity of invariant manifolds should not be enough to guarantee the qualitative similarity of the renormalization group equation to the original equation.

What Chiba demonstrated is that if the original system has a normally hyperbolic invariant manifold, then the renormalization group equation preserves it. Although for continuous dynamical systems, the relation between the normal hyperbolicity and C^1-structural stability seems not known, probably, we can conjecture that if the original equation is C^1-structurally stable (at least near its invariant manifold), then its renormalization group equation preserves the invariant manifolds of the original system.

9.19 Some technical comments as to Chiba's theory

The relation between the solution $A_R(t)$ to the renormalization group equation and the renormalized perturbation series solution $y(t, 0, A_R(t))$ of the original equation is given by (here maximally the same notations are used as in Section 3.7) the

[124] The explanation may oversimplify and may not do justice to the original theory, so those who are seriously interested in the proof should read the original paper.

[125] M. Hirsch, C. Pugh and M. Shub, "Invariant manifolds," Bull. Amer. Math. Soc. **76**, 1015 (1970), and R. Mañé, "PERSISTENT MANIFOLDS ARE NORMALLY HYPERBOLIC," Trans. Amer. Math. Soc. **246**, 271 (1978). These papers discuss the C^1-closeness of the invariant manifolds.

map α_t defined as $\alpha_t(A) = \sum A_i e_i(t) + \eta(t, A)$ (thus $\alpha_t(A_R(t)) = y(t, 0, A_R(t))$; see (**??**)). Here, we are interested in truncated solutions to some power of ε. Thus, we make a truncated version of α_t by truncating η. The differential equation governing $\alpha_t(A_R(t))$ (both α_t and $A_R(t)$ are truncated) is the equation V' governing the truncated renormalized perturbation series. A technical complication is that the truncated $\alpha_t(A_R(t))$ is generally explicitly time-dependent ($\alpha_t(x)$ is t-dependent), so the invariant set of V' must be considered in the 'space-time' (i.e., (t, y)-space); what (1) asserts is that the invariant set of the truncated original equation V' considered in the (t, y)-space and the direct product of time and the invariant set of the renormalization group equation are diffeomorphic.

(2) should not be a surprise to physicists. The difference between the equation V' governing $y(t, 0, A_R(t))$ (appropriately truncated) and the original equation V must be bounded, since renormalized result is bounded uniformly. Thus, their closeness should be obvious. Here, we need the closeness of the derivatives as well. The approximate solution and the exact solution are differentiable (actually C^1). Therefore, derivatives needed to calculate the derivatives of the vector fields are all bounded (if the appropriate derivatives of formula with respect to y exist). Thus, the equation governing $y(t, 0, A_R(t))$ (appropriately truncated) and the original equation are C^1-close.

The final step is (3); since the equation V' governing $y(t, 0, A_R(t))$ (appropriately truncated) and the original equation V are C^1-close, if normal hyperbolicity (see the next note) of the invariant manifold may be assumed, then Fenichel's theorem concludes the demonstration, if both V and V' are autonomous, but the equation V' governing $y(t, 0, A_R(t))$ (appropriately truncated) is not generally autonomous. Chiba overcame this problem as in (1); to consider the systems in space-time. Fenichel's theorem can be applied there, and the C^1-closeness of the invariant sets is established.

Thus, the invariant manifolds of the renormalization group equation and of the original equation are diffeomorphic.

9.20 Normal hyperbolicity and Fenichel's theorem

Let us consider a (continuous time) dynamical system defined on a subset U of a vector space with a (compact) invariant manifold M. M is assumed to have its unstable and stable manifolds, and the tangent space is decomposed as $T_M U = TM \oplus E^s \oplus E^u$, where E^s is the stable bundle and E^u the unstable bundle. M is normal hyperbolic, if the flow along the manifold M is 'slower' than the flows in the stable and unstable manifolds. The illustration (Fig. 9.1) is basically the same in Fenichel's original paper.

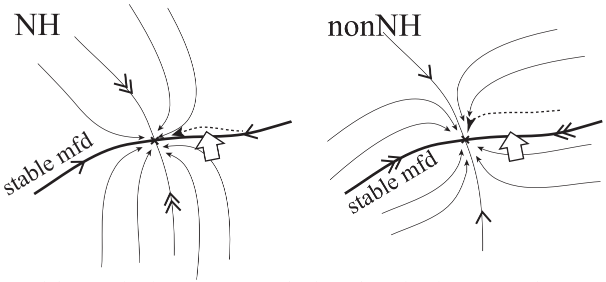


Figure 9.1: Why normal hyperbolicity is needed. The thick curve denotes the invariant manifold and its stable manifold is illustrated. The big white arrow denotes perturbation. 'NH' denotes the normal hyperbolic case, where the flow on the stable manifold is 'faster' (double arrows) than the flow on the invariant manifold. That is, normal direction is 'more' hyperbolic. 'nonNH' denotes the hyperbolic but not normally hyperbolic case. In the NH case the perturbation cannot disrupt the smoothness of the invariant manifold near 'x,' but that is not the case for the nonNH case; a cusp could be formed.

Theorem 3.8.1 [Fenichel][126] Let X be a C^r-vector field ($r \geq 1$) on $\mathbb{R}^n$. Let M be a normally hyperbolic invariant manifold of X without boundary. Then, for any C^r-vector field Y in a certain C^1-neighborhood of X is a Y-invariant manifold C^r-diffeomorphic to M. This diffeomorphism is C^1-close to the identity.

[126] N. Fenichel, "Persistence and smoothness of invariant manifolds for flows," Indiana Univ. Math. J., **21**, 193 (1971).

10 Lecture 10: Classical mechanics: review

Newton realized that if we know the current position x and velocity $\dot{x}$ of a point mass, its future (and its past) is completely determined. This is called the Newton-Laplace principle of determinacy. This means its acceleration $\ddot{x}$ is determined by x and $\dot{x}$. The resultant ODE is called Newton's equation of motion.

Since every graduate student should be very familiar with the practical use of classical mechanics, in this lecture, I concentrate on topics that are not very widely known or not so much stressed:
(1) How to construct a variational principle,
(2) When the variational principle is actually a minimization principe,
(3) A succinct demo of the Jacobi identity for Poisson brackets.
(4) How Schrödinger 'used' the Hamiltonian principle to derive his equation.

10.1 Newton-Laplace Principle of Determinacy[127]

The principle asserts that the state (= point in the phase space) of a mechanical system (= everything from Newton's and Laplace's point of view) at any fixed moment of time t uniquely determines all of its (future and past) states:

From $x(t_0)$ and $v(t_0) = \dot{x}(t_0)$, $(x(t), v(t))$ for all t is uniquely determined.

In particular, we can calculate the acceleration as

$$\ddot{x} = f(x, \dot{x}, t). \tag{10.1}$$

This is known as Newton's equation of motion. With time-reversal symmetry there is no first order derivatives in the equation

$$\ddot{x} = f(x, t). \tag{10.2}$$

Thus, to describe the system mechanics is to provide f, the force (experimentally). As noted in the preface to *Principia*, for Newton to find f for various phenomena was the core physics. This idea hindered the kinetic theory of gasses to explain the gas pressure.

The equivalence of Newton's equation of motion and the principle of determinacy is shown by the unique existence theorem of the solution for the ODE (see **3.18**).

[127] The following paper proposes to use the halting problem to deny the existence of Laplace's demon: Josef Rukavicka, Rejection of Laplace's Demon. Am Math Month 121 498 (2014). Basically, the question is: what is the significance of undecidable questions in this context?

Thus it is not unconditional, but as long as the motion is sufficiently smooth the equivalence is guaranteed.

Newton introduced the concept of 'force' and established the law of universal gravitation (with superposition principle).

10.2 Determinacy implies predictability?

One of the key issues of nonlinear dynamics is to make it clear that the answer to the question is negative: even though deterministic you cannot predict the future of the system, because the indefiniteness (error) in the intial condition could be exponentially magnified within a short time span.

There are, however, actually, more serious reasons why determinacy cannot generally imply predictability. One is already mentioned in the footnote of **10.1**: for example, the calculation needed to predict the future may not end. I do not know whether we can make a natural-looking ODE example for this. In this case whether computation can actually produce a number or not is the issue; we cannot even predict whether the computer will eventually give the answer or not. The other case of unpredictability is that the computer can indeed produce numbers, but their reliability (i.e., the size of the error bar) are never guaranteed.

10.3 Variational principle

The fundamental equation of mechanics is (10.2), but why is this form? Is there any deeper reason (yes, rational reason) for the Creator to choose this law?[128] Somehow, the law should be 'optimized', a natural route to variational principles. In elementary classical mechanics we have already learned Lagrange's principle, but here let us construct the variational principle from (10.2). We use

Theorem [Veinberg][129] Suppose

[128] This is exactly the 'naturalness' question (used in high energy physics). C. Lanczos, *The variational principle of mechanics* (University of Toronto Press, 3rd ed., 1960) Preface says, "There is hardly any other branch of the mathematical sciences in which abstract mathematics speculation and concrete physical evidences go so beautifully together and complement each other so perfectly." It is a good lesson to know that mathematical beauty does not guarantee the correctness of a theory in natural science. It is very often the case that what we believe natural is not actually naturally realized in Nature; to recognize this may be a sign of a true progress.

[129] See R. W. Atherton and G. M. Homsy, On the existence and formulation of variational principle for nonlinear differential equations, Studies Appl. Math. **LIV**, 31 (1975). For ODE there is a newer paper: I. A. Anderson and G. Thompson, The inverse problem of the calculus of variation for ordinary differential equations, Memoir AMS **98**, Number 473 (1991).

(1) N is an operator from a Hilbert space $\mathcal{H}$ into its conjugate space,
(2) N has a linear Gateau derivative[130] $DN(u, h)$ at every point of the ball $B = \{u \,|\, \|u - u_0\| < r\}$ and for any $h \in \mathcal{H}$,
(3) The scalar product $\langle h_1, DN(u, h_2)\rangle$ is continuous at every point of the ball B for any $h_1, h_2 \in \mathcal{H}$.
Then, a necessary and sufficient condition for $N(u) = 0$ to be the Euler-Lagrange's equation of a variational principle in the ball B is the symmetry

$$\langle h_1, DN(u, h_2)\rangle = \langle h_2, DN(u, h_1)\rangle. \tag{10.3}$$

The variational functional $F(u)$ is given by

$$F(u) = -\int dt \int_0^1 d\lambda\, u(t) N(\lambda u(t)). \tag{10.4}$$

You should have realized a perfect parallelism between the condition for a force $\boldsymbol{F}$ to be conservative: curl$\boldsymbol{F} = 0$ and the above theorem.

10.4 Application to Newton's equation of motion

(10.3) implies (exercise!) that f cannot depend on $\dot{x} = v$ (that is, imposing the variational principle enforces time-reversal symmetry) and the force must be conservative. Under these conditions we obtain (exercise!) the usual result we know: the variational functional is A called the action:

$$A = \int dt\, L, \tag{10.5}$$

with

$$L = T - V, \tag{10.6}$$

called the Lagrangian, where T is the kinetic energy and V the potential energy. The equation of motion obtained from $\delta A = 0$ (the action principle) reads (Lagrange's equation of motion), as you know well,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0. \tag{10.7}$$

[130] ⟪**Gateau derivative**⟫ This is a functional-derivative counterpart of the directional derivative, and is also called the weak derivative. Let $F : X \to Y$, where X and Y are normed space. Then,

$$DF(x, h) = \frac{d}{dt}F(x + th)\bigg|_{t=0}.$$

Remark on the Lagrangian:
(0) Landau-Lifshitz, *Mechanics* Chapter 1 is the best practical introduction.
(1) L is not unique; we may add any total derivative wrt time. This allows more general transformations than the ordinary coordinate changes (called canonical transformations (Lecture 13)).
(2) T is always a quadratic form of $\dot{x}$.

10.5 Action minimum principle

The variational principle itself does not care whether A is minimum or not along the actual motion, but the founding fathers of the principle clearly expected 'minimization.' This is actually true as long as the path is not too long. Precisely put, until the stationary curve hits the conjugate point, the minimum principle is true.

10.6 Conjugate point

Consider two actual trajectories going through a point A making a small angle with each other. If these two trajectories cross with each other at B, it is called a conjugate point of A (see Fi.g 10.1).

If the final point is reached from the initial point before reaching its conjugate point, the action is actually minimum.

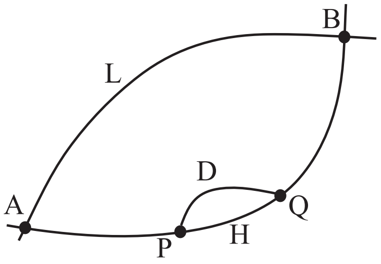


Figure 10.1: A conjugate point. Here, trajectories L and H both satisfy the variational principle.

[Demo] In Fig. 10.1 trajectories L and H both satisfy the variational principle, starting from A with different directions, and then cross for the first time at B (a conjugate point of A). Suppose APHQB is not an actual trajectory. If PHQ is not the actual trajectory, then there must be an actual one PDQ. Since we can choose Q sufficiently close to P,[131] **10.7** tells us the action along PDQ

[131] Suppose PHQ is not an actual path. Take R on this path. If PR is an actual path, we can choose replace PHQ with RHQ. If PR is not actual, there must be a 'bypass' which is the actual path.

is smaller than that along PHQ. Then, $\delta^2 A < 0$ on PDQ, contradicting the assumption that $\delta^2 A = 0$ along PHQ.

10.7 Locally, action principle is minimum principle

We first rewrite the action principle in the form of Maupertuis (1698-1759)'s principle (= the action principle on the constant energy E surface):

$$A = \int_{t_0}^{t_1} dt\,(T - V) = \int_{t_0}^{t_1} dt\,(2T - E) = 2\int_{t_0}^{t_1} dt\,T - E(t_1 - t_0). \tag{10.8}$$

This means we have only to consider the first term as the action (denoted as A'). Since T is quadratic wrt $\dot{q}$, where q is the spatial coordinate, we may write $T = A_{ij}\dot{q}_i\dot{q}_j/2 = A_{ij}dq_i dq_j/2dt^2$ (summation convention implied), so

$$dt = \sqrt{A_{ij}dq_i dq_j/2T} \tag{10.9}$$

Therefore,

$$A' = 2\int_{t_0}^{t_1} dt\,T = \int_{s=t_0}^{s=t_1} ds\,\sqrt{2(E - V)}\sqrt{A_{ij}q_i'(s)q_j'(s)}. \tag{10.10}$$

Now, we must study $\delta^2 A'$. We vary the trajectory as $q \to q + \delta q$. Choose the time range $[t_0, t_1]$ sufficiently small so the variation δq is much smaller than $\delta q'$. Therefore, we have only to consider the second $\sqrt{\cdots}$ in (10.10). This is a (positive definite) quadratic form, so its stationary value must be minimum.

10.8 Legendre transformation $L \to H$

We know Lagrange's equation of motion means the following 1-form:

$$dL = p d\dot{q} - \frac{\partial V}{\partial q}dq, \tag{10.11}$$

where q are spatial coordinates, and p the momenta. Now we apply the following Legendre transformation

$$H = \sup_{\dot{q}}[p\dot{q} - L] = T + U. \tag{10.12}$$

H is called the Hamiltonian and

$$dH = \dot{q}dp + \frac{\partial V}{\partial q}dq \tag{10.13}$$

implies the Hamilton's equation of motion (= the canonical equation of motion)

$$\dot{q} = \frac{\partial H}{\partial p}, \ \frac{\partial V}{\partial q} = \frac{\partial H}{\partial q} = -\dot{p}. \tag{10.14}$$

10.9 Hamilton's principle.

Since $L = \sum p_i \dot{q}_i - H$ (**10.8**), the action principle **10.4** may be rewritten as

$$\delta \int \left[\sum_i p_i \dot{q}_i - H \right] dt = 0. \tag{10.15}$$

Regarding p and q to be independent variables, we obtain directly the canonical equation of motion (10.14) (use integration by parts). The resultant variational principle (10.15) is called *Hamilton's principle.*

10.10 Poisson bracket.

Let f and g be differentiable phase functions (functions of q and p). We introduce the *Poisson bracket* $[f, g]_{PB}$[132] as

$$[f, g]_{PB} \equiv \frac{\partial f}{\partial q}\frac{\partial g}{\partial p} - \frac{\partial g}{\partial q}\frac{\partial f}{\partial p} = \sum_i \left(\frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial q_i}\frac{\partial f}{\partial p_i} \right). \tag{10.16}$$

10.11 Canonical equation of motion in terms of Poisson brackets

(10.14) reads

$$\dot{q}_i = [q_i, H]_{PB}, \quad \dot{p}_i = [p_i, H]_{PB}. \tag{10.17}$$

Thus, the canonical equations of motion for q and p have become symmetric. (cf. Heisenberg's equation of motion in QM)

10.12 Properties of Poisson bracket

Note the following general relations:

[132] or simply, $[f, g]$ when quantum and classical mechanics do not appear simultaneously.

(i) $[f,g]_{PB} = -[g,f]_{PB}$.
(ii) $[f,g+h]_{PB} = [f,g]_{PB} + [f,h]_{PB}$.
(iii) $[cf,g]_{PB} = c[f,g]_{PB}$, where c is a constant.
(iv) [*Jacobi's identity*] $[f,[g,h]_{PB}]_{PB} + [g,[h,f]_{PB}]_{PB} + [h,[f,g]_{PB}]_{PB} = 0$.

(i)-(iv) imply that the Poisson bracket defines a *Lie algebra* structure for the set of the differentiable phase functions. (i)-(iii) are easy to show. (iv) implies that $[\ ,\ ]_{PB}$ is not associative. Usually showing (iv) requires almost brute force lengthy calculation, but see at the end of this item.

Notice further that
(v) $[fg,h]_{PB} = f[g,h]_{PB} + [f,h]_{PB}g$.
(vi) If f and g depend on a parameter α differentiably, then $d[f,g]_{PB}/d\alpha = [df/d\alpha, g]_{PB} + [f, dg/d\alpha]_{PB}$.
(vii) Let F be a function of phase functions f_i. Then $[F,g]_{PB} = (\partial F/\partial f_i)[f_i,g]_{PB}$.

Notice that for any differentiable phase function h, we can define a one parameter group defined by $df/d\alpha = [f,h]_{PB}$ (actually this is called an infinitesimal canonical transformation and h its generator; see **13.4**). Use this to (vi) (compute $d[f,g]/d\alpha$), and we get Jacobi's identity (iv) immediately.

10.13 Integral of motion; conservation of energy

An *integral of motion* Q is a phase function which is time-independent. (10.17) implies

$$[Q,H]_{PB} = 0, \tag{10.18}$$

if Q does not have any explicit time dependence.

Obviously,

$$[H,H]_{PB} = 0. \tag{10.19}$$

This is the conservation of (mechanical) energy (H must be t-independent).

10.14 Poisson brackets of various quantities

We can easily demonstrate

$$[p_i,p_j]_{PB} = 0, \quad [q_i,q_j]_{PB} = 0, \quad [q_i,p_j]_{PB} = \delta_{ij}. \tag{10.20}$$

Here (i,j,k) is a cyclic permutation of $(1,2,3)$. We also have for angular momentum L_i

$$[L_i,L_j]_{PB} = L_k, \quad [L_i,p_j]_{PB} = p_k, \quad [L_i,q_j]_{PB} = q_k, \tag{10.21}$$

and other Poisson brackets between angular momentum components and phase variables are zero. More generally, for a differentiable phase function F,

$$[F, p_i]_{PB} = \frac{\partial F}{\partial q_i}, \quad [F, q_i]_{PB} = -\frac{\partial F}{\partial p_i}. \tag{10.22}$$

10.15 Hamilton-Jacobi's equation.

Let us consider the action A defined in **10.4** as a function of the end time t and the end position $q_i = q_i(t)$:

$$A(q, t) = \int_{t_0}^{t} L(q(s), \dot{q}(s), s)ds, \tag{10.23}$$

where L is the Lagrangian. Obviously,

$$\frac{dA}{dt} = L(q_i(t), \dot{q}_i(t), t) = p_i(t)\dot{q}_i(t) - H, \tag{10.24}$$

where H is the Hamiltonian (10.12). Hence,

$$dA = \sum_i p_i dq_i - Hdt, \tag{10.25}$$

that is,

$$\frac{\partial A}{\partial q_i} = p_i, \quad \frac{\partial A}{\partial t} = -H. \tag{10.26}$$

Since H is a function of q_i, p_i, and t, we obtain a closed equation for A

$$\frac{\partial A}{\partial t} + H\left(q_i, \frac{\partial A}{\partial q_i}, t\right) = 0. \tag{10.27}$$

This is called the *Hamilton-Jacobi equation*. Note that it does not contain the momentum coordinates. For example, for a particle of mass m in the potential V, the Hamilton-Jacobi equation reads

$$\frac{\partial A}{\partial t} + \frac{1}{2m}\left\{\left(\frac{\partial A}{\partial x}\right)^2 + \left(\frac{\partial A}{\partial y}\right)^2 + \left(\frac{\partial A}{\partial z}\right)^2\right\} + V(x, y, z) = 0. \tag{10.28}$$

10.16 Schrödinger's "Quantization as eigenvalue problem I"[133]

Schrödinger starts with the Hamilton-Jacobi equation (10.27) in the following separated form:

$$H\left(q, \frac{\partial A_0}{\partial q}\right) = E, \tag{10.29}$$

with $A = -Et + A_0$, where A_0 is called Hamilton's principal function. He introduces ψ as

$$A_0 = \hbar \log \psi. \tag{10.30}$$

Schrödinger use the symbol K instead of $\hbar$. Thus, the equation (10.29) now reads

$$(\nabla\psi)^2 - \frac{2m}{\hbar^2}(E - V)\psi^2 = 0. \tag{10.31}$$

Instead of this equality, he replaces the quantization condition with the following variational problem; he says in a footnote that he is aware that this formulation is not necessarily unique:

$$\delta J = \delta \int d^3\boldsymbol{x} \left[(\nabla\psi)^2 - \frac{2m}{\hbar^2}(E - V)\psi^2\right] = 0. \tag{10.32}$$

The resultant equation is (assuming that the wave function vanishes at infinity) the time-independent Schödinger equation.

In this paper, he goes on to solve the hydrogen atom completely (he acknowledges H. Weyl's help). The ad hoc introduction of the Schrödinger equation as outlined above is fully justified by recovering Bohr's energy levels.

[133] Quantisierung als Eigenwertproblem I, Ann. Phys(4) 79, 361-376 (1926): received on January 27, 1926

11 Lecture 11. Determination of motion

Here, we discuss (analytically) solvable classical mechanics problems. First, we understand the completely integrable cases geometrically (Liouville-Arnold theorem), and then describe such systems in terms of the action-angle variables.

Then, in Lecture 12, a recent integration technique (the Lax pair approach) is introduced. It will be illustrated with the Toda lattice. To discuss the method historically, we go back to Korteweg-de Vries (KdV) equation. [This naturally leads us to solitons, but I do not go into this vast topic.]

11.1 Determination of motion

Here the word 'determine' implies that we can construct a map that allows us to transform the dynamical flow in the phase space into a simple flow ('laminar flow' on a torus just like the rectifiability theorem **3.19**): (Q, P) (P is constant and $Q = Pt + c$) as we will see below. That is, to rectify the flow while preserving the structure of Hamiltonian dynamical systems (i.e., by a canonical transformation). If this can be done, the system is said to be integrable.

Thus, we need at least a rudimentary familiarity to the canonical transformation.

Needless to say, not all the systems allow such transformations (only integrable systems). Poincaré's recognition that the three (celestial) body problem is not integrable in this sense was the birth of modern classical mechanics/study of dynamical systems.

11.2 Canonical transformation

The transformation $T : (q, p) \to (Q, P)$ that preserves the form of the canonical equation of motion (10.14) is called a *canonical transformation*.[134] Thus, the transformation T to be canonical implies that the equation of motion in terms of the new variables reads

$$\frac{dQ_i}{dt} = \frac{\partial K}{\partial P_i}, \ \frac{dP_i}{dt} = -\frac{\partial K}{\partial Q_i}, \tag{11.1}$$

where $K(Q, P) = H(q(Q, P), p(Q, P))$ is the Hamiltonian in terms of the new variables.

[134] Corresponding to unitary transformations in quantum mechanics.

11.3 Generator of canonical transformation

For (11.1) to hold, we must have the action principle in terms of the new variables (Q, P).

Let (q, p) be the canonical coordinates and (Q, P) the new coordinates (they are phase functions). Then, the Hamilton's principle reads

$$\delta \int (P\dot{Q} - K)dt = 0. \tag{11.2}$$

Thus, we demand (see the remark below)

$$\delta \int \Big[(p\dot{q} - H) - (P\dot{Q} - K)\Big]\, dt = 0. \tag{11.3}$$

This means that the difference of these two integrals can be constant, so

$$(pdq - Hdt) - (PdQ - Kdt) = dF, \tag{11.4}$$

or

$$dF = pdq - PdQ + (K - H)dt \tag{11.5}$$

must be path-independent. That is, dF must be exact.[135] Thus, there is a function $F = F(q, Q, t)$ such that

$$dF = \sum_i p_i dq_i - \sum_i P_i dQ_i + (K - H)dt, . \tag{11.6}$$

F is called the *generator* of the canonical transformation $T : (q, p) \to (Q, P)$. (11.6) gives

$$\frac{\partial F}{\partial q} = p,\ \frac{\partial F}{\partial Q} = -P,\ \frac{\partial F}{\partial t} = K - H. \tag{11.7}$$

Solving these equations we can construct the canonical transformation T. In particular, if F is time-independent, then $K = H$ is obtained by replacing q and p in H in terms of Q and P.

Remark: For the system described by q and p to satisfy the canonical equation of motion, much more general transformations are allowed, but the form with the generator discussed above is the most convenient, especially because the Hamiltonian is virtually preserved if the generator is time-independent. Thus, when we say

[135] If form θ can be written as an (external) derivative of another form ω as $\theta = d\omega$, θ is called an exact form. A form θ that satisfies $d\theta = 0$ is called a closed form. Obviously, an exact form is a closed form.

'canonical transformation' we restrict ourselves to this type of transformations.

Applying a sort of Legendre transformation to generators, we can construct different (perhaps more convenient) transformations:

$$d(F + PQ) = pdq + QdP + (K - H)dt. \tag{11.8}$$

Thus, replacing F with $G = F + \sum_i P_i Q_i$, we obtain

$$dG = \sum_i p_i dq_i + \sum_i Q_i dP_i + (K - H)dt. \tag{11.9}$$

11.4 Complete integrability

Let $F_1, \cdots, F_n$ ($F_1 = H$) be pairwise involutive (i.e., $[F_i, F_j]_{PB} = 0$) smooth functions on $2n$-mfd M. If F_i's are functionally independent (or ∇F_i's are linearly independent on a dense set of M), then,
(1) $M_f = \{x \mid F_i(x) = f_i\}$ is diffeomorphic to $T^k \times \mathbb{R}^{n-k}$.
(2) The Hamiltonian flow on M_f takes the following form with the coordinate system on $T^k \times \mathbb{R}^{n-k}$: $\varphi_1, \cdots, \varphi_k$ mod 2π, $y_1, \cdots, y_{n-k}$

$$\dot{\varphi}_i = \omega_i, \;\; \dot{y}_j = c_j, \tag{11.10}$$

where ω and c are constants.

11.5 Slightly general form of integrability[136]

For a n-dimensional Hamiltonian system suppose there are n first integrals $F_1, \cdots, F_n$ (i.e., $[H, F_i] = 0$) such that

$$[F_i, F_j] = \sum_k c_{ij}^k F_k, \tag{11.11}$$

where c_{ij}^k are constants. If
(1) on $M_f = \{x \mid F_i(x) = f_i\}$ F_i are functionally independent
(2) $\sum_k c_{ij}^k f_k = 0$.
(3) The Lie algebra $\mathcal{A}$ of linear combination $\sum \lambda_i F_i$ is solvable.[137]

[136] Theorem 1 Chapter 4 of V. I. Arnold, V. V. Kozlov and A. I. Neishtadt, Mathematical aspects of classical and celestial mechanics in *Dynamical systems III* ed. V. I. Arnold (Springer, 1988).

[137] **⟪Solvable Lie algebra⟫** A Lie algebra $\mathfrak{g}$ is solvable, if

$$\mathfrak{g} \supset [\mathfrak{g}, \mathfrak{g}] \supset [[\mathfrak{g}, \mathfrak{g}], [\mathfrak{g}, \mathfrak{g}]] \to \{0\}$$

.

Then, the Hamiltonian dynamical system is solvable by quadrature. p107

11.6 Liouville-Arnold's theorem

This is a special case of 11.4 when the motion is bounded (confined in a finite space; M is compact). In this case 11.4 reads:
Let $F_1, \cdots, F_n$ ($F_1 = H$) be pairwise involutive (i.e., $[F_i, F_j]_{PB} = 0$) smooth functions on $2n$-mfd M. If F_i's are functionally independent (or ∇F_i's are linearly independent on a dense set of M), then
(1) $M_f = \{x \mid F_i(x) = f_i\}$ is diffeomorphic to T^n.
(2) The Hamiltonian flow on M_f takes the following form with the coordinate system on T^n: $\varphi_1, \cdots, \varphi_n$ mod 2π

$$\dot{\varphi}_i = \omega_i, \tag{11.12}$$

where ω are constants.

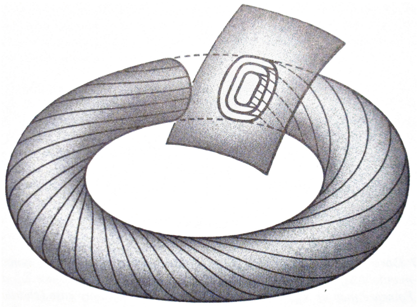


Figure 11.1: Arnold-Liouville foliated torus [Fig. 8.3.2 of Abraham & Marsden]

[Demo] Consider n Hamiltonian dynamical systems on M each of which is governed by F_j ($j \in \{1, \cdots, n\}$) as its Hamiltonian. We introduce a canonical coordinate system (q, p) for the phase space. The canonical equation of motion (s is the time variable) governed by the Hamiltonian F_j reads

$$\frac{dq}{ds} = \frac{\partial F_j}{\partial p}, \ \frac{dp}{ds} = -\frac{\partial F_j}{\partial q}. \tag{11.13}$$

Since F_j are involutive, the tangent vectors parallel to the flows for these n systems commute. Let us check this:[138]

$$[F_{1,p}\partial_q - F_{1,q}\partial_p, F_{2,p}\partial_q - F_{2,q}\partial_p] = [F_{1,p}\partial_q, F_{2,p}\partial_q] + [F_{1,q}\partial_p, F_{2,q}\partial_p] - [F_{1,p}\partial_q, F_{2,q}\partial_p] - [F_{1,q}\partial_p, F_{2,p}\partial_q] \tag{11.14}$$

$$\begin{aligned}
[F_{1,p}\partial_q, F_{2,p}\partial_q] &= (F_{1,p}F_{2,pq} - F_{2,p}F_{1,pq})\partial_q. && (11.15)\\
[F_{1,q}\partial_p, F_{2,q}\partial_p] &= (F_{1,q}F_{2,qp} - F_{2,q}F_{1,qp})\partial_p. && (11.16)\\
[F_{1,p}\partial_q, F_{2,q}\partial_p] &= F_{1,p}F_{2,qq}\partial_p - F_{2,q}F_{1,pp}\partial_q && (11.17)\\
[F_{1,q}\partial_p, F_{2,p}\partial_q] &= F_{1,q}F_{2,pp}\partial_q - F_{2,p}F_{1,qq}\partial_p && (11.18)\\
& && (11.19)
\end{aligned}$$

Collecting the terms containing ∂_q, we get

$$F_{1,p}F_{2,pq} - F_{2,p}F_{1,pq} + F_{2,q}F_{1,pp} - F_{1,q}F_{2,pp} = -\partial_p(F_{1,q}F_{2,p} - F_{1,p}F_{2,q}) = \partial_p[F_2, F_1]_{PB} = 0. \tag{11.20}$$

Similarly, collecting the terms containing ∂_p, we get

$$F_{1,q}F_{2,qp} - F_{2,q}F_{1,qp} - F_{1,p}F_{2,qq} + F_{2,p}F_{1,qq} = \partial_q(F_{1,q}F_{2,p} - F_{2,q}F_{1,p}) = \partial_q[F_1, F_2]_{PB} = 0. \tag{11.21}$$

Thus, the Lie bracket (11.14) vanishes.[139] This implies that the time evolution groups due to $\{F_i\}$ commute. Thus there are n independent directions.

The compact manifold M_f must be invariant under the n-dimensional Abelian group defined by the time evolutions due to $\{F_i\}$. This implies that M_f is diffeomorphic to T^n.

11.7 Action-angle variables[140]

If M is compact for a completely integrable system, then M_f is diffeomorphic to T^n.
(1) The small nbh of M_f in the ambient symplectic mfd M is diffeomorphic to $D \times T^n$, where D is a small domain in $\mathbb{R}^n$.
(2) In $D \times T^n$ there exists canonical coordinates φ mod $2\pi$and I ($I \in D$, $\phi \in T^n$ in which F_k's depend only on I.

[138] We are checking $[X, Y]f = 0$ for any differentiable f. That is, Two paths going from $(a, b) \to (A, B)$, that is, $(a, b) \to (A, b) \to (A, B)$ and $(a, b) \to (a, B) \to (A, B)$ give the same result. Notation: $\partial_x f = f_{,x}$ is used.

[139] We have shown that $[F, G]_{PB} = 0$ implies that the Hamiltonian flows defined by F and by G determine independent (commutative) time evolutions.

[140] Nekhoroshev proved a version when the involution relation holds for a subset of $\{F_j\}$ (Arnold DS III p117).

I are called action variables and φ angle variables. The completely integrable Hamiltonian H has the form $H = H(I)$. Consequently, the equation of motion reads

$$\dot{I} = 0, \quad \dot{\varphi} = \omega(I) = \frac{\partial H}{\partial I}. \tag{11.22}$$

Thus I labels the invariant tori. If the Hessian of $H(I)$ is nondegenerate, the systems is said to be non-degenerate.

11.8 Demonstration of 11.7[141]

Thanks to the Liouville-Arnold theorem **11.6** in the nbh of the torus $M_f \equiv T^n$, we can take as coordinates the functions $I_i = F_i$ and the angles φ_i. Since the differentials dF_i are linearly independent, we can use I and ϕ as a coordinate system for $D \times T^n$. We have to show that I, φ system is a canonical coordinate system. We know $[I_i, I_j]_{PB} = 0$ (involutive invariants). $[\varphi_j, I_k]_{PB}$ is a constant on M_f according to **11.6**, so it is a function of I only. The Jacobi identity applied to I_k, φ_i and φ_j implies (here $[\ ,\ ]$ implies $[\ ,\ ]_{PB}$)

$$[I_k, [\varphi_i, \varphi_j]] + [\varphi_i, [\varphi_j, I_k]] + [\varphi_j, [I_k, \varphi_i]] = 0, \tag{11.23}$$

so $[I_k, [\varphi_i, \varphi_j]]$ is a function of I only. Since I and φ describe the system dynamics, the Poisson bracket determinant $\det[I_k, \varphi_i]$ must not be zero. Therefore, we can solve the following equation

$$[I_k, [\varphi_i, \varphi_j]] = \sum_s [I_k, \varphi_s] \frac{\partial}{\partial \varphi_s} [\varphi_i, \varphi_j] \tag{11.24}$$

to conclude that $\partial[\varphi_i, \varphi_j]/\partial\varphi_s$ is a function of I only. Therefore, we may write

$$[\varphi_i, \varphi_j] = \sum_s f^s_{ij}(I)\varphi_s + g_{ij}(I). \tag{11.25}$$

Since the derivatives of φ must be single-valued, when φ_i and φ_j go around the torus, the Poisson bracket should also return to the original value. Therefore, φ_s dependence should not exist: $f^s_{ij}(I) = 0$.

We change $I \to J$ to make $[J_i, \varphi_k] = \delta_{jk}$. Let us check the possibility.

$$[I_i, \varphi_j] = \sum_k \frac{\partial I_i}{\partial J_k} [J_k, \varphi_j] = \frac{\partial I_i}{\partial J_j}. \tag{11.26}$$

[141] The proof here paraphrases with details the proof on p115 of Arnold, DS III. A general version for the situation **11.5** also holds (p118 Th10).

The consistency condition reads

$$\frac{\partial}{\partial J_s}[I_i,\varphi_j]=\frac{\partial}{\partial J_j}[I_i,\varphi_s] \iff \sum_k [I_k,\varphi_s]\frac{\partial}{\partial I_k}[I_i,\varphi_j]=\sum_k [I_k,\varphi_j]\frac{\partial}{\partial I_k}[I_i,\varphi_s], \tag{11.27}$$

but this follows from the Jacobi identity for I_i, φ_j nd φ_k:

$$[I_i,[\varphi_j,\varphi_k]]+[\varphi_j,[\varphi_k,I_i]]+[\varphi_j,[I_i,\varphi_k]]=0 \Rightarrow [\varphi_j,[\varphi_k,I_i]]+[\varphi_j,[I_i,\varphi_k]]=0. \tag{11.28}$$

Thus, we can obtain J such that $[J_i,\varphi_j]=\delta_{ij}$.

The remaining task[142] is to guarantee $[\varphi_i,\varphi_j]=0$. If this is not the case, set $\psi_i=\varphi_i+f_i(J)$. Then, we must solve the following equation for f:

$$[\psi_i,\psi_j]=[\varphi_i,\varphi_j]-\frac{\partial f_i}{\partial J_j}+\frac{\partial f_j}{\partial J_i}=0. \tag{11.29}$$

Consider formally

$$q=\sum_{i<j}[\varphi_i,\varphi_j]dJ_i\wedge dJ_j=\sum_{i<k}\left(\frac{\partial f_i}{\partial J_k}-\frac{\partial f_k}{\partial J_i}\right)dJ_i\wedge dJ_k=d\sum_i f_i dJ_i, \tag{11.30}$$

so $dq=0$ is the consistency (solvability) condition. This follows from the closedness of the symplectic 2-form $d\varphi\wedge dJ$ (which is a shorthand of $\sum_i d\varphi_i\wedge dJ_i$; that is, the invariance of (13.26) in **13.11**): $d(d\varphi\wedge dJ)=0$; we can compute this 2-form as

$$\sum_i d\varphi_i\wedge dJ_i=\sum_{i,j}\frac{\partial\varphi_i}{\partial J_j}dJ_j\wedge dJ_i=\sum_{i<j}\left(\frac{\partial\varphi_i}{\partial J_j}-\frac{\partial\varphi_j}{\partial J_i}\right)dJ_j\wedge dJ_i, \tag{11.31}$$

but note that

$$\frac{\partial\varphi_i}{\partial I_j}-\frac{\partial\varphi_j}{\partial I_i}=[\varphi_i,\varphi_j]. \tag{11.32}$$

Thus, q is a symplectic 2-form, so $dq=0$.

11.9 Calculation of action variables

Let $q.p$ be the symplectic coordinates in $\mathbb{R}^{2n}$, and $\gamma_1,\cdots,\gamma_n$ are the fundamental cycles of $M_f\equiv T^n$. Since $pdq-Id\varphi$ is closed, the difference

$$\oint_{\gamma_s} pdq-\oint_{\gamma_s} Id\varphi=\oint_{\gamma_s} pdq-2\pi I_s \tag{11.33}$$

[142] $[J_i,J_k]=0$, because $J=J(I)$ and I does not depend on φ.

is a constant. Since the action variables can be determined up to additive constants, this means

$$I_s = \frac{1}{2\pi}\oint_{\gamma_s} p dq. \tag{11.34}$$

11.10 Action for harmonic oscillator

Let the Hamiltonian be

$$H = \frac{1}{2}(a^2p^2 + b^2q^2). \tag{11.35}$$

Then, M_E specified by $H = E$ is an ellipse

$$\frac{p^2}{2E/a^2} + \frac{q^2}{2E/b^2} = 1. \tag{11.36}$$

Its area is $2\pi E/ab$. The angular frequency is $\omega = ab$ as can be seen from $\dot{p} = -b^2q$, $\dot{q} = a^2p \Rightarrow \ddot{q} = -(ab)^2q$. Thus,

$$I(E) = \frac{1}{2\pi}\oint p dq = E/ab = E/\omega. \tag{11.37}$$

11.11 Adiabatic invariants

Let $G(q, I)$ be the generating function of the canonical transformation $(q, p) \to (\omega, I)$ (see (13.4)) for a completely integrable system:

$$\frac{\partial G}{\partial q} = p, \frac{\partial G}{\partial I} = \omega. \tag{11.38}$$

Assume that the system is subjected to a small perturbation $\lambda(t)$. Then, the above canonical transformation now depends on time as well (through $\lambda(t)$), so the new Hamiltonian reads to order $O[\lambda]$

$$K = H(I) + \frac{\partial G}{\partial t} = H(I) + \Lambda\dot{\lambda}(t), \tag{11.39}$$

where $\Lambda = \partial G/\partial\lambda$. Therefore, the perturbed Hamilton's equation reads

$$\dot{I} = -\frac{\partial K}{\partial \omega} = \frac{\partial \Lambda}{\partial \omega}\dot{\lambda}(t). \tag{11.40}$$

Lam is a bounded function, so it must be a multiple periodic function. Therefore its derivative must have the vanishing time average. Therefore, it λ change very slowly, $\dot{I}$ must be zero. That is, I is invariant to order $\dot{\lambda}$. Such a quantity is called an adiabatic invariant.

12 Lecture 12. Lax pair and Toda lattice

The story in this lecture goes as follows chronologically.

Scott-Russel observed (1834) the first solitary wave propagation along the Union Canal, Scotland (and also he confirmed it experimentally) (see **12.6**). Then, Rayleigh and Bousinesque explained it theoretically (∼1870), and later Korteweg and his student de Vries derived the equation governing the solitary waves (the KdV equation; 1895).

Fermi had been interested in the foundation of statistical mechanics, and the availability of Maniac I in Los Alamos allowed him to pursue the equilibration process of nonlinear lattice systems (Fermi-Pasta-Ulam problem; ∼1955; see **12.5**). This study taught us that nonlinearity is not enough to establish thermal equilibrium due to persistent recurrence.

Zabusky and Kruskal studied a continuum approximation of the system (= the KdV equation) and discovered solitons (1965).[143] This led to the Gardener, Greene, Kruskal and Miura discovery of the relation between solitons and the Schrödinger equation.[144] This led to the general theory of the Lax pair (1968; see **12.1**).

The explanation of recurrence in the FPU problem in terms of solitons was wrong, but a lattice system with solitons was later discovered (the Toda lattice, see **12.3**), which can be reduced in a certain limit to the KdV equation.

12.1 Lax pair

Suppose

$$\dot{x} = f(x) \tag{12.1}$$

can be expressed as

$$\dot{L} = [A, L], \tag{12.2}$$

where A and L are square matrices whose components are functions of x.

Proposition [Lax] The eigenvalues of L are first integrals of (12.1).
[Demo] Let us solve (12.2) for L: for sufficiently small s (notice that A may be time-dependent)

$$L(t+s) = e^{sA}L(t)e^{-sA} + o[s]. \tag{12.3}$$

[143] "Interaction of "solitons" in a collisionless plasma and the recurrence of initial states", PRL **15**, 240 (1965).

[144] "Method for solving the Korteweg-de Vries equation" PRL **19**, 1095 (1967).

Therefore,

$$\det[L(t+s)-\lambda] = \det[L(t)-\lambda] + o[s] \tag{12.4}$$

That is, the characteristic equation of $L(t)$ is time-independent.

Remark The L-A pair has been found for almost all problems of classical mechanics previously integrated by other methods.

12.2 Euler's equation

Euler's equation reads

$$\dot{\boldsymbol{M}} = \boldsymbol{M} \times \boldsymbol{\omega}. \tag{12.5}$$

The Lax pair representation is

$$L = \begin{pmatrix} 0 & M_1 & -M_2 \\ -M_1 & 0 & M_3 \\ M_2 & -M_3 & 0 \end{pmatrix}, \ A = \begin{pmatrix} 0 & \omega_1 & -\omega_2 \\ -\omega_1 & 0 & \omega_3 \\ \omega_2 & -\omega_3 & 0 \end{pmatrix}. \tag{12.6}$$

Confirm the Lax pair representation.

12.3 Toda lattice

Consider a chain consisting of point masses with the nearest intereaction only given by the following potential

$$\phi(r) = \frac{a}{b} e^{-br} + ar. \tag{12.7}$$

If the position of the nth point mass is x_n, the equation of motion reads

$$m\frac{d^2 x_n}{dt^2} = a\left[e^{-b(x_n - x_{n-1})} - e^{-b(x_{n+1}-x_n)}\right]. \tag{12.8}$$

so $r_n = x_{n+1} - x_n$ obeys

$$m\frac{d^2 r_n}{dt^2} = 2a\left[e^{-br_n} - \frac{1}{2}\left(e^{-br_{n+1}} + e^{-br_{n-1}}\right)\right]. \tag{12.9}$$

A dimensionless form reads

$$\dot{Q}_n = P_n, \ \ \dot{P}_n = e^{-(Q_n - Q_{n-1})} - e^{-(Q_{n+1}-Q_n)}. \tag{12.10}$$

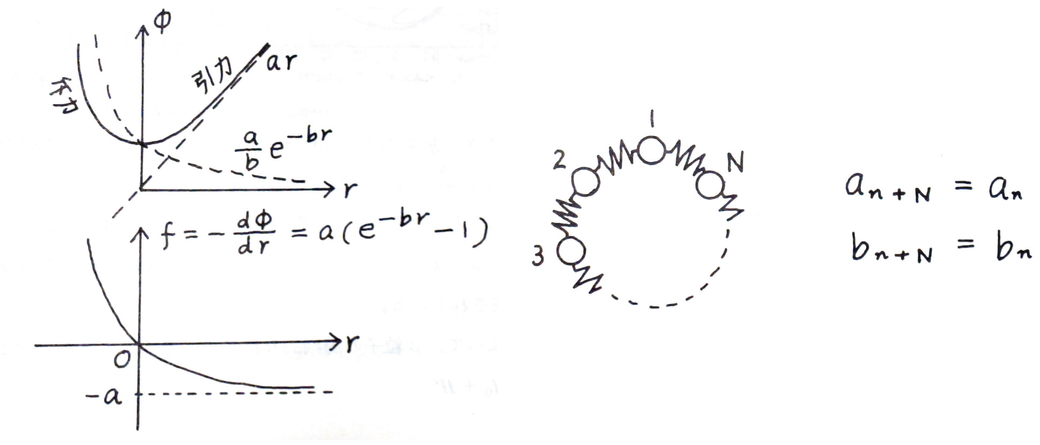


Figure 12.1: Toda lattice after M Toda

Now, introduce (these variables are not canonical variables)

$$a_n = \frac{1}{2}e^{-(Q_{n+1}-Q_n)/2},\ b_n = \frac{1}{2}P_n. \tag{12.11}$$

Then the dimensionless form reads (with a cyclic boundary condition $a_{n+N} = a_n$, $b_{n+N} = b_n$.

$$\begin{aligned} \dot{a}_n &= a_n(b_n - b_{n+1}), & (12.12)\\ \dot{b}_n &= 2(a_{n-1}^2 - a_n^2). & (12.13) \end{aligned}$$

This can be cast into the Lax pair form with

$$L = \begin{pmatrix} b_1 & a_1 & 0 & \cdots & \cdots & 0 & a_N \\ a_1 & b_2 & a_2 & 0 & \cdots & \vdots & 0 \\ 0 & a_2 & b_3 & \ddots & \ddots & \vdots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & b_{N-2} & a_{N-2} & 0 \\ 0 & 0 & \cdots & 0 & a_{N-2} & b_{N-1} & a_{N-1} \\ a_N & 0 & \cdots & \cdots & 0 & a_{N-1} & b_N \end{pmatrix}. \tag{12.14}$$

and

$$A = \begin{pmatrix} 0 & -a_1 & 0 & \cdots & \cdots & 0 & a_N \\ a_1 & 0 & -a_2 & 0 & \cdots & \vdots & 0 \\ 0 & a_2 & 0 & \ddots & \ddots & \vdots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 & -a_{N-2} & 0 \\ 0 & 0 & \cdots & 0 & a_{N-2} & 0 & -a_{N-1} \\ -a_N & 0 & \cdots & \cdots & 0 & a_{N-1} & 0 \end{pmatrix}. \tag{12.15}$$

However, it is not straightforward to check that all the eigenvalues are independent invariants.

12.4 From Toda to KdV[145]

(12.16) may be rewritten by space-time scaling as

$$\frac{d^2 r_n}{dt^2} = 2e^{-r_n} - e^{-r_{n+1}} - e^{-r_{n-1}}. \tag{12.16}$$

In terms of interaction forces $f_n = e^{-r_n} - 1$ this equation reads

$$\frac{d^2}{dt^2} \log(1 + f_n) = f_{n+1} + f_{n-1} - 2f_n. \tag{12.17}$$

Let us scale time and the function as ($h \in (0, 1]$)

$$t = \tau/h^3, \quad f_n = h^2 u_n(\tau). \tag{12.18}$$

(12.17) reads

$$\frac{d^2}{d\tau^2} \log(1 + h^2 u_n) = \frac{1}{h^4}(u_{n+1} + u_{n-1} - 2u_n). \tag{12.19}$$

Let us write $u_n(\tau) = u(hn, \tau)$. We introduce $x = nh$. Next, we observe this system from a moving coordinates:

$$x \to y = x - \left(\frac{1}{h^2} - h^2\right)\tau, \tag{12.20}$$

[145] Noriko Saito, “A transformation connecting the Toda lattice and the KdV equation,” J Phys Soc Japan **49**, 409 (1980).

Notice that

$$\frac{\partial}{\partial \tau}\bigg|_x u(x,\tau) = \left(\frac{\partial}{\partial \tau}\bigg|_y + \frac{\partial y}{\partial \tau}\frac{\partial}{\partial y}\right) u(y,\tau) = \left(\frac{\partial}{\partial \tau}\bigg|_y - \left(\frac{1}{h^2} - h^2\right)\frac{\partial}{\partial y}\right) u(y,\tau) \tag{12.21}$$

Therefore, the equation of motion reads[146]

$$\left(\frac{\partial}{\partial \tau} - \left(\frac{1}{h^2} - h^2\right)\frac{\partial}{\partial y}\right)^2 \log(1+h^2u(y,\tau)) = \frac{1}{h^4}(u(y+h,\tau)+u(y-h,\tau)-2u(y,\tau)). \tag{12.22}$$

We consider $h \to 0$ limit. The LHS reads

$$\left(\frac{\partial^2}{\partial \tau^2} - 2\left(\frac{1}{h^2} - h^2\right)\frac{\partial^2}{\partial \tau \partial y} + \left(\frac{1}{h^2} - h^2\right)^2\frac{\partial^2}{\partial y^2}\right)\left(h^2u(y,\tau) - \frac{1}{2}h^4u^2(y,\tau)\right) \tag{12.23}$$

$$\to -2\frac{\partial^2 u}{\partial \tau \partial y} + \frac{1}{h^2}\frac{\partial^2 u}{\partial y^2} - \frac{1}{2}\frac{\partial^2 u^2}{\partial y^2}. \tag{12.24}$$

The RHS reads

$$\frac{1}{h^4}(u(y+h,\tau)+u(y-h,\tau)-2u(y,\tau)) \to \frac{1}{h^2}\frac{\partial^2 u}{\partial y^2} + \frac{1}{12}\frac{\partial^4 u}{\partial y^4}. \tag{12.25}$$

Thus, we have arrived at

$$\frac{\partial}{\partial y}\left[-2\frac{\partial u}{\partial \tau} - \frac{1}{2}\frac{\partial u^2}{\partial y} - \frac{1}{12}\frac{\partial^3 u}{\partial y^3}\right] = 0. \tag{12.26}$$

[ ] must be a function of τ only. Let us assume that the 'wave' is localized in space (i.e., if we chase it, it looks like a localize disturbance). Thus, [ ] must be zero. We have arrived at

$$\frac{\partial u}{\partial \tau} + \frac{1}{2}u\frac{\partial u}{\partial y} + \frac{1}{24}\frac{\partial^3 u}{\partial y^3} = 0. \tag{12.27}$$

This is the Korteweg-de Vries (KdV) equation.

Actual experiment of a solitary wave:
https://www.youtube.com/watch?v=w-oDnvbV8mY&frags=wn

[146] Saito starts from this equation; for $h = 1$ this is just the Toda lattice.

12.5 Fermi-Pasta-Ulam problem

Everybody knows that for a harmonic lattice no thermal equilibrium is realized. Thus, to study the foundation of thermal physics Fermi was interested in the effect of nonlinearity. Since analytic study is out of question, he utilized the Los Alamos computer Maniac I. The model is a chain of point masses interacting with the neighbor points with the potential

$$\phi(r) = \frac{\kappa}{2}r^2 + \frac{\kappa\alpha}{3}r^3. \tag{12.28}$$

The equation of motion is

$$\begin{aligned} m\frac{d^2y_n}{dt^2} &= \phi'(y_n - y_{n-1}) - \phi'(y_{n+1} - y_n) && (12.29)\\ &= \kappa(2y_n - y_{n+1} - y_{n-1}) + \kappa\alpha((y_n - y_{n-1})^2 - (y_n - y_{n+1})^2). && (12.30)\end{aligned}$$

The numerical experiments exhibited almost complete recurrence of the intial condition.[147] Thus, it was clear that nonlinearity is not enough to thermalize the system.

FPU simulation (the quadratic case); you can observe recurrence:
https://www.youtube.com/watch?v=0WS548JX6D8

Zabusky and Kruskal studied its continuum limit. Assuming the lattice spacing is h, and we consider small h limit.

$$y_{n\pm1} = y_n \pm h\frac{\partial y_n}{\partial x} + \frac{h^2}{2}\frac{\partial^2 y_n}{\partial x^2} \pm \frac{h^3}{3!}\frac{\partial^3 y_n}{\partial x^3} + \frac{h^4}{4!}\frac{\partial^4 y_n}{\partial x^4} \tag{12.31}$$

We see

$$2y_n - y_{n+1} - y_{n-1} = h^2\frac{\partial^2 y_n}{\partial x^2} + \frac{h^4}{12}\frac{\partial^4 y_n}{\partial x^4}, \tag{12.32}$$

and

$$\begin{aligned} (y_n - y_{n-1})^2 - (y_n - y_{n+1})^2 &= (y_{n+1} - y_{n-1})(2y_n - y_{n+1} - y_{n-1}) && (12.33)\\ &= 2h\frac{\partial y_n}{\partial x}\left(h^2\frac{\partial^2 y_n}{\partial x^2} + \frac{h^4}{12}\frac{\partial^4 y_n}{\partial x^4}\right). && (12.34)\end{aligned}$$

Therefore, (12.30) becomes (with $\varepsilon = 2\alpha h$ and $c_0^2 = \kappa h/m$)

$$\frac{\partial^2 y}{\partial t^2} = c_0^2\left(1 + \varepsilon\frac{\partial y}{\partial x}\right)\frac{\partial^2 y}{\partial x^2} + \frac{c_0^2\alpha h^2}{12}\frac{\partial^4 y}{\partial x^4}. \tag{12.35}$$

[147] As to the experiment and the story surrounding it, see T. Dauxois, Fermi, Pasta, Ulam and a *mysterious* lady, arXiv:0801.1590.

Now, consider only the wave propagating to the right: new variables are $\xi = x - c_0 t$ and $\tau = t$. We have

$$\frac{\partial}{\partial x} = \frac{\partial \xi}{\partial x}\frac{\partial}{\partial \xi} + \frac{\partial \tau}{\partial x}\frac{\partial}{\partial \tau} = \frac{\partial}{\partial \xi},\ \frac{\partial}{\partial t} = \frac{\partial}{\partial \tau} - c_0 \frac{\partial}{\partial \xi} \tag{12.36}$$

This gives

$$\frac{\partial^2}{\partial t^2} - c_0^2 \frac{\partial^2}{\partial x^2} = \left(\frac{\partial}{\partial \tau} - c_0 \frac{\partial}{\partial \xi}\right)^2 - c_0^2 \frac{\partial^2}{\partial \xi^2} = \frac{\partial^2}{\partial \tau^2} - 2c_0 \frac{\partial}{\partial \tau}\frac{\partial}{\partial \xi}. \tag{12.37}$$

Thus (12.35) now reads

$$\frac{\partial^2 y}{\partial \tau^2} - 2c_0 \frac{\partial^2 y}{\partial \tau \partial \xi} = \varepsilon c_0^2 \frac{\partial y}{\partial \xi}\frac{\partial^2 y}{\partial \xi^2} + \frac{c_0^2 \alpha h^2}{12}\frac{\partial^4 y}{\partial \xi^4}. \tag{12.38}$$

We scale time as $s = \varepsilon \tau$. Then,

$$\varepsilon^2 \frac{\partial^2 y}{\partial s^2} - 2\varepsilon c_0 \frac{\partial^2 y}{\partial s \partial \xi} = \varepsilon c_0^2 \frac{\partial y}{\partial \xi}\frac{\partial^2 y}{\partial \xi^2} + \frac{c_0^2 \alpha h^2}{12}\frac{\partial^4 y}{\partial \xi^4}. \tag{12.39}$$

Ignoring the highest order in ε, and assuming h^2 and ε are comparable, we get

$$\frac{\partial^2 y}{\partial s \partial \xi} = -\frac{1}{2} c_0 \frac{\partial y}{\partial \xi}\frac{\partial^2 y}{\partial \xi^2} - \frac{c_0 \alpha h^2/\varepsilon}{12}\frac{\partial^4 y}{\partial \xi^4}. \tag{12.40}$$

Define $u = \partial y / \partial \xi$. We get

$$\frac{\partial u}{\partial s} = -\frac{1}{2} c_0 u \frac{\partial u}{\partial \xi} - \frac{c_0 \alpha h^2/\varepsilon}{12}\frac{\partial^3 u}{\partial \xi^3}. \tag{12.41}$$

Notice that this can be rewritten as

$$\frac{\partial u}{\partial s} + Bu \frac{\partial u}{\partial \xi} + \frac{\partial^3 u}{\partial \xi^3} = 0 \tag{12.42}$$

for any $B > 0$. First scale u so that the ratio of the coefficients of the second and the third terms is B, then scale time. Thus, the KdV equation was obtained. Its numerical solution gave solitons.

12.6 What is KdV?

The Korteweg-de Vries equation has the following form

$$\frac{\partial u}{\partial t} + 6u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0. \tag{12.43}$$

This describes the propagation of waves in a shallow channel observed by an observer following the wave. It was motivated by the observation (and subsequent experiments) by John Scott Russell in 1834 along the Union Canal (Scotland) [John Scott Russell, Report on Waves, Report of the 14th Meeting of the British Association for the Advancement of Science, (1844), pp.311-390]

> I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped—not so the mass of water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour, preserving its original figure some thirty feet long and a foot to a foot and a half in height. Its height gradually diminished, and after a chase of one or two miles I lost it in the windings of the channel. Such, in the month of August 1834, was my first chance interview with that singular and beautiful phenomenon which I have called the Wave of Translation.

KdV can be written in the Lax form with

$$L = -\frac{d^2}{dx^2} + u(x,t), \quad A = -4\frac{d^3}{dx^3} + 3u\frac{d}{dx} + 3\frac{d}{dx}u. \tag{12.44}$$

Thus, eigenvalues of the Schrödinger operator are invariant.

The problem to find the potential (the inverse-scattering problem) and solving KdV are closely related (this leads to a very large research field).

3 soliton solution

https://www.youtube.com/watch?v=mqwi8UHYwJA https://www.youtube.com/watch?v=VFM48pSLwGc

Sine wave breaking into solitons

https://www.youtube.com/watch?v=agteGpbhEaE

12.7 Related equations

The equation with a constant v

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0 \tag{12.45}$$

has a general solution $u(t, x) = f(x - vt)$ where f is an arbitrary differentiable function. If $v > 0$, this describes a translational motion to the right with speed v.

The following nonlinear equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0 \tag{12.46}$$

may be locally (in space-time) understood, if we look at (12.45), as illustrated in Fig. 12.2. The position with height u moves to the right with speed u:

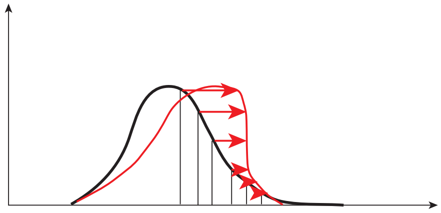


Figure 12.2: A short term evolution of u governed by (12.46).

As is clear, the solution u ceases to exist beyond some time as a differentiable function.

To prevent this singularity from being produced, there are two ways to modify (12.46). Both ways add a higher order spatial derivatives which adds dissipation (even orders) or dispersion (odd orders) to (12.46).

A famous example is Burgers' equation:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \tag{12.47}$$

which has a diffusion terms (viscosity effect) that kills sharp edges and prevent the emerging singularity. The KdV equation is the simplest dispersive case, where the speed of a plane wave depends on the wavelength (shorter wavelength waves travel relatively slower, so the 'wave front' is always dominated by longer wavelength waves; consequently the front is never sharp).

Incidentally, Burger's equation can be solved analytically with the aid of the Cole-Hopf transformation

$$u = -2\nu\frac{\partial}{\partial x}\log\phi = -2\nu\frac{1}{\phi}\frac{\partial \phi}{\partial x} \tag{12.48}$$

Then, we obtain a diffusion equation,

$$\frac{\partial \phi}{\partial t} = \nu \frac{\partial^2 \phi}{\partial x^2}, \tag{12.49}$$

which may be solved analytically.

This can be shown as follows: (12.47) may be rewritten as

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\nu \frac{\partial u}{\partial x} - \frac{1}{2}u^2\right). \tag{12.50}$$

Thus,

$$-2\nu \frac{\partial}{\partial x}\left(\frac{1}{\phi}\frac{\partial \phi}{\partial t}\right) = \frac{\partial}{\partial x}\left(-2\nu^2 \frac{\partial}{\partial x}\frac{1}{\phi}\frac{\partial \phi}{\partial x} - \frac{2\nu^2}{\phi^2}\left(\frac{\partial \phi}{\partial x}\right)^2\right). \tag{12.51}$$

That is

$$-2\nu \frac{\partial}{\partial x}\left(\frac{1}{\phi}\frac{\partial \phi}{\partial t}\right) = -2\nu^2 \frac{\partial}{\partial x}\left(-\frac{1}{\phi}\frac{\partial^2 \phi}{\partial x^2}\right). \tag{12.52}$$

This means

$$\frac{1}{\phi}\left(\frac{\partial \phi}{\partial t} - \nu \frac{\partial^2 \phi}{\partial x^2}\right) = f(t). \tag{12.53}$$

Assuming that ϕ is constant for large $|x|$ (u vanishes there), we get (12.49).

12.8 Quick derivation of KdV from scratch[148]

If you read the original papers, or recent systematic expositions, the derivation via systematic expansion with respect to the smallness of the vertical displacements of water looks quite tedious and lengthy. Therefore, here we will 'cheat,' following the logic outlined in a recent textbook of classical physics.[149]

The strategy is to use the linearized Euler equation to obtain the nontrivial dispersion relation, and then the nonlinear transport ignored in the original Euler equation is recovered with the aid of the conservation of water.

[148] I am pretty sure that the KdV may be obtained as an RG equation just as almost all other 'named' equations. Publishable.

[149] The following is the logic in K. S. Thorne and R. D. Blandford, *Modern classical physics* (Princeton UP 2017) p852 made palatable.

12.9 Dispersion relation of shallow water waves

Consider the disturbance of surface height of a horizontal channel of depth h_0. Our starting point is always the Navier-Stokes equation

$$\rho\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v}\cdot\nabla\boldsymbol{v} = -\eta\Delta\boldsymbol{v} - \nabla P + \boldsymbol{F}, \tag{12.54}$$

with the incompressibility $\nabla\cdot\boldsymbol{v} = 0$, where ρ is the density, P the pressure and $\boldsymbol{F}$ the external force (we assume here gravity only, so it is a potential force). As long as the fluid speed is slow, we may drop the nonlinear term, and also we may ignore the viscosity term (i.e., we use the linearized Euler equation). In this case curl$\boldsymbol{v}$ is time independent, so if there is no vorticity initially, the flow is a potential flow, so we may introduce the velocity potential ϕ:

$$\boldsymbol{v} = \nabla\phi. \tag{12.55}$$

The incompressibility means ϕ must be a harmonic function,

$$\Delta\phi = 0 \tag{12.56}$$

and the linearized Euler equation reads

$$\nabla\left(\frac{\partial\phi}{\partial t} + P/\rho + gz\right) = 0. \tag{12.57}$$

This means

$$\frac{\partial\phi}{\partial t} + P/\rho + gz = f(t). \tag{12.58}$$

If we are interested in a local disturbance, $f(t) = 0$.[150]

Let $\zeta(x,t)$ be the surface shape relative to the free surface at $z = 0$, where the pressure is constant P_0:

$$P_0 = -\rho\zeta g - \rho\frac{\partial\phi}{\partial t}. \tag{12.59}$$

This gives a boundary condition:

$$\frac{\partial\zeta}{\partial t}g + \frac{\partial^2\phi}{\partial t^2} = 0. \tag{12.60}$$

Since $\partial\zeta/\partial t = v_z \simeq \partial\phi/\partial z$; this is admissible since ζ is small. We have an equation for ϕ

$$\left(\frac{\partial\phi}{\partial z} + \frac{1}{g}\frac{\partial^2\phi}{\partial t^2}\right)_{z=\zeta} = 0. \tag{12.61}$$

[150] This is what Landau did. We could instead redefine ϕ to absorb f without spoiling $\boldsymbol{v} = \nabla\phi$.

However, ζ is small, so we may drop it. Thus, we have the following set of equation at the liquid surface:

$$\Delta\phi = 0, \tag{12.62}$$

$$\left(\frac{\partial\phi}{\partial z} + \frac{1}{g}\frac{\partial^2\phi}{\partial t^2}\right)_{z=0} = 0. \tag{12.63}$$

At the bottom $z = -h_0$ of the channel we must impose that condition that $v_z = 0$.

Let us solve the above boundary value problem for ϕ that is periodic:

$$\phi = f(z)\cos(kx - \omega t). \tag{12.64}$$

$\Delta\phi = 0$ reads

$$f''(z) - k^2 f = 0, \tag{12.65}$$

whose general solution is $f(z) = Ae^{kz} + Be^{-kz}$. Its derivative must vanish at the bottom $z = -h_0$:

$$kAe^{-kh_0} - kBe^{kh_0} = 0. \tag{12.66}$$

Therefore, ϕ must have the following form:

$$\phi = A\cosh k(z + h_0)\cos(kx - \omega t). \tag{12.67}$$

This must satisfy the boundary condition (12.63)

$$gk\sinh(kh_0)\cos(\omega t) - \omega^2\cosh(kh_0)\cos(\omega t) = 0. \tag{12.68}$$

Thus, we have obtained the dispersion relation:

$$\omega^2 = gk\tanh kh_0. \tag{12.69}$$

Let us assume that kh_0 is not large (shallow water wave[151]).

$$\omega^2 = gk\left(kh_0 - \frac{1}{3}(kh_0)^3\right) + \cdots = gh_0k^2\left(1 - \frac{1}{3}h_0^2k^2\right) + \cdots, \tag{12.70}$$

or

$$\omega = \sqrt{gh_0}k\left(1 - \frac{1}{6}h_0^2k^2\right). \tag{12.71}$$

[151] In the opposite limit we can compute the group velocity as $(1/2)\sqrt{g/k}$: long wavelength wave propagates fast.

This means that a propagating wave ζ obeys (cf $\omega \to i\partial_t$, $k \to -i\partial_x$)

$$\frac{\partial \zeta}{\partial t} + \sqrt{gh_0}\frac{\partial \zeta}{\partial x} + \frac{1}{6}h_0^2\frac{\partial^3 \zeta}{\partial x^3} = 0. \tag{12.72}$$

We cannot accurately reconstruct the nonlinear term (the original form of the second term on the LHS) from this.

12.10 Nonlinear transport term

To study the nonlinear term we restart from Euler's equation$(/\rho)$ and also the mass conservation equation:

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + g\frac{\partial h}{\partial t} = 0, \tag{12.73}$$

$$\frac{\partial h}{\partial t} + \frac{\partial vh}{\partial x} = 0, \tag{12.74}$$

where $h = h_0 + \zeta$. That is, now we choose the bottom to be the origin of the height. Let us define $gh = \eta$. (12.74) read

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \frac{\partial \eta}{\partial t} = 0. \tag{12.75}$$

$$\frac{\partial \eta}{\partial t} + v\frac{\partial \eta}{\partial x} + \eta\frac{\partial v}{\partial x} = 0, \tag{12.76}$$

If we wish to combine these two equations into one,[152] v must be linearly combined with $\sqrt{\eta}$ (dimensional homogeneity requirement). Let us introduce $Z = \sqrt{\eta}$.

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + 2Z\frac{\partial Z}{\partial t} = 0. \tag{12.77}$$

$$2\frac{\partial Z}{\partial t} + 2v\frac{\partial Z}{\partial x} + Z\frac{\partial v}{\partial x} = 0, \tag{12.78}$$

Therefore,

$$\frac{\partial (v-2Z)}{\partial t} + v\frac{\partial (v-2Z)}{\partial x} - Z\frac{\partial (v-2Z)}{\partial t} = 0 \tag{12.79}$$

or

$$\frac{\partial (v-2Z)}{\partial t} + (v-Z)\frac{\partial (v-2Z)}{\partial x} = 0 \tag{12.80}$$

[152] The motivation for this is to find a conserved quantity that may be followed from a moving coordinate.

This implies that $v - 2\sqrt{gh}$ is constant, if we observe it from the moving coordinate with speed $v - \sqrt{gh}$.

Suppose the wave propagation starts from a quiet surface with height (depth) h_0 with $v = 0$. Initially, $v - 2\sqrt{gh} = -2\sqrt{gh_0}$. Since this must be constant, $v = 2\sqrt{gh} - 2\sqrt{gh_0}$. Putting this into (12.74), we obtain

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}\sqrt{g}(h^{3/2} - hh_0^{1/2}) = \frac{\partial h}{\partial t} + (3\sqrt{gh} - 2\sqrt{gh_0})\frac{\partial h}{\partial x} = 0. \tag{12.81}$$

Since

$$3\sqrt{g(h_0+\zeta)} - 2\sqrt{gh_0} = \sqrt{gh_0} + \frac{3}{2}\sqrt{\frac{g}{h_0}}\zeta, \tag{12.82}$$

the equation for ζ reads

$$\frac{\partial \zeta}{\partial t} + \left(\sqrt{gh_0} + \frac{3}{2}\sqrt{\frac{g}{h_0}}\zeta\right)\frac{\partial \zeta}{\partial x} = 0 \tag{12.83}$$

Now, observe this from a moving coordinate with speed $\sqrt{gh_0}$. We arrive at

$$\frac{\partial \zeta}{\partial t} + \frac{3}{2}\sqrt{\frac{g}{h_0}}\zeta\frac{\partial \zeta}{\partial x} = 0. \tag{12.84}$$

Augmenting this with the higher order dispersion relation (or combining (12.85) with (12.72)) we finally obtain

$$\frac{\partial \zeta}{\partial t} + \frac{3}{2}\sqrt{\frac{g}{h_0}}\zeta\frac{\partial \zeta}{\partial x} + \frac{1}{6}h_0^2\frac{\partial^3 \zeta}{\partial x^3} = 0, \tag{12.85}$$

the Korteweg-de Vries equation.

13 Lecture 13. Canonical transformation

We have already used canonical transformations, but here important facts are collected (point transformation, infinitesimal canonical transformation, Lagrange's bracket, various invariants, Liouville's theorem).

It is important to recognize that time evolution is a canonical transformation.

Classical canonical transformations do not usually correspond to quantum unitary transformations.

13.1 Canonical transformation with generators (review)

The transformation $T : (q,p) \to (Q,P)$ that preserves the form of the canonical equation of motion (10.14) is called a *canonical transformation*. As discussed in **11.3**, we discuss only the canonical transformations with generators F

$$dF = \sum_i p_i dq_i - \sum_i P_i dQ_i + (K-H)dt. \tag{13.1}$$

(13.1) gives

$$\frac{\partial F}{\partial q} = p,\ \frac{\partial F}{\partial Q} = -P,\ \frac{\partial F}{\partial t} = K-H. \tag{13.2}$$

Solving these equations, we can construct the canonical transformation T. In particular, if F is time-independent, then $K = H$ is obtained by replacing q and p in H in terms of Q and P.

Applying a sort of Legendre transformation to generators, we can construct different (perhaps more convenient) transformations:

$$d(F+PQ) = pdq + QdP + (K-H)dt. \tag{13.3}$$

Thus, replacing F with $G = F + \sum_i P_i Q_i$, we obtain

$$dG = \sum_i p_i dq_i + \sum_i Q_i dP_i + (K-H)dt. \tag{13.4}$$

13.2 Canonical transformations make a group

Let us write the generator of the canonical transformation T_i as F_i. The generator of $\prod_i T_i$ is given by $\sum_i F_i$, so the totality of the canonical transformation for a given system makes a(n abelian) group (called the canonical transformation group).

13.3 Point transformations

Through a general canonical transformation position and momentum coordinates are usually mixed up. When we use the spatial coordinate change (say, from the Cartesian to the spherical), no mix-up occurs. Thus, $q \to Q$, $p \to P$ can be described by a subset (subgroup) of canonical transformations called point transformations.

Perhaps, the most convenient generator is the original $G = F + PQ$:

$$dG = p_i dq_i + P_i dQ_i, \tag{13.5}$$

because usually we know q and $Q(q,t)$. Since dG is exact, to obtain G we may use a convenient path: q is constant. Therefore,

$$G = p_i q_i. \tag{13.6}$$

If we write q in terms of Q,

$$P = \frac{\partial G}{\partial Q} \tag{13.7}$$

gives the canonical momentum in the new coordinate system in terms of the old momentum variables. This may be the most mechanical way to get the momentum in the new coordinate system.

Examples:
(1) CM coordinates: $(q_1, q_2) \to Q_1 = (q_1+q_2)/2, Q_2 = q_1 - q_2$. Since $q_1 = Q_1 + Q_2/2$, $q_2 = Q_1 - Q_2/2$, so (13.6) reads

$$G = p_1(Q_1 + Q_2/2) + p_2(Q_1 - Q_2/2). \tag{13.8}$$

Therefore,

$$P_1 = \frac{\partial G}{\partial Q_1} = p_1 + p_2,\ \ P_2 = \frac{\partial G}{\partial Q_2} = \frac{1}{2}(p_1 - p_2). \tag{13.9}$$

(2) $(x,y,z) \to (r,\theta,\varphi)$: $x = r\sin\theta\cos\varphi$, $y = r\sin\theta\sin\varphi$, $z = r\cos\theta$. The generator reads

$$G = p_x r\sin\theta\cos\varphi + p_y r\sin\theta\sin\varphi + p_z r\cos\theta. \tag{13.10}$$

Its inverse may be generated by

$$G = p_r\|x\| + p_\theta \mathrm{Tan}^{-1}\frac{\sqrt{x^2+y^2}}{z} + p_\varphi \mathrm{Tan}^{-1}\frac{y}{x}. \tag{13.11}$$

(3) Rotation. We can use an orthogonal transformation to describe it as $x'^i = a^i_j x^j$ (with the summation convention). The generator is $G = a^i_j p_i x^j = p_i x'^i$.
(4) Translation: $x'^k = x^k + \Delta x^k$. Its generator is $G = P(x + \Delta x)$. This gives $P = p$, of course.
(5) Mirror image wrt x^1: $x'^1 = -x^1$, $x'^2 = x^2$, $\cdots$. Its generator is $G = -x_1 P_1 + x^2 P_2 + \cdots$..

13.4 Infinitesimal canonical transformation

A canonical transformation $(q, p) \to (Q, P)$ is said to be an infinitesimal canonical transformation, if $Q - q$ and $P - p$ are infinitesimal.

Let G be its generator:

$$dG = pdq + QdP \tag{13.12}$$

Notice that

$$d(G - qP) = (p - P)dq + (Q - q)dP, \tag{13.13}$$

so $G - qP$ may be written as εS, where ε is an infinitesimal parameter. Therefore, any infinitesimal canonical transformation has the following generator

$$G(q, P) = qP + \varepsilon S(q, p), \tag{13.14}$$

where P in S has been replaced by p because of ε in front of it. S is called the generator of the infinitesimal canonical transformation. We have

$$p = \frac{\partial G}{\partial q} = P + \varepsilon \frac{\partial S}{\partial q}, \tag{13.15}$$

$$Q = \frac{\partial G}{\partial P} = q + \varepsilon \frac{\partial S}{\partial p}. \tag{13.16}$$

That is,

$$Q = q + \varepsilon \frac{\partial S}{\partial p}, \; P = p - \varepsilon \frac{\partial S}{\partial q}. \tag{13.17}$$

13.5 Infinitesimal canonical transformation of mechanical variables

A mechanical variable $F = F(q, p)$ changes, according to the infinitesimal canonical transformation with generator S, as

$$F(Q, P) - F(q, p) = \varepsilon \left(\frac{\partial F}{\partial q} \frac{\partial S}{\partial p} - \frac{\partial F}{\partial p} \frac{\partial S}{\partial q} \right) = \varepsilon [F, S]_{PB}. \tag{13.18}$$

13.6 Time evolution is canonical transformation

We know the time evolution of a mechanical variable according to the natural motion of the system obeys

$$\frac{dF}{dt} = [F, H]_{PB}. \tag{13.19}$$

$\varepsilon \to \delta t$ and $S \to H$ in (13.18) just gives this equation. Thus we may conclude that time evolution is a canonical transformation.

13.7 One form of Noether's theorem

Suppose the Hamiltonian is invariant under the infinitesimal canonical transformation with the generator W that changes a mechanical variable $f \to f + df$. Then (13.18) implies

$$dH = df[H, W] = 0. \tag{13.20}$$

That is, W is an invariant of motion.

13.8 What is the relation between classical canonical and quantum unitary transformations?

As point transformations tell us, any coordinate transformation is canonical classically. However, it is well known that the canonical quantization using the correspondence between the commutator and the Poisson bracket must choose (basically) the Cartesian coordinates; if the formalism is naively applied to the system described with the aid of the spherical coordinates, the resultant quantum mechanics are not at all equivalent. Thus, it is clear that the usual canonical quantization and canonical transformations are usually not compatible.

If a transformation gives an equivalent quantum mechanical system, and if there is a corresponding classical system, there is a corresponding canonical transformation. However, the converse is not usually true.

13.9 Lagrange bracket

Let f and g be differentiable mechanical quantities. The Lagrange bracket between them is defined as

$$(f,g) = \frac{\partial q}{\partial f}\frac{\partial p}{\partial g} - \frac{\partial q}{\partial g}\frac{\partial p}{\partial f}. \tag{13.21}$$

13.10 Poisson vs Lagrange brackets

Let $2n$ canonical variables be denoted as $\{u_j\}_{j=1}^{2n}$. Then

$$\sum_k [u_k, u_i]_{PB}(u_k, u_j) = \delta_{ij}. \tag{13.22}$$

This is shown by explicit calculation:[153]

$$\sum_k \left(\frac{\partial u_k}{\partial q}\frac{\partial u_i}{\partial p} - \frac{\partial u_k}{\partial p}\frac{\partial u_i}{\partial q}\right)\left(\frac{\partial q}{\partial u_k}\frac{\partial p}{\partial u_j} - \frac{\partial p}{\partial u_k}\frac{\partial q}{\partial u_j}\right) = \frac{\partial u_i}{\partial p}\frac{\partial p}{\partial u_j} + \frac{\partial u_i}{\partial q}\frac{\partial q}{\partial u_j} = \frac{\partial u_i}{\partial u_j}. \tag{13.23}$$

13.11 Integral invariant

Take a smooth closed curve C in the phase space (q,p). If a canonical transformation $(q,p) \to (Q,P)$ is applied to this curve, we get a closed curve C' in the phase space

[153] If you do not like the following shorthand, do explicit calculation with summation convention. For example, one cross term reads

$$\frac{\partial u_k}{\partial p_f}\frac{\partial u_i}{\partial q_f}\frac{\partial q_g}{\partial u_k}\frac{\partial p_g}{\partial u_j} = \frac{\partial q_q}{\partial p_f}\frac{\partial u_i}{\partial q_f}\frac{\partial p_g}{\partial u_j} = 0$$

The surviving terms read

$$\frac{\partial u_k}{\partial q_f}\frac{\partial u_i}{\partial p_f}\frac{\partial q_g}{\partial u_k}\frac{\partial p_g}{\partial u_j} + \frac{\partial u_k}{\partial p_f}\frac{\partial u_i}{\partial q_f}\frac{\partial p_g}{\partial u_k}\frac{\partial q_g}{\partial u_j} = \frac{\partial q_g}{\partial q_f}\frac{\partial u_i}{\partial p_f}\frac{\partial p_g}{\partial u_j} + \frac{\partial p_g}{\partial p_f}\frac{\partial u_i}{\partial q_f}\frac{\partial q_g}{\partial u_j} = \delta_{fg}\left(\frac{\partial u_i}{\partial p_f}\frac{\partial p_g}{\partial u_j} + \frac{\partial u_i}{\partial q_f}\frac{\partial q_g}{\partial u_j}\right)$$

That is,

$$\sum_f \left(\frac{\partial u_i}{\partial p_f}\frac{\partial p_f}{\partial u_j} + \frac{\partial u_i}{\partial q_f}\frac{\partial q_f}{\partial u_j}\right) = \frac{\partial u_i}{\partial u_j}.$$

which is the sum over all the independent variables (recall u are understood as a function of (q,p); just as $\frac{\partial F(G_i(x))}{\partial x} = \frac{\partial F}{\partial G_i}\frac{\partial G_i}{\partial x}$)

spanned by (Q, P). Since $pdQ - PdQ = dF$,[154]

$$\oint_C pdq - \oint_{C'} PdQ = 0. \tag{13.24}$$

This is called a relative integral invariant. Applying the Stokes' theorem, we get

$$\int_A dqdp = \int_A' dQdP \tag{13.25}$$

where $\partial A = C$ and $\partial A' = C'$. This is called the absolute integral invariant. This is true for any area A floating in the phase space.

Suppose A has a coordinate system (u, v). Then, on A q and p are functions of (u, v). After the canonical transformation Q and P depend on (u, v) through the original variable, so (13.26) reads (with the summation convention)

$$\int_A \frac{\partial(q^r, p^r)}{\partial(u, v)} dudv = \int_A' \frac{\partial(Q^r, P^r)}{\partial(u, v)} dudv \tag{13.26}$$

This is true for any A, so we must conclude that

$$\frac{\partial(q^r, p^r)}{\partial(u, v)} \tag{13.27}$$

is invariant (do not forget the summation convention).

13.12 Lagrange brackets are invariant of canonical transformation

The invariance of (13.27) implies

$$\frac{\partial(q^r, p^r)}{\partial(u, v)} dudv = dq^r d'p^r - d'q^r dp^r \tag{13.28}$$

is invariant. Here, d and d' are independent infinitesimal changes. Therefore, Lagrange brackets are invariant.

As we will see soon the converse is true: if a transformation keeps Lagrange brackets invariant, it is a canonical transformation.

[154] pdQ is a shorthand notation of $\sum_r p_r dQ_r$ as usual.

13.13 Poisson brackets are invariant of canonical transformation

(13.22) implies that if Lagrange bracket is invariant, then so is Poison bracket. Thus, a necessary and sufficient condition for $(q,p) \to (Q,P)$ is canonical is that

$$[P,P]_{PB} = [Q,Q]_{PB} = 0, [Q,P]_{PB} = 1. \tag{13.29}$$

13.14 Invariance of Lagrange bracket implies canonical nature of the transformation

Looking at **13.11**, we see we have only to show that $pdq - PdQ$ is exact. That is, there is a function W such that

$$dW = pdq - PdQ. \tag{13.30}$$

If there is such W, it must satisfy, as a function of P and Q

$$\frac{\partial W}{\partial P} = p\frac{\partial q}{\partial P}, \ \frac{\partial W}{\partial Q} = p\frac{\partial q}{\partial P} - P. \tag{13.31}$$

Also we need $\partial^2 W/\partial P\partial Q = \partial^2 W/\partial Q\partial P$: we have

$$\frac{\partial^2 W}{\partial Q\partial P} = \frac{\partial p}{\partial Q}\frac{\partial q}{\partial P} + p\frac{\partial^2 q}{\partial Q\partial P}, \tag{13.32}$$

$$\frac{\partial^2 W}{\partial P\partial Q} = \frac{\partial p}{\partial P}\frac{\partial q}{\partial Q} + p\frac{\partial^2 q}{\partial P\partial Q} - 1. \tag{13.33}$$

Equating these formulas, we get

$$(Q,P) = 1. \tag{13.34}$$

That is,[155] if we have the invariance of Lagrange's brackets, then $(q,p) \to (Q,P)$ is canonical.

13.15 Liouville's theorem

The phase volume is invariant under canonical transformations. That is

$$\int_V d^n p d^n q = \int_{V'} d^n P d^n Q, \tag{13.35}$$

[155] Needless to say, we must also demand $(P,P) = (Q,Q) = 0$ as well for dW to be exact. Check this.

where V is a ($2n$-)subset of the phase space, and V' its image due to the canonical transformation $(q,p) \to (Q,P)$. To demonstrate this we have only to show that following Jacobian to be ± 1:

$$J = \frac{\partial(q,p)}{\partial(Q,P)}. \tag{13.36}$$

We compute this as

$$J = \frac{\partial(q,p)}{\partial(Q,p)}\frac{\partial(Q,p)}{\partial(Q,P)} = \left.\frac{\partial q}{\partial Q}\right|_p \left.\frac{\partial p}{\partial P}\right|_Q = \left.\frac{\partial p}{\partial P}\right|_Q \Big/ \left.\frac{\partial Q}{\partial q}\right|_p. \tag{13.37}$$

Now, in terms of the generator $dG = pdq - QdP$, we have

$$\left.\frac{\partial Q}{\partial q}\right|_p = \frac{\partial^2 G}{\partial q \partial P}, \tag{13.38}$$

$$\left.\frac{\partial p}{\partial P}\right|_Q = \frac{\partial^2 G}{\partial P \partial q}. \tag{13.39}$$

Thus, $J = 1$.

13.6 implies that the phase volume is invariant of motion.

13.16 Integration by quadratures

Integration by quadrature is the search of the solutions by a finite number of algebraic operations (including inversion of functions) and 'quadratures' = calculation of integrals of known functions. T

If n first integrals are commutative (the classical involution case), the system is solvable by quadrature.

13.17 Jacobi's method of complete integral[156]

Jacobi showed that for a system with n degrees of freedom of motion if we can find a complete solution (solution with $n+1$ arbitrary and independent constants) for the Hamilton-Jacobi equation, we can determine the motion completely.

The Hamilton-Jacobi equation has the following form

$$\frac{\partial A}{\partial t} + H = 0, \tag{13.40}$$

[156] See Landau-Lifshitz Section 47.

so a complete solution can have the form $A = C + f(q, \alpha, t)$, where C and α (n-vector) are the arbitrary constants. We can introduce a canonical transformation whose generator is f that makes α as the new momentum (f corresponds to G)

$$p_i = \frac{\partial f}{\partial q_i}, \beta_j = \frac{\partial f}{\partial \alpha_j}, \ K = \frac{\partial f}{\partial t} + H = \frac{\partial A}{\partial t} + H = 0. \tag{13.41}$$

Thus, the new coordinates are β. Since the Hamiltonian in this coordinate system vanishes, the canonical equation of motion tells us that β are constants. Solving n equations $\beta = \frac{\partial f}{\partial \alpha}$, we can determine q as a function of t.

13.18 Separation of variables

If the action A (or called Hamilton's principal function) may be written as a sum of two functions without common coordinates: $A = A_1(a_1, t) + A_2(q_2, t)$, we can separately find complete solutions. Especially when A_2 depends only on one coordinate q_n, we say this coordinate is separated from the rest, and the 1D problem may be solved.

14 Lecture 14. Celestial mechanics

14.1 Two-body summary

Everybody should be very familiar with the two celestial body dynamics. Here I mention some topics that may not appear in elementary expositions.

⟨⟨**Bertrand's theorem**⟩⟩ Suppose U is a spherically symmetric potential. For a motion with non-zero angular momentum close to a circle to have a closed orbit, U must be harmonic or gravitational.[157]

⟨⟨**Collision and extension**⟩⟩ If the angular momentum is zero, the particles can collide. The orbit may be, however, uniquely extended beyond collision.[158]

14.2 Necessary condition for stability

The general n-body problem consists of n point masses (m_1, r_1), $\cdots$, (m_n, r_n) attracting one another according to the law of gravity. The total kinetic energy is

$$K = \frac{1}{2}\sum_i m_i \dot{r}_i^2, \tag{14.1}$$

and the potential energy U is

$$U = -\sum_{i<j} \frac{m_i m_j}{|r_i - r_j|}. \tag{14.2}$$

We describe the system from the inertial frame and the origin of the position coordinates is the center of mass.

We say the system is stable, if
(a) No collision: $|r_i - r_j| > 0$ for all i and j for all t.
(b) No escape: there is $c > 0$ such that $|r_i| < c$ for all i and t.

The fate of a three body system is very hard to predict (Fig. 14.1)

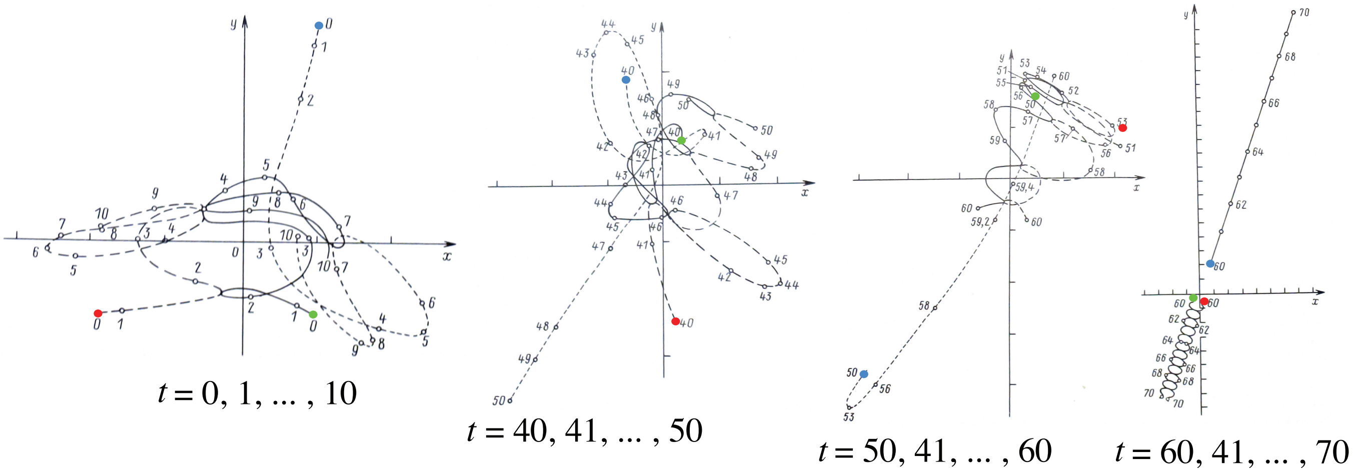
14.3 Jacobi's necessary condition for non-escape[159]

If there is no escape nor collisions, the total energy of the system must be negative.

[157] S. A. Chin, A truly elementary proof of Bertrand's theorem, arXiv:1411.7057 [physics.class-ph] (2014).

[158] Arnold III p56

[159] Arnold III p59



$t = 0, 1, \ldots, 10$ $t = 40, 41, \ldots, 50$ $t = 50, 41, \ldots, 60$ $t = 60, 41, \ldots, 70$

Figure 14.1: Fate of a three body system [Fig. 12-15 of Arnold III]

Remark For $n > 2$, this is not sufficient as seen in Fig. 14.1.
[Demo]
We use Lagrange's formula:*[160] for $I = \sum m_i r_i^2$ (the moment of inertia) and the total energy E

$$\ddot{I} = 4E - 2U. \tag{14.3}$$

Since $U < 0$, $E \geq 0$ implies $\ddot{I} > 0$, so $I(t)$ must be convex. Therefore, it cannot be

[160]

$$\ddot{I} = \sum_i m_i(2\dot{r}_i^2 + 2r_i\ddot{r}_i) = 4T - 2\sum_i r_i \sum_{j \neq i} \frac{m_i m_j}{|r_i - r_j|^3}(r_i - r_j)$$

(Note that $\nabla(1/|x|) = -(1/|x|^2)(x/|x|)$, because $\nabla|x|^2 = 2|x|\nabla|x| = 2x$). Therefore,

$$\begin{aligned} \ddot{I} &= 4T - \sum_i r_i \sum_j \frac{m_i m_j}{|r_i - r_j|^3}(r_i - r_j) - \sum_j r_j \sum_i \frac{m_i m_j}{|r_i - r_j|^3}(r_j - r_i) \\ &= 4T - 2\sum_{i,j} \frac{m_i m_j}{|r_i - r_j|^3}(r_i - r_j)^2 = 4T + 2U = 4E - 2U. \end{aligned}$$

bounded from above. Note that*[161]

$$I\sum_i m_i = \sum_{i<j} m_i m_j |r_i - r_j|^2 + \left(\sum_i m_i r_i\right)^2. \tag{14.4}$$

The center of mass is fixed (say, 0), so the unbounded I means the increase of the mutual distance without bound.

If there is no collision and U is bounded from below, the virial theorem[162] may be applied to the time average of U: $\langle U\rangle = 2E$. Therefore, $E < 0$.

14.4 Collisions[163]

If there is a simultaneous n-body collision I vanishes. If $I(t) \to 0$ as $t \to t_0$, the the total angular momentum must be zero.

For binary collisions the motion can be smoothly extended beyond collisions. This means that for a three body problem with nonzero angular momentum, the motion is well defined for all t.[164]

14.5 Three-body problem and its reduction

For the $n = 3$ case (the three-body problem case) the original equation of motion is $3 \times 3 \times 2 = 18$ first order equations. The problem can be reduced to a problem of 6 first order differential equations (Lagrange 1772).
(i) The location and the velocity of the center of mass are cyclic coordinates (= the coordinate that do not explicitly appear in the equations), so we may regard them

[161]

$$I\sum_i m_i = \sum_{i,j} m_i m_j r_i^2 = \frac{1}{2}\sum_{i,j} m_i m_j (r_i^2 + r_j^2) = \frac{1}{2}\sum_{i,j} m_i m_j [(r_i - r_j)^2 + 2r_i r_j].$$

[162] ⟪**Virial theorem**⟫ Notice that the long-time average of the derivative of a bounded function f(t) vanishes: $(1/T)\int_0^T f'(s)ds \to 0$. Consider pq. If the phase space for the system is bounded, then the time average of $d(pq)/dt$ vanishes. Therefore, for U which is a homogenous function of degree -1

$$p\dot{q} + q\dot{p} = 2T - q\frac{\partial U}{\partial q} = 2T + U = 2E - U.$$

Therefore, $\langle U\rangle = 2E$.

[163] Arnold III p59

[164] details can be found on p60-61 of Arnold III.

to be known. Thus 12 equations remain.
(ii) The total angular momentum is conserved. Thus 9 equations remain.
(iii) We perform the elimination of the nodes. The rotation as a whole may be regarded as known; actually this is done with (ii). Thus, 8 equations remain.
(iv) The total energy is conserved: 7 equations now.
(v) Now, eliminate t from the equation. The resultant set contains 6 first order equations.

The actual procedures and formulas are explained in detail on p343-347 of Whittaker.

14.6 Restricted problem of three body

If the third body (planetoid P) has an infinitesimal mass moving in the plane of the motion of the other two bodies S (Sun) and J (Jupiter) under their influence, the three-body problem is called the restricted problem of three bodies. The Hamiltonian governing the motion of P is given by

$$H = \frac{1}{2}(U^2 + V^2) - \frac{m_1}{\mathrm{SP}} - \frac{m_2}{\mathrm{JP}}. \tag{14.5}$$

SP ad JP are time-dependent, so H is not a constant of motion.

We introduce a moving coordinate system Fig. 14.2 whose origin is the CM of S and J and whose x-axis is from O to J. O-x is chosen to span the plane on which S and J always sit. Let n be the angular speed of SJ.

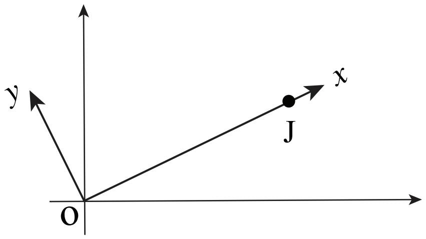


Figure 14.2: Moving coordinates for the restricted problem

Then, the original position coordinates (X, Y) of P may be related to (x, y) as

$$X = x\cos nt - y\sin nt \tag{14.6}$$
$$Y = x\sin nt + y\cos nt. \tag{14.7}$$

Now, we wish to canonical transform the coordinate system from the inertial (X, Y) (the old (q, p)) to the moving (x, y) (the new (Q, P)). We use $W = pq - F$, where F

is the 'most' standard one introduced in **13.1**

$$dW = PdQ - Kdt + qdp + Hdt = udx + vdy + XdU + YdV + (H - K)dt. \quad (14.8)$$

Since dW is exact, we can integrate this as $W = pq$ (i.e., in terms of the old variables). Expressing q in terms of the new coordinates, we get

$$W = U(x \cos nt - y \sin nt) + V(x \sin nt + y \cos nt). \quad (14.9)$$

Note that

$$X = \frac{\partial W}{\partial U},\ Y = \frac{\partial W}{\partial V},\ u = \frac{\partial W}{\partial x},\ v = \frac{\partial W}{\partial y}. \quad (14.10)$$

Since our new coordinates are time-dependent, the new Hamiltonian ahas an extra term:*[165]

$$\frac{\partial W}{\partial t} = H - K,\ \text{or}\ K = H - \frac{\partial W}{\partial t} = \frac{1}{2}(u^2 + v^2) + n(uy - vx) - F, \quad (14.11)$$

where

$$F = \frac{m_1}{\mathrm{SP}} + \frac{m_2}{\mathrm{JP}} \quad (14.12)$$

is a function of x and y only, so note that K is time independent. K = const is an integral and called the Jacobian integral. If the sum of the masses of S and J to be chosen unity, we may rewrite F as

$$F = \frac{1-\mu}{\mathrm{SP}} + \frac{\mu}{\mathrm{JP}}. \quad (14.13)$$

If $\mu = 0$, we ignore the effect of J. Thus, μ could be regarded as a perturbation parameter.

A nice restricted three-body simulation video is:
https://www.youtube.com/watch?v=jarcgP1rRWs

[165]

$$\frac{\partial W}{\partial t} = U(-nx \sin nt - ny \cos nt) + V(nx \cos nt - ny \sin nt).$$

Also we have

$$u = \frac{\partial W}{\partial x} = U \cos nt + V \sin nt,\ v = \frac{\partial W}{\partial y} = -U \sin nt + V \cos nt.\ (*)$$

Therefore,

$$\frac{\partial W}{\partial t} = x(-nU \sin nt + nV \cos nt) + y(-nU \cos nt - nV \sin nt) = n(xv - yu).$$

Obviously, $U^2 + V^2 = u^2 + v^2$, since $(*)$ is an orthogonal transformation.

14.7 Bruns' theorem

Theorem [Bruns 1887] The classical integrals (energy, momentum and angular momentum) are the only independent algebraic integrals of the problem of three bodies.

A proof can be found on p359-377 of Whittaker. The proof fully utilizes the peculiarity of the three-body problem.

14.8 Poincare's theorem 'denying' integrability of perturbed systems

Suppose we have a completely integrable system, whose Hamiltonian is given in terms of action variables only as $H_0(I)$ (with non-resonance condition satisfied, i.e., the Hessian of $H_0(I)$ is non-singular). The perturbation term $H_1(\theta, I)$ is analytic in I and the angle variable θ. The total Hamiltonian reads

$$H(\theta, I, \mu) = H_0(I) + \mu H_1(\theta, I). \tag{14.14}$$

Then, under a natural condition, there is no first integral of motion other than H itself that are analytic in μ.

Poincare proved in 1889 this theorem for the system with two degrees of freedom (as in the restricted three-body problem).

Suppose there is a first integral of motion Φ

$$\Phi = \Phi_0 + \mu\Phi_1 + \mu^2\Phi_2 + \cdots. \tag{14.15}$$

The key observation is that Φ_0 must be a function of H_0, which follows from the observation that Φ_0 is a function of I only. If Φ depends on μ smoothly, perhaps it is not a wild guess that this structure must be preserved for Φ (i.e., Φ must be a function of H only). A proof is given in **14.10-14.12**.[166]

14.9 Significance of Poincare's negative result

Bruns proved[167] in 1887 that for the three-body problem algebraic integrals of motion are exhausted by the classical integrals of motion (3 CM coordinates− t × velocity, 3 angular momentum components, 3 momentum components and energy), so no simplification further than accomplished by Lagrange long ago (see **14.5**). Then,

[166] based on H. Yoshida 5.1 in Y. Ohnuki and H. Yoshida, *Mechanics* (Iwanami 1994).

[167] E. T. Whittaker, *A treatise on the analytical dynamics of particles and rigid bodies* (Cambridge, 1937 [4th edition]) p358-377.

Poincare proved even for the restricted simpler problem there is no generally smooth (wrt μ) integration possible. Thus, almost all the people lost hope in solving the celestial mechanics in a closed form.

However, do not forget that Poincare did not show that for fixed values of μ analytic Φ exists (a less smooth Φ may exist).

14.10 Φ_0 is a function of I only

Let us Fourier-expand Φ_0:

$$\Phi_0(\theta, I) = \sum_{k\in\mathbb{Z}^2} \phi_k(I) e^{ik\cdot\theta}. \tag{14.16}$$

The Fourier expansion of $[H_0(I), \Phi_0]_{PB} = 0$ reads

$$\frac{\partial}{\partial I} H_0(I) \cdot \frac{\partial}{\partial \theta}\Phi_0 = \nabla H_0(I) \cdot \sum_{k\in\mathbb{Z}^2} (ik)\phi_k(I)e^{ik\cdot\theta} = 0. \tag{14.17}$$

Therefore, for all k

$$\nabla H_0(I) \cdot k\phi_k(I) = 0. \tag{14.18}$$

This must be true for any I, so its derivative wrt I must vanish. Thus,

$$\mathrm{Hess}(H_0(I))k\phi_k(I) + \nabla H_0(I) \cdot k\nabla\phi_k(I) = 0. \tag{14.19}$$

If $\phi_k(I) \neq 0$, then (14.18) implies $\nabla H_0(I) \cdot k = 0$, so $\mathrm{Hess}(H_0(I))k$ must vanish. But since the Hessian is non-singular, this is true only for $k = 0$. Thus, only $\phi_0(I) \neq 0$. That is, Φ_0 cannot depend on θ.

14.11 Φ_0 is a function of H_0 only

This is demonstrated under a complicated condition whose significance is unclear to me. So let us assume this. That is (recall that our system has 2 degrees of freedom),

$$\frac{\partial(H_0, \Phi_0)}{\partial(I_1, I_2)} = 0. \tag{14.20}$$

This implies[168] the existence of a function ψ such that $\Phi_0 = \psi \circ H_0$.

[168] This conclusion follows from (14.20) if everything is smooth enough; of course, we have assume everything is holomorphic, so the argument is OK. Since Jacobian is the volume ratio of $dI_1 d_I 2$ and $dH_0 d\Phi_0$, (14.20) implies dH_0 and $d\Phi_0$ are parallel.

14.12 Conclusion of proof: Φ is a function of H only

Take ψ in **14.11** and make $\Phi - \psi(H)$. This is $\Phi_0 - \psi(H_0) = 0$ for $\mu = 0$. Therefore, we may write

$$\Phi(\theta, I; \mu) - \psi(H(\theta, I; \mu)) = \mu \Phi^{(1)}(\theta, I; \mu), \tag{14.21}$$

which is an integral of motion as well. Expanding this,

$$\Phi^{(1)} = \Phi_0^{(1)} + \mu \Phi_1^{(1)} + \cdots, \tag{14.22}$$

we see that $\Phi_0^{(1)}$ is a function of H_0 (a function of I only is a function of H_0):

$$\Phi_0^{(1)} = \psi^{(1)}(H_0), \tag{14.23}$$

so we can repeat that argument as

$$\Phi^{(1)} - \psi^{(1)}(H) = \mu \Phi^{(2)}(\theta, I; \mu). \tag{14.24}$$

Thus,

$$\begin{aligned}
\Phi &= \psi(H) + \mu \Phi^{(1)} && (14.25) \\
&= \psi(H) + \mu(\psi^{(1)}(H) + \mu \Phi^{(2)}) && (14.26) \\
&= \psi(H) + \mu \psi^{(1)}(H) + \mu^2 \Phi^{(2)} && (14.27) \\
\cdots & && (14.28) \\
&= \psi(H) + \mu \psi^{(1)}(H) + \mu^2 \psi^{(2)}(H) + \cdots && (14.29)
\end{aligned}$$

This implies that if Φ can be obtained perturbatively, it is not an independent invariant of motion.

14.13 Asteroids and gaps

Due to Jupiter they could not form a planet (inside the so-called frost line). The total mass is about 4% of Moon and 40% of the total mass is concentrated in Ceres and Vesta.
Introductory video (this is for kids, but good)

https://www.youtube.com/watch?v=iy19nHTVLEY

All known asteroids

https://www.youtube.com/watch?v=vfvo-Ujb_qk

There are several major gaps in the distribution of asteroids called Kirkwood gaps (discovered 1866 by D Kirkwood 1814-1895). He correctly explained their origin in terms of motional resonance with Jupiter (e.g., 3:1 at 2.5 AU; see Fig. 14.3).

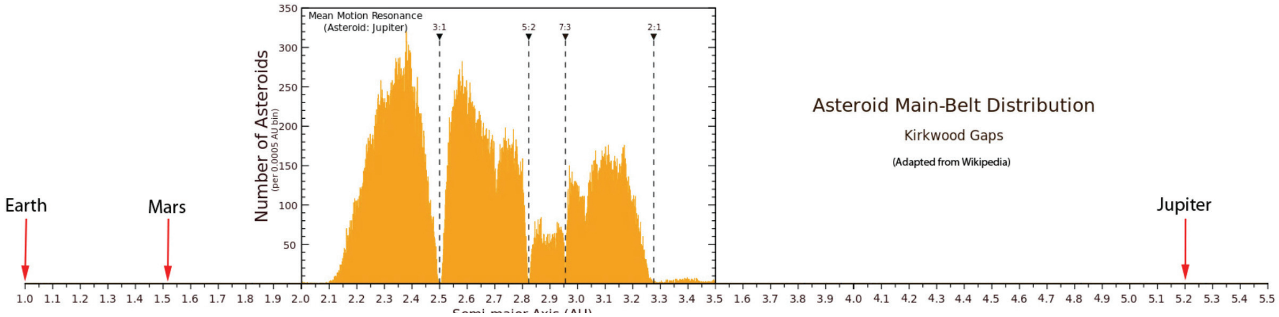


Figure 14.3: Asteroid belt and Kirkwood gaps

14.14 Special solutions of restricted three-body problem

Needless to say, many people tried to find special solutions (orbits) for the restricted three-body problem. There are 5 equilibrium points in the moving coordinate system (x, y). Three of them are along the SJ axis (Euler's linear solutions), but they are not stable. The remaining two are Lagrange's regular triangle solutions: the points are at the apexes of the regular triangles on the rotational plane (the plane spanned by (x, y)) whose one edge is the SJ segment.

Are the Lagrange points stable? For this to be stable it is not very hard to show that $\mu(1-\mu) < 1/27$ (or $\mu < 0.0385\cdots$). To show the stability actually, we must show the existence of small invariant tori surrounding these orbits. Thus, it is related to KAM and highly nontrivial, but except for three values of $\mu < 0.0385\cdots$ their stability was shown by Arnold.

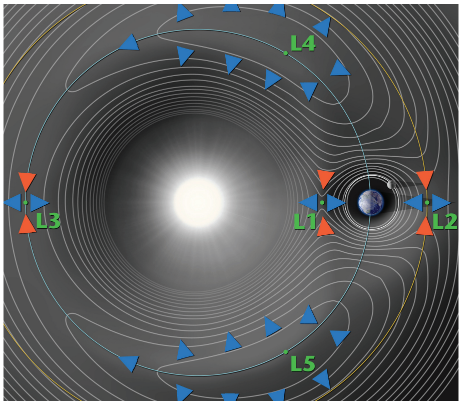


Figure 14.4: A gravitational potential contour plot showing Earth's Lagrangian points; L4 and L5 are above and below the planet, respectively. [Wikipedia: Jupiter Trojan]

The equilibrium points may not be so hard to find from the gravitational potential plot (Fig. 14.4, although this is for the earth). Lagrange's points are potential max points. Then, how can a particle stay near the hill tops? Physically speaking, this is due to Coriolis' force.

We actually observe asteroids around Lagrange's equilibrium points along Jupiters orbit (the so-called Trojan and Greek asteroid groups; the Trojan 'camp' trails Jupiter). The Hilda group asteroids are strong resonance with Jupiter. A good illustration of Trojan (green), Greek (red), and Hilda (white) families of asteroids is:

https://people.duke.edu/~ng46/borland/hilda%20family.gif.[169]

Here the Hilda is separated:

https://www.youtube.com/watch?v=yt1qPCiOq-8

14.15 Lagrange points for Earth and Moon collect dust particles (Kordylewski dust cloud)[170]

The L4 and L5 points of the Earth and Moon might be empty due to the gravitational perturbation of the Sun. However, in 1961, the Polish astronomer, Kazimierz Kordylewski found two bright patches near the L5 point, which might refer to an accumulation of interplanetary particles. However, many astronomers assume that these dust clouds do not exist, because the gravitational perturbation of the Sun, solar wind, and other planets may disrupt the stabilizing effect of the L4 and L5 Lagrange points of the Earth and Moon.

Using ground-born imaging polarimetry, the paper II presents new observational evidence for the existence of the KDC around the L5 point of the Earth-Moon system.

14.16 Did asteroids cause mass extinctions?

The biggest mass extinction (the end Permian MS) is certainly not due to asteroids. I am not so convinced by the physicists' asteroid theory of the KPg mass extinction (65.5 MaBP; Cretaceous-Paleogene mass extinction 'due to' an impact at Chicxulub). It is sure that an asteroid hit had a severe effect. However, this does not mean the asteroid hit was the main cause; it could well be the last straw. Remember that big scale mass extinctions are always with drastic sea level changes as seen in Fig.

[169] from DrBill's Astronomy Web Site http://hildaandtrojanasteroids.net

[170] J. Sliz-Balogh, A. Barta and G. Horvath, Celestial mechanics and polarization optics of the Kordylewski dust cloud in the Earth-Moon Lagrange oint L5 Pat I and II, Month Notices R. Astronom Soc, 480 5550, 482 762 (2018).

14.5. How can asteroid hit be predicted by the sea level changes? It should be fair to claim that an asteroid can cause havoc only when the biosphere is strained severely already.

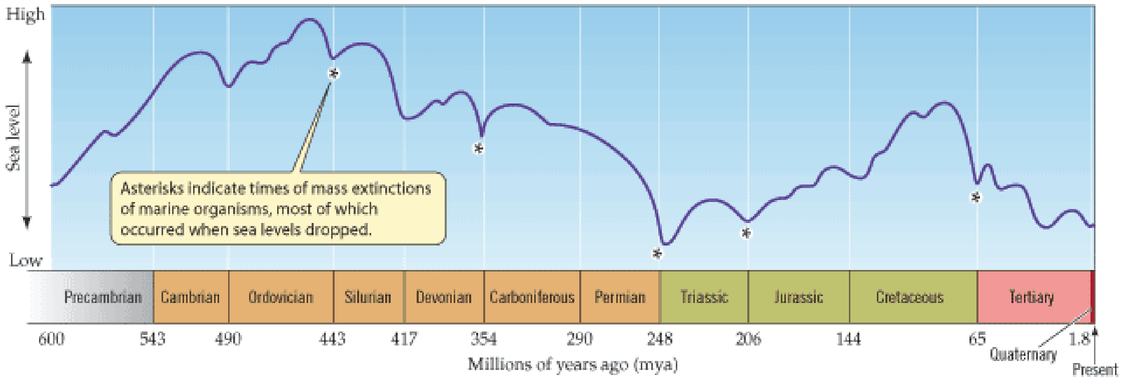


Figure 14.5: Mass extinctions and see level changes

14.17 Rings

Saturn detail:

https://www.youtube.com/watch?v=ENwQ7-qLlrA

15 Lecture 15. Siegel and KAM

There is almost no hope to solve celestial n ($n > 2$) body problem analytically in closed forms, the only remaining analytic hope is perturbative. Perturbative approaches generally have a grave difficulty called the small divisor problem. If the unperturbed periodic orbit is sufficiently 'irrational' (the Diophantine condition), this difficulty may be overcome. How to do this systematically is the key issue and is the core of its solution is proposed by Kolmogorov.

Here, (1) we consider perturbation around a center: when are the majority of invariant periodic orbits survive? [Poincaré-Siegel problem], then go to (2) the Kolmogorov-Arnold-Moser (KAM) theorem, asserting the survival of numerous invariant tori under perturbation. (1) is given with full demonstration details to taste the strategy used in (2). (2) is complicated, so only an outline of the demonstration is given.

15.1 Perturbative solution of motion

We know what we mean by solving the motion of a Hamiltonian system. It is to devise a smooth canonical transformation that 'linearizes the equation of motion (see **11.1**). What happens if the system is not solvable in the sense we already discussed? If the nonlinear part of the dynamics is small (only a perturbation), a natural idea is to construct a transformation near identity to get rid of this nonlinearity. This is the basic idea of perturbation. Especially because the results due to Poincare and others (**24.4**, **14.8**) that (almost) dashed the hope of solving celestial mechanics by quadrature, we desperately need perturbative approaches.

If we try to implement this strategy, we almost immediately encounter a grave difficulty called the problem of small divisors. Thus, first let us taste the difficulty and how to overcome it with a 'simple'[171] complex map problem: Find $x \to y$ that can linearize the original ODE in terms of x that has small nonlinear term g (that is $dg/dx|_{x=0} = 0$):

$$\dot{x} = Ax + g(x) \;\Rightarrow\; y = Ay, \tag{15.1}$$

where A is a diagonal matrix (for simplicity). This is the famous Poincaré-Siegel linearization problem (answers are **15.8** and **15.12**).

[171] The map problem is simple, but the solution is anything but simple as you will see.

15.2 Formal linearization around singular point

Consider an ode (if needed let us complexify it 4.7)

$$\dot{x} = f(x), \tag{15.2}$$

where $x \in \mathbb{R}^n$ and f is real analytic with $f(0) = 0$. The linear term is assumed to be diagonalized as Ax, where $A = [\alpha_1, \cdots, \alpha_n]$.

Let us consider the formal power series transformation

$$x = u(y) = y + u^2(y) + \cdots + u^k(y) + \cdots, \tag{15.3}$$

where u^k is a degree k homogeneous polynomial of y. (15.2) reads componentwisely as

$$\dot{x}_j = \alpha_j x_j + \sum_{|k|\geq 2} f_j^k x^k. \tag{15.4}$$

We have used the Hadamard notation.[172] (15.3) reads in this notation

$$x_j = y_j + \sum_{|k|\geq 2} u_j^k y^k. \tag{15.5}$$

15.3 Formal transformation to the second order

If we solve (15.3) of (15.5) for y we obtain

$$y_j = x_j - \sum_{|k|=2} u_j^k x^k + \cdots. \tag{15.6}$$

Here the higher order terms are very complicated and are suppressed. The original ODE in terms of y can be obtained by differentiating this as[173]

$$\dot{y}_j \overset{(15.6)}{=} \dot{x}_j - \sum_{|k|=2} u_j^k (k\cdot\alpha) x^k + \cdots \tag{15.7}$$

[172] ⟪**Hadamard notation**⟫ For $x = (x_1, \cdots, x_n)$ and $\alpha = (\alpha_1, \cdots, \alpha_n)$, $x^\alpha = \prod_{j=1}^n x_j^{\alpha_j}$. Also we write for $k = (k_1, \cdots, k_n) \in \mathbb{N}^n$, $k! = \prod_{j=1}^n k_j!$. Very often $|k| = \sum_j k_j$. If $f : \mathbb{R}^n \to \mathbb{R}$ and differentiable, the multivariate Taylor expansion reads

$$f(x+y) = \sum_{k\in\mathbb{N}^n} \frac{1}{k!} y^k \frac{d^k}{dx^k} f(x).$$

[173] $dx^k = d(\sum_j x_j^{k_j}) = \sum_j k_j dx_j (x_j^{k_j-1}) = \sum_j \alpha_j x_j k_j (x_j^{k_j-1}) dt = \sum_j \alpha_j k_j (x_j^{k_j}) dt = (a\cdot k) x^k dt$. More formally, $dx^k = (k\cdot dx) x^{k-1} = (Ak) x^k$.

$$\overset{(15.4)}{=} \sum_j \alpha_j x_j + \sum_{|k|=2} f_j^k x^k - \sum_{|k|=2} u_j^k (k\cdot\alpha) x^k + \cdots \tag{15.8}$$

$$\overset{(15.5)}{=} \sum_j \alpha_j \left(y_j + \sum_{|k|=2} u_j^k y^k \right) + \sum_{|k|=2} f_j^k y^k - \sum_{|k|=2} u_j^k (k\cdot\alpha) y^k + \cdots \tag{15.9}$$

$$= \sum_j \alpha_j y_j + \sum_{|k|=2} \left[\alpha_j u_j^k + \ f_j^k - u_j^k (k\cdot\alpha) \right] y^k + \cdots \tag{15.10}$$

$$= \sum_j \alpha_j y_j + \sum_{|k|=2} \left\{ f_j^k + \left[\alpha_j - (k\cdot\alpha) \right] u_j^k \right\} y^k + \cdots . \tag{15.11}$$

Therefore, if we can set

$$u_j^k = f_j^k / (k\cdot\alpha - \alpha_j), \tag{15.12}$$

the second order terms have been eliminated. For this to be possible we need a non-resonance condition for $|k| = 2$

$$\alpha_j - k\cdot\alpha \neq 0. \tag{15.13}$$

15.4 Non-resonance condition

As we will see the non-resonance condition is crucial, so let us state the condition clearly:

For $\alpha \in \mathbb{C}^n$ If the following set $\Gamma(\alpha) = \emptyset$, we say α is non-resonant:

$$\Gamma(\alpha) = \{ j \in \{1, 2, \cdots, n\},\ k \in \mathbb{N}^n \,|\, \alpha_j - k\cdot\alpha = 0 \}. \tag{15.14}$$

15.5 Formal transformation to order *k*

Replacing '2' with K in (15.5) as

$$x_j = y_j + \sum_{|k|=K} u_j^k y^k + \cdots, \tag{15.15}$$

we can repeat the computation in **15.3** to get the following instead of (15.11)

$$\dot{y}_j = \sum_j \alpha_j y_j + \sum_{2\le|k|<K} f_j^k y_j^k + \sum_{|k|=K} \left[f_j^k + (\alpha_j - (k\cdot\alpha))\, u_j^k \right] y^k + \cdots \tag{15.16}$$

Therefore, under the non-resonance condition (15.13), we can eliminate the order K terms.

Notice that in (15.15) $\cdots$ terms are not needed. Thus we can get the following obvious theorem:

15.6 Finite order elimination of nonlinear terms

For any $M \in \mathbb{N}^+$ we can make a polynomial transformation of order M

$$x = u(y) = y + u^2(y) + \cdots + u^M(y) \tag{15.17}$$

such that the original ODE (15.2) with the non-resonance condition **15.4** can be transformed into the following form:

$$\dot{y}_j = \alpha_j y_j + \sum_{|k|>M} g_j^k y^k. \tag{15.18}$$

15.7 What happens if $|\boldsymbol{k}|$ is very large?

If $|k|$ is very large, then $\alpha_j - k \cdot \alpha$ can be very close to zero (imagine a lattice and then try to draw a line not hitting any lattice point).

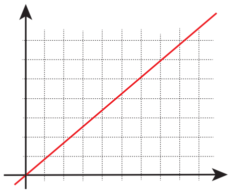


Figure 15.1: Suppose $\alpha_1 > 0$ and $\alpha_2 < 0$. $\alpha_1 k_1 + \alpha_2 k_2 = 0$ (red line) can be very close to a positive integer lattice points for large $|k|$ even if $\alpha_1/\alpha_2 \notin \mathbb{Q}$. In this case Poincaré's condition is trivially violated.

Thus, non-resonance alone cannot guarantee that u_k^j is of manageable size. Poincaré proved that if the convex hull of $\{\alpha_1, \cdots, \alpha_n\}$ does not contain the origin on the complex plane, and if α satisfies the non-resonance condition, then we can iterate the procedure in **15.6** indefinitely. That is,

15.8 Poincaré's lemma

For an ODE with the linear portion diagonalized and the nonlinear part denoted as

g:

$$\dot{x} = Ax + g(x), \tag{15.19}$$

if the convex hull of the eigenvalues do not contain the origin, and the non-resonance condition is satisfied, then there is an analytic transformation $x = y + u(y)$ with $du/dy|_{y=0} = 0$ such that

$$\dot{y} = Ay. \tag{15.20}$$

Its proof is through an 'honest' construction of the majorizing series that is convergent (tedious but straightforward).

15.9 Significance of the convex hull condition

Let $\alpha_1, \cdots, \alpha_n \in \mathbb{C}$ and its convex hull $[\alpha_1, \cdots, \alpha_n]$ does not contain the origin. Then, for any α in this convex hull there is $\delta > 0$ such that $|k \cdot \alpha| \geq \delta|k|$ for any $k \in \mathbb{N}^n$. Thus, what we worried in **15.7** never happens.

This is easy to see. Consider a unit vector $k/|k|$. Then $|k \cdot \alpha|/|k|$ is the length of the projection of the vector $(\alpha_1, \cdots, \alpha_n)$ onto $k/|k|$. Thus, this cannot be smaller than the distance between $[\alpha_1, \cdots, \alpha_n]$ and the origin.

15.10 What happens if Poincare condition fails?

Siegel demonstrated the following theorem:

Theorem [Siegel] For

$$\dot{x} = Ax + f(x), \tag{15.21}$$

where (as above) $A = [\alpha_1, \cdots, \alpha_n]$ is diagonal and f is analytic and $f'(0) = 0$. If there is $\gamma > 0$ such that for any $k \in \mathbb{Z}^n$ ($|k| \neq 0$)

$$|k \cdot \alpha| > \gamma |k|^{-n}, \tag{15.22}$$

then (15.21) is analytically transformed to $y = Ay$.

Notice (see **15.11**) that $\{\alpha_1, \cdots, a_n\}$ satisfying (15.22) is with full measure (exception is measure zero) wrt the usual Lebesgue measure m on $\mathbb{R}^{2n}$ when we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$.

15.11 Diophantine approximation

Lemma. For almost all $\alpha \in \mathbb{C}^n$[174] we can take a positive number γ such that for all

[174] This means: regarding α as a $2n$-real vector, it is almost sure with respect to the Lebesgue measure m on $\mathbb{R}^{2n}$.

$k \in \mathbb{Z}^n$ ($|k| \neq 0$)

$$|k \cdot \alpha| > \gamma/|k|^n. \tag{15.23}$$

[Demo]
Choose an arbitrary $R > 0$. Consider the totality Σ_γ of α for which (15.23) does not hold (i.e., (15.23) is untrue for any choice of $\gamma > 0$) and $|\alpha_i| < R$ for $i \in \{1, \cdots, n\}$. To estimate $m(\Sigma_\gamma)$ we consider the condition for α:

$$|k \cdot \alpha| \leq \gamma/|k|^n \tag{15.24}$$

for each shell of k: $K \leq |k| < K + 1$; it is an $n-1$-sphere of radius K with shell thickness 1. There are $\sim (2K)^{n-1}$ vectors. For each such k, the projection of α must be smaller than $\sim 1/K^{n+1}$ according to (15.24), so the α must be in the slab of thickness $\sim 2\gamma/K^{n+1}$ perpendicular to k and the remaining $n-1$ dimensions are with span $2R$. Thus, the total number of α satisfying (15.24) for $|k| \sim K$ is bounded by

$$\sim (K)^{n-1} \times 2\gamma/K^{n+1} \times (2R)^{n-1} = C(R)\gamma/K^2, \tag{15.25}$$

where $C(R)$ is an R dependent constant. Thus, we have an upper bound

$$m(\Sigma_\gamma) \leq C(R)\gamma \sum_K K^{-2}. \tag{15.26}$$

Therefore, the infimum of this wrt γ is zero. Since R is arbitrary, we have shown the lemma.

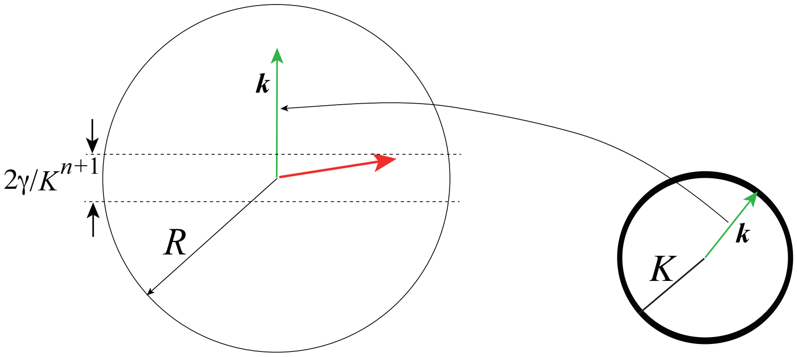


Figure 15.2: (15.25) illustrated

15.12 Siegel's stability theorem for conformal maps

We should continue to study the ODE considered by Poincaré, but here, we go to a simpler problem with the same difficulty.[175]

Let $S : \mathbb{C} \to \mathbb{C}$ be defined as $z \to f(z)$:

$$f(z) = \lambda z + g(z), \tag{15.27}$$

where $\lambda \neq 0$, $g(0) = 0$ which is holomorphic in $z < r$

We say the origin is stable for the system S, if and only if for any neighborhood of the origin U, there is a neighborhood of the origin $V \subset U$ such that $S^n V \subset U$ for all $n \in \mathbb{Z}$.

Siegel proved:

Theorem: If $|\lambda| = 1$ and $|\lambda^q - 1|^{-1} < C_0 q^2$ for $C_0 > 0$ and for any $q \in \mathbb{N}^+$ (the Diophantine condition), then S is stable.

15.13 Strategy to demonstrate Siegel's theorem

The strategy of the proof here (due to Moser) following Kolmogorov's idea is as follows:
First we show
Lemma: S is stable iff
(1) $|\lambda| = 1$,
(2) There is a holomorphy u in a neighborhood of the origin such that $\lim_{|z|\to 0} u(z)/z \to 1$ (asymptotically an identity) and

$$u(\lambda\zeta) = f(u(\zeta)). \tag{15.28}$$

That is, the variable change due to $z = u(\zeta)$ converts the original into $\lambda\zeta$:

$$\begin{tikzcd} z \arrow[r, "S"] & f(z) \\ \zeta \arrow[u, "u"] \arrow[r] & \lambda\zeta. \arrow[u, "u"] \end{tikzcd} \tag{15.29}$$

Thus, the theorem requires to solve (15.28). Instead of solving this, that is,

$$u(\lambda\zeta) = \lambda u(\zeta) + g(u(\zeta)) \tag{15.30}$$

at once, we solve a partially linearized version:

$$u_1(\lambda\zeta) = \lambda u_1(\zeta) + g(\zeta). \tag{15.31}$$

[175] Polya said: if you cannot understand a problem, there must be a simpler problem you cannot understand. Find it.

This u_1 cannot render the original equation to $\lambda\zeta_1$ even if $z = u_1(\zeta_1)$; actually,

$$u_1(\lambda\zeta_1 + \Phi^{(1)}(\zeta_1)) = f(\lambda u_1(\zeta_1)). \tag{15.32}$$

That is,

$$\begin{tikzcd}
z \arrow[r, "S"] & f(z) \\
\zeta_1 \arrow[u, "u_1"] \arrow[r] & \lambda\zeta_1 + \Phi^{(1)}(\zeta_1). \arrow[u, "u_1"']
\end{tikzcd} \tag{15.33}$$

Rewrite the original equation in terms of the new variable $\zeta_1 = u_1(z)$. The nonlinear term is reduced. That is, if $g < \varepsilon$, then, roughly speaking, $|\Phi^{(1)}| < \varepsilon^2$. Repeating this procedure rapidly decreases the size of the nonlinear term.

$$\begin{tikzcd}
z \arrow[r, "S"] & f(z) \\
\zeta_1 \arrow[u, "u_1"] \arrow[r, "S_1"] & \lambda\zeta_1 + \Phi^{(1)}(\zeta_1) \arrow[u, "u_1"'] \\
\zeta_2 \arrow[u, "u_2"] \arrow[r, "S_2"] & \lambda\zeta_2 + \Phi^{(2)}(\zeta_2) \arrow[u, "u_2"'] \\
\cdots \arrow[u, "u_3"] & \cdots. \arrow[u, "u_3"']
\end{tikzcd} \tag{15.34}$$

Thus, $u = \cdots \circ u_n \circ u_{n-1} \circ \cdots \circ u_2 \circ u_1$.

15.14 Preparatory lemma

$S : \mathbb{C} \to \mathbb{C}$ is stable iff for $\forall U$ neighborhood of $z = 0$ there is a simply connected neighborhood V $(\subset U)$ of origin such that $SV = V$.

[Demo] $\Leftarrow$ Trivial.
$\Rightarrow$ According to the definition of stability $\exists W \mathrel{-\!\!\in} S^nW \subset U$ for $\forall n \in \mathbb{Z}$. Define

$$\tilde{V} = \cup_{n\in\mathbb{Z}} S^nW. \tag{15.35}$$

Take a connected component V of $\tilde{V}$ containing $z = 0$. Then, any curve C connecting any point in V and the origin inside V satisfies $S^nC \subset V$ (i.e., different connected components do not map between them by S). Therefore, $SV = V$.

If V is not simply connected, repeat the above procedure all over starting from a smaller U. Suppose the resultant V is not simply connected even if U is very small. This, however, contradicts the assumption that S is holomorphic around $z = 0$, because there must be a neighborhood on which S looks very close to a rigid rotation (= linearized version). Thus, we can get the desired V.

15.15 Poof of Lemma in 15.13

Our key lemma already mentioned in **15.13** is:
Lemma: S is stable iff
(1) $|\lambda| = 1$
(2) There is a holomorphy u in a neighborhood of the origin such that $\lim_{|z|\to 0} u(z)/z \to 1$ (asymptotically an identity) and

$$u(\lambda\zeta) = f(u(\zeta)). \tag{15.36}$$

[Demo]
$\Leftarrow$ is obvious. Take a small disk D and $u(D) = V$.
$\Rightarrow$ According to the preparatory lemma **15.14**, stability implies the existence of $V \mathrel{-\!\!\in} SV = V \ni 0$, where V is simply connected (i.e., topologically equivalent to a disk). Then the Riemann mapping theorem guarantees the existence of a conformal map u from V to a disk D of radius ρ such that $u(0) = 0$ and

$$u(\zeta) = \zeta + b_2\zeta^2 + \cdots. \tag{15.37}$$

$u^{-1} \circ S \circ u : D \to D$ must be a rigid rotation of D, so there is μ ($|\mu| = 1$) and

$$u^{-1} \circ S \circ u(\zeta) = \mu\zeta. \tag{15.38}$$

$\mu = \lambda$, so $|\lambda| = 1$.

Now, we can start the strategy outlined in **15.13**.

15.16 If g is small, $u_1(z) - z$ must be small

If g, which satisfies $g(0) = g'(0) = 0$, in (15.27) is small in a neighborhood of the origin, u_1 constructed as an approximation to u should not be very far from the identity. This is shown in the Lemma below.

More precisely,
Lemma: Let g be holomorphic on a disk $|\zeta| < r$ and $|g| < \varepsilon$ with $g(0) = g'(0) = 0$. Define u_1 as the solution to

$$u_1(\lambda\zeta) = \lambda u(\zeta) + g(u_1(\zeta)). \tag{15.39}$$

Then, u_1 is holomorphic in $|\zeta| < r$ and on $|\zeta| < r(1-\theta)$ ($\theta \in (0,1)$)

$$|u_1(\zeta) - \zeta| < 2C_0\varepsilon/\theta^3, \tag{15.40}$$

where C_0 is a positive constant.

> Technically, to show that u_n are all close to identity, we must show $\Phi^{(n)}$ are all small. It is more convenient to show that u_1 solving (15.39) should not be very far from the identity, if g', instead of g, is small. This is given as a corollary **15.17**.

[Demo of Lemma]
Since g is holomorphic in $|\zeta| < r$, we can expand it as

$$g(\zeta) = \sum_{k\geq 2} g_k \zeta^k \tag{15.41}$$

with $|g_k| < \varepsilon/r^k$ (the Cauchy bound). Let us formally expand u_1 as

$$u_1 = \zeta + \sum_{k\geq 2} u_k \zeta^k. \tag{15.42}$$

Putting this into (15.39), we obtain

$$\sum_{k\geq 2} u_k[(\lambda\zeta)^k - \lambda\zeta^k] = \sum_{k\geq 2} g_k \zeta^k. \tag{15.43}$$

Thus, we must solve

$$[\lambda^k - \lambda]u_k = g_k. \tag{15.44}$$

That is (notice that we potentially have a small denominator difficulty, which is overcome by the Diophantine condition),

$$u_k = g_k/[\lambda^k - \lambda]. \tag{15.45}$$

We can show that u_1 is well defined and its deviation from the identity is bounded by ε by an explicit calculation as follows.

$$|u_1(\zeta) - \zeta| = \left|\sum_{k\geq 2}[\lambda^k - \lambda]^{-1} g_k \zeta^k\right| \leq \sum_{k\geq 2} |\lambda^{k-1} - 1|^{-1}|g_k||\zeta|^k \tag{15.46}$$

$$\leq \sum_{k\geq 2} C_0(k-1)^2 \varepsilon r^{-k}|\zeta|^k < \sum_{k\geq 2} C_0(k-1)^2 \varepsilon (1-\theta)^k, \tag{15.47}$$

where we have used the Cauchy bound for g_k and the Diophantine condition. Notice that

$$\sum_{k\geq 2}(k-1)^2 x^2 \sim \frac{d^2}{dx^2}\sum x^k \sim \frac{d^2}{dx^2}\frac{1}{1-x} \sim (1-x)^{-3}, \tag{15.48}$$

so (set $x = 1 - \theta$ to use this identity)

$$|u_1(\zeta) - \zeta| < \varepsilon\alpha C_0/\theta^3, \tag{15.49}$$

where α is a positive constant (which can be chosen as 2 with a bit more detailed calculation).

15.17 If g' is small, $u_1(z) - z$ must be small

If g, which satisfies $g(0) = g'(0) = 0$, in (15.27) and if its derivative g' is small in a neighborhood of the origin, u_1 constructed as an approximation to u should not be very far from the identity.

More precisely,
Lemma: Let g be holomorphic on a disk $|\zeta| < r$ and $|g'| < \varepsilon$ with $g(0) = g'(0) = 0$. Define u_1 as the solution to

$$u_1(\lambda\zeta) = \lambda u(\zeta) + g(u_1(\zeta)). \tag{15.50}$$

Then, u_1 is holomorphic in $|\zeta| < r$ and in $|\zeta| < r(1 - \theta)$ ($\theta \in (0, 1)$)

$$|u_1(\zeta) - \zeta| < 2C_0\varepsilon r/\theta^3, \tag{15.51}$$

where C_0 is a positive constant.
[Demo]
This should not be hard to guess, since $|g|$ must be of order $r \times |g'| < r\varepsilon$. A formal demonstration may be as follows. From (15.50) we get

$$\lambda u_1'(\lambda\zeta) = \lambda u_1'(\zeta) + g'(\zeta). \tag{15.52}$$

Let $v = \zeta u_1'(\zeta)$. We have

$$\lambda v(\lambda\zeta) = \lambda v(\zeta) + \zeta g'(\zeta). \tag{15.53}$$

Since $|\zeta g'(\zeta)| < \varepsilon|\zeta|$ in $|\zeta| < r$, the lemma in **15.16** tells us that in $|\zeta| < r(1-\theta)$

$$|v(\zeta) - \zeta| < 2C_0|\zeta|/\theta^3. \tag{15.54}$$

That is,

$$|u_1' - 1| < 2C_0\varepsilon/\theta^3. \tag{15.55}$$

Thus,

$$|u_1 - \zeta| = \left|\int_0^\zeta (u_1' - 1)d\zeta\right| < 2C_0 r\varepsilon/\theta^3. \tag{15.56}$$

15.18 (15.34) is a commutative diagram

To construct u we must be able to show that all the 'squares' in (15.34) must commute. That is, for example, u_1 and u_1^{-1} must be well defined in a neighborhood of the origin. We have already shown that u_1 is well defined in $|\zeta| < r(1-\theta)$.

To demonstrate u_1^{-1} is well-defined, we show that $u_1(\zeta)$ is topologically the same as ζ. To show this we use Rouche's theorem:[176]

Lemma. Choose ε small enough to satisfy for $\theta \in (0, 1/4)$

$$2C_0\varepsilon < \theta^4. \tag{15.57}$$

Then, u_1^{-1} is well defined in $|z| < r(1-2\theta)$ and

$$u_1^{-1}(\{z \,|\, |z| < r(1-2\theta)\}) \subset \{\zeta \,|\, |\zeta| < r(1-\theta)\}. \tag{15.58}$$

[Demo]
(15.57) and (15.56) tells us that for $v(\zeta) = u_1(\zeta) - \zeta$

$$|v(\zeta)| \le r\theta. \tag{15.59}$$

If $|z| < r(1-2\theta)$ and $|\zeta| < r(1-\theta)$, then this implies that

$$|v(\zeta)| \le |\zeta| - |z| \le |\zeta - z|. \tag{15.60}$$

Since $|\lambda| = 1$, this implies that the equations $\lambda\zeta + v(\zeta) = u_1(\zeta) = z$ and $\zeta = z$ have the same zeros. Thus, u_1^{-1} is well defined in $|z| < r(1-2\theta)$ and its image is in $|\zeta| < r(1-\theta)$.

15.19 S_1 is well defined close to the origin

We can show

Lemma: $S_1(\zeta) = u_1^{-1} \circ f \circ u_1 = \lambda\zeta + \Phi^{(1)}(\zeta)$ in (15.34) is well defined in $|\zeta| < r(1-4\theta)$ if $\varepsilon < \theta$.

[Demo]
Since $|f| \le |\zeta| + |g|$ and $|g| < r\varepsilon$, if $|g'| < \varepsilon$

$$|f| \le |\zeta| + |g| \le |\zeta| + r\varepsilon \le |\zeta| + r\theta, \tag{15.61}$$

because we assume $\varepsilon < \theta$.

The image of f must be in the domain of u_1^{-1}, i.e., $|z| < r(1-2\theta)$. Therefore,

[176] ⟪**Rouche's theorem**⟫ Let D be a simply connected open set. Suppose f and g are non-constant holomorphic functions on $[D]$ (closure), and $|f| > |g|$ on ∂D. Then, the number of zeros of f and $f + g$ agree in D.

the domain of f must be restricted to $|z| < r(1-3\theta)$. Then, (15.58) implies that the domain of u_1 must be restricted to $|\zeta| < r(1-4\theta)$.

15.20 $\Phi^{(1)}$ is small and we can iterate the above argument indefinitely

If we can show that $\Phi^{(1)}$ is sufficiently small, we can repeat the above argument for u_2, and ad infinitum. We can show (we wish to use an analogue of **15.17** to show that u_2 is close to identity)

Lemma: $|{\Phi^{(1)}}'(\zeta)| < C_1\varepsilon^2/\theta^4$, where $C_1 < 3C_0$ in $|\zeta| > r(1-5\theta)$ with $0 < \varepsilon < \theta < 1/5$.

[Demo]

$$u_1(\lambda\zeta + \Phi^{(1)}) = f(u_1) = \lambda u_1(\zeta) + g(u_1), \tag{15.62}$$

so (recall $v = u_1 - \zeta$ and $v(\lambda\zeta) = \lambda v(\zeta) + g(\lambda\zeta)$)

$$\lambda\zeta + \Phi^{(1)} + v(\lambda\zeta + \Phi^{(1)}) = \lambda\zeta + \lambda v(\zeta) + g(\lambda\zeta + v(\zeta)). \tag{15.63}$$

Therefore,

$$\Phi^{(1)} = \lambda v(\zeta) - v(\lambda\zeta + \Phi^{(1)}) + g(\lambda\zeta + v(\zeta)), \tag{15.64}$$

but $\lambda v(\zeta) = v(\lambda\zeta) - g(\lambda\zeta)$ gives

$$\Phi^{(1)} = v(\lambda\zeta) - v(\lambda\zeta + \Phi^{(1)}) + g(\lambda\zeta + v(\zeta)) - g(\lambda\zeta). \tag{15.65}$$

Using the mean value theorem

$$|v(\lambda\zeta) - v(\lambda\zeta + \Phi^{(1)})| \le \sup|v'| \sup|\Phi^{(1)}| \le \frac{2\varepsilon C_0}{\theta^3} \sup|\Phi^{(1)}| < \theta \sup|\Phi^{(1)}| < \frac{1}{5}\sup|\Phi^{(1)})|. \tag{15.66}$$

From this and (15.65)

$$|\Phi^{(1)}| \le \frac{1}{5}\sup|\Phi^{(1)}| + |g(\lambda\zeta + v(\zeta)) - g(\lambda\zeta)| \tag{15.67}$$

or

$$\frac{4}{5}\sup|\Phi^{(1)}| \le \sup|g(\lambda\zeta + v(\zeta)) - g(\lambda\zeta)|. \tag{15.68}$$

Again we use the mean-value theorem

$$\frac{4}{5}\sup|\Phi^{(1)}| \le \sup|g'| \sup|v(\zeta)| < \varepsilon\frac{2C_0\varepsilon}{\theta^3}r. \tag{15.69}$$

Thus we get for $|\zeta| < r(1-4\theta)$

$$\sup|\Phi^{(1)}| < \frac{5C_0\varepsilon^2}{2\theta^3}r. \tag{15.70}$$

Its derivative may be estimated following Cauchy in a somewhat smaller domain $|\zeta| < r(1-5\theta)$ as

$$|\Phi^{(1)\prime}(\zeta)| < C_1\varepsilon^2/\theta^4. \tag{15.71}$$

15.21 Summary of a recursion step

Thus we have completely constructed the single step for the recursion scheme ((15.33) realized) [Please ignore strange arrowheads at the ends of the formulas]

$$\begin{tikzcd}
z \arrow[r, "S"] & \lambda z + g(z) & |g'(z)| < \varepsilon \text{ in } |z| < r \curlywedge \\
\zeta_1 \arrow[u, "u_1"] \arrow[r, "S_1"] & \lambda\zeta_1 + \Phi^{(1)}(\zeta_1) \arrow[u, "u_1"] & |\Phi^{(1)}(\zeta)| < C_1\varepsilon^2/\theta^4 \text{ in } |\zeta| < r(1-5\theta). \curlywedge
\end{tikzcd} \tag{15.72}$$

or more generally

$$\begin{tikzcd}
\zeta_n \arrow[r, "S_n"] & \lambda\zeta_n + \Phi^{(n)}(\zeta_n) & |\Phi^{(n)\prime}(\zeta_n)| < \varepsilon_n \text{ in } |\zeta_n| < r_n \curlywedge \\
\zeta_1 \arrow[u, "u_{n+1}"] \arrow[r, "S_{n+1}"] & \lambda\zeta_{n+1} + \Phi^{(n+1)}(\zeta_{n+1}) \arrow[u, "u_{n+1}"] & |\Phi^{(n+1)}(\zeta_{n+1})| < C_1\varepsilon_n^2/\theta_n^4 = \varepsilon_{n+1} \text{ in } |\zeta_{n+1}| < r_n(1-5\theta_n) = r_{n+1}. \curlywedge
\end{tikzcd} \tag{15.73}$$

15.22 Convergence of the recursive transformations

For the recursive transformations to converge we must at least demand $\varepsilon_n \to 0$ and $r_n \to r_\infty > 0$. Thus, $\sum \theta_n$ must be convergent. Moreover,

$$U_n = u_n \circ u_{n-1} \circ \cdots \circ u_1 \tag{15.74}$$

must be uniformly convergent in $|\zeta| < r_\infty$. Or equivalently

$$U_n' = u_n' u_{n-1}' \cdots u_1' \tag{15.75}$$

must be uniformly convergent. Since

$$|U_n'| \le \prod_{k=1}^{n} (1 + |\Phi^{(n)\prime}|), \tag{15.76}$$

if

$$\sum |\Phi^{(n)\prime}| \le \sum \varepsilon_n \tag{15.77}$$

converges, we are done!

Thus the requirements are

$$\sum \varepsilon_n \quad < \quad \infty, \tag{15.78}$$

$$\sum \theta_n \quad < \quad \infty \tag{15.79}$$

with $0 < \varepsilon_n < \theta_n < 1/5$. This is satisfied if we choose $\theta_n = (1/5)c^{-n}$ for $c > 1$. Indeed,

$$\varepsilon_{n+1} = \frac{C_1}{\theta_n^4}\varepsilon_n^2 = \frac{C_1}{5}c^{4n}\varepsilon_n^2 < C_2^{n+1}\varepsilon_n^2 \tag{15.80}$$

Define $\varepsilon_n = C^{n+2}\varepsilon_n$. Then $\varepsilon'_{n+1} < (\varepsilon'_n)^2$, so choose $\varepsilon'_0 \leq 1$ and all the requirements for θ_n are met.

15.23 Setup of the simplest version of KAM theorem

Let $H(\theta, I, \varepsilon)$ be a near integrable Hamiltonian; for $\varepsilon = 0$

$$H_0(I) = H(\theta, I, 0). \tag{15.81}$$

That is, I is invariant, and determines an invariant T^n as asserted by the Liouville-Arnold theorem. We wish to see whether the invariant T^n with $I = I_0$ survives the perturbation. Since we may add any constant to I, we choose $I_0 = 0$. We consider I close to $I_0 = 0$: $B = \{I \,|\, |I| < b\}$ for some small $b > 0$. We assume $H(\theta, I, \varepsilon)$ is holomorphic on $\mathcal{M} \times (-\varepsilon_0, \varepsilon_0)$. We write the zeroth and the first order terms as

$$H(\theta, I, 0) = E + \omega \cdot I + Q(I), \tag{15.82}$$

where $Q = O[I^2]$ in the $I \to 0$ limit.

We assume

(1) ω satisfies a Diophantine condition **15.11**

$$|k \cdot \omega| \geq \frac{\alpha}{|k|^\tau} \tag{15.83}$$

with $\tau > n - 1$.

(2) $H(\theta, I, 0)$ is non-degenerate in the sense that $\det(\partial^2 Q/\partial I_i \partial I_j) \neq 0$.

15.24 KAM theorem

With the setup **15.23** there is a holomorphic canonical transformation $\phi : T^n \times$

$B^* \times (-\varepsilon_*, \varepsilon_*) \to \mathcal{M}$, where $B^* \subset B$ and $0 < \varepsilon_* < \varepsilon_0$, such that the canonical transformation $(\theta, I) \to (\theta', I')$ transforms $H(\theta, I, \varepsilon)$ as

$$K(\theta', I'.\varepsilon) = E_*(\varepsilon) + \omega \cdot I' + Q_*(\theta', I', \varepsilon), \tag{15.84}$$

where $Q_* = O[I'^2]$ in the $I' \to 0$ limit. Here, ϕ is an identity for $\varepsilon = 0$.

This implies that the canonical equation of motion reads

$$\begin{aligned} \frac{d\theta'}{dt} &= \frac{\partial K}{\partial I'} = \omega + \frac{\partial Q_*}{\partial I'}, & (15.85) \\ \frac{dI'}{dt} &= -\frac{\partial K}{\partial \theta'} = -\frac{\partial Q_*}{\partial \theta'}. & (15.86) \end{aligned}$$

Notice that the derivatives of Q_* for $I' = 0$ vanish because $Q_* = O[I'^2]$. This means that the 'holomorphically deformed' $I = 0$ torus survives and the motion on it is just the original (almost) periodic flows. This torus is called a KAM torus.

15.25 Strategy of proof of KAM

The idea is just as explained for Siegel's theorem. Let the perturbed Hamiltonian reads

$$H = H_0 + \varepsilon P - 0, \tag{15.87}$$

where $H_0 = E_0 + \omega I + Q(I)$ with $Q = O[I^2]$.

We construct a canonical transformation $\phi_1 : (\theta, I) \to (\theta', I')$ such that

$$H_1 = H \circ \phi_1 = K_1 + \varepsilon^2 P_1, \tag{15.88}$$

where

$$K_1 = E_1(\varepsilon) + \omega I' + Q_1(\theta', I', \varepsilon), \tag{15.89}$$

with $Q_1 = O[I'^2]$.

This is accomplished by removing the θ-dependent constant portion and the first order in I of P_0 of order ε choosing ϕ_1.

If we repeat the same strategy, we get

$$H_2 = H_1 \circ \phi_2 = K_2 + \varepsilon^4 P_2. \tag{15.90}$$

Of course we must show that $\phi_n \circ \cdots \circ \phi_2 \circ \phi_1$ converges.

15.26 Canonical transformation that can remove $O[\varepsilon]$ terms

We introduce $(\theta, I) \to (\theta', I')$ with the generator $G(\theta, I')$

$$dG = Id\theta + \theta' dI' + (K - H)dt, \tag{15.91}$$

but $G(\theta, I')$ is time independent, so $K = H(I(\theta', I'), \theta(\theta', I'))$.

We wish to remove I'-independent θ' dependence (to keep I' invariant) and the $O[I']$ terms that cannot be written as $\omega \cdot I'$ to order ε. We assume the following form:

$$\begin{aligned} I &= \frac{\partial G}{\partial \theta} = I' + \varepsilon\beta(\theta, I'), & (15.92)\\ \theta' &= \frac{\partial G}{\partial I'} = \theta + \varepsilon a(\theta). & (15.93) \end{aligned}$$

If ε is sufficiently small, we can invert the second equation as

$$\theta = \varphi(\theta', \varepsilon). \tag{15.94}$$

Thus we may set (b is a constant vector)

$$G = \theta \cdot I' + \varepsilon[\theta \cdot b + s(\theta) + a(\theta) \cdot I'] \tag{15.95}$$

We see that

$$\beta(\theta, I') = b + \frac{\partial}{\partial \theta} s(\theta) + \frac{\partial a}{\partial \theta} I'. \tag{15.96}$$

Notice that we can choose s and a integration over T^n to vanish (by subtracting appropriate constants tat we can ignore from G).

15.27 Transformation of Hamiltonian

Let us write

$$H(\theta, I, \varepsilon) = H(\theta, I, 0) + \varepsilon P(\theta, I, \varepsilon) = E + \omega \cdot I + Q(I) + \varepsilon P(\theta, I, \varepsilon). \tag{15.97}$$

First, we change $I \to I'$:

$$\begin{aligned} H(\theta, I, \varepsilon) &= H(\theta, I' + \varepsilon\beta(\theta, I'), \varepsilon) & (15.98)\\ &= E + \omega \cdot I' + Q(\theta, I') + \varepsilon \left[\omega \cdot \beta + \frac{\partial Q}{\partial I'}\beta + P(\theta, I', 0)\right] + \varepsilon^2 P'(\theta, I', \varepsilon), & (15.99) \end{aligned}$$

where P' denotes all the remaining terms.

Since we do not care for $O[I'^2]$, but we wish to rewrite the rest of $O[\varepsilon]$ as constant. Let us expand $O[\varepsilon]$ in powers of I':

$$\begin{aligned}
\omega \cdot \beta + \frac{\partial Q}{\partial I'}\beta + P(\theta, I', 0) &= \omega \cdot \left(b + \frac{\partial s}{\partial \theta} + \frac{\partial a}{\partial \theta} I'\right) + \frac{\partial Q}{\partial I'}\left(b + \frac{\partial s}{\partial \theta} + \frac{\partial a}{\partial \theta} I'\right) + P(\theta, I', 0) \qquad (15.100)\\
&= \omega \cdot (b + \frac{\partial s}{\partial \theta}) + P(\theta, 0, 0) \qquad (15.101)\\
&\quad + \left[\omega \cdot \frac{\partial a}{\partial \theta} + (b + \frac{\partial s}{\partial \theta})\frac{\partial^2 Q}{\partial I' I'} + \frac{\partial P}{\partial I'}\right] I' + O[I'^2] \qquad (15.102)
\end{aligned}$$

15.28 Removing θ-dependence from energy

The I' independent term in (15.102) is

$$\omega \cdot b + D_\omega s(\theta) + P(\theta, 0, 0), \qquad (15.103)$$

where

$$D_\omega = \omega \cdot \frac{\partial}{\partial \theta}. \qquad (15.104)$$

Therefore, to remove the θ dependence, we must choose s so that

$$\omega \cdot b + D_\omega s(\theta) + P(\theta, 0, 0) = \omega \cdot b + P_0 \qquad (15.105)$$

where P_0 is the average value of $P(\theta, 0, 0)$. Thus, we must solve

$$D_\omega s(\theta) = -(P(\theta, 0, 0) - P_0), \qquad (15.106)$$

or we must show the well-definedness of D_ω^{-1}. Here, we encounter the small devisor problem.

As shown below **15.31-15.33**, since the average of $P(\theta, 0, 0) - P_0$ over T^n vanishes, s is well-defined.

15.29 Removing the $O[I']$ term

We wish to eliminate

$$D_\omega a + (b + \frac{\partial s}{\partial \theta})\frac{\partial^2 Q}{\partial I' I'} + \frac{\partial P}{\partial I'}. \qquad (15.107)$$

Since we wish to determine a the average of the terms other than $D + w\alpha$ over T^n must vanish. First, we choose b to satisfy this condition. This is possible due to the non-degeneracy condition (Hess $\neq 0$). Then, a can be computed just as s.

15.30 Full transformation

The remaining task is to replace θ in $H(\theta, I' + \varepsilon\beta(\theta, I'), \varepsilon)$ with θ' (see (15.94)).

15.31 Well-definedness of D_{ω}^{-1}: formal solution

We wish to solve

$$D_{\omega}u = \omega \cdot \frac{\partial u}{\partial \theta} = f. \tag{15.108}$$

To solve this formally, u and f are Fourier expanded as

$$u(\theta) = \sum_{k \in \mathbb{R}^n} u_k e^{ik\theta}, \quad f(\theta) = \sum_{k \in \mathbb{R}^n} f_k e^{ik\theta}, \tag{15.109}$$

where

$$u_k = \frac{1}{(2\pi)^n} \int_{T^n} d\theta \, u(\theta) e^{-ik\theta}, \quad f_k = \frac{1}{(2\pi)^n} \int_{T^n} d\theta \, f(\theta) e^{-ik\theta}. \tag{15.110}$$

Therefore, (15.108) gives

$$ik\omega u_k = f_k \tag{15.111}$$

For this to be solvable $f_0 = 0$ (i.e., f averaged over T^n is zero. Therefore, the general solution to (15.108) is given be

$$u(\theta) = u_0 + \sum_{k \in \mathbb{R}^n \setminus \{o\}} \frac{f_k}{i\omega k} e^{ik\theta}. \tag{15.112}$$

Thus, if the average of u over T^n vanishes (i.e., $u_0 = 0$), the solution to (15.108) is formally unique.

15.32 Well-definedness of D_{ω}^{-1}: convergence of formal solution

We assume that the real analytic function on T^n is analytic in a 'strip' T_{ξ}^n containing the real axis:

$$T_{\xi}^n = \{\theta \in \mathbb{C}^n \,|\, |\mathrm{Im}\theta_j| < \xi\}/(2\pi\mathbb{Z})^n \tag{15.113}$$

If f is analytic on T_{ξ}^n, then

$$|f_k| \leq \|f\| e^{-|k}\xi, \tag{15.114}$$

where $\|f\|$ is the max of $|f$ on T^n_ξ. Its demonstration is in **15.33**.

Now, we assume the Diophantine condition

$$|k\omega| \geq \alpha |k|^\tau. \tag{15.115}$$

Then,

$$|u_k| = \left|\frac{f_k}{i\omega k}\right| \leq \alpha^{-1}\|f\|e^{-|k|\xi}|k|^\tau. \tag{15.116}$$

Using this estimate the maximal convergence of u is shown.

15.33 Cauchy estimate of $|f_k|$

$$f_k = \frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta)e^{-ik\theta} \tag{15.117}$$

The integration path may be shifted to the imaginary direction so that jth component is shifted by $i\xi\mathrm{sgn}(k_j)$ ($\theta_j \to \theta_j + i\xi\mathrm{sgn}(k_j)$). Then, $ik\theta \to ik\theta - \xi|k|$ (we stick to the Hadamard notation).

$$\begin{aligned}
f_k &= \frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta)e^{-ik\theta} && (15.118)\\
&= \frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta_j + i\xi\mathrm{sgn}(k_j))e^{-ik_j(\theta_j + i\xi\mathrm{sgn}(k_j))} && (15.119)\\
&= \frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta_j + i\xi\mathrm{sgn}(k_j))e^{-ik\theta-\xi|k|}. && (15.120)
\end{aligned}$$

Therefore,

$$\begin{aligned}
|f_k| &= e^{-\xi|k|}\left|\frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta_j + i\xi\mathrm{sgn}(k_j))e^{-ik\theta}\right| && (15.121)\\
&\leq e^{-\xi|k|}\frac{1}{(2\pi)^n}\int_{T^n} d\theta\, \left|f(\theta_j + i\xi\mathrm{sgn}(k_j))e^{-ik\theta}\right| && (15.122)\\
&\leq e^{-\xi|k|}\frac{1}{(2\pi)^n}\int_{T^n} d\theta\, \|f\| = \|f\|e^{-|k|\xi}. && (15.123)
\end{aligned}$$

16 Lecture 16. General picture of Hamiltonian systems

16.1 Completely integrable systems

We know the Liouville-Arnold theorem (**11.6**): if a system is completely integrable, we have a canonical transformation converting it to a system described by the action-angle variables that parameterize 'orderly' trajectories on T^n.
Example: the Toda system: three point masses interacting with the Toda potential (**12.3**)[177],

$$H = \frac{1}{2}(p_x^2 + p_y^2) + \frac{1}{24}\left[\exp(2y + 2\sqrt{3}x) + \exp(27 - 2\sqrt{3}x) + \exp(-4y)\right] - \frac{1}{8}. \quad (16.1)$$

We know how to convert this into the Lax pair representation All the trajectories are periodic or almost periodic (see Fig. 16.1[178]).

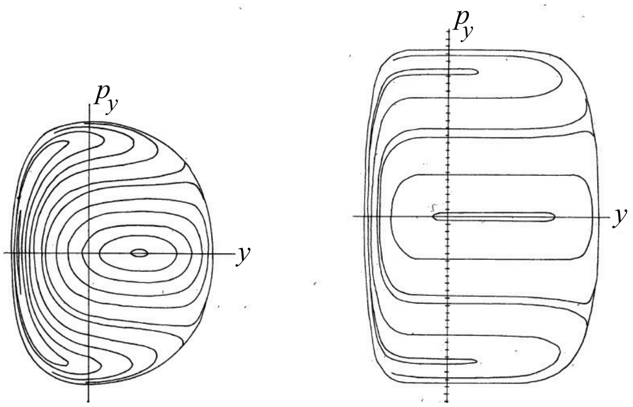


Figure 16.1: Left: A surface of section for the Toda Hamiltonian (16.1) at energy $E = l$. The scales on both axes are the same. The outermost oval crosses the positive p_y-axis at 1.3 and the positive y-axis at 1.4; Right: At energy $E = 256$. Here the scales on two axes are not the same. The outermost oval crosses the positive y-axis at $y = 4$ and the positive p_y-axis at $p_y = 22.6$. All the allowed phase space is surveyed in these figures, because outside the outermost curves $p_x^2 < 0$. [Fig. 1,2 of FORD et al., PTP 50 1547 (1973)]

[177] Its polynomial expansion truncated at the third order (the Henon system) is not integrable.

[178] Joseph FORD, Spotswood D. STODDARD and Jack S. Turner, "On the Integrability of the Toda Lattice," Prog Theor Phys 50 1547 (1973).

16.2 Near integrable systems

The Henon system is obtained by expanding and truncating the Toda Hamiltonian (16.1) as

$$H = \frac{1}{2}(p_x^2 + p_y^2) + \frac{1}{2}(x^2 + y^2) + x^2 y - \frac{1}{3}y^3. \tag{16.2}$$

This describes a motion in the potential that is something like a symmetric monkey saddle. A perturbative approach dissects H into H_0 and H_1 as

$$H_0 \quad = \quad \frac{1}{2}(p_x^2 + p_y^2) + \frac{1}{2}(x^2 + y^2), \tag{16.3}$$

$$\varepsilon H_1 \quad = \quad = \varepsilon\left(x^2 y - \frac{1}{3}y^3\right). \tag{16.4}$$

Comparison between the formal perturbation series (truncated at third order) and numerical results is in Fig. 16.2.[179]

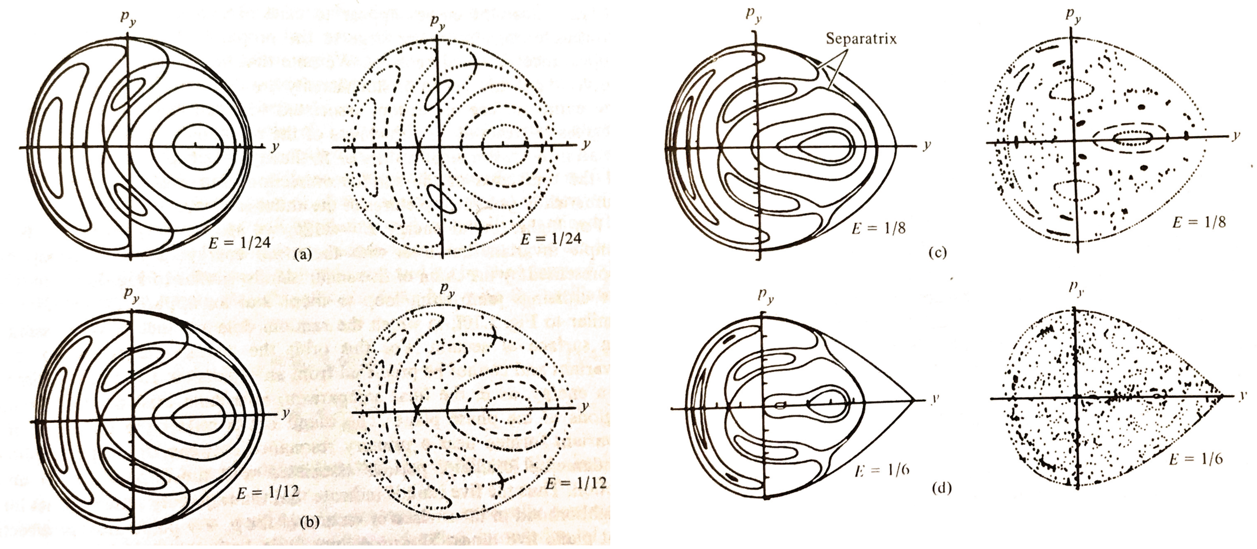


Figure 16.2: Perturbation fails even for small nonintegrable perturbation: [Figs originally from Gustavson Ast J 71 670 (1966).]

The comparison tells us that perturbation series fail to converge.

[179] F. G. Gustavson "On Constructing Formal Integrals of a Hamiltonian System Near an Equilibrium Point," Astronom. J 71 670 (1966). Here, the figure is from A J Lichtenberg andL A Lieberman, *Regular and Stochastic motion* (Springer, Applied Mathematical Science 38, 1983). Fig. 1.13.

16.3 What Poincaré realized when perturbation changes systems qualitatively

The picture Poincaré conceived may look like Fig. 16.3.

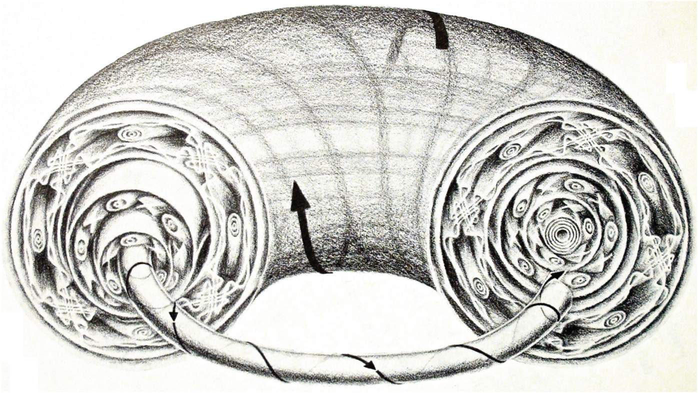


Figure 16.3: What Poincaré conceived (after Arnold) [Fig. 8.3.3 of]

> "On January 21, 1889, King Oscar II turned 60. On the same day, the king announced the winner of the mathematics prize that had been established in his name. All treatises had been submitted anonymously, supplied with a title and an envelope containing the author's name, The envelops were now opened with great ceremony. The winner was Henri Poincaré." (p377[180])
>
> "Poincaré's treatise was a comprehensive attempt to answer the question: "Is the solar system stable?" "(p378) However, he discussed the restricted three-body problem. "Mittag-Leffler had informed Hermite that he and Weierstrass thought Poincaré should be awarded the prize... Hermite, who had received copies of the various submissions, agreed. But even though they could immediately see and acknowledge the quality of Poincaré's work, not everything was equally easy to understand. Weierstrass had sought clarification on several points, and Poincaré had sent several extensive addenda. Hermite remarked that, as usual, it was difficult to understand Poincaré—that it was his style to spring over details and leave the reader to fill in the gaps. $\cdots$ According to Hermite, Poincaré was a seer to whom the truth revealed itself in a brilliant light.
>
> Mittag-Leffler's plan was to present the winning submissions in *Acta*[181] in October 1889,

[180] A. Stubhaug, *Gösta Mittag-Leffler, a man of conviction* (translated by Tiina Nunnally, Springer 2010; Norwegian original 2007).

[181] Acts Mathematica

and Phragmén, who acted as the editorial secretary, was handed the big job of editing the treatise. When printing began in July, Phragmén discovered several passages in Poincaré's work that seemed quite obscure. Mittag-Leffler immediately pointed this out in a letter to Poincaré, who upon further study found that in another place in the treatise he had made more serious errors. But Poincaré did not report this to Mittag-Leffler until December, and by that time fifty copies of *Acta* had already been printed. These were sent out to subscribers and booksellers, mostly in the Nordic countries, but also to England, France, and Germany. Mittag-Leffler feared that the error would prove ruinous for the reputation of both Poincaré and the journal, as well as Oscar II's entire involvement. He immediately sent a letter to all of these subscribers and booksellers, asking them to return the copies they had received. He explained that certain corrections were necessary." (p379)

Mittag-Leffler "asked Poincaré to pay for the already printed treatise. Poincaré agreed without protest, and his total cost for the new printing was eventually calculated to be 3,500 kronor—1,000 kronor more than the prize money he had received." (p380)

"Poincaré's work totaled 270 pages—the note and addenda filled 93 pages. What was new and brilliant about Poincaré's work eventually overshadowed any talk of errors or corrections." (p380).

16.4 Resonance vs nonresonance

The following example is related to heating ions in confined plasmas (Fig. 16.4).[182] A constant magnetic field and a perpendicularly propagating electrostatic wave with frequency several times the ion cyclotron frequency are applied to charged particles.

Consider a single ion with mass m and charge q in the fields

$$\boldsymbol{B} = B_0\boldsymbol{e}_z, \;\; \boldsymbol{E} = E_0\boldsymbol{e}_y\cos(ky-\omega t). \tag{16.5}$$

These fields are given by the potentials

$$\boldsymbol{A} = -B_0 y\boldsymbol{e}_x, \; \phi = -(E_0/k)\sin(ky-\omega t); \tag{16.6}$$

thus the Hamiltonian H is given by

$$H = \frac{1}{2m}[(p_x+qB_0y)^2+p_y^2] - \frac{qE_0}{k}\sin(ky-\omega t). \tag{16.7}$$

[182] Original: C F F Karney Princeton thesis: Stochastic Ion Heating By a Lower Hybrid Wave (1978). See C F F Karney and A Ber, Stochastic Ion Heating by a Perpendicularly Propagating Electrostatic Wave, PRL **39**, 550 (1977): "The motion of an ion in the presence of a constant magnetic field and a perpendicularly propagating electrostatic wave with frequency several times the ion cyclotron frequency is shown to become stochastic for fields satisfying ...". The exposition here follows C. F. F. Karney, "Stochastic ion heating by a lower hybrid wave" Phys. Fluid **21**, 1584 (1978).

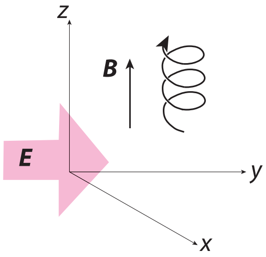


Figure 16.4:

Scaling variables appropriately,[183] the Hamiltonian reads

$$H = \frac{1}{2}[(p_x + y)^2 + p_y^2] - \alpha \sin(y - vt), \tag{16.8}$$

where $v = \omega/\Omega$ and $\alpha = qkE_0/m\Omega^2 = (E_0/B_0)/(\Omega/k)$.

Next, we wish to remove the time dependence. The canonical transformation with the following generator for $(x, p) \to (X, P)$

$$G = (P_x - vt)x + P_y(y - vt + Px) \tag{16.9}$$

would do this:

$$X = x + p_y,\ P_x = p_x + vt,\ Y = y + p_x,\ P_y = p_y, \tag{16.10}$$

and the new Hamiltonian K is given by

$$K = H - vX + \frac{1}{2}(Y^2 + P_y^2) - vX - \alpha \sin(Y - P_x). \tag{16.11}$$

Note that $Y - y$ is a constant of motion p_x, since (16.8) does not depend on x. From (16.8) $\dot{x} = p_x + y$ and p_x is invariant, so we may set $\dot{x} = y$.

Finally, we introduce the action-angle variables through the following generating function

$$F = \frac{1}{2}Y^2 \cot \phi_1 + X\phi_2. \tag{16.12}$$

Thus, (cf **11.9**)

$$P_x = \frac{\partial F}{\partial X},\ P_y = \frac{\partial F}{\partial Y},\ I_1 = -\frac{\partial F}{\partial \phi_1},\ I_2 = -\frac{\partial F}{\partial \phi_2} \tag{16.13}$$

[183] $t \to \Omega t, x \to kx, p \to pk/m\Omega$, where $\Omega = qB_0/m$.

or

$$P_x = \phi_2,\ P_y = Y\cot\phi_1,\ I_1 = Y^2/\sin^2\phi_1,\ I_2 = -X. \tag{16.14}$$

Thus, we get

$$P_x = \phi_2,\ X = -I_2,\ P_y = (2I_1)^{1/2}\cos\phi_1,\ Y = (2I_1)^{1/2}\sin\phi_1. \tag{16.15}$$

The Hamiltonian reads

$$K = I_1 + vI_2 - \alpha\sin[(2I_1)^{1/2}\sin\phi_1 - \phi_2]. \tag{16.16}$$

Depending on the amplitude of the perturbation α, we see what can happen in Fig. 16.5.

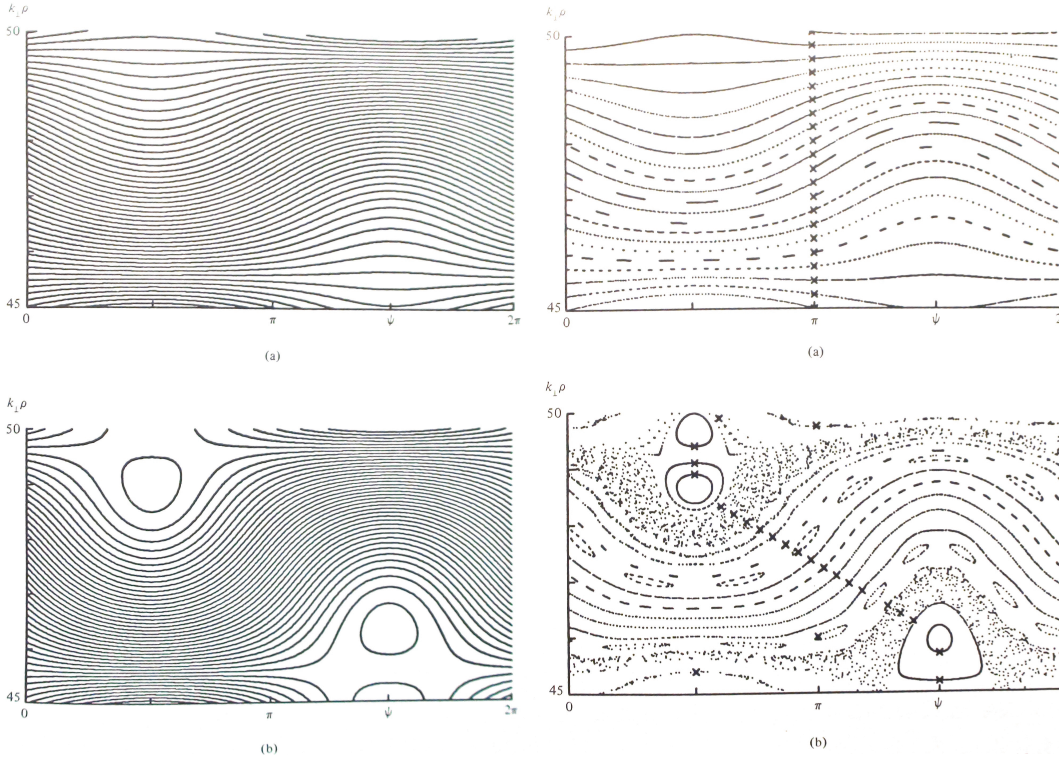


Figure 16.5: [Fig. 2.4, 2.5 of LL p83,4; originals from Karney's thesis]

As can be seen from the figure, if there is no resonance or near resonance in the system analytic results are reliable, but the analytically obtained invariant curves are actually non-existence globally. This causes the ions to heat up.

16.5 Hamiltonian systems and area preserving maps

The Poincaré map of a Hamiltonian system is area preserving. This is due to the invariance of the integral of PdQ under a canonical transformation (see **13.11**), esp., under dynamics. Therefore, it is highly interesting to study an area preserving map from $\mathbb{R}^2$ into itself.

16.6 Twist map

If the system is integrable, the trajectories are on T^2, so the map reads

$$\theta_{n+1} = \theta_n + 2\pi\alpha(J), \tag{16.17}$$

where J is the action variable specifying a torus (integral of motion). The map generally twists the cross section because of $\alpha(J)$.

16.7 General observation about perturbation results

The twist map need not be in terms of J and θ. We can use the usual xy-coordinates as

$$x_{n+1} = x_n \cos\psi - y_n \sin\psi, \tag{16.18}$$
$$y_{n+1} = x_n \sin\psi + y_n \cos\psi, \tag{16.19}$$

where $\psi = 2\pi\alpha$, α being the rotation number. If $\alpha = r/s$, where $r, s \in \mathbb{N}$ and r and s are mutually incongruent, then any point on the circle $\alpha(J) = r/s$ is on a periodic orbit with period s.

If α is irrational, then the circle is filled with a dense orbit.

The KAM-type theorem tells us that the orbits with the rotation number 'irrational enough' survive, but rational orbits are destroyed.

16.8 Poincaré-Birkhoff's theorem

If, in **16.7**, $\alpha = r/s$, after perturbation, most point on the circle is no more periodic, but there still remain even multiple of s fixed points (that is, the s-periodic orbit bifurcates into a period $2ns$ for some $n \in \mathbb{N}^+$) orbit (See Fig. 16.6).

[Demo] We assume the original unperturbed circle is bounded from inside and from outside by KAM curves (surfaces). Between these two KAM curves lies the rational circle for $\alpha = r/s$. This curve is deformed by perturbation, but after s iterations the angular coordinate is unchanged, so it can deform in the radial direction. However, the map is area preserving (**16.5**), so the deformed curve must have even crossings with the unperturbed circle.

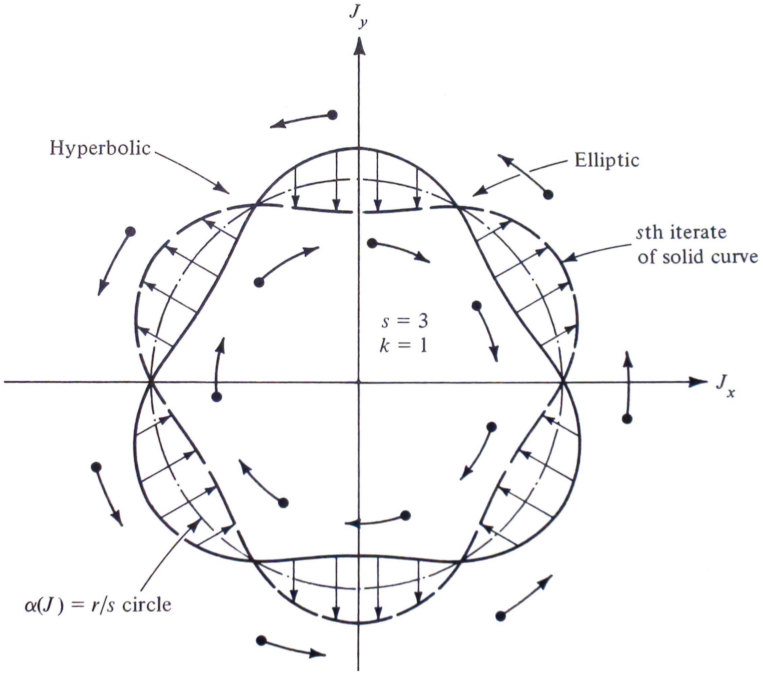


Figure 16.6: For the case $s = 3$ six (6) fixed points could exist after perturbation. [Fig. 3.3 of LL p169]

16.9 Newly formed elliptic and hyperbolic fixed points

As can be seen from Fig. 16.6, a multiple of s of new hyperbolic and elliptic fixed points are formed.

We have ns chain of hyperbolic fixed points. If integrable, we have heteroclinic orbits connecting them, and the remaining orbits are just 'laminar.' as we see for harmonic oscillators.

However, generally, we cannot avoid heteroclinic crossing, causing horseshoes as illustrated in Fig. 16.7.

An actual illustration is furnished by Henon's quadratic twist map:

$$x_{n+1} = x_n \cos\psi - (y_n - x_n^2)\sin\psi, \tag{16.20}$$

$$y_{n+1} = x_n \sin\psi + (y_n - x_n^2)\cos\psi. \tag{16.21}$$

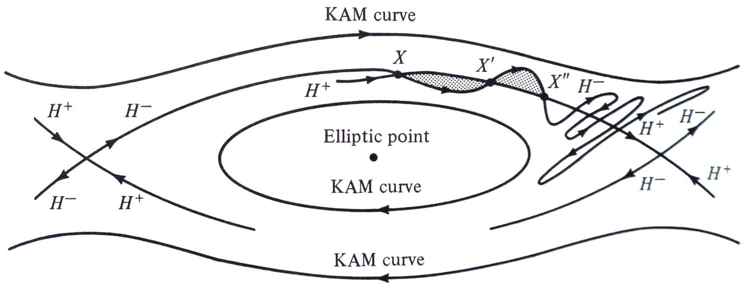


Figure 16.7: Generally we cannot avoid heteroclinic orbits, leading to horseshoe dynamics. [Fig. 3.4a of LL p171]

In Fig. 16.8 $\psi = 2\pi\alpha$ with $\alpha = 0.2114$. The first island or heteroclinic chain is exhibited.

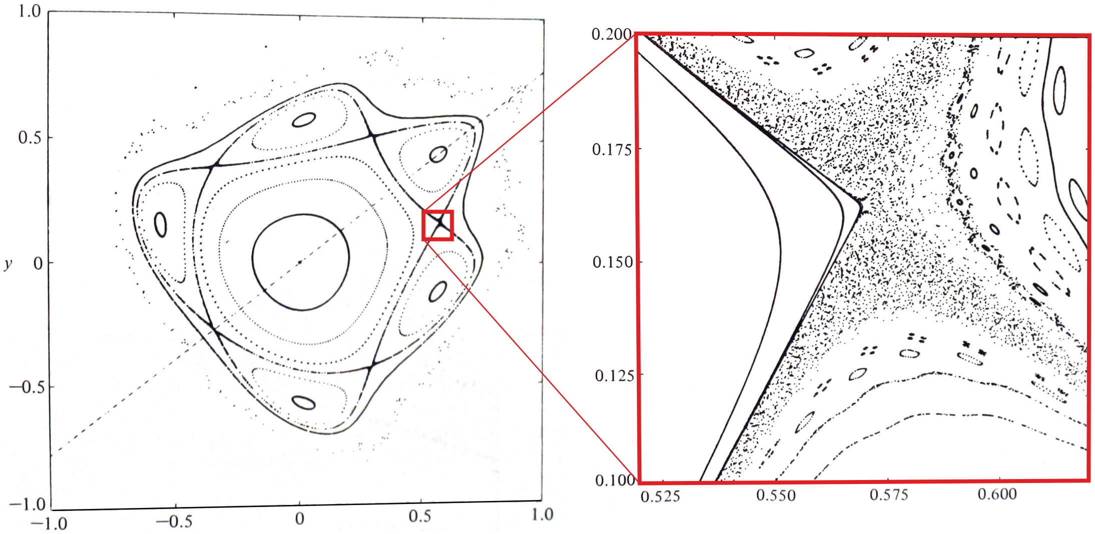


Figure 16.8: Henon quadratic twist map [Fig. 3.6 of LL]

16.10 Standard map or Chirikov-Taylor map

If a completely integrable system is perturbed as

$$H(J, \theta) = H_0(J) + \varepsilon H_1(J, \theta). \tag{16.22}$$

The Poincaré map corresponding to (16.17) must have the following form:

$$J_{n+1} \quad = \quad J_n + \varepsilon f(J_n, \theta_n), \tag{16.23}$$

$$\theta_{n+1} = \theta_n + 2\pi\alpha(J_n) + \varepsilon g(J_n, \theta_n). \tag{16.24}$$

For many interesting cases f does not depend of J and g may be ignored, if J_n in α is replaced by J_{n+1}:

$$J_{n+1} = J_n + \varepsilon f(\theta_n), \tag{16.25}$$

$$\theta_{n+1} = \theta_n + 2\pi\alpha(J_{n+1}). \tag{16.26}$$

Introducing $2\pi\alpha(J) = I$, assuming that the torus does not warp too much and then expanding $2\pi\alpha(J_n + \varepsilon f)$ around $\varepsilon = 0$, ignoring the J dependence of the derivative, we have a linearized equation:

$$I_{n+1} = I_n + Kf(\theta_n), \tag{16.27}$$

$$\theta_{n+1} = \theta_n + I_{n+1}, \tag{16.28}$$

where K is called the stochastic parameter. If we assume further $f(\theta) = \sin\theta$, we get the standard map (or the Chrikov-Taylor[184] map)

$$I_{n+1} = I_n + K\sin\theta_n, \tag{16.29}$$

$$\theta_{n+1} = \theta_n + I_{n+1}, \tag{16.30}$$

This may be interpreted as a simplified model of particle accelerator by a localized oscillating electric field.[185]

K dependence overview:
https://www.youtube.com/watch?v=PgBzZ6CcyPY

16.11 Fermi acceleration problem

The problem of a ball bouncing between a fixed and an oscillating wall is called the Fermi acceleration problem (Fog. 16.9L), because it was first examined by Fermi as a possible mechanism of accelerating cosmic ray particles.

The original moving wall problem can be solved completely, but the problem can be simplified without spoiling its essence by assuming that the wall does not move but

[184] Chirikov (see Phys. Rep. 50 263 (1979)) and Greene used this from to stud the transition to chaos.

[185] $K = (\sqrt{5} - 1)/2$ gives the last KAM surface.

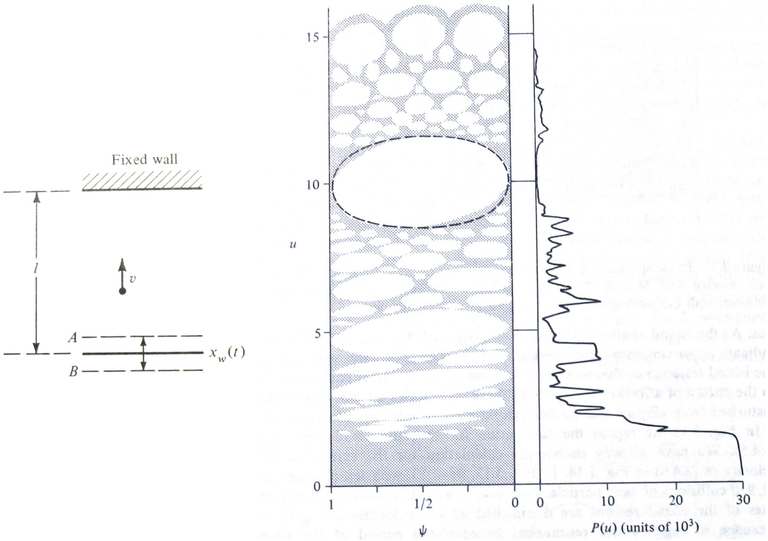


Figure 16.9: Left: Fermi acceleration problem; Right: Phase space (ψ, u) and velocity distribution $P(u)$. $M = 10$ in (16.32) with 7 typical initial conditions. The broken ellipse is the result of secular perturbation theory. [Fig. 3.11a, 3.12 of LL]

somehow gives a kick to the ball. The successive normalized ball speed u_n and the wall phase ψ_n (Sawtooth-like oscillation assumed) obey

$$u_{n+1} = |u_n + \psi_n - 0.5|. \tag{16.31}$$

$$\psi_{n+1} = \psi_n + M/u_n \ (\mathrm{mod}\ 1). \tag{16.32}$$

where $M = l/16a$ and $u = MTv/2l$, where l is the wall space, a the oscillation amplitude, T the wall oscillation period in the original problem before simplification. The absolute sign corresponds to the velocity reversal due to reflection. The phase portrait is in Fig. 16.9R.[186]

Saw-tooth standard map:
http://demonstrations.wolfram.com/TheSawtoothStandardMap/

16.12 Fermi problem and standard map

If the wall moves sinusoidally, then (16.32) now becomes the following form (the period of the wall is now scaled to 2π)

$$u_{n+1} = |u_n + \sin\psi_n|, \tag{16.33}$$

[186] M A Lieberman and A J Lichtenberg, Stochastic and adiabatic behavior of particles accelerated by periodic forces, Phys Rev A 5 1872 (1978).

$$\psi_{n+1} = \psi_n + 2\pi M/u_n. \tag{16.34}$$

Then, we linearize the above equation near the fixed point. Such a fixed point is given by $2\pi M/u_1$ being 2π times positive integers. Introduce $\Delta u_n = u_n - u_1$. The second equation of (16.34) becomes

$$\psi_{n+1} = \psi_n + \frac{2\pi M}{u_1} - \frac{2\pi M}{u_1^2}\Delta u_n. \tag{16.35}$$

Here, $2\pi M/u_1$ may be ignored. $\theta_n = \psi_n - \pi$ is used to rewrite the equations as

$$\Delta u_{n+1} = \Delta u_n - \sin\theta_n, \tag{16.36}$$

$$\theta_{n+1} = \theta_n - \frac{2\pi M}{u_1^2}\Delta u_n. \tag{16.37}$$

Then, introduce $I_n = -2\pi M \Delta u_n/u_1^2$, and setting $K = 2\pi M/u_1^2$, we have arrived at the standard map:

$$I_{n+1} = I_n + K\sin\theta_n, \tag{16.38}$$

$$\theta_{n+1} = \theta_n + I_n. \tag{16.39}$$

As can be seen from the derivation, the standard map with various K can mimic the original system around various fixed points as illustrated in Fig. 16.10:

16.13 Arnold diffusion

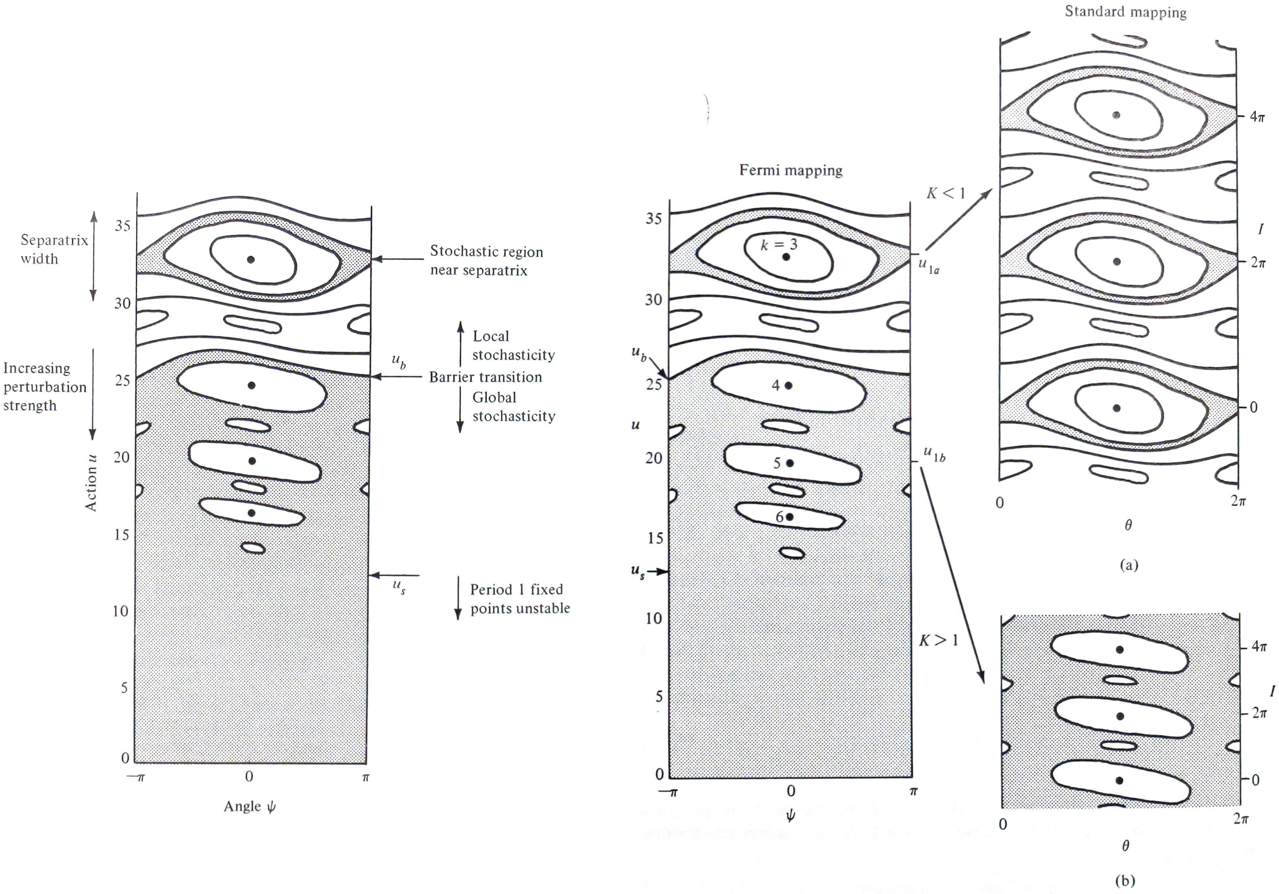


Figure 16.10: Fermi model-standard map corresponodence [Fig. 4,1, 2 of LL]

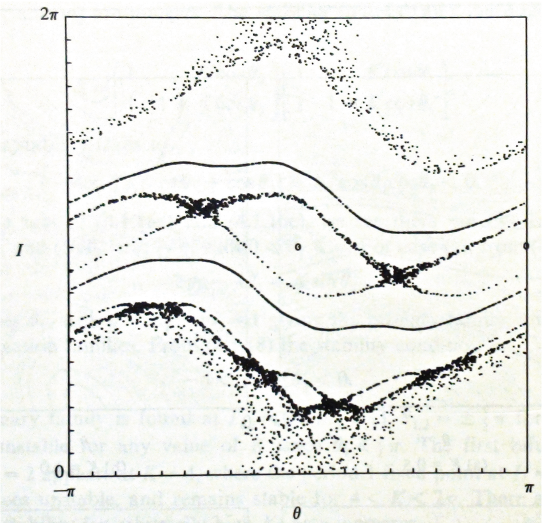


Figure 16.11: Arnold diffusion can be seen away from the fixed point [Fig. 4.4 of LL]

17 Lecture 17. Billiards

N. Chernov and R. Markarian, *Chaotic Billiards* (Mathematical Surveys and Monograph 127, AMS, 2006).

17.1 Billiard

Let Q be a compact subset of E^d (d-Euclidean space)[187] with piecewise smooth boundary. The time evolution T^t of a particle traveling according to the following rules is called a billiard in Q:
(1) A single particle moves with speed 1 along the geodesic (= straight line) if it is away from ∂Q.
(2) When the particle hits the boundary, the specular reflection is assumed.
Technically, we assume that at almost all points in Q the particle hits ∂Q within a finite time except for zero measure direction subsets of S^{d-1}.[188]

The phase space of the system is $M = \{(q,v)\,|\,q \in Q, v \in S^{d-1}\}$. The ordinary d-Lebesgue measure × the Riemann volume of S^{d-1} is an invariant measure (= phase volume) of the billiard.

You could imagine Q as a room whose wall(s) ∂Q is made of mirrors. The dynamics we wish to study is the geometrical optics i the room.
cf. Illumination problem:

https://www.youtube.com/watch?v=xhj5er1k6GQ&frags=wn

17.2 Ambrose-Kakutani representation

Since the particle of a billiard travels at speed 1 between collisions with ∂Q, we can describe the dynamics of the particle with the initial condition $x = (q,n)$ ($q \in \partial Q$ and $n \in S^{d-1}$) as a sequence of collisions with ∂Q. Thus we can describe the dynamics by a map T from ∂M ($= \partial Q \times S^{d-1}$) into itself and the needed time for the travel between successive collisions: $\tau(x)$. Here, T is defined by $x_{k+1} = Tx_k$, where $x_k = (q_k, n_k)$ ($q_k \in \partial Q$, $n_k \in S^{d-1}$) is the position and the velocity immediately after

[187] A mathematically more general definition uses d-Riemannian manifold, instead.

[188] ∂Q can contain non-differentiable subsets (called the singular components of the boundary), but we ignore them. We will not define dynamics if the particle hits these subsets.

the k-th collision (see Fig. 17.1).

Notice that this representation corresponds to the suspension of the Poincaré map.

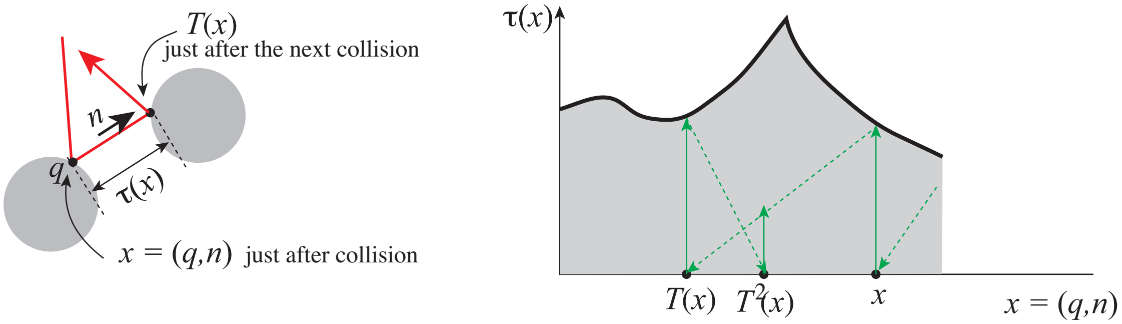


Figure 17.1: Ambrose-Kakutani representation

17.3 Polygonal or polyhedral billiards

If Q is a polygon, then we can consider a map $\sigma : S^1 \to S^1$, the directional portion of the Ambrose-Kakutani representation of the billiard. We can define a reflection at the i-th boundary component (edge or surface hyperplane) as a map $\sigma_i : S^{d-1} \to S^{d-1}$. n collisions may be expressed as a map $\sigma_{i_n}\sigma_{i_{n-1}} \cdots \sigma_{i_2}\sigma_{i_1}$. The totality of such chains define a subgroup G_Q of the isometry of S^{d-1}. If this is a finite group, the system is not ergodic. For example, if Q is a triangle and all the angles are commensurate, then G_Q is finite.[189] If G_Q is finite, the billiard is not ergodic (because S^{d-1} will never be densely traversed).

17.4 Billiards in polyhedra cannot be chaotic

If Q is a polyhedron, there cannot be any exponential spreading of the trajectories, so it cannot be chaotic (its Kolmogorov-Sinai entropy is zero as we will see later).

For an arbitrary polyhedron (or even polygon) the ergodicity of billiards is not generally understood.
Periodic orbits?: https://www.youtube.com/watch?v=AGX0cLbHaog

17.5 Square billiard

Suppose Q is a square (or a flat T^2). Then, we can visualize the motion with a line

[189] An analogous assertion holds for any polygon.

on the universal covering space (Fig. 17.2)

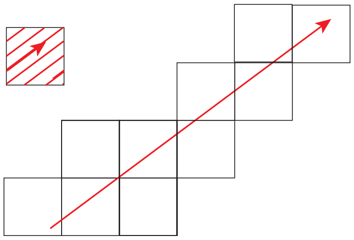


Figure 17.2: A trajectory of a square billiard on the universal covering space

Thus, if the slope is rational, the trajectory is periodic, but if irrational, the trajectory densely cover Q (Weyl's theorem), BUT it is not ergodic, since the velocities make a finite point set.

17.6 Billiards with smooth ∂Q

If Q is a 2-disk (thus, $\partial Q = S^1$), then all the successive reflection points are equally spaced and the angle of reflection is constant. There is a caustic (Fig. 17.3): a caustics is a smooth closed curve γ such that if one segment of the trajectory is tangent to γ all other segments of the same trajectory are also tangent to γ.

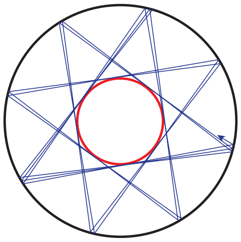


Figure 17.3: A caustic of a circular billiard

Lazutkin showed that if ∂Q is convex and smooth enough (say, C^5), then the totality of caustics is with positive measure (and obviously the system is not ergodic).

Remark If there is a caustics, then there is Dirichlet eigenvalues for $\Delta\psi = \lambda\psi$ that are localized near the caustic (e.g., whisper modes). Shnirelman proved that ψ spreads all over Q in the $\lambda \to \infty$ limit, if the billiard is ergodic.[190]

Fun video: Elliptic pool table https://www.youtube.com/watch?v=3WHBlPvK3Ek&

[190] Bunimovich p156.

frags=wn

In general anything seems to happen:
cardioid billiard: https://www.youtube.com/watch?v=eQfh0gaU4NE

17.7 Dispersing or Sinai billiards[191]

If the boundary ∂Q is inwardly convex at its regular points, the billiard is called a dispersing (or Sinai) billiard.
How dispersive can be seen from:
https://www.youtube.com/watch?v=C1iuNH_99v8 .

Its comparison with a circular billiard is interesting: https://www.youtube.com/watch?v=dI4WuafBF-w&index=4&list=PL2wfI-9_pR8AvNGx0wpfE07fyS-giW20b&frags=wn

The simplest example is $Q = T^2 \backslash$ 2-disk (Fig. 17.4)

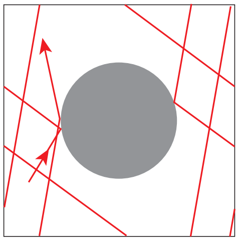


Figure 17.4: Sinai billiard table

Unfortunately, there is no good video for the simplest billiard.

Theorem. Sinai billiards are ergodic and K.

What is K? Roughly,....
There is a finite partition of its phase space $\{A_i\}$ such that almost all the points in M and their coding $x \to \{x_k\}$ according to the rule $T^n(x) \in A_i as x_n = i$ is one-to-one correspondent, the system is called a K-system.

17.8 Why Sinai billiards are chaotic

Here, we use the word 'chaotic' intuitively to mean that the spread of the velocity vectors increases exponentially in time on the average. At each reflection with ∂Q

[191] B. A. Friedman, *The Billiard Problem* (UIUC Physics Thesis, 1985) contains a friendly introduction to the topic.

spread the directions, and, generically speaking, the number of collisions increases linearly with time, so we can expect exponential spreading of the propagation directions on the average in time. Thus, we can expect the system is chaotic.

We can be more quantitative.

17.9 Local time evolution of velocity curvature by propagation

Take a smooth curve passing through $x_0 \in Q$ and assume it is the curve perpendicular to the trajectories. The curvature $\kappa(x_0) = d\varphi/dr$ describes how spread the velocities are (notice that $1/\kappa$ is the curvature radius).

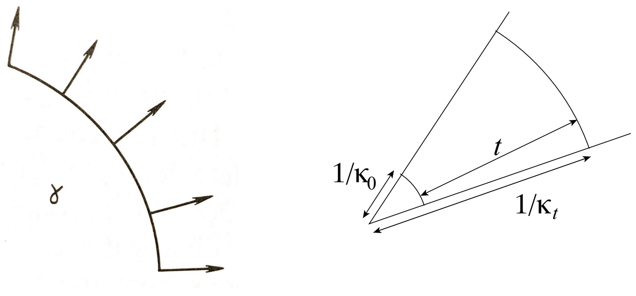


Figure 17.5: A propagation front and the velocities; here γ is a curve in M (not in Q) consisting of a curve $\tilde{\gamma}$ in Q + the unit velocities in ther normal directions

If there is no collision between time 0 and t,

$$\frac{1}{\kappa(x_t)} = t + \frac{1}{\kappa(x_0)}, \tag{17.1}$$

so

$$\kappa(x_t) = \frac{\kappa(x_0)}{1 + t\kappa(x_0)}. \tag{17.2}$$

The expansion rate of the base curve $\tilde{\gamma}$ is $1 + t\kappa(0)$.

17.10 Local time evolution of velocity curvature by collision

If there is a collision at time τ, this curvature changes discontinuously between $\tau - 0$ and $\tau + 0$:

$$\kappa(x_{\tau+0}) = \kappa(x_{\tau-0}) + \frac{2k}{\cos\varphi_\tau}, \tag{17.3}$$

where k is the curvature of the colliding surface and φ_τ is the incidence angle. This may be derived as explained below, but it is basically the well-known lens formula $1/f = 1/d_{\mathrm{o}} + 1/d_{\mathrm{i}}$. For the near optical axis rays, $\varphi = 0$ and $f = -2(1/k)$ (the twice radius of the mirror; $-$ because the mirror is convex);

$1/d_{\rm o} = \kappa(x_{\tau-0})$ (real object) and $1/d_{\rm i} = -\kappa(x_{\tau+0})$ (virtual image).

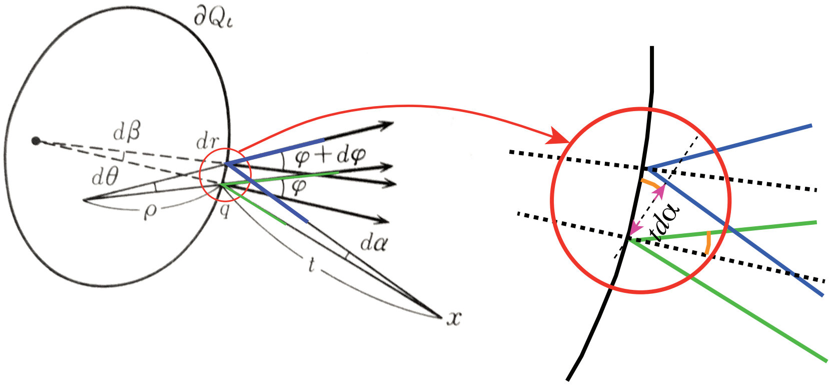


Figure 17.6: A propagation front and the velocities

In Fig. 17.6 the relation between $d\beta$ and $d\alpha$ is

$$\frac{1}{k}d\beta = dr = \frac{t(-d\alpha)}{\cos\varphi_\tau}, \tag{17.4}$$

where α is measured clockwisely while β counterclockwisely, so we need the minus sign. We see

$$d\varphi = 2d\beta - d\alpha, \tag{17.5}$$

because (i) if $d\alpha = 0$, then changing the mirror direction by $d\beta$ changes the reflection angle by $2d\beta$, (ii) if $d\beta = 0$, changing the incidence angle by $d\alpha$ the reflection angle changes by $-d\alpha$. Therefore,

$$d\varphi = 2\frac{kt(-d\alpha)}{\cos\varphi} - d\alpha. \tag{17.6}$$

We know

$$\kappa(x_{\tau-0}) = 1/t = \frac{-1}{\cos\varphi}\frac{d\alpha}{dr}, \quad \kappa(x_{\tau+0}) = \frac{1}{\cos\varphi}\frac{d\varphi}{dr}, \quad k = \frac{d\beta}{dr}. \tag{17.7}$$

Thus, (17.6) reads

$$d\varphi = 2kdr - d\alpha, . \tag{17.8}$$

so we finally obtain (17.3).

17.11 Expansion between successive collisions

The expansion rate of the curve $\tilde{\gamma}$ is, as already seen in **17.9**, $1+\tau(x_n)k(x_n)$, where $\tau(x_n)$ is time between nth and $n+1$th collisions, and $k(x_n)$ is the curvature just after the nth collision. Notice that from (17.3),

17.12 Relation to geodesics on negative curvature surfaces

The geodesic trajectories near the saddle spread, as is illustrated in

`https://www.youtube.com/watch?v=3u2SJKxJhh8&frags=wn` after about 11:20

Thus, the geodesic trajectories in the negative curvature Riemannian manifold are exponentially spreading.

Sinai billiards are without any curvature, but what happens at the collision points is just what happens at saddles as illustrated in Fig. 17.7

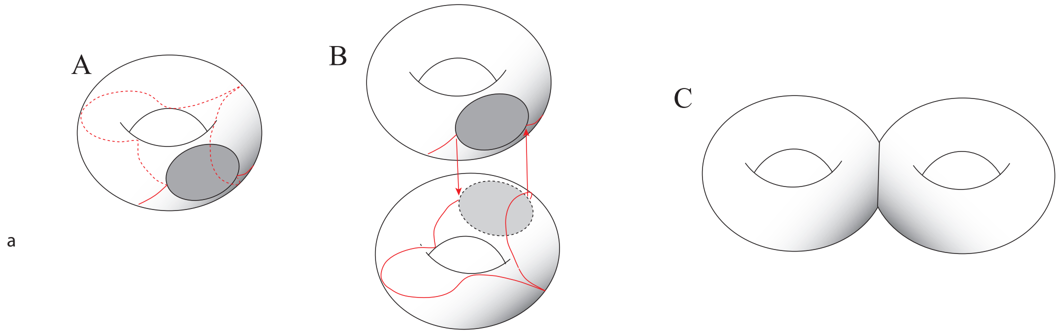


Figure 17.7: 'Saddle' dynamics near ∂Q; here the illustration uses the 'standard' Sinai billiard. A is the torus with a disk scatterer. You can imagine that the particle goes to the 'backside' of the torus upon collision. At the next collision the particle reappears to this side of the world. B: Then, take out the second surface, and connect them to make a 'coupled' T^2. The particles moves 'straight' until it hits the connection ring R; if the ring is 'mollified a bit, it is just a saddle.

17.13 Billiards with focusing elements

Bunimovich proved that billiards on $Q \subset \mathbb{R}^2$ are K, if
(1) ∂Q consists of line segments and parts of circles.
(2) There is no pair of focusing components that are a part of the same circle.
(3) For each circular segment, its completed circle must be in Q.

For example, Qs illustrated in Fig. 17.8 satisfy the above conditions:"

The most famous example is the Bunimovich stadium:

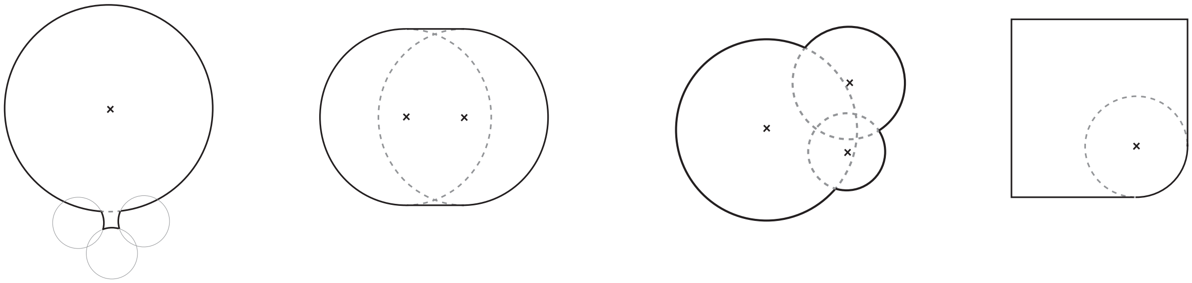


Figure 17.8: Examples of chaotic focusing billiard tables

https://blogs.ams.org/visualinsight/2016/11/15/bunimovich-stadium/.

Remark: the eigenfunction of $\Delta\psi$ on Q are illustrated as
Quantum: https://www.dhushara.com/DarkHeart/QStad/QStad.htm
Bunimovich mitosis http://community.wolfram.com/groups/-/m/t/977013
First 500 eigenmodes: https://youtu.be/3voKV4az0Sk

17.14 Invariant measure of billiards

Since billiards were motivated by Krylov to understand the foundation of statistical mechanics, their probabilistic properties have been of their main interest. There are many different invariant measures (the probability law preserved by dynamics), but the most interesting one is the ones absolutely continuous wrt the Lebesgue measures on the phase space.

Here, I do not discuss what the measure is: it is roughly something like volume. 'Absolute continuity' of a measure ν wrt μ means $\nu(A) = 0$ whenever $\mu(A) = 0$ for any (measurable) set.

From now on we discuss only billiards on 2-dimensional tables Q. The phase space is $\Gamma = Q \times S^1$, where $Q \subset \mathbb{R}^2$. Let us introduce the position coordinates (q_1, q_2) on Q and the angle variable φ to specify the particle velocity during its free time. It should be clear that

$$d\mu = \frac{1}{2\pi|Q|} dq_1 dq_2 d\varphi \tag{17.9}$$

is invariant; this is simply a classical particle system with a constant speed.

17.15 Invariant measure for Ambrose-Kakutani representation

We could introduce an invariant measure on the space shaded in Fig. 17.1 Right. The coordinates we use are q, n and σ, where σ is the vertical coordinate (expressing the travel distance between successive collisions). Here q may be the coordinate along ∂Q and n may be represented by an angle φ wrt the outward normal on ∂Q. The case of the single disk Sinai billiard on T^2 is illustrated in Fig. 17.9.

Here the small box represents dq_1dq_2. From Fig. 17.9 we see

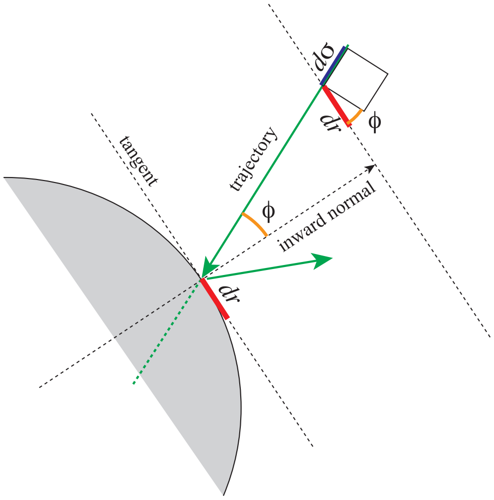


Figure 17.9: The coordinates of the Ambrose-Kakutani representation of 'the' Sinai billiard; notice that $\varphi \in [-\pi/2, \pi/2]$.

$$dq_1dq_2 = d\sigma dr \cos\varphi. \tag{17.10}$$

Therefore,

$$d\mu \propto \cos\varphi \, d\varphi d\sigma dr. \tag{17.11}$$

We must compute the normalization constant, if μ is a probability measure. For the standard Sinai billiard, the phase volume is $2\pi|Q|$ ($|Q| = 1 - \pi r^2$, if the torus area is 1 and the disc radius is r), so this is the normalization constant.

17.16 Mean free time

The average of $\tau(x)$ for $x \in \Gamma$ is the mean free time. Notice that integral of $d\sigma$ between successive collisions is the free time $\tau(q, n)$. Therefore, for the standard Sinai billiard

$$2\pi|Q| = \int_0^{|\partial Q|} dr \int_{-\pi/2}^{\pi/2} \cos\varphi \, d\varphi \int_0^{\tau(r,\varphi)} d\sigma = \int_0^{|\partial Q|} dr \int_{-\pi/2}^{\pi/2} \cos\varphi \, \tau(r,\varphi) d\varphi \tag{17.12}$$

The mean free time is[192]

$$\langle\tau\rangle = \frac{\int_0^{|\partial Q|} dr \int_{-\pi/2}^{\pi/2} |\cos\varphi|\tau(r,\varphi)d\varphi}{\int_0^{|\partial Q|} dr \int_{-\pi/2}^{\pi/2} |\cos\varphi|d\varphi} \tag{17.13}$$

The denominator is $2\partial Q$. Therefore, we have arrived at

$$\langle\tau\rangle = 2\pi|Q|/2|\partial Q| = (1-\pi r^2)/2r. \tag{17.14}$$

17.17 Abramov formula for the loss of information The Kolmogorov-Sinai entropy (information loss rate per time) h of a billiard may be expressed as

$$h = H/\langle\tau\rangle, \tag{17.15}$$

where H is the entropy change per collision, and $\langle\tau\rangle$ is the mean free time. The formula is called the Abramov formula. That is, a mean field idea works.

It is not hard to see why (17.15) holds. Notice that between collisions there is no loss of information; information is lost only at collisions. Let H_j be the loss of information at the jth collision. Then, its average is H. Let τ_j be the free time before the jth collision. Then, $T = \sum \tau_j$ is needed for the information loss of $\sum H_j$, so

$$h = \frac{\sum H_j}{\sum \tau_j} = \frac{H}{\langle\tau\rangle}. \tag{17.16}$$

17.18 Loss rate of information of billiards

Take a small square as in Fig. 17.10 Left. Suppose the length of its edge is the minimum length that we can discern. Thus we can tell whether a point is in it or not: this answer can be obtained by a single yes-no question. Thus, we get one bit of information.

After a single application of the map T (in our context T) the square is stretched in the unstable manifold direction and compressed in the stable manifold direction.

If we know that the system is in the square now, after applying T we know the system point is near the unstable manifold; more precisely, we know the system is

[192] Here, we assume that the particle can go any where on the table = ergodicity; this needs a proof.

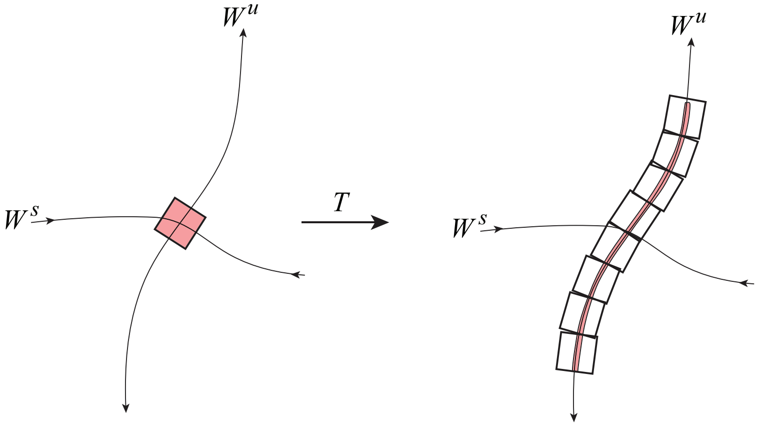


Figure 17.10: Local loss of information due to expansion

in one box which cover the stretched square as in Fig. 17.10 Right. To locate the system as accurately as one time step before, we should choose one box out of 8 boxes. We need 3 bits of information. That is, T dissipated 3 bits of information; our knowledge about the system is lost by 3 bits per one mapping step.

Thus, the log (or $\log_2$) of the expansion rate of the unstable manifold must be the information loss rate, whose 'ensemble average'[193] is called the Kolmogorov-Sinai (KS) entropy h_T, as we will see later.

17.19 Kolmogorov-Sinai entropy of Sinai billiard: introduction

Let $r = \psi(\varphi)$ be the unstable manifold. This is mapped by $T : (r, \varphi) \to (r_1, \varphi_1)$ to $\varphi_1 = \psi_1(r_1)$. Then, the expansion rate m can be computed as

$$m = \frac{\|(dr_1, d\varphi_1)\|}{\|(dr, d\varphi)\|} = \left|\frac{\partial r_1}{\partial r}\right| \frac{\|(1, \partial\varphi_1/\partial r_1)\|}{\|(1, \partial\varphi/\partial r)\|}. \tag{17.17}$$

Let us write μ as the normalized phase volume (invariant measure) (**??**) in **17.14**. Then,

$$\begin{aligned} h_T &= \int d\mu \log m && (17.18)\\ &= \int d\mu \log\left|\frac{dr_1}{dr}\right| + \int d\mu \log\|(1, \partial\varphi_1/\partial r_1)\| - \int d\mu \log\|(1, \partial\varphi/\partial r)\| && (17.19) \end{aligned}$$

[193] average over the invariant probability measure.

$$= \int d\mu \log \left| \frac{dr_1}{dr} \right| . \tag{17.20}$$

We must calculate the derivative: note, however, dr_1/dr is the derivative along the unstable manifold (i.e., φ also changes, when r is changed according to $d\varphi = (d\psi/dr)dr$.

To calculate h_T, we need much more details of the Sinai billiard. See calculations up to **17.19**.

17.20 Domain of *T*

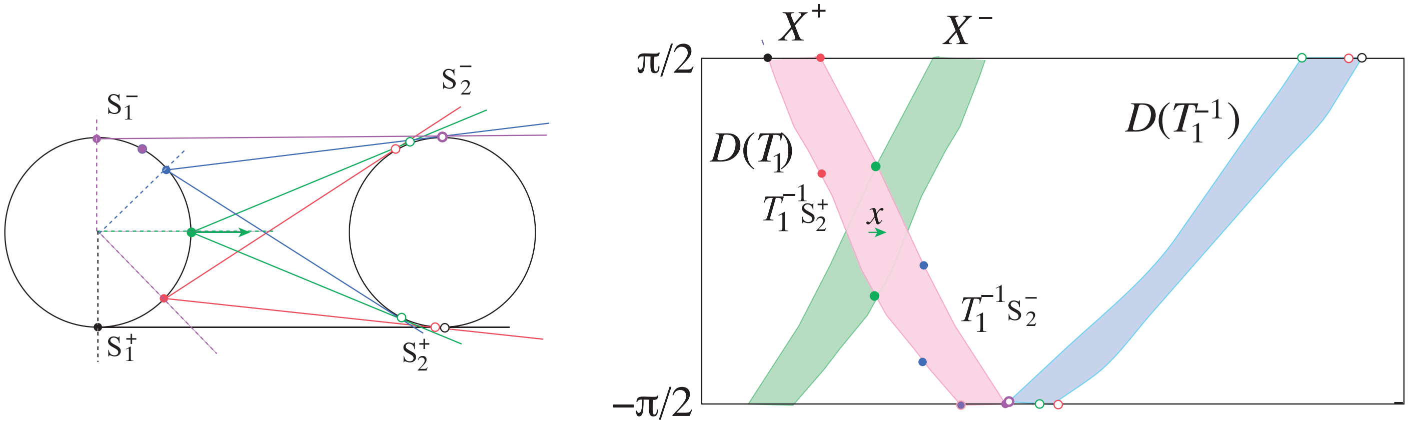


Figure 17.11: The 'Poincaré' map T for the Sinai billiard.

Fig. 17.11 The 'Poincaré' map $T$for the Sinai billiard.
Right: The rectangle shows the boundary ∂M of the phase space (i.e., ∂Q and the velocity vectors (just after collisions into Q direction) there in S^1. The pink band is the domain for T hitting '2' (from '1', '1' and '2' may be the same disk (different sides) shown Left; the colored dots correspond to the starting sites on '1'.
The purple strip is the domain of T^{-1} coming back to '1' from '2'.
The green strip is the domain of T^{-1} from '1', which is definable by the pink strip with the mirroring its velocities with respect to the normal directions (time reversal transformation).
The green arrow x denotes point in ∂M indicated by the green dot ($\in \partial Q$) and the direction denoted by the green arrow on Left.

The invariant curve going through x in Fig. 17.11

17.21 How T transforms curves in ∂Q

Consider a curve $\varphi = \psi(r)$ in ∂M to understand how T maps this. After T let us write this curve becomes $\varphi_1 = \psi_1(r_1)$. We can write

$$\varphi_1(r, \varphi) = \psi_1(r_1(r, \varphi)), \tag{17.21}$$

Its derivatives are obtained easily with the chain rule and the results summarized in **17.25** must lie in the strips. One in the pink strip must be the unstable manifold of T through x (i.e., $W^+(x)$) and that in the green strip the stable manifold of x.

17.22 Unstable manifold of T

In contrast to the smooth dynamical systems, billiards are riddled with breaking due to collisions.[194] There is a transversal stable manifold that looks similar.

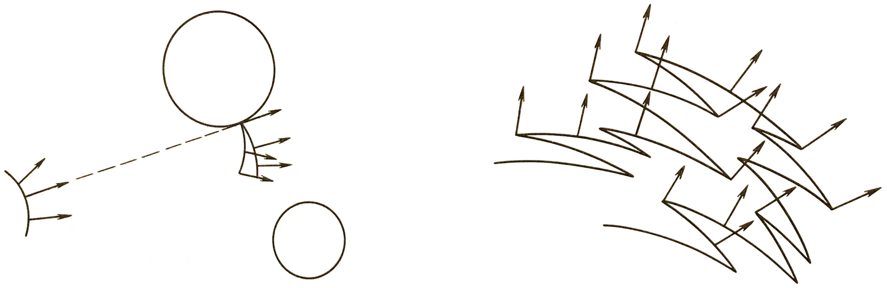


Figure 17.12: Unstable manifold of a Sinai billiard; the left figure explains why there are break points. [Fig. 7, 8 of DS II Bunimovich]

17.23 Change of the next collision by position change

Let us calculate one step $x \to x_1$ in detail.

First, let us study the changes in r_1 and φ_1 when r is changed by dr (Fig. 17.13).

If we keep φ and change the first collision position slightly by dr, due to the change of the normal directions of the reflecting surface, the direction of the reflected

[194] Actually, it is usually made of many continuous components. Here, only one such component is illustrated.

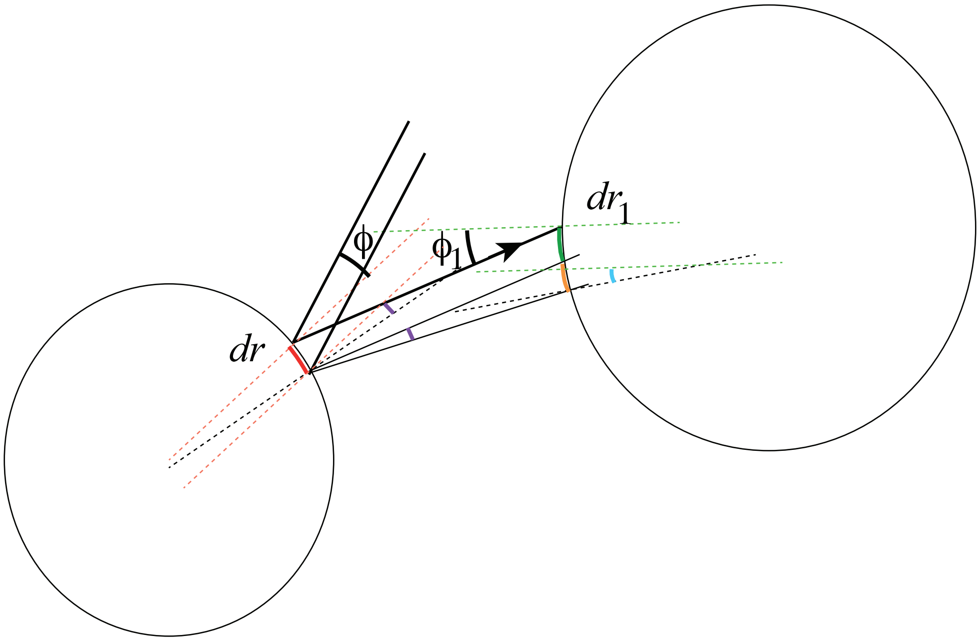


Figure 17.13: A propagation front and the velocities: change due to r

trajectory changes its angle by kdr (the purple angle). It takes τ fo the particle to reach the next collision point from the last collision, so

$$kdr\tau, \tag{17.22}$$

where k is the curvature of the boundary at the first collision point, is the displacement of the collision point perpendicular to the incidence direction. However, this direction makes angle φ_1 with the tangent direction of the new colliding surface. Thus, we need the length of te orange segment, which is

$$kdr\tau/\cos\varphi_1. \tag{17.23}$$

dr_1 has the green portion due to ther simple parallel displacement of the trajectory due to dr. The displacement distance is $dr\cos\varphi$. Thus, when this is projected on the second surface the displacement must be

$$dr\frac{\cos\varphi}{\cos\varphi_1}. \tag{17.24}$$

Thus the answer for dr_1 is the green and orange segments added, but we must worry about the sign. If dr is in the counterclockwise direction, dr_1 must be clockwise, so they should have opposite signs:

$$dr_1 = -dr\left(\frac{\cos\varphi}{\cos\varphi_1} + \frac{\tau k}{\cos\varphi_1}\right). \tag{17.25}$$

The change of the incidence angle to the second body consists of two parts: the blue angle due to the change of the normal direction due to displacement dr_1 and the purple angle due to the change of the direction of the incidence ray (due to dr on the first body). The former is $k_1 dr_1$, where k_1 is the curvature at the new collision point. The purple portion is obviously kdr, which reduces φ_1. Therefore,

$$d\varphi_1 = -k_1 dr \left(\frac{\cos\varphi}{\cos\varphi_1} + \frac{\tau k}{\cos\varphi_1} \right) - kdr. \tag{17.26}$$

We also should consider the change in τ. This is clearly

$$d\tau = dr_1 \sin\varphi_1 - dr \sin\varphi. \tag{17.27}$$

17.24 Change of the next collision by angle change

Next, let us study the changes in r_1 and φ_1 when φ is changed by $d\varphi$ (Fig. 17.14).

If we keep r and change the first collision incidence angle slightly by $d\varphi$, the direction of the reflected trajectory changes with the same amount but in the opposite direction. It takes τ fo the particle to reach the next collision point from the last collision, so the particle to reach the next collision point from the last collision, so $d\varphi\tau$ is the displacement of the collision point perpendicular to the incidence direction. However, this direction makes angle φ_1 with the tangent direction of the new colliding surface. Thus, the green length $= dr_1$ is

$$dr_1 = -\tau d\varphi / \cos\varphi_1. \tag{17.28}$$

The incidence angle to the new boundary changes by the red angle due to the change of the direction of the trajectory, but the collision point is displaced by dr_1, so the normal direction changes by the orange angle: $k_1 dr_1$. Both contribute in the same direction, so

$$d\varphi_1 = \frac{k_1\tau}{\cos\varphi_1} d\varphi - d\varphi = -d\varphi \left(1 + \frac{k_1\tau}{\cos\varphi_1} \right). \tag{17.29}$$

We also have

$$d\tau = \tan\varphi_1 d\varphi. \tag{17.30}$$

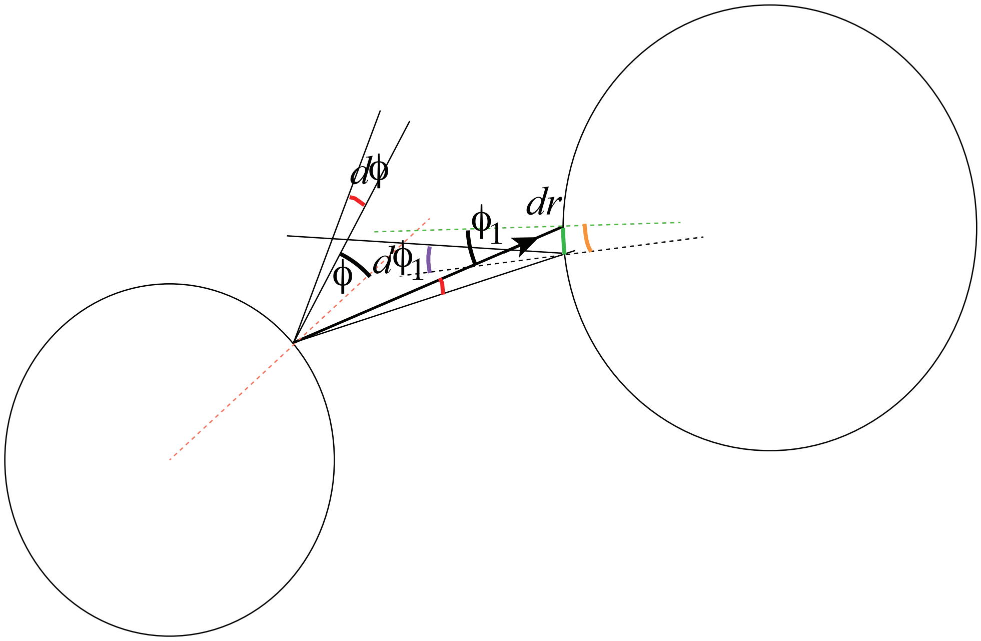


Figure 17.14: A propagation front and the velocities: change due to r

17.25 Summary of derivatives

We may summarize the results of **17.23** and **17.24** as follows:

$$\begin{pmatrix} \frac{\partial r_1}{\partial r} & \frac{\partial r_1}{\partial \varphi} \\ \frac{\partial \varphi_1}{\partial r} & \frac{\partial \varphi_1}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} -\left[1+\frac{\tau k}{\cos\varphi}\right]\frac{\cos\varphi}{\cos\varphi_1} & -\frac{\tau}{\cos\varphi_1} \\ -k_1\left[1+\frac{\tau k}{\cos\varphi}\right]\frac{\cos\varphi}{\cos\varphi_1}+k & -\left[1+\frac{\tau k_1}{\cos\varphi_1}\right] \end{pmatrix}. \tag{17.31}$$

17.26 Kolmogorov-Sinai entropy of Sinai billiard: calculation

To compute h_T we need $\partial r_1/\partial r$. We differentiate $r_1 = r_1(r, \psi(r))$ (recall the remark at the end of **17.19**. Since $r_1 = r_1(r, \psi(r))$:

$$\frac{dr_1}{dr} = \frac{\partial r_1}{\partial r} + \frac{\partial r_1}{\partial \varphi}\frac{d\psi}{dr}. \tag{17.32}$$

Also, since

$$\varphi_1(r, \psi(r)) = \psi_1(r_1(r, \psi(r))), \tag{17.33}$$

$$\frac{\partial \varphi_1}{\partial r} + \frac{\partial \varphi_1}{\partial \psi}\frac{d\psi}{dr} = \frac{d\psi_1}{dr_1}\left(\frac{\partial r_1}{\partial r} + \frac{\partial r_1}{\partial \psi}\frac{d\psi}{dr}\right). \tag{17.34}$$

Therefore, we can solve

$$\frac{d\psi_1}{dr_1} = \frac{\frac{\partial \varphi_1}{\partial r} + \frac{\partial \varphi_1}{\partial \varphi}\frac{d\psi}{dr}}{\frac{\partial r_1}{\partial r} + \frac{\partial r_1}{\partial \varphi}\frac{d\psi}{dr}}. \tag{17.35}$$

(17.32) reads with (17.25) and (17.28)

$$\begin{aligned} \frac{dr_1}{dr} &= -\left(\frac{\cos\varphi}{\cos\varphi_1} + \frac{\tau k}{\cos\varphi_1}\right) + \frac{\tau}{\cos\varphi_1}\frac{d\psi}{dr} \qquad (17.36)\\ &= -\frac{\cos\varphi}{\cos\varphi_1}\left(1 + \frac{\tau}{\cos\varphi}\left(k + \frac{d\psi}{dr}\right)\right). \qquad (17.37) \end{aligned}$$

We must compute $d\psi/dr$.

$$\begin{aligned} \frac{d\psi_1}{dr_1} &= \frac{-k_1\left(\frac{\cos\varphi}{\cos\varphi_1} + \frac{\tau k}{\cos\varphi_1}\right) - k - \left(1 + \frac{k_1\tau}{\cos\varphi_1}\right)\frac{d\psi}{dr}}{-\left(\frac{\cos\varphi}{\cos\varphi_1} + \frac{\tau k}{\cos\varphi_1}\right) - \frac{\tau}{\cos\varphi_1}\frac{d\psi}{dr}} \qquad (17.38)\\ &= k_1 + \frac{k + \frac{d\psi}{dr}}{\left(\frac{\cos\varphi}{\cos\varphi_1} + \frac{\tau k}{\cos\varphi_1}\right) + \frac{\tau}{\cos\varphi_1}\frac{d\psi}{dr}} \qquad (17.39)\\ &= k_1 + \frac{k + \frac{d\psi}{dr}}{\frac{\cos\varphi}{\cos\varphi_1} + \frac{\tau}{\cos\varphi_1}\left(k + \frac{d\psi}{dr}\right)} \qquad (17.40)\\ &= k_1 + \frac{\cos\varphi_1}{\cos\varphi}\frac{k + \frac{d\psi}{dr}}{1 + \frac{\tau}{\cos\varphi}\left(k + \frac{d\psi}{dr}\right)} \qquad (17.41)\\ &= k_1 + \frac{\cos\varphi_1}{\cos\varphi}\frac{1}{\frac{\tau}{\cos\varphi} + \frac{1}{k + \frac{d\psi}{dr}}}. \qquad (17.42) \end{aligned}$$

Notice that this is a recurrence relation:

$$\frac{d\psi_n}{dr_n} = k_n + \frac{\cos\psi_n}{\cos\psi_{n-1}}\frac{1}{\frac{\tau_{n-1}}{\cos\psi_{n-1}} + \frac{1}{k_{n-1} + \frac{d\psi_{n-1}}{dr_{n-1}}}}. \tag{17.43}$$

The result of this recursion gives a continued fraction expression of $d\psi/dr$ required in (17.32). We can fairly accurately evaluate this numerically.

17.27 H in the small R limit

Friedman et al.[195] actually evaluated and conjectured $-H/\log R \to 2$ in the small R limit. Actually, they proved this value to be in $(1.5, 2]$ and numerically obtained the upper limit value 2.[196]

17.28 Billiards with scatterers with finite potentials or soft billiards

Baldwin[197] considered an 'optical' system or the muffin-tin potential system with potential U (measured in kinetic energy):

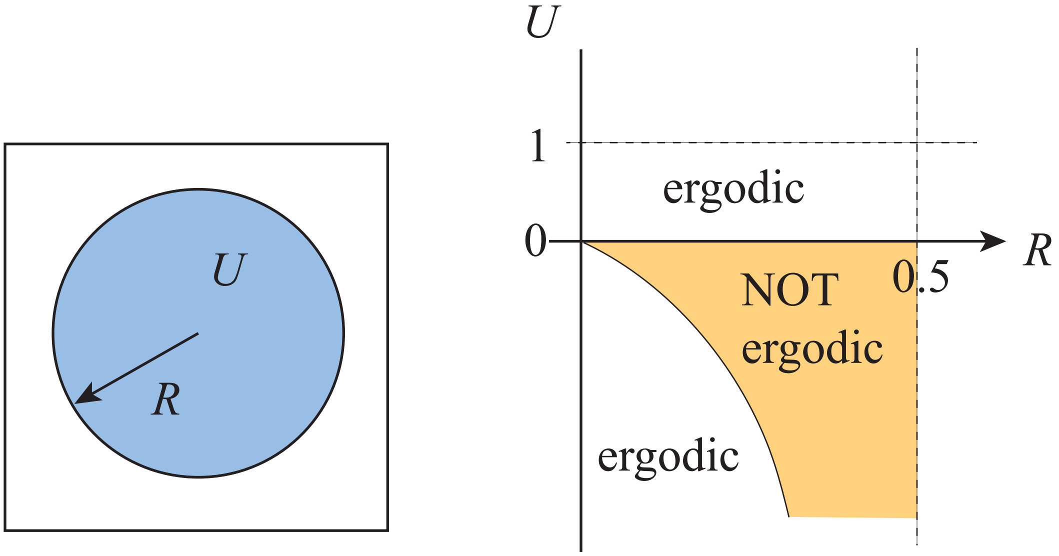


Figure 17.15: The table with U and the phase diagram; the non-ergodicity of the colored region is proved, but ergodicity of the remaining regions is a conjecture based on simulations. The lower bondary curve of the colored region is $U = 1 - 1/(1 - 2R)^2$.

In the nonergodic region we see elliptic fixed points.

[195] B. A. Friedman, Y. Oono and I. Kubo, PRL 52 709 (1984). See B. A. Friedman, UIUC Physics thesis 1985.

[196] $d(d-1)$ is the general formula: N Chernov, Entropy Values and Entropy Bounds (2006) is a good review): people.cas.uab.edu/~mosya/papers/hb.pdf.

[197] P. R. Baldwin, Soft billiard system, Physica D 29, 321 (1988). See P. R. Baldwin, UIUC Physics thesis 1987.

18 Lecture 18. Introductory examples of chaos

18.1 K. Ito's model of great earthquakes

Island arcs such as the Japanese Archipelago consist of blocks spaced by faults. Each block is being pulled down by a subducting ocean plate and accumulates strain energy (Fig. 18.1). When this energy reaches a threshold, the block breaks (a great earthquake occurs) and the stored strain energy is quickly reset to its lowest value; the model chooses this value as the energy origin.

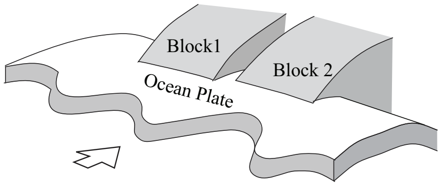


Figure 18.1: The great earthquake model. The subducting ocean plate pulls two blocks down. The blocks accumulate strain energy and at some point they break apart from the plate and earthquakes occur. Then, the blocks return to the original positions.

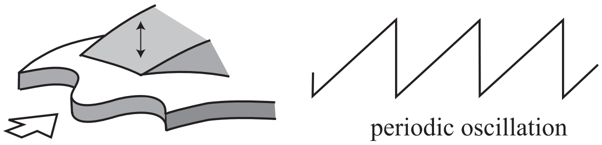


Figure 18.2: If there is only one block, it is a simple model of the relaxation oscillator.

If there is no interaction among blocks, earthquakes occur periodically (Fig. 18.2). This is probably the simplest model of the relaxation oscillator. See a traditional example `https://www.youtube.com/watch?v=JUOmNSRtHMM&frags=pl%2Cwn`. Periodic earth quakes do exist.[198]

If there are more than one blocks, the earthquake occurring in one block must affect the surrounding blocks. Ito thought that the earthquake produces cracks in the neighboring blocks that decelerate the accumulation rate of the strain energy.

The simplest nontrivial case is the two-block case. The following rules of time evolution can model the idea depicted above. Let u_i be the strain energy stored in block i ($i = 1, 2$).

(1) The rate of increase of the strain energy is initially b, which is assumed to be a positive number.

[198] Kelvin R. Berryman et al. Major Earthquakes Occur Regularly on an Isolated Plate Boundary Fault, Science 336 1690 (2012.

(2) If u_i reaches 1, an earthquake occurs in block i. Subsequently,

(2a) u_i is reset to 0, and the rate of increase of strain energy is also reset to b.

(2b) The rate of increase of the strain energy in the other block without the occurrence of the earthquake becomes b^{-1}; if the rate is already with this value, it is maintained as b^{-1}.

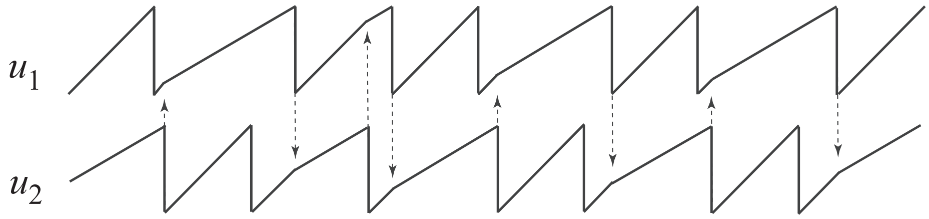


Figure 18.3: A typical behavior of stored energies u_1 and u_2 in two coupled blocks 1 and 2, respectively. This is the case of mutual hindrance $b > 1$. The earthquake in one block decelerates the energy increase in the other block. Vertical arrows denote hindering effects on strain energy accumulation. (Except for the beginning stage of the system behavior, the rates of energy increase of the blocks never coincide.)

A typical behavior under the above rule is in Fig. 18.3. If the rate of increase of the strain energy $b > 1$, we see apparently complicated behavior.

18.2 Trajectories on T^2

The time evolution of the two blocks can be illustrated as a trajectory on $T^2 = [0,1] \times [0,1]$ (with periodic boundary conditions), if we plot u_1 on the horizontal axis and u_2 at the same moment on the vertical axis (Fig. 18.4).

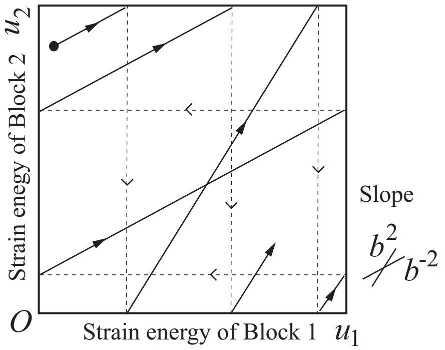


Figure 18.4: The typical time dependence of the stored energy in the blocks may be depicted as a trajectory on $[0,1] \times [0,1]$. Horizontal jumps correspond to the earthquakes in block 1 and the vertical jumps those in block 2.

The rule of the game may be illustrated as in Fig. 18.5. We can write the rules in formulas, but no new insight is obtained by doing so. For example, a portion of the trajectory with slope less than 1 on the square implies that the energy increase rate in block 1 is larger than that for block 2.

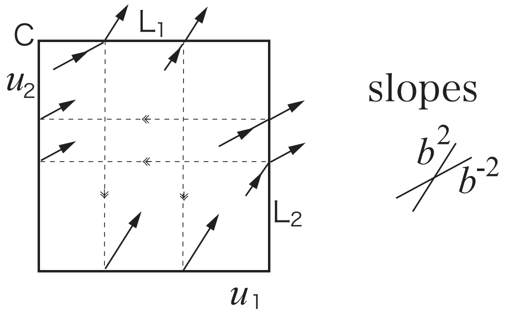


Figure 18.5: The rules for the motion: after a jump how the slope changes is specified. If the point reaches L_1 (the top edge; an earthquake in block 2), it jumps down to the point following the broken line, and then start running with slope b^2 (because u_2 increases at rate b and u_1 at rate b^{-1}); If the point reaches L_2 (the right edge), it jumps to the left according to the broken line and start running with slope b^{-2} ($b > 1$ assumed). If the opposite edges of the square are glued, we can make a torus, and we obtain a continuous trajectory on the torus (as seen in Fig. 18.4).

18.3 Trajectories on universal covering space of T^2

Instead of gluing the opposite edges of the square to make the torus, we may tessellate many copies of the square to make the so-called universal covering space.[199] In this space (i.e., instead of returning to itself, moving on to next tiles) we can clearly see what is going on (Fig. 18.6).

In Fig. 18.6 Right three trajectories starting from closely located initial points are depicted.[200] Earthquakes occur in block 1 (resp. block 2) when the trajectory crosses the vertical (resp. horizontal) lines. If $b > 1$, two initially close trajectories come apart exponentially in time. The great earthquake model is with $b > 1$, so according to this model long time prediction of earthquakes is impossible.

[199] To obtain elementary knowledge of topology I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry* (Undergraduate Texts in Mathematics) (Springer 1976; original 1967) is the best.

[200] The distance between the two adjacent points increases as $e^{t\lambda}$ as a function of time on the average, where λ is called the Lyapunov exponent (will be discussed in a later lecture). In the present example, it can be computed as λ'/τ, where λ' is the exponent defined for the corresponding discrete system in Fig. 18.11 and τ is the average sojourn time on a single tile. To estimate λ' may not be easy, but it is obviously positive as can be trivially seen from the reduced map in Fig. 18.12.

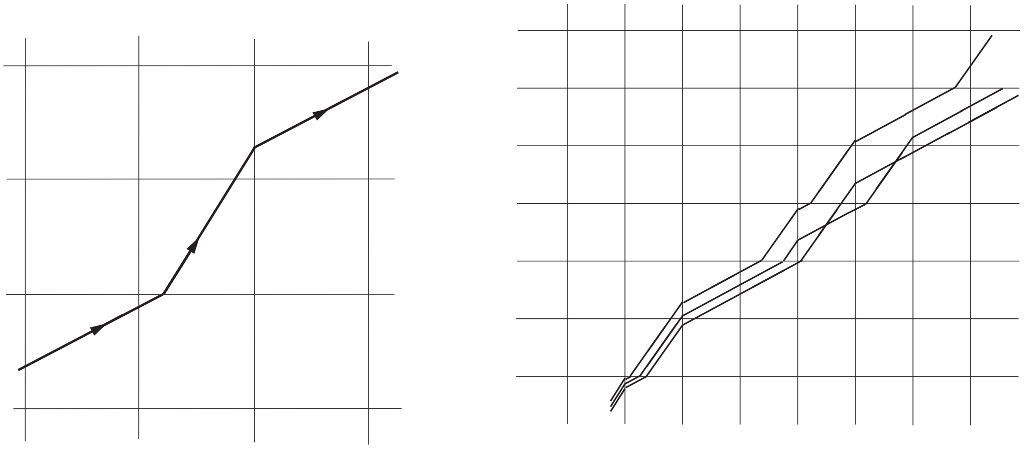


Figure 18.6: **Left**: A typical trajectory on the so-called universal covering space that is made by tessellating the squares instead of forming a torus. The crossings with the vertical lines correspond to earthquakes in block 1 and those with the horizontal lines to those in block 2. **Right**: Locally, trajectories depart from each other exponentially in time. (Around the lattice points something complicated may happen, but this is a global problem.) (To prove that the trajectories separate from each other exponentially on the average, the easiest way is to use the one-dimensional map shown Fig. 18.12.)

18.4 As a vector field on two $\boldsymbol{T}^2$

If Fig. 18.4 is understood as a description of a two-dimensional motion, we must conclude that the trajectory follows the vector field v_1 in Fig. 18.7 at one time, and then v_2 at another. The history up to the point determines on which flow (vector field) the trajectory follows. A two-valued vector field dependent on the history is not a very convenient object.

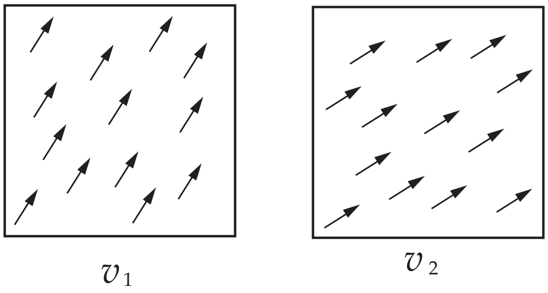


Figure 18.7: In Fig. 18.4 at a particular point on the torus the trajectory runs in one of the two directions depending on the history up to the point. That is, the trajectory runs on the torus according to one of the two vector fields, v_1 or v_2, depending on the history.

To avoid this we prepare two tori, on one of which is v_1 and on the other v_2, and then connect these two tori according to the connection rule consistent with Fig. 18.5 as shown in Fig. 18.8.

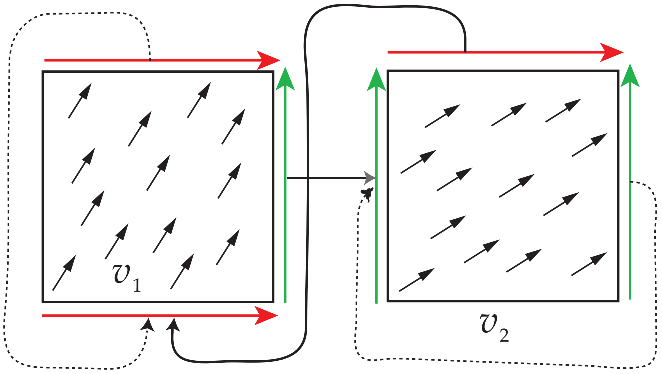


Figure 18.8: The univalent vector field independent of the history on the connected two copies of the original torus. The two vector fields prepared in Fig. 18.7 are connected according to the rules consistent with the trajectory connection rules in Fig. 18.5(i). How to glue the tori is shown by straight colored arrows with large arrowheads (the directions denote correct orientations to glue).

18.5 Why unpredictability?

As can be seen from Fig. 18.8, something special happens when a crossover between two tori occurs. To understand what happens there, it is convenient to glue only the edges connected by broken arrow curves in Fig. 18.8 as A to B in Fig. 18.9 instead of completing the two tori. The trajectories coming into the connection edge from left in Fig. 18.8 (along the short horizontal arrow) have smaller angles with the connection edge than the ones going out to right as clearly illustrated in Fig. 18.9C, so we see the trajectories are spread and then are inserted into the trajectories on the right cylinder. That is, the event happening at the connection is akin to card shuffling. Thus, we understand intuitively why the system 'produces randomness' despite its deterministic nature.

It is clear that the apparently random behavior observed in Fig. 18.3 is caused by the intrinsic nature of this *deterministic dynamical system* (= a system whose behavior is uniquely determined if its past is known) and that external noise has nothing to do with it. If we knew the trajectory precisely without any error, we would not lose any information even if there is an expansive tendency of the trajectory bundle. However, we can never know very small scales. This unknowable is amplified by the expansion of the trajectory bundle and then is fixed into the system behavior by the insertion occurring at the connection to the other cylinder, as illustrated in Fig.

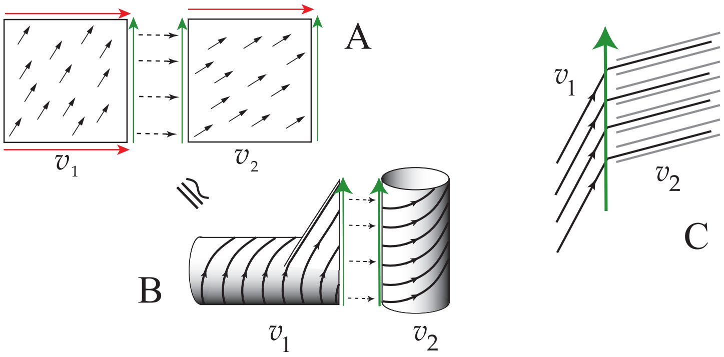


Figure 18.9: Let us pay attention to the connection from v_1 to v_2 only (the broken arrows in A, which is almost a copy of Fig. 18.8). Gluing only one pair of edges in A, we obtain B. The 'screw' part of B corresponds to the inside of the squares in A, where the trajectories run parallelly. In order to make easy to observe the crossover from the left to the right square (or torus), a cut is introduced in the left cylinder that does not affect the parallel trajectories and the edge to be glued is made straight. What happens at the connection between v_1 and v_2 is equivalent to spreading out the trajectory spacings and then inserting trajectories into the right cylinder as illustrated in C; it reminds us of shuffling cards.

18.9C. Consequently, we feel the behavior of this system random;[201] Recall the KS entropy computation in the Sinai billiard in **??**.

18.6 Another illustration of randomization

If we deform the system slightly further, we can map it to a system that must be familiar to chaos aficionados. If we squish the cylinders in Fig. 18.9B on which the motion is 'spiral,' we obtain Fig. 18.10.

The figure reminds us of the famous *Lorenz model* that will be discussed in the two subsequent lectures.

18.7 Correspondence to discrete time system

A trajectory in the universal covering space is piecewise linear as we have seen in Fig.

[201] As can be seen from this example, nonlinearity is not needed to expand small scales. Nonlinearity is, however, usually needed to contain the system within a finite range despite local linear expansion. Therefore, if the phase space is intrinsically compact, then we might be able to say nonlinearity is not absolutely necessary for chaos. Actually, a typical chaotic system is given by linear maps from a torus onto itself. The decisive paper on this topic is R. L. Adler and B. Weiss, "Similarity of automorphisms of the torus," Memoir. Am. Math. Soc. **98** (1970).

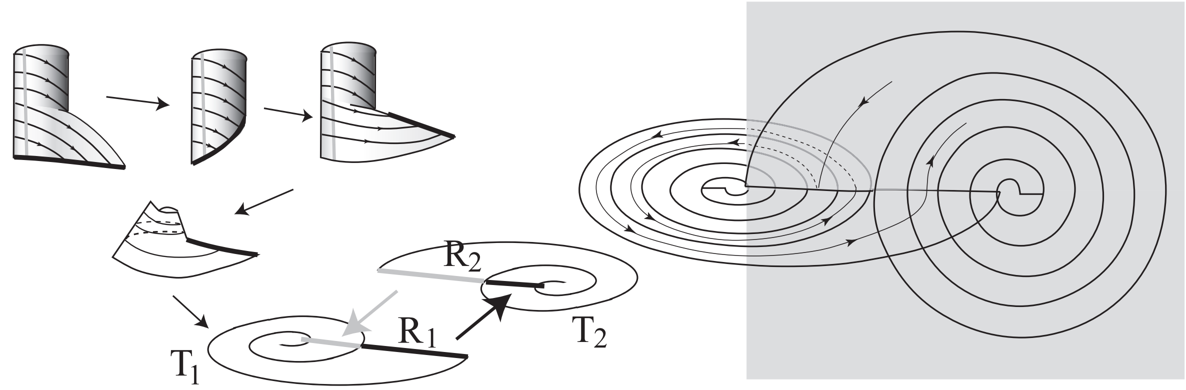


Figure 18.10: Further topological acrobat. The portions where the cylinders are connected are the same as Fig. 18.9B, but the ways to extend the trajectory ends are different. After running the spiral portion, trajectories are inserted into the other spiral. If we dovetail these two sheets with spiral trajectories, we obtain the rightmost figure. T$_1$, and T$_2$ correspond to the tori and R$_1$ and R$_2$ are connecting edges, where insertion and fixation of the expansion outcomes occur. The difference from Fig. 18.9B is only that the spiral on the cylinders becomes that on the disks, so, similarly as before, we can see how expanded microscopic details influence decisively the world around our scale.

18.6, so we must be able to record a trajectory only by recording its breaking points. We have only to pay attention to the points where the trajectory crosses the lattice lines in the universal covering space, and convert the continuous time system to a discrete time system (however, it is not a simple discretization such as observing a system periodically with stroboscopic light). More explicitly, in Fig. 18.5(i) (copied in Fig. 18.11) if we record the distance x measured along the edges from the corner C when the trajectory crosses L$_1$ or L$_2$, we can map a trajectory to a sequence $\{x_i\}$. This correspondence is one to one.[202] That is, from $\{x_i\}$ we can reconstruct the original continuous trajectory, because we know the speed of the point along the trajectory.

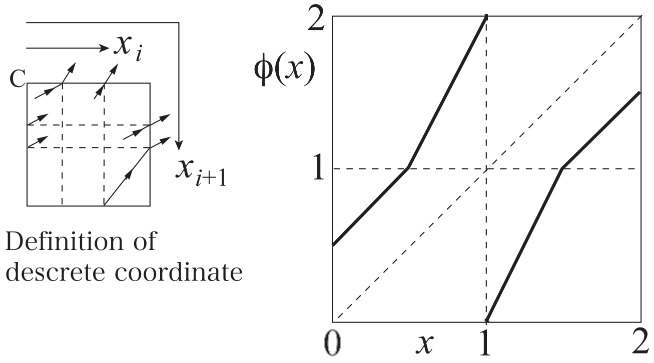
If the sequence $\{x_i\}$ can be reconstructed by a certain recursive rule, the system may be described more simply. Here, a 'recursive rule' means a rule that can give the next time state in terms of the current state (corresponding to the equation of motion in classical mechanics). Actually, this sequence $\{x_i\}$ is determined as a solution to an initial value problem of a (nonlinear) difference equation

$$x_{n+1} = \phi(x_n), \tag{18.1}$$

where $\phi : [0, 2] \to [0, 2]$ is given by the graph in Fig. 18.11

This discrete model can be, as explained in Fig. 18.12, folded into a unimodal

[202] Here we ignore the 'loose ends' of the trajectories between the lattice crossings.



Definition of descrete coordinate

Figure 18.11: If the successive lattice crossing positions of a trajectory is described in terms of the distance measured along the edges of the square from the top left corner C, we obtain a one-dimensional map $[0,2] \to [0,2]$: $x_{i+1} = \phi(x_i)$.

piecewise linear map from $[0,1]$ into itself.[203] Folding implies identifying two points in $[0,2]$. Therefore, the original behavior cannot be reconstructed from the reduced system. However, the correspondence is simple, so we can learn various things about the original system from the simplified system.

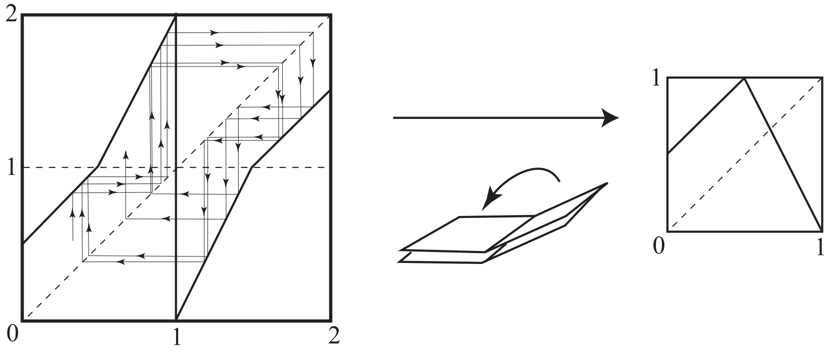


Figure 18.12: A reduced discrete map may be constructed by 'folding' the original map. Such unimodal piecewise linear maps are mathematically thoroughly understood. On the left the way to chase a discrete history is illustrated

[203] Unimodal piecewise linear maps are thoroughly studied in the following papers: Sh. Ito, S. Tanaka and H. Nakada, "On unimodal linear transformations and chaos I, II," Tokyo J. Math. **2**, 221, 241 (1979).

19 Lecture 19. Lorenz system: Introduction

Warning
After switching the OS (to Mojave) perhaps security has become tighter, and I cannot run many of the copied demos (due to 'unsupported plug-in').

Introductory video
The following video is an leisurely introduction to the Lorenz system (Chaos Chapter 7), giving a nice overview of what we should understand mathematically:

https://www.youtube.com/watch?v=aAJkLh76QnM&frags=pl%2Cwn

We look at the key portions of Lorenz's historic paper, "Deterministic nonperiodic flow" (1963); here the basic equation derived by Saltzman is outlined **19.2**. Lorenz numerically established his system is intrinsically nonperiodic. We watch various behaviors of the Lorenz equation. Notice that to reduce the original system to a simpler system (e.g., time discrete dynamical system) is crucial (Fig. 19.3). You see very similar structures we already encountered in a simpler Ito's system in the preceding Lecture. Then, we see a similar system describing magnetic reversal of the earth (the Rikitake model **19.7**).

19.1 Lorenz's motivation

> Certain hydrodynamical systems exhibit steady-state flow patterns, while others oscillate in a regular periodic fashion. Still others vary in an irregular, seemingly haphazard manner, and, even when observed for long periods of time, do not appear to repeat their previous history.

This is the first paragraph of Introduction to "Deterministic nonperiodic flow" by E N Lorenz.[204]

> The short-range weather forecaster, however, is forced willy-nilly to predict the details of the large scale turbulent eddies—the cyclones and anticyclones which continually arrange themselves into new patterns. Thus there are occasions when more than the statistics of irregular flow are of very real concern.

[204] J Atmos Sci., 20, 130 (1963).

As a system to study Lorenz chose a fluid layer heated from below (the Rayleigh-Benard convection problem):

> Rayleigh (1916) studied the flow occurring in a layer of fluid of uniform depth H, when the temperature difference between the upper and lower surfaces is maintained at a constant value ΔT. Such a system possesses a steady-state solution in which there is no motion, and the temperature varies linearly with depth. If this solution is unstable, convection should develop: In the case where all motions are parallel to the xz-plane, and no variations in the direction of the y-axis occur, the governing equations may be written (see Saltzman, 1962):
>
> $$\frac{\partial}{\partial t}\nabla^2\psi = -\frac{\partial(\psi, \nabla^s\psi)}{\partial(x,z)} + \nu\nabla^4\psi + g\alpha\frac{\partial\theta}{\partial x}. \tag{19.1}$$
>
> $$\frac{\partial}{\partial t}\theta = -\frac{\partial(\psi,\theta)}{\partial x, z)} + \frac{\Delta T}{H}\frac{\partial\psi}{\partial z} + \kappa\nabla^2\psi. \tag{19.2}$$
>
> Here ψ is a stream function for the two-dimensional motion, θ is the departure of temperature from that occurring in the state of no convection, and the constants g, α, ν, and κ denote, respectively, the acceleration of gravity, the coefficient of thermal expansion, the kinematic viscosity, and the thermal conductivity.

Here, let us try to understand Saltzman's equation from basic physics.

19.2 Saltzman's equation[205]

For the system considered in **19.1** the basic equations are the incompressible Navier-Stokes equation

$$\frac{\partial}{\partial t}\boldsymbol{v} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\nabla P + g\alpha T\boldsymbol{e}_z + \nu\nabla^2\boldsymbol{v} \tag{19.3}$$

with $\nabla\cdot\boldsymbol{v} = 0$ and the equation governing the thermal advection for T (temperature departure)

$$\frac{\partial}{\partial t}T + (\boldsymbol{v}\cdot\nabla)T = \kappa\nabla^2 T. \tag{19.4}$$

From (19.3) we can derive the vorticity equation (vorticity $\omega = \mathrm{curl}\,\boldsymbol{v}$) as

$$\frac{\partial\omega}{\partial t} + (\boldsymbol{v}\cdot\nabla)\omega = (\omega\cdot\nabla)\boldsymbol{v} + g\alpha\frac{\partial T}{\partial x} + \nu\nabla^2\omega. \tag{19.5}$$

[205] B. Saltzman, "Finite amplitude free convection as an initial value problem-I," J. Atmos. Sci., 19, 329 (1962). From the Acknowledgement of Lorenz's paper, "The writer is indebted to Dr. Barry Saltzman for bringing to his attention the existence of nonperiodic solutions of the convection equations."

Here, we assume the flow is translationally symmetric along the y-axis, so $\boldsymbol{v} = (u, w)$ is a 2-vector. The incompressibility means there is ψ (streaming function) such that $\operatorname{curl}\psi = (u, w)$[206]

$$u = -\frac{\partial\psi}{\partial z}, w = \frac{\partial\psi}{\partial x}. \tag{19.6}$$

The vortex equation (19.5) reads, since

$$\omega = (0, \partial_x w - \partial_z u, 0) = (0, \nabla^2\psi, 0),\ \boldsymbol{v}\cdot\nabla(A, B) = -\frac{\partial\psi}{\partial z}\frac{\partial A}{\partial x} + \frac{\partial\psi}{\partial x}\frac{\partial B}{\partial z} = \frac{\partial(A, B)}{\partial(x, z)}, \tag{19.7}$$

$$\frac{\partial}{\partial t}\nabla^2\psi - \frac{\partial(\psi, \nabla^2\psi)}{\partial(x, z)} = g\alpha\frac{\partial T}{\partial x} + \nu\nabla^4\psi. \tag{19.8}$$

Now, $T = \theta + \frac{\Delta T}{H}z$ is introduced to (19.8) and (19.4), and we get the equations in **19.1**.

19.3 'Derivation' of Lorenz equation

Saltzman derived a set of ODE by expanding ψ and θ in Fourier series. Solving the ODE (quoted from Lorenz, ibid.):

> He then obtained time-dependent solutions by numerical integration. In certain cases all except three of the dependent variables eventually tended to zero, and these three variables underwent irregular, apparently nonperiodic fluctuations.

Lorenz then thought that if only the three surviving coefficients are kept from the start by 'drastically' truncating the expansion as

$$a(1 + a^2)^{-1}\kappa^{-1}\psi = X\sqrt{2}\sin(\pi a x/H)\sin(\pi z/H), \tag{19.9}$$

$$\pi R_c^{-1}\Delta T^{-1}\theta = Y\sqrt{2}\cos(\pi a x/H \sin(\pi z/H) - Z\sin(2\pi z/H), \tag{19.10}$$

the same results should be obtained. The outcome is the Lorenz equation:

$$\dot{X} = -\sigma X + \sigma Y, \tag{19.11}$$

$$\dot{Y} = rX - Y - XZ, \tag{19.12}$$

$$\dot{Z} = XY \quad - bz, \tag{19.13}$$

[206] Let $\omega = -udz + wdx$. Then, incompressibility means $d\omega = 0$, so the converse of Poincaré's lemma (= closedness means exactness) means there is ψ such that $\omega = d\psi$.

where r, σ, τ are positive constants and 'dot' means the derivative wrt the dimensionless time τ.[207]

Remark: How 'qualitatively' good is (19.12) for understanding the original system? This is not a very relevant question in the present conceptual study, but if you can solve a PDE as a set of a small number of ODEs, it is numerically still very advantageous. This question is related to various 'finite dimensional attractors' of PDE systems. See (**??**) as a general approach.

19.4 What Lorenz established

The phase volume of the space spanned by (X, Y, Z) shrinks (i.e., div $\boldsymbol{v} < 0$):

$$\frac{\partial \dot{X}}{\partial X} + \frac{\partial \dot{Y}}{\partial Y} + \frac{\partial \dot{Z}}{\partial Z} = -(\sigma + b + 1). \tag{19.14}$$

The origin is a fixed point. If $r < 1$ it is a sink. Note that $r = 1$ means $R_a = R_c$ where the advection starts. If $r > 1$ it is a hyperbolic saddle with one unstable direction.

For $r > 1$ there are two more fixed points: $X = Y = \pm\sqrt{b(r-1)}, Z = r - 1$. If $\sigma < b + 1$ they are still sinks (one stable direction + two more stable but spiraling plane), corresponding to steady convections. If $\sigma > b + 1$, then it has a plane of instability (Fig. 19.1).

Fig. 19.1 What Lorenz observed numerically

Fig. 1. Numerical solution of the convection equations. Graph of Y as a function of time for the first 1000 iterations (upper curve), second 1000 iterations (middle curve), and third 1000 iterations (lower curve).

Fig. 2. Numerical solution of the convection equations. Projections on the XY-plane and the YZ-plane in phase space of the segment of the trajectory extending from iteration 1400 to iteration 1900. Numerals "14," "15," etc., denote positions at iterations 1400, 1500, etc. States of steady convection are denoted by C and C'.

Fig. 3. Isopleths of X as a function of Y and Z (thin solid curves), and isopleths of the lower of two values of X, where two values occur (dashed curves), for approximate surfaces formed by all points on limiting trajectories. Heavy solid curve, and extensions as dotted curves, indicate natural boundaries of surfaces.

Let us see the actual solution.

`https://media.pearsoncmg.com/aw/ide/idefiles/media/JavaTools/lrnzdscv.`

[207] $\tau = \pi^2(1 + a^2)\kappa t/H$, $\sigma = \nu/\kappa$ (the Prandtl number), $r = R_a/R_c$ and $b = 4/(1 + a^2)$. $R_a = g\alpha H^3 \Delta T/\nu\kappa$ is the Rayleigh number, and $R_c = \pi^4 a^{-2}(1 + a^2)^2$ is the critical Rayleigh number.

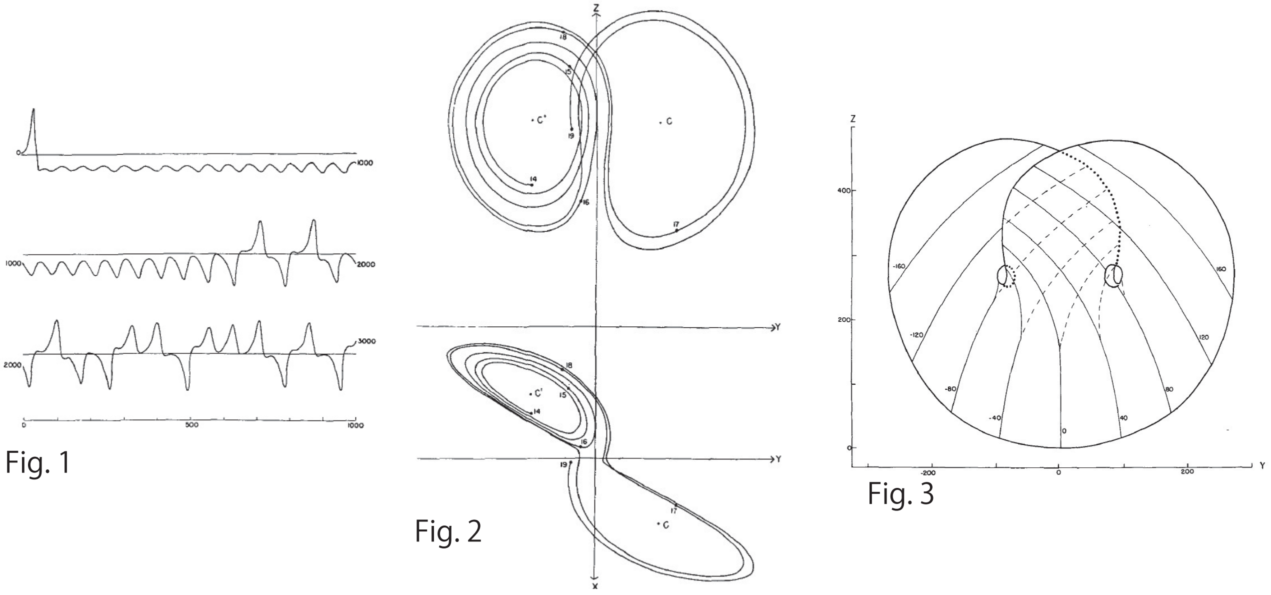


Figure 19.1: What Lorenz found numerically [Fig. 1-3 of Lorenz JAS 20 130 (1963).]

html

The initial demo uses the initial conditions in the paper. You can see the solutions sensitively depend on the initial choices. After that you can choose the intial x by tapping the panel (Fig. 19.2).

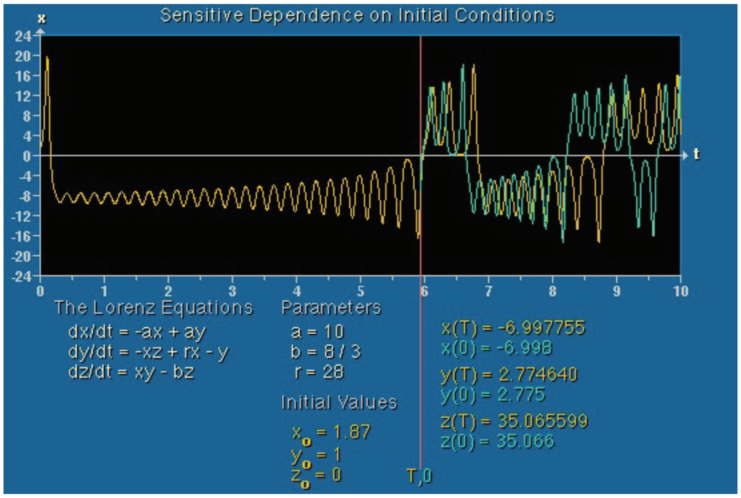
As to the global feature about Fig. 3 in Fig. 19.1 Lorenz said:

> Thus, within the limits of accuracy of the printed values, the trajectory is confined to a pair of surfaces which appear to merge in the lower portion of Fig. 3. The spiral about C lies in the upper surface, while the spiral about C' lies in the lower surface. Thus it is possible for the trajectory to pass back and forth from one spiral to the other without intersecting itself.

He observes further

> Returning to Fig. 2, we find that the trajectory apparently leaves one spiral only after exceeding some critical distance from the center. Moreover, the extent to which this distance is exceeded appears to determine the point at which the next spiral is entered; this in turn seems to determine the number of circuits to be executed before changing spirals again.

Therefore, he collected successive local max values $\{M_n\}$ of Z and plotted M_{n+1} against M_n (Fig. 19.3).



The Lorenz Equations
dx/dt = -ax + ay
dy/dt = -xz + rx - y
dz/dt = xy - bz

Parameters
a = 10
b = 8 / 3
r = 28

Initial Values
x_o = 1.87
y_o = 1
z_o = 0

x(T) = -6.997755
x(0) = -6.998
y(T) = 2.774640
y(0) = 2.775
z(T) = 35.065599
z(0) = 35.066

Figure 19.2:

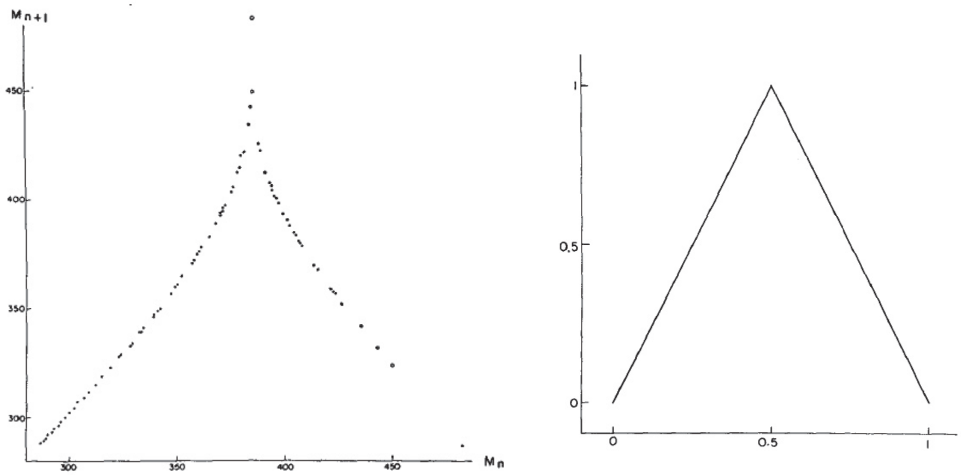


Figure 19.3: Lorenz map [Fig. 4,5 of Lorenz JAS 20 130 (1963).]

Fig. 19.3 Lorenz map
Left: Corresponding values of relative maximum of Z (abscissa) and subsequent relative maximum of Z (ordinate) occurring during the first 6000 iterations. The right figure (the tent map) is proposed to understand the system conceptually.

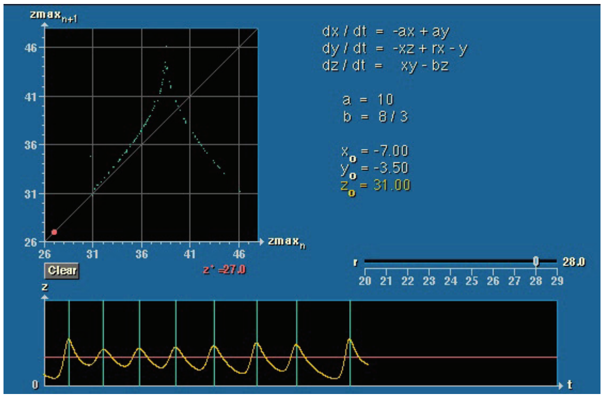
Lorenz discussed the sensitive dependence of the system to disturbance with the aid of the tent map.

In Conclusion of his paper Lorenz writes:

When our results concerning the instability of nonperiodic flow are applied

> to the atmosphere, which is ostensibly nonperiodic, they indicate that prediction of the sufficiently distant future is impossible by any method, unless the present conditions are known exactly. In view of the inevitable inaccuracy and incompleteness of weather observations, precise very-longrange forecasting would seem to be non-existent.

Let us watch a live Lorenz map. You must be a bit patient:
Lorenz map:
https://media.pearsoncmg.com/aw/ide/idefiles/media/JavaTools/lrnzzmax.html (Fig. 19.4).



dx / dt = -ax + ay
dy / dt = -xz + rx - y
dz / dt = xy - bz

a = 10
b = 8 / 3

x_o = -7.00
y_o = -3.50
z_o = 31.00

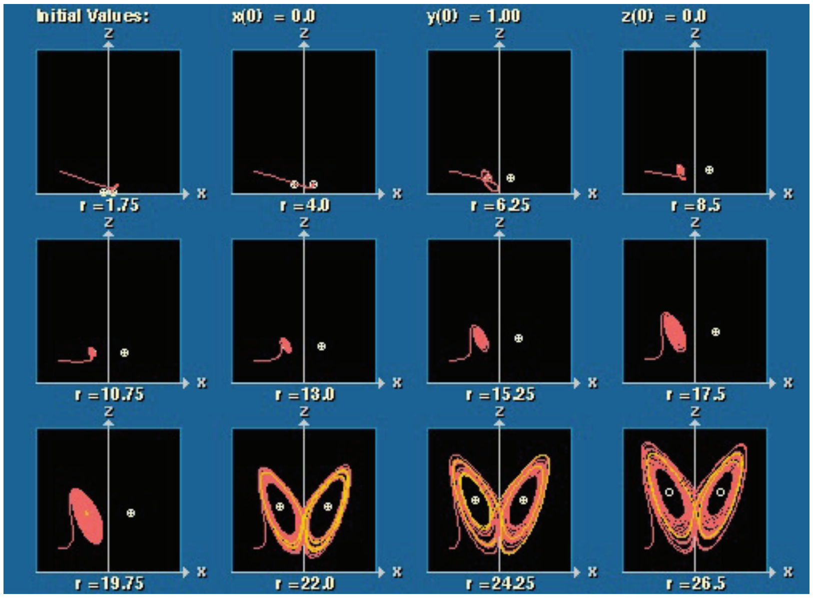
r 28.0

Figure 19.4:

The r dependence of the phase portrait discussed a bit above may be glimpsed from the following panels:
https://media.pearsoncmg.com/aw/ide/idefiles/media/JavaTools/lrnzpgrd.html (Fig. 19.5).

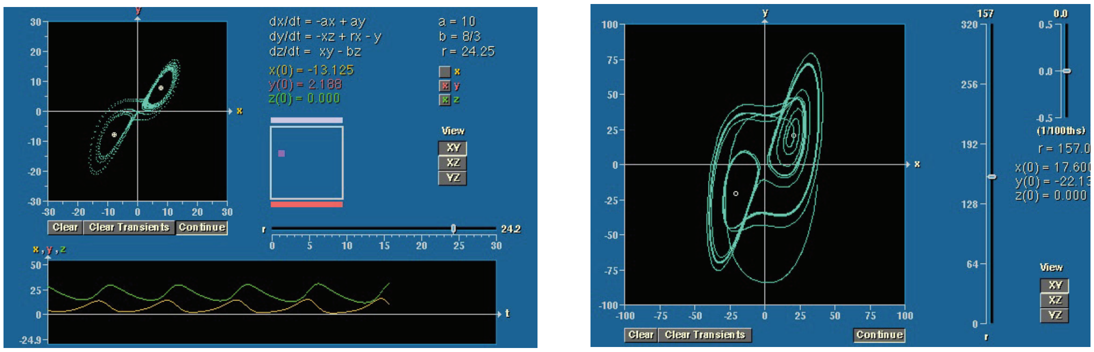
19.5 Illustration of Lorenz system behaviors

Now, you can scan r to watch what happens (see Fig. 19.6L):
https://media.pearsoncmg.com/aw/ide/idefiles/media/JavaTools/lrnzphsp.html
For wider range of r (XZ-view is popular):
https://media.pearsoncmg.com/aw/ide/idefiles/media/JavaTools/lrnzr320.html
You can slide the r scale to choose its value, and then tap the panel. The intial



Initial Values: x(0) = 0.0 y(0) = 1.00 z(0) = 0.0

Figure 19.5: "ω-limit set(s) may be guessed. You can choose the intial condition on the panel position is where you tap the panel.



dx/dt = -ax + ay
dy/dt = -xz + rx - y
dz/dt = xy - bz

a = 10
b = 8/3
r = 24.25

x(0) = -13.125
y(0) = 2.188
z(0) = 0.000

r = 157.0
x(0) = 17.600
y(0) = -22.13
z(0) = 0.000

Figure 19.6: Right: wider range.

The following is an interactive visualization in which you can change r, σ and b:

http://www.malinc.se/m/Lorenz.php

You can watch basically all the behaviors in Fig. 19.5. To see how the system attractors evolve you may have to be very patient.

A 3D trajectory is visualized with a stereoplot:

https://www.youtube.com/watch?v=-tpRZCnoih0

Fate of small change in the initial conditions:

https://www.youtube.com/watch?v=FYE4JKAXSfY 3 point animation (esp. after 50 sec)

You can zoom in:
https://www.ibiblio.org/e-notes/webgl/lorenz_model.html

19.6 Physical model mimicking Lorenz

Watch Chaos Chapter 8 beyond about 6 min:
https://www.youtube.com/watch?v=SlwEt5QhAGY&frags=wn

A schematic model is here:
https://www.youtube.com/watch?v=M-94UxoZD2M&frags=pl%2Cwn

19.7 Earth dynamo reversal

The magnetic field of the earth reverses from time to time, and gave a decisive proof of plate tectonics.

A simple model to explain this reversal is the Rikitake model:

Rikitake, T., Oscillations of a system of disk dynamos, Proc. Cambr. Phil. Soc., 54 89 (1958).

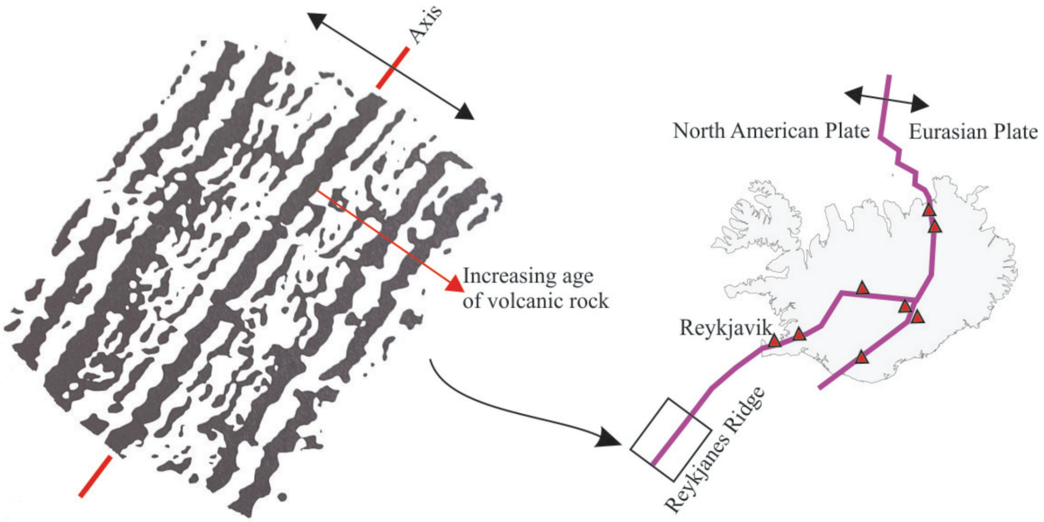
Tsuneji Rikitake, Nonsteady Geomagnetic Dynamo Models, Geophys J Internat 35 277 (1973).

$$\begin{aligned} \dot{x} &= -\nu x + zy, & (19.15)\\ \dot{y} &= -\nu y + (z-a)x, & (19.16)\\ \dot{z} &= 1 - xy. & (19.17) \end{aligned}$$

This system exhibits just the Lorenz-like attractors.

http://demonstrations.wolfram.com/RikitakeModelOfGeomagneticReversal/

What is a lesson to learn as a researcher? Whatever a model or a system you study is, study it throughly with conceptual questions.



Mid-ocean spreading along Reykjanes Ridge, depicted by magnetic 'stripes'. Black patterns are normal magnetic polarity (compass points North), and white are reversed polarity. The age of successive magnetic reversals increases with distance from the active ridge. The inset map (right) shows the complexity of the spreading ridge where it transects Iceland. Red triangles are active volcanoes on Iceland.

Image credit: Reykjanes magnetic pattern from Heirtzler, J.R., Le Pichon, X. and Baron, J.G. (1966). Magnetic anomalies over the Reykjanes Ridge. *Deep Sea Research 13: 427-433.*

Figure 19.7: magnetic reversals recorded near Iceland

20 Lecture 20. Lorenz system: Advanced topics

20.1 Knotted orbits[208,209]

Periodic orbits need not be without any knot. The Lorenz model exhibits tons of such orbits.[210]

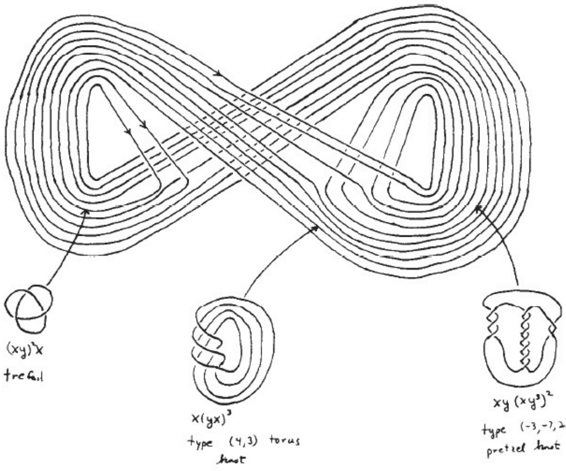


Figure 20.1: Schematic representation of coexisting knotted orbits [Fig. 1.1 of BIRMAN and WILLIAMS, Topology 22 47 (1983)]

Theorem [Franks and Williams][211]A smooth flow on S^3 with positive topological entropy[212] must possess periodic closed orbits in infinitely many different knot type equivalence classes.

Actually, Ghrist[213] proved that there is a structurally stable flow on S^3 that con-

[208] ⟪**Knot**⟫ A knot is am embedding of S^1 into $\mathbb{R}^3$. It is knotted if the image is not homotopic to S^1.

[209] Incidentally, if a protein is stretched by pulling both the ends in th opposite direction, generally, the resultant polypeptide chain should be knotted. It is known that actually, knotted proteins are very rare.

[210] JOAN S. BIRMAN and R. F. WILLIAMS, KNOTTED PERIODIC ORBITS IN DYNAMICAL SYSTEMS I: LORENZ'S EQUATIONS, Topology 22 47 (1983).

[211] J. Franks and R. F. Williams, Entropy and knots, TAMS 291 241 (1985).

[212] Topological entropy is, practically, the largest KS entropy allowed to the system.

[213] R. Ghrist, Branched two-manifolds supporting all links, Topology 36 423 (1997).

tains all knot types as periodic orbits.

If the Lorenz system is periodically perturbed, we can even have links of periodic orbits.[214]

20.2 Geometrical Lorenz model[215]

A very fruitful approach was undertaken, independently, by V. Afraimovich, V. Bykov, L. Shirnikov [2], and by J. Guckenheimer and R. Williams [3, 4]. Based on the behavior observed in the previous Lecture (Fig. 20.2 Leftmost[216]), they exhibited a list of geometric properties such that any flow satisfying these properties must contain a "strange attractor,"[217] with orbits converging to it being sensitive to initial conditions. Most important for the general theory, they proved that such flows do exist in any manifold with dimension 3. These examples came to be known as geometric Lorenz models.

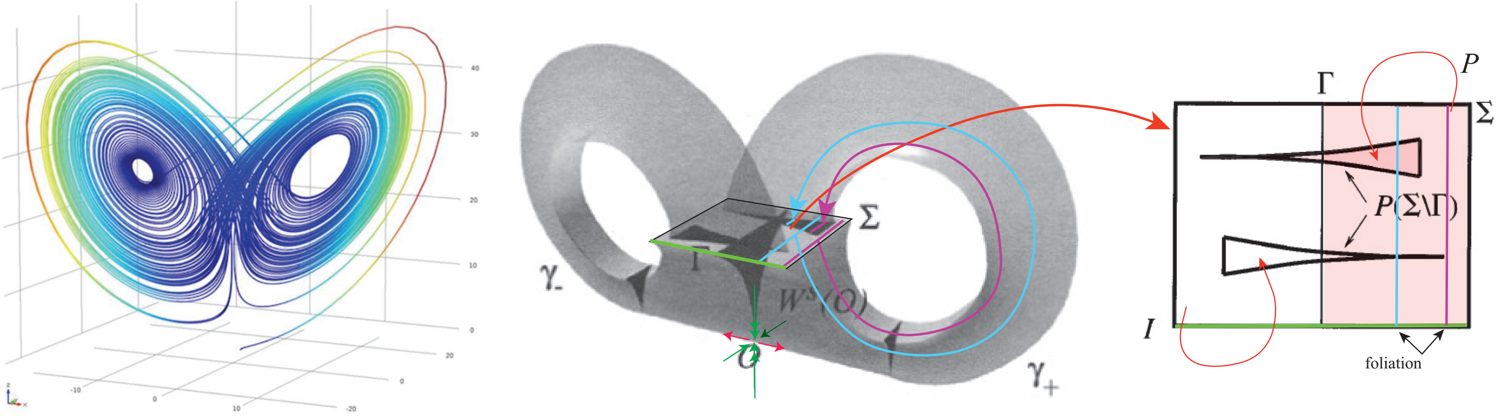


Figure 20.2: Geometric Lorenz model. The return map P maps $\Sigma \setminus \Gamma$ to Σ. I is the bottom segment (the green segment) of Σ. [Fig. 6, 8 of Viana Math Intel. 22(3) 7 (2000)]

In the above figure $\gamma_\pm$ (and $\{O\}$ (fixed point)) combined is the unstable manifold. The vertical direction is the (strong) stable direction. The weak stable direction is

[214] Y. Aizawa and T. Uezu, Topological aspects in chaos and in 2^k-period doubling cascade, Prog. Theor. Phys. 67 982 (1982).

[215] M Viana, "What's new on Lorenz strange attractors?" Math Intel. 22(3) 7 (2000).

[216] From https://ubisafe.org/explore/attractor-clipart-lorenz-attractor/.

[217] Here, the word is merely used to indicate an attracting set containing non-periodic orbits.

orthogonal to these manifolds. The stable manifold is spanned by these stable directions close to O.

The cross section is in the right of Fig. 20.2. We can define a return map P which is strongly contracting because of (19.14).

For the geometrical model, also the following foliation is assumed for P: Σ is foliated roughly transversally to I such that P is compatible with the foliation. That is, if z and z' ($\neq z$) are on the same leaf, so are $P(z)$ and $P(z')$. Furthermore, the distance between $P^k(z)$ and $P^k(z')$ shrinks to zero as $k \to \infty$. This means a leaf γ_1 is mapped inside some leaf γ_2. Since we can specify a leaf by its 'x' coordinate of the crossing point with I, P defines a map f from I into itself: $f : I \to I$. We assume

$$|f(x) - f(x')| \geq \tau |x - x'| \tag{20.1}$$

with $\tau > 1$. The graph of f looks like Fig. 20.3 (cf. Fig. 18.11).

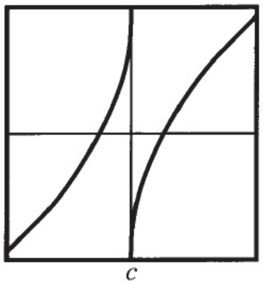


Figure 20.3: Interval maps related to the geometrical Lorenz model. Recall Fig. 18.11. [Fig. 10 of Viana Math Intel. 22(3) 7 (2000)]

The existence of a strange attractor containing O was proved for the geometric model. It has a topologically transitive orbit (if $\tau > \sqrt{2}$).

[1] V. S. Afraimovich, V. V. Bykov, and L. P. Shil'nikov. On the appearance and structure of the Lorenz attractor. Dokl. Acad. Sci. USSR, 234 336 (1977).

[2] J. Guckenheimer and R. F. Williams. Structural stability of Lorenz attractors. Publ. Math. IHES, 50 59-72 (1979).

[3] R.F. Williams. The structure of the Lorenz attractor. Publ Math. IHES, 50 73 (1979).

20.3 Lorentz attractor exists[218]

Theorem [Tucker 1998] For the classical parameters, the Lorenz equations (19.12)

[218] W. Tucker, The Lorenz attractor exists. PhD thesis, Univ. Uppsala, 1998. Text and program codes available at `www.math.uu.se/~warwic/`.; W. Tucker, The Lorenz attractor exists. C. R. Acad. Sci. Paris, 328, Serie I, Mathematique: 1197-1202 (1999).

support a robust strange attractor.

Tucker's computer-assisted proof is a combination of two main ingredients:[219]
(I) He uses rigorous numerics to find a cross-section Σ and a region N in Σ such that orbits starting in N always return to it in the future. After choosing reasonable candidates, Tucker covers N with small rectangles, as in Figure 20.4, and estimates the forward trajectories of these rectangles numerically, until they return to Σ. His computer program also provides rigorous bounds for the integration errors, good enough so that he can safely conclude that all of these rectangles return inside N. This proves that the equations do have some sort of attractor.

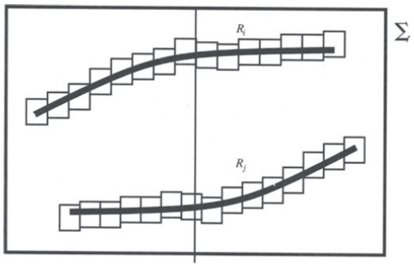


Figure 20.4: Attractor construction by Tucker [Fig. 4 of of Viana Math Intel. 22(3) 7 (2000)]

(II) Normal form theory comes in to avoid the accumulation of integration errors when trajectories are close to the equilibrium sitting at the origin.

20.4 Pseudo-orbit tracing property[220]

Here we consider a map $T : M \to M$, where M is a metric space. $\{x_n\}$ is a δ-pseudo-orbit if

$$\|Tx_n - x_{n+1}\| < \delta. \tag{20.2}$$

We say the orbit $\{T^n x\}$ ε-traces the pseudo-orbit $\{n_n\}$, if

$$\|T_n y - x_n\| < \varepsilon. \tag{20.3}$$

We say T has a pseudo-orbit tracing property (POTP), if for any ε (> 0) we can find δ (> 0) such that δ-pseudo-orbit can be ε-traced by an orbit $\{T^n y\}$.

[219] M Viana, "What's new on Lorenz strange attractors?" Math Intel. 22(3) 7 (2000).

[220] There is a general theorem that only the B systems can have POTP. That is, only if the system is very chaotic, POTP holds. [A necessary and sufficient condition for a system to have POTP is that it is isomorphic to a Markov subshift (Kubo Th 3.16). A Markov subshift is isomorphic to a Bernoulli shift (essentially the Ornstein theorem to be discussed later).]

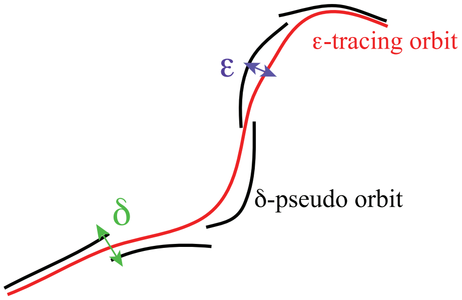


Figure 20.5: A pseudo orbit with a tracing property

This property is practically important. Pseudo-orbit tracing property implies that for a numerically obtained orbit with error δ there is a true orbit close to it for all time. If a system lacks POTP, then it is questionable that numerical simulation can always tell something correct about the system.

POTP for continuous dynamical system can be defined similarly (actually, the illustration in Fig. 20.5 is for this case).

We know chaotic billiards have POTP. Thus, if you make a reasonable numerical simulation, there is a true orbit starting somewhere. However, POTP never guarantees that pseudo orbits correctly sample the true orbits, so if you are interested in the statistical behavior of chaotic systems, you must not rely on POTP.

20.5 Lorenz system is not ε-traceable

The precise statement is that the geometrical Lorenz model lacks POTP. This is proved by showing the map $f : I \to I$ defined in **20.2** lacks POTP. f satisfies (see Fig. 20.3)

$$f(c+0) = 0, f(c-0) = 1, \tag{20.4}$$

$$f' > \lambda \geq 1, \tag{20.5}$$

$$f(0) < c < f(1), \tag{20.6}$$

$$\lim_{x\to c} f'(x) \to \infty. \tag{20.7}$$

Theorem [Komuro][221] $f : I \to I$ has the POTP if and only if $f(0) = 0$ and $f(1) = 1$.

The above theorem means that

[221] M. Komuro, Lorenz attractor do not have the pseudo-orbit tracing property, J Math Soc Jpn 37 489 (1985).

Theorem. Let M be a 3-dimensional compact manifold. Then the set of vector fields with the strong[222] POTP is not dense in $\mathcal{X}^2(M)$.

It is worth noting the following empirical fact. Usually, if a system exhibits ‘chaotic behavior’ in a certain parameter range, increasing the numerical accuracy in simulating the system shrinks the chaotic parameter range. The Lorenz system is known to be the opposite (at least in certain parameter ranges).[223]

20.6 Lorenz template

A template is the union of strips, and each strip is a copy of the standard 2-flow box $[0, 1] \times [0, 1]$ with flow parallel to the z edge. Where the strips meet, along branches, the vectors coincide so that a unique semi-flow is determined. The copying homeomorphism stretches the $x = 1$ end so that the resulting flow, where defined, is expanding. Two or more of these strips are assembled into a template, so that the resulting flow is well defined in the positive direction(, but at the branches it is not well defined for the negative time direction).[224]

An example is the Lorenz template (Fig, 20.6). We can understand the periodic orbits of the Lorenz system.

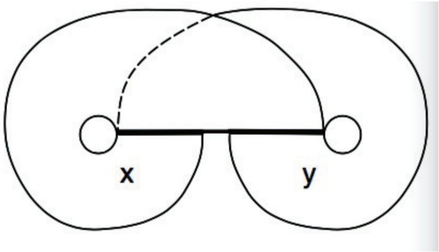


Figure 20.6: Lorenz template [Fig. 2 of Williams BAMS 35 145 (1998)]

In figure 20.6, the template consists of two strips, x and y, so that the flow on the left side passes around to the left and back down to the branch. Similarly, the flow on the right passes around to the right behind the strip x and back down to the branch. There is a middle portion at each branch, called a gap, at which the orbits leave the template. Orbits which leave are no longer of interest to us. Note that the periodic orbits never leave. The dark horizontal line is a branch; above it two planar

[222] For time continuous systems monotone time variable change is also allowed for the definition of POTP. If this time ‘dilation is restricted to the scale range between $1 \pm \varepsilon$ globally, we say it is the strong POTP. Thus, numerically, we need strong POTP.

[223] M. Komuro, private communication (1979).

[224] R. F. Williams, The universal template of Ghrist, Bull AMS 35 145 (1998).

pieces come together where they are tangent to each other. Thus each point on a branch lies in two smooth disks which coincide below the branch, but are disjoint above. Each branch in a template is homeomorphic to the unit interval [0,1] and called a branch line.

As can be seen from Fig. 18.10, if there is no gap, the Lorenz template is exactly the spiral figure in this figure. For example, there is a one-to-one correspondence between periodic orbits of the two templates.

20.7 Shimada's Ising representation of Lorenz attractor[225]

The Lorenz attractor has two wings, so one rotation in one wing is assigned +1 spin and the other −1 to code the sequence. Since the orbits are deterministic, this coding may be interpreted as the coding of the initial position:

$$x \to s_1 s_2 \cdots, s_n \cdots, \quad \text{where } s_i \in \{-1, 1\}. \tag{20.8}$$

This map can also be understood as a map from x to $b \in [0,1]$, if we regard (−1 as 0 and +1 as 1 and read the sequence as the binary expansion of a number in $[0,1]$)

$$b = \sum_{k=1}^{\infty} \left(\frac{s_k + 1}{2}\right) 2^{-k}. \tag{20.9}$$

This is illustrated in Fig. 20.7

The spin configuration converges to a stationary distribution (Fig. 20.8 Left) which exhibits a self-similar structure due to the translational symmetry (time-stationarity) of the original system and the coding scheme. We can construct a 1D Ising Hamiltonian to reproduce this distribution as

$$H = -\sum_{i,j} J(|i-j|) s_i s_j. \tag{20.10}$$

The coupling constant decays exponentially as in Fig.20.8Right.

The entropy per spin mus t be the KS entropy of the Poincarémap . The latter may be estimated from the expansion rate of the nearby orbits (= Lyapunov characteristic number; we will discuss later). Numerically, Shimada confirmed the agreement.

[225] I Shimada, Gibbsian distribution on the Lorenz attractor," Prog Theor Phys 62 61 (1979).

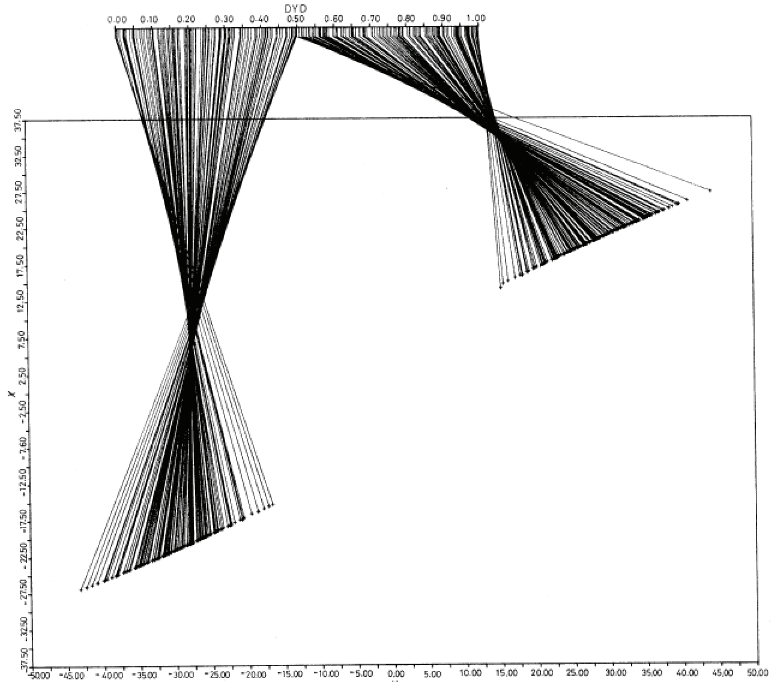


Figure 20.7: Correspondence $x \to b$ [Fig. 1 of Shimada PTP 62 61 (1979)]

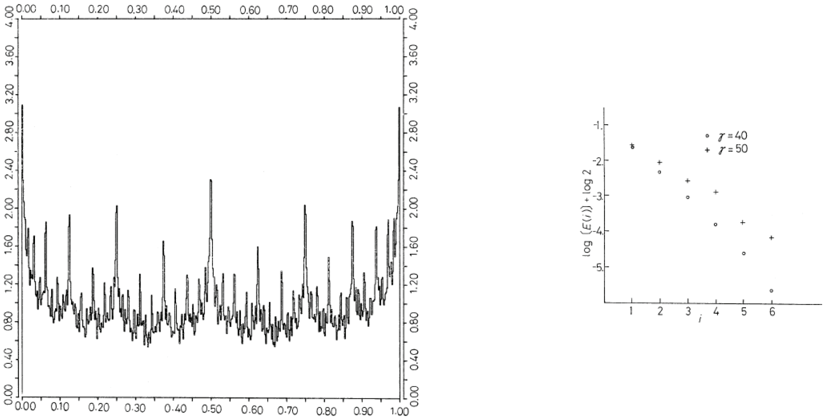


Figure 20.8: Left 9 spin stationary configuration distribution; spin configurations are described as b. Right: The coupling constant corresponding to the configurations (γ is our r) [Fig. 2 of Shimada PTP 62 61 (1979)]

20.8 Galerkin method

Reduction of PDE to a finite number of coupled ODEs may be very useful numerically as well as summarizing the system behavior in a finite dimensional space. A popular strategy is to use an appropriate orthogonal function set.

Consider a functional equation in a 'spatial domain' D:

$$\frac{d}{dt}u = N[u], \tag{20.11}$$

where N is a general nonlinear operator that can contain spatial derivatives. Take an appropriate orthonormal basis $\mathcal{B} = \{\phi_n(x)\}$ (assumed as a real function set) for the set of functions on D (considered as a Hilbert space). Expand u as

$$u(t,x) = \sum_j c_j(t)\phi_j(x), \tag{20.12}$$

and introduce it into (20.11)

$$\sum_j \frac{\partial c_j}{\partial t}\phi_j(x) = N\left[\sum_j c_j(t)\phi_j(x)\right]. \tag{20.13}$$

Then, project this onto ϕ_k:

$$\frac{\partial c_k}{\partial t} = \int_D dx\, \phi_k(x) N\left[\sum_j c_j(t)\phi_j(x)\right]. \tag{20.14}$$

To make this system manageable, we choose a finite subset $\mathcal{B}'$ of $\mathcal{B}$. The resultant set of (generally nonlinear) ODEs defines a Galerkin approximation to (20.11).[226]

One practical approach is to Fourier expand physically convenient ON function set $\{\phi_k\}$ and use the result to compute (20.14). For example, Yahata used this method to reduce the Taylor flow problem to 32 dimensions.[227]

[226] This is related to the principal component analysis.

[227] H. Yahata, Temporal Development of the Taylor Vortices in a Rotating Fluid, PTP Supp 64 176 (1978); he chose the number of functions to reproduce experimental results (say, bifurcations given by Harry Swinney). Some people try to choose $\mathcal{B}'$ (or $\mathcal{B}$) to minimize the errors, but they actually use heavy numerical solutions. However, you might have to do this optimization for only one state (or a parameter set) of the system and could use the result for 'neighbor' states, so one-time heavy investment may pay.

20.9 Use of inertial manifold in conjunction with Galerkin method[228]

To be added (maybe).

[228] C. FOIAS, M. S. JOLLY, I. G. KEVREKIDIS, G. R. SELL and E. S. TITI, ON THE COMPUTATION OF INERTIAL MANIFOLDS PL A131 433 (1988).

21 Lecture 21. Strange attractors

21.1 Landau's view of turbulence

Turbulence in fluid dynamics is a phenomenon in which all the space-time scales of apparently random fluctuations simultaneously appear in the fluid motion. It is believed that this is governed by the (incompressible) Navier-Stokes equation, a deterministic system.[229] Naturally, Landau considered the problem as the couplings between numerous destabilized modes (Fourier components of the velocity vector). Let A be a complex mode. It obeys a time-dependent Landau-Ginzburg equation

$$\frac{dA}{dt} = \gamma A - aA|A|^2 + Q, \tag{21.1}$$

where γ changes its sign when the mode linearly destabilizes, a is a positive constant and Q denotes the mode coupling terms. Thus, Landau and many statistical physicists thought the problem is closely related to critical phenomena, only the cascade of the driving proceed in the opposite direction, from large (stirring scale) to small (viscous dissipation scale).[230] The resultant quasi periodic motion with numerous modes of different frequencies is Landau's interpretation of turbulence.

21.2 Ruelle and Takens point of view[231]

Ruelle and Takens pointed out that quasi periodic motion is not the usual 'chaotic' motion we observe in dissipative systems, so Landau's picture must be wrong. They pointed out that if the dimension of dynamics is too low, no complicated motion is generically possible (quoting Peixoto's theorem which will be discussed later). They pointed out that in any neighborhood of a parallel flow on T^4 (i.e., 4 mode quasi periodic motion) there is a flow with a 'strange attractor.'[232] An example may be

[229] We will not discuss this equation. Its derivation from particle mechanics is not mathematically justifiable generally, and we do not know whether the equation is well-posed or not. Furthermore, its mathematical nature is quite sensitive to seemingly benign modifications. For example, if the viscosity is velocity gradient dependent (as physically natural), then the unique existence of the solution of the modified Navier-Stokes equations may be proved almost trivially.

[230] It turned out that the energy flow in the opposite direction (causing intermittency) turned out to be crucial as well.

[231] D Ruelle and F Takens, On the nature of turbulence, CMP 20 167 (1971). D. Ruelle, *Chance* tells us that they could not publish the paper (by rejections), so he decided to publish it to the journal for which he was on the editorial board.

[232] The original proposal is an attractor that is not a manifold, and its cross section is a Cantor set.

constructed as follows.

Make a diffeomorphism that maps a solid 2-torus (donut) into itself as illustrated in Fig. 21.1.

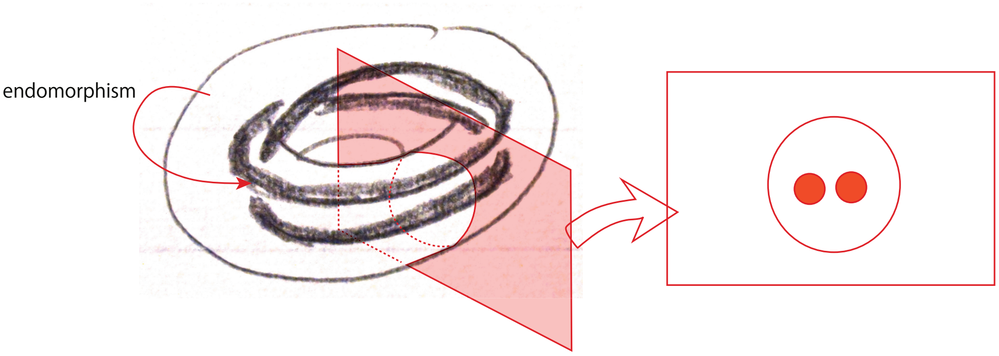


Figure 21.1:

Then, we can suspend (recall Ambrose-Kakutani **17.2**) this diffeomorphism as a flow in 4-space. By a local surgery of the parallel flow on T^4, we can embed this suspended flow into the flow on T^4 smoothly. Notice that this surgery can be as local as one wishes (i.e., the donut in Fig. 21.1 can be indefinitely small). We may conclude that in any neighborhood of T^4 quasi periodic flow is a flow with an attractor that is neither a fixed point nor a periodic orbit.

21.3 3-flow is enough to have "strange attractors"[233]

If a surgery of T^3 parallel flow (3-mode quasi periodic flow) exists, then strange attractors can exist in 3-space flows. Therefore, to show this, the key point is to make a 2D 'analogue' of Fig. 21.1, say, from a disk into itself. Such a map had been constructed by Plykin[234] (Fig. 21.2[235]):

[233] S Newhouse, D Ruelle, and F Takens, Occurrence of strange axiom A attractors near quasi periodic flows on T^m, $m \geq 3$, CMP 64 35 (1978).

[234] R V Plykin Source and sink of A-diffeomorphism of surfaces, Math Sbor 94 233 (1974).

[235] Newhouse et al., proposed different examples as well in their paper.

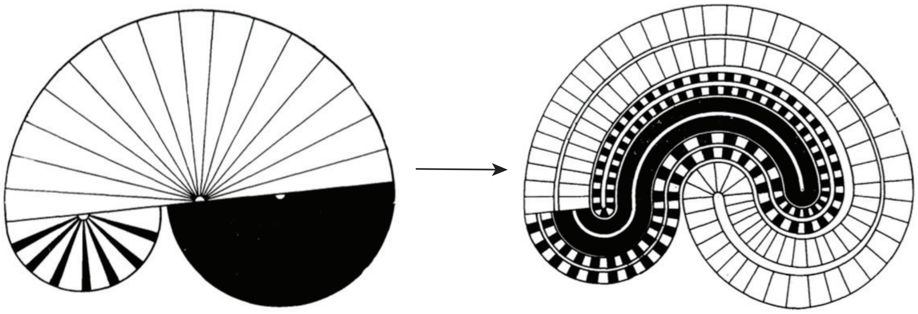


Figure 21.2: Plykin's map from a disk into itself [Fig. 1 of Plykin S Math Sbor 94 233 (1974)]

21.4 Can we really find 'chaos' indefinitely close to 3 or 4 mode quasiperiodic systems?

After reading Ruelle-Takens, I numerically studied whether it is easy to find an example, and did not see any positive result (before 1979). Later Grebogi, Ott and Yorke did an extensive study:[236]

"The results reported here suggest that if arbitrarily small smooth perturbations exist which destroy three-frequency quasiperiodicity, then they must have to be very delicately chosen and are thus unlikely to occur in practice. In particular, we believe that for a fixed typical."

21.5 Strange attractors

Here, a definition by Palis and Takens[237] is given.

For a diffeomorphism $\phi : M \to M$ a positive orbit $\{\phi^n(x)\}_{n\in\mathbb{N}}$ $(x \in M)$ is sensitive or chaotic, if there is a positive constant C (> 0) such that for any $q \in \omega(x)$ and for any $\varepsilon > 0$ $\exists n_1, n_2, n \in \mathbb{N}^+$ such that $\|\phi^{n_1}(x) - q\| < \varepsilon$, $\|\phi^{n_2}(x) - q\| < \varepsilon$ and $\|\phi^{n_1+n}(x) - \phi^{n_2+n}(x)\| > C$.

A compact set $A \subset M$ is a strange attractor, if there is an open set U with a measure zero subset $N \subset U$ such that $\forall x \in U \setminus N$, $\omega(x) = A$ and its positive orbit

[236] C Grebogi, E Ott and J A Yorke, Are Three-Frequency Quasiperiodic Orbits to Be Expected in Typical Nonlinear Dynamical Systems?, PRL 51 339 (1983)

[237] J Palis and F Takens, *Hyperbolicity & sensitive chaotic dynamics at homoclinic bifurcations* (Cambridge studies in advanced mathematics 35, Cambridge UP, 1993) p8-9. A more general discussion can be found in Section 7.2 of this book.

is chaotic.[238]

Perhaps, however, Bunimovich's definition may be easier to understand, although he does not call it a strange attractor and his definition is measure-theoretic:
(1) An invariant set A is an attractor, if there exists a neighborhood U_0 of A ($\subset U_0$) such that $U_t = T_t U_0 \subset U_0$ for $t > 0$ and $\cap_t U_t = A$.
(2) For any absolutely continuous measure μ_0 on U_0, its time evolved version μ_t weakly converges to an invariant measure λ on A that does not depend on μ_0, and $\{T_t, \lambda, A\}$ is mixing.

[238] That is, almost surely the points in U are attracted to A, and their orbits are chaotic. The exception set N is required because even for hyperbolic attractors dense set of points are attracted to periodic orbits.

22 Lecture 22. Interval maps

22.1 Interval maps = interval endomorphism

Let I be an interval (often closed). A map $f : I \to I$ is called an interval map, or an interval endomorphism. Starting from $x \in I$, we can define (at least a one-sided) sequence $\{f^n(x)\}_{n\in\mathbb{N}}$.footnote$f^n(x) = (f \circ f \circ \cdots \circ f)(x) = f(f(f \cdots (f(x)) \cdots))$. n f's show up in each expression. We have already encountered with a nontrivial examples in Figs. 18.12, 19.4 and 20.3. In these cases the relations of these maps to the original higher dimensional (often time continuous) systems are 'natural' (not very artificially contrived), so from the maps we can learn a lot about the original systems as Lorenz demonstrated.

The word 'chaos' was introduced into mathematics (and subsequently into physics) by Li and Yorke (see **22.2**). The simplest nontrivial example may be $x_{n+1} = 2x_n$ (mod 1) defined on $[0, 1]$ (see **22.4**).

Utida[239] used such a discrete systems to describe the population dynamics of insects (although empirically 'cyclic fluctuations' were observed, no simulation results were reported):

$$P_{n+1} = P_n \left(\frac{1}{b + cP_n} - \sigma \right), \tag{22.1}$$

where P_n is the population of the generation n and b, c, σ are non-negative parameters. The existence of complicate behavior in nonlinear difference equations was reviewed by May[240], who discussed

$$P_{n+1} = \lambda P_n (1 + P_n)^{-\beta} \tag{22.2}$$

and compared with some experimental and observational results (Fig. 22.1).

[239] S Utida, Population fluctuation, an experimental and theoretical approach, Cold Spring Harbor Symposia on Quantitative Biology, 22 139 (1957).

[240] R M May Simple mathematical models with very complicated dynamics, Nature 261 459 (1976)

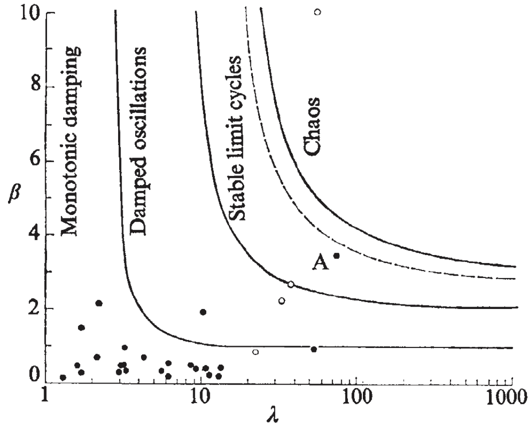


Figure 22.1: The solid lines demarcate the stability domains for the density dependence parameter β and the population growth rate λ in (22.2); the dashed line shows where 2-point cycles give way to higher cycles of period 2^n. The solid circles come from analyses of life table data on field populations, and the open circles from laboratory populations [Fig. 6 of May N 261 459 (1976)]

22.2 Period three implies 'chaos'[241]

Stimulated by Lorenz's map 19.3, Li and Yorke proved the following famous theorem. **Theorem**. Let J be an interval and let $F : J \to J$ be continuous. Assume there is a point $a \in J$ for which the points $b = F(a)$, $c = F^2(a)$ and $d = F^3(a)$, satisfy

$$d \le a < b < c \text{ (or } d \ge a > b > c).$$

Then

T1: for every $k = 1, 2, \cdots$ there is a periodic point in J having period k.

Furthermore,

T2: there is an uncountable set $S \subset J$ (containing no periodic points), which satisfies the following conditions:
(A) For every $p, q \in S$ with $p \ne q$,

$$\limsup_{n\to\infty} |F^n(p) - F^n(q)| > 0 \tag{22.3}$$

and

$$\liminf_{n\to\infty} |F^n(p) - F^n(q)| = 0. \tag{22.4}$$

[241] T-Y Li and J A Yorke, Period three implies chaos. Amer Math Month 82 985 (1975).

(B) For every $p \in S$ and periodic point $q \in J$,

$$\limsup_{n\to\infty} |F^n(p) - F^n(q)| > 0. \tag{22.5}$$

The authors added:
REMARKS. Notice that if there is a periodic point with period 3, then they will be satisfied.

The uncountable set S is called a 'scrambled set.' Their definition of chaos is the existence of a scrambled set:

A system exhibits the Li-Yorke chaos, if the system has a scrambled set.

We know S is generally non-measurable, and if measurable, it is measure zero (i.e., the inner measure of S is always zero).[242]

Thus, observable chaos (numerically detectable chaos) cannot be characterized as LI-Yorke chaos. Therefore, in these lecture notes, we do not demonstrate the theorem and will adopt a more natural definition of chaos.

22.3 Observability

Definition [Observability] We say a set B is observable with respect to a given dynamical system (f, M), if the totality of the points on the trajectories that can reach B has a positive Lebesgue measure. In other words, B is observable, if B has a basin with a positive Lebesgue measure or the set $\{x : \exists n \geq 0, f^n(x) \in B, x \in \Gamma\}$ has a positive Lebesgue measure.

In short, if you throw at the collection of what you wish to observe, and if you can hit what you wish to see with a finite probability, it is observable.

[242] Y Baba, I Kubo and Y Takahashi, Li-Yorke's scrambled sets have measure zero, Nonlinear Anal 25 1611 (1996). Smital [A chaotic function with some extremal properties, PAMS 87 54 (1983)] constructed a scrambled set S of outer measure 1 for the tent map and mentioned also that measurable scrambled sets have measure zero for this case.

If a map is only continuous (not differentiable), then there are examples of scrambled sets that are measurable and with positive measure: see, for example, I. Kan, "A chaotic function possessing a scrambled set with positive Lebesgue measure," Proc. Amer. Math. Soc. **92**, 45 (1984) or J. Smital, "A chaotic function with a scrambled set of positive Lebesgue measure," Proc. Amer. Math. Soc. **92**, 50 (1984).
V. J. Lopez, "Paradoxical functions on the interval," Proc. Amer. Math. Soc., 120, 465 (1994) proves the following: If a map f from an interval I to I is expansive, then the dynamical system cannot have a measure positive scrambled set. However, if in $I \times I$, x and y are both in the scrambled set, then the totality $\mathrm{Ch}(f)$ of $\{x, y\}$ is always measurable. Furthermore, if f is expansive and its derivative is piecewisely Lipshitz, then $\mathrm{Ch}(f)$ has a positive measure.

22.4 The simplest genuine chaotic dynamical system[243]

The purpose of this unit is to give a preview and outline of our logic with the aid of perhaps the simplest example of chaos.

A map T from $[0,1]$ into itself is defined as

$$Tx = 2x \mod 1. \tag{22.6}$$

If we use the symbol $\{r\}$ that extracts the fractional part of a real number r, we may write $Tx = \{2x\}$ (i.e., multiply 2 and then remove the integer part). Important points are summarized in Fig. 22.2.

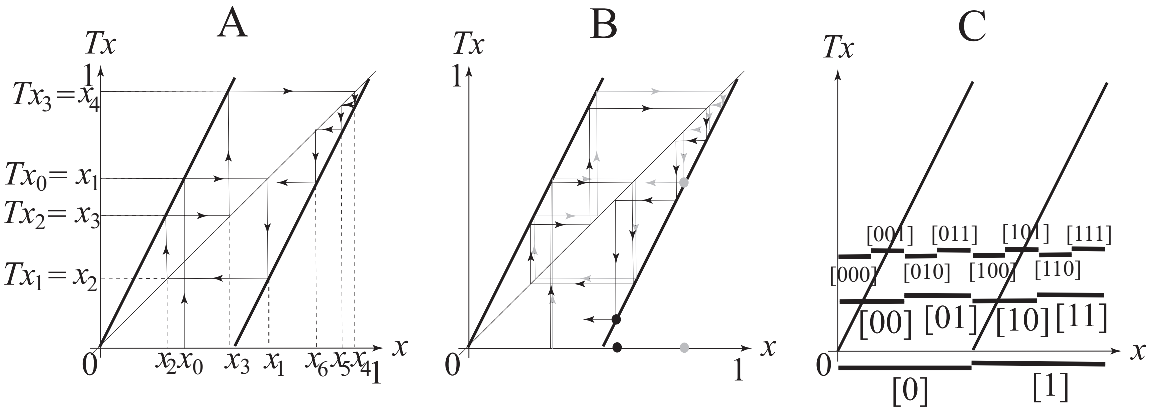


Figure 22.2: A simple map $Tx = \{2x\}$ that produces chaos.

Fig. 22.2 A simple map $Tx = \{2x\}$ that produces chaos

A: ⟪**How to chase history graphically**⟫ A method to chase the trajectory which is determined by the initial condition x_0 on the graph is illustrated. The broken line diagonal denotes $y = x$ where the values on the horizontal and vertical axes coincide. Oblique thick parallel lines denote the graph of $y = Tx$. If an initial condition x_0 is given on the horizontal axis, look vertically upward to find the point on the graph of T. Its vertical coordinate is x_1. To find x_2, we must find x_1 on the horizontal axis, and then $Tx_1 = x_2$ may be obtained just as before. To this end, with the aid of the diagonal, we can fold the vertical axis onto the horizontal axis and locate x_1 on the latter. Therefore, if we chase the vertical or horizontal lines with an arrow we can successively find $x_2, x_3, \cdots$.

B: ⟪**Exponential separation of nearby trajectories**⟫ Doubling of the gap is shown due to each

[243] YO TNW Appendix 2.1A

application of T between the gray and black trajectories that are initially very close. The positions after 6 applications of T are denoted by the gray and black small disks on the horizontal axis.
C: ⟪**Coding of trajectory**⟫ How to convert a trajectory into a 01 symbol sequence is illustrated. The interval $[0,1]$ is divided into two and each is named [0] or [1]?[0] is further subdivided into [00] and [01]. T maps [00] onto [0] and [01] onto [1]. If the subdivision of [10] is named [100] or [101] as in the figure, T maps [100] onto [00] and [101] onto [01]. If we recursively use this prescription, each point in $[0,1]$ becomes correspondent to a particular 01 infinite sequence. This is nothing but the binary expansion of a number in [0,1] (however, $[111\cdots]$ is not identified with [0]).

⟪**History is determined by initial condition**⟫ If an initial condition is given, an indefinitely long sequence $\{x_n\}$ may be constructed as $x_1 = Tx_0$, $x_2 = Tx_1 = T^2x_0, \cdots$; the future is perfectly determined by x_0. How to chase this trajectory graphically is illustrated in Fig. 22.2A.

⟪**Chaos directly connects the unknown world to our world**⟫ Fig. 22.2B illustrates how chaos expands the world of microscopic scales and connect it to the world we can directly observe. In this example, the microscopic world is doubled every time T is applied. Therefore, although the system we consider at present is a deterministic system, we lose our predictive power. Roughly speaking, for large n x_n becomes indistinguishable from a random number.

⟪**Coding trajectories or correspondence to number sequence**⟫ To see the random nature of chaotic trajectories more explicitly, their discrete coding $s_0, s_1, \cdots$ is introduced in terms of symbols s_n that take 0 or 1. It is a rather obvious transformation in this case; the interval $[0,1]$ is divided into two intervals, $[0,1/2]$ and $(1/2,1]$, and we call them, [0] and [1], respectively. Generally, $[s_0s_1,\cdots s_n]$ is defined as a set $\{x : T^kx \in [s_k] \text{ for } k = 0,\cdots n\}$, where s_k are 0 or 1. In Fig. 22.2C,[244] we see that if we apply T once to the interval, say, [011], it is mapped onto [11].
One more application of T to it coincide [1] (it is easy to see this, if we chase the end points as explained in Fig. mod2mapA). That is, [011] is a bundle of trajectories for which x_0 is in [0], x_1 is in [1] and x_2 is in [1] (Such a bundle of trajectories is called a cylinder set.).

⟪**How prediction becomes difficult**⟫ If we follow the way to construct small intervals in Fig. 22.2C, we see that, for example, the bundle of trajectories named as [0011010011] (there are 10 digits) is mapped by T successively as [011010011] → [11010011] → [1010011] → [010011] → [10011] → $\cdots$ → [011] → [11] → [1](digits are lost one by one from the left end). One more application of T makes the trajectories 'all over' $[0,1]$. The lesson we have learned is that even if the initial condition is in

[244] The reader might worry about to which interval the boundary points belong, but in this example, no careful assignment of symbols is important.

the interval [0011010011] of width $1/2^{10} \sim 10^{-3}$, after 10 consecutive T map applications we will totally lose the location information. Perhaps, the reader might guess that we will do much better if we specify the initial point (the cylinder set) more accurately. If we wish to keep some knowledge of the initial position for 20 seconds, we need the accuracy of unrealistic $1/2^{20} \sim 10^{-6}$. In short, sooner or later we will fail to predict the behavior of the system.

Even if we fail to make any prediction, it does not imply the end of the world. Beyond the predictable time range what determines x_n? The system is deterministic, so it is determined by the far right portion of the 01 sequence obtained by coding of the initial condition, which we can never know beforehand. Isn't it virtually the same as an arbitrary 01 sequence? Then, after a while, chaotic behavior would be indistinguishable from the head-tail sequence obtained by tossing a coin. This is the intuitive meaning of the definition of chaos given later.

⟪Chaos is a random deterministic behavior⟫ After all, it is a natural idea that the essence of chaos is randomness with a tint of initial condition effects due to determinism. However, without carefully reflecting on the concept of randomness, we cannot express this intuition precisely. After a rather 'heavy' preparation, eventually we will arrive at the conclusion that chaos is a phenomenon that deterministic trajectories exhibit randomness.

⟪Quantitative correspondence of chaos and randomness⟫ The reader may think the above argument provides only a qualitative characterization of chaos. However, there is a way to quantify randomness, which allows us to make a quantitative correspondence between chaos and randomness. This quantification is realized through quantifying the needed information to predict the future with a predetermined fixed accuracy. This is not hard in terms of the model being considered. How the information we know at the initial time becomes insufficient to describe the system behavior can be seen almost explicitly from Fig. 22.2C and loss of digits from the 'cylinder sets' due to the application of T. The information is lost by 1 bit every time T is applied. Suppose we wish to predict the position of the point at time t in the future with the same accuracy we describe the system now ($t = 0$). The information we must prepare now increases by 1 bit, if we push the future time t further to $t + 1$. The increasing rate of the needed information (1 bit per unit time in the present example) is called the Kolmogorov-Sinai entropy (already discussed informally). On the other hand, to describe an arbitrary 01 sequence we of course need 1 bit per digit. That is, the needed information to describe a trajectory is 1 bit per unit time. This equality of the amounts of information needed to predict the future and to describe the trajectory is a general assertion of Brudno's theorem (**24.4**). Thus, we may conclude that our definition of chaos is right on the mark.

22.5 Chaos in the logistic map

The logistic map: $[0,1] \to [0,1]$ is given by

$$F(x) = 4x(1-x). \tag{22.7}$$

We can define a dynamical system on $[0,1]$ as $x_{n+1} = F(x_n)$. Run the system for some intial conditions x_0. You will find quite erratic sequences $\{x_k\}_{k=0}^{\infty}$.

There is a very important warning: whether we can have an explicit analytical expression for x_n as a function of x_0 and n and whether the system exhibits chaotic behaviors are generally not logically related. Indeed, for the logistic map we have an explicit formula

$$x_n = \sin^2[2^n \text{Arcsin}\sqrt{x_0}]. \tag{22.8}$$

You might think that these sequences can be predicted for any far future time thanks to such analytic expressions, but try to evaluate this for three digits with $n = 100$.

There are many other examples.[245]

22.6 Coin-tossing, deterministic or not

Throwing a coin many times, we can construct a sequence of heads and tails. If we denote 'head' by 1 and 'tail' by 0, we have a 01 sequence. In this case, the phase space is $\Gamma = \{0,1\}$ and the time set is $\boldsymbol{T} = \mathbb{N}^+ \equiv \{1, 2, \cdots\}$ (i.e., discrete). As its path space Ω we may choose the totality of one-sided 01 infinite sequences $\Omega = \{0,1\}^{\mathbb{N}^+}$, because heads and tails can appear in any order. The resultant dynamical system (σ, Ω) is called the coin-tossing process. Let $\omega \in \Omega$. If $\omega(n)$ happens to be 0, this implies that the n-th outcome is a tail. Even if we know a history of this system up to time n: $\omega(1), \cdots, \omega(n)$, no one can predict anything beyond. Therefore, "the system is not deterministic."

22.4 already mentioned that the sequences we can get from 'the simplest' chaotic map $Tx = \{2x\}$ correspond (one-to-one) to all the possible outcomes of a 'coin tossing process.' You might say one is deterministic, and the other stochastic. However, there is a basic question: Can you distinguish deterministic and non-deterministic systems from their trajectories alone?

It is possible to regard the difference between the deterministic and non-deterministic dynamical systems is due to the difference of view points. Take a discrete dynamical system (one-sided), whose history starting from $\omega(0)$ $\omega \in \mathcal{D}$ may be written as

[245] K. Umeno, "Method of constructing exactly solvable chaos," Phys. Rev. E **55**, 5280 (1997).

$\omega(0)\omega(1)\omega(2)\cdots$. $\omega(t)$ is the state at time t. In our simplest example. The whole sequence $\omega(0)\omega(1)\omega(2)\cdots$ is just the binary expansion of the initial position $x \in [0,1]$, perfectly deterministic, although, of course, you cannot tell even $\omega(1)$ from $\omega(0)$ alone.

Now, let us define the shift operator (or simply, *shift*) $\sigma : \mathcal{D} \to \mathcal{D}$ as

$$(\sigma\omega)(t) = \omega(t+1). \tag{22.9}$$

The shift is a vehicle to experience the history in the chronological order. Suppose, for example, the phase space is $\Gamma = \{0,1\}$ and its one possible history ω starting from $\omega(0) = 0$: $001010011101001101\cdots$, where the leftmost number is interpreted as the state we observe at present. Its time evolution is

$$\begin{aligned}
\omega &= 001010011101001101\cdots && (22.10)\\
\sigma\omega &= 010100111010011010\cdots && (22.11)\\
\sigma^2\omega &= 101001111100110101\cdots. && (22.12)
\end{aligned}$$

The currently observed state evolves step by step. The state $\omega(t+1)$ may not be determined by the states up to time t: $\cdots, \omega(t-1), \omega(t)$, so for the observer who is observing the current and the (recorded) past states only the system behavior does not look deterministic.

Instead of interpreting the shift as a vehicle to experience a history chronologically, however, if we interpret ω as a whole history (chronicle), the shift maps one chronicle to another: $\sigma\omega = \omega_1$, where ω_1 gives the chronicle one time unit ahead of ω; indeed, for all time $t \in \boldsymbol{T}$ $\omega_1(t) = \omega(t+1)$. σ is understood as a map from $\mathcal{D}$ into itself; it defines a discrete dynamical system on the path space: $(\sigma, \mathcal{D})$. (22.10)-(22.12) illustrated its time evolution. Here, we observe the history not around present, but observe histories from God's point of view. Notice that ω completely determines ω_1. The dynamical system $(\sigma, \mathcal{D})$ is obviously deterministic.

22.7 Actual coin tossing

J. STRZALKO, J. GRABSKI, A. STEFANSKI, P. PERLIKOWSKI AND T. KAPITANIAK, Understanding coin-tossing, Math Intel 32(4), 54 (2010) illustrates the basin of attraction of the actual coin.

Fig. 22.3 Basins of attraction indicating the face of the coin which is up after the n-th collision: (a) n = 0, (b) n = 3, (c) n = 10, heads and tails are indicated in black and white, respectively. The initial conditions are $x_0 = y_0 = 0$; $\dot{x}_0 = \dot{y}_0 = \dot{z}_0 = 0$; $\varphi_0 = \psi_0 = 0$; $\theta_0 = 7$ deg, $\omega_{\zeta_0} = 0$, $\omega_{\eta_0} = 40.15$ rad/s.
Right: 3-dimensional model of the coin and its orientation in space.

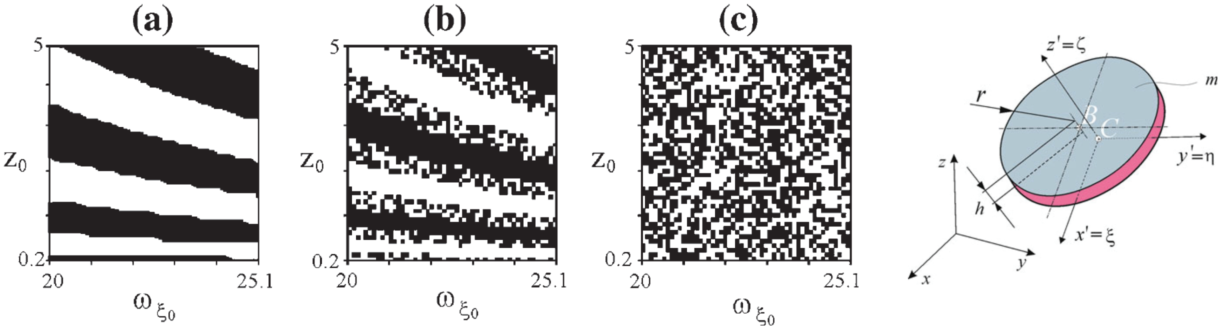


Figure 22.3: [Fig. 1,3 of STRZALKO et al., Math Intel 32(4), 54 (2010)]

22.8 No true randomness without quantum mechanics?

As you have realized, the randomness in a deterministic systems is totally in the initial conditions. Therefore, whether we can sample the intial condition 'totally randomly' or not is a crucial question as to the (un)predictability of chaos. Some people[246] assert that without quantum mechanics (so, essentially, the Born rule) there cannot be any true randomness.

Is this assertion meaningful?

22.9 Use of 'symbolic dynamics'

As we will discuss later we study a dynamical system with the aid of a certain coding system for the system *(called a symbolic dynamics). $\{\sigma, \{0,1\}\}$ is the symbolic dynamical expression (in this case very faithful: isomorphic) of $\{T, [0,1]\}$, where $Tx = \{2x\}$. Thus, we study symbol sequences, so the relevance of information coded in the history is obvious. In our simple example, all the symbols are equally probable, so are all the words (finite sequences of symbols). However, this is not generally the case, so we must consider 'natural' probabilities associated with words. Thus, we are injecting a new ingredient in our study of nonlinear systems, measures.

22.10 Definition of chaos[247]

Let f be a map from the phase space M into itself (i.e., an endomorphism of Γ).[248]

[246] For example, T. Sagawa (of (quantum) information thermodynamics). When I heard this assertion from him for the first time, I thought it is a deep idea. Is it really so?

[247] The following several units are taken from TNW Chapter 2.

[248] In this definition, maps are not understood measure-theoretically; they are pointwise transfor-

Choose $n \in \mathbb{N}$ and construct f^n that applies f n-times,[249] and restrict it to its invariant set $A \subset M$ ($f^{-n}(A) = A$). If it is isomorphic to the shift dynamical system $(\sigma, \{0,1\}^{\mathbb{N}})$, then we say the dynamical system (f, M) exhibits chaos (or the system is chaotic). That is, if φ is a one to one map and the following diagram becomes commutative,[250] we say the dynamical system (f, M) exhibits chaos.

$$\begin{tikzcd}
A \arrow[r, "f^n"] \arrow[d, "\varphi"] & A \arrow[d, "\varphi"] \\
\{0,1\}^{\mathbb{N}} \arrow[r, "\sigma"] & \{0,1\}^{\mathbb{N}}
\end{tikzcd}$$

In words, if an appropriate one-to-one coding scheme φ of points (i.e., states) in A (we can decode uniquely the sequences) allows to transform the original dynamical system (restricted to A) into the full shift on two symbols, we say the dynamical system is chaotic.

Roughly speaking, if the behavior of a dynamical system (restricted on an invariant set) is coded with symbols 0 and 1, and the result cannot be distinguished from the totality of the outcomes of the coin-tossing process, we wish to say the system is chaotic or is a chaotic dynamical system.

A continuous time dynamical system is chaotic, if a discrete dynamical system constructed from the original system in a 'natural' fashion is chaotic in the above sense.

The chaos defined here is sometimes called formal chaos, because there is no guarantee of its observability **22.3**. For chaos of one-dimensional map systems, the observability of chaos and the existence of an absolutely continuous invariant measure seem to be equivalent.

22.11 How good is the definition of chaos given above?

It is not easy to find criteria for a definition to be good, but consistency with intuition, close connection to fundamental concepts, equivalence to definitions based on very different points of view, etc., may be counted among them. Our definition relies on our intuition: as long as we assume that "(apparent) random behavior is fundamentally important characteristic of chaos," consistency with intuition is built into the definition. Randomness must be a fundamental concept, so the definition

mations.

[249] $f^n(x) = (f \circ f \circ \cdots \circ f)(x) = f(f(f \cdots (f(x)) \cdots))$. n f's show up in each expression.

[250] This means in the present context that the results obtained by following various combinations of arrows never yield different results.

has a close connection with a fundamental concept. This can be seen further from the theorem we demonstrate later. It is also important to check the consistency and relations with other definitions of chaos to confirm the naturalness of the definition.

22.12 Sensitive dependence on initial conditions

As can be seen from (22.8) and as emphasized by Ruelle, Guckenheimer[251] and others, magnification of small effects was regarded as a fundamental significance of chaos (the "butterfly effect"). Intuitively, this is easy to understand in terms of (one-sided) shifts, because, in contrast to the digits on the far right that describes minute differences in states, the digits near the left end correspond to global differences. Thus, movement of the digits to the left by shift corresponds to magnification of small structures. Two initial states whose codes are different only in digits on the far right are very close in the phase space, and time evolution magnifies the difference.

Sensitive dependence on initial conditions, however, does not necessarily imply that the dynamical system is chaotic. This must be obvious to those who know the roulette; the ball jumps around awhile, but eventually it settles down to a fixed point (the system is a multiply stable system). If several attractors coexist, the ultimate fate of the system is determined by which basin its initial condition lies. Even if the long-time (eventual) behavior of the system is not chaotic, if the boundaries between basins are extremely complicated, then sensitive dependence on initial conditions can exist.[252] Obviously, roulettes, dice, and coins must be (at least approximately) such systems. Sensitive dependence on initial conditions is not enough to characterize chaotic dynamical behavior.[253]

22.13 Apparently different definitions of chaos

S.-H. Li proposed a revised version of scrambled set called ω-scrambled set to make the Li-Yorke chaos equivalent to our chaos.[254]

The most popular definition of chaos at present may be due to Devaney.[255] If we take into account S.-H. Li's result,[256] this definition may be stated as

[251] J. Guckenheimer, "Sensitive dependence on initial conditions for one-dimensional maps," Commun. Math. Phys., **70**, 133 (1979).

[252] See, for example, H. E. Nusse and J. A. Yorke, "Basin of attraction", Science **271**, 1376 (1996).

[253] However, if we require sensitivity to perturbation at most instants may characterize chaotic systems.

[254] S.-H. Li, "ω-chaos and topological entropy," Trans. Amer. Math. Soc. **339**, 243 (1993).

[255] R. Devaney, *An introduction to chaotic dynamical systems* (Benjamin/Cummings, 1986).

[256] S.-H. Li, "Dynamical properties of the shift maps on the inverse limit space," Ergodic Theor. Dynam. Syst. **12**, 95 (1992). See also J. Banks, J. Brooks, G. Cairns, G. Davis and P. Stacey, "On

follows. Let (f, X) be a discrete time dynamical system, and D ($\subset X$) be a closed invariant set (i.e., $f^{-1}(D) \supset D$). If the following two conditions hold, the dynamical system exhibits chaos.
(D1) $f|_D$ (the restriction of f to D) is topologically transitive on D (i.e, $f|_D$ is surjective on D and has an orbit dense in D),
(D2) The totality of the periodic orbit of f is dense in D.
D is called a chaos set. The set A in **22.10** is a chaos set. S.-H. Li showed the equivalence of this definition and our definition, if the phase space is a compact metric space.

Already in 1968[257] Alekseev defined quasirandom dynamical systems in terms of a Markov chain with a positive Kolmogorov-Sinai entropy. It is an example that Russian dynamical systems study was far ahead of the Western counterpart (actually, the use of entropy to classify dynamical systems was the Russian starting point of the 'modern' dynamical systems study). The term 'chaos' may have been good for popularization of the concept, but the term 'quasirandom' summarizes the essence.

22.14 Period $\neq 2^n$ implies chaos

Theorem[258] Let I be a finite interval and $F : I \to I$ be a continuous endomorphism. F exhibits chaos if and only if F has a periodic orbit whose period is not equal to the power of 2.

Remark We have already mentioned the Li-Yorke chaos **22.2**, but the chaos in the above theorem is distinct from the Li-Yorke chaos. The most important distinction is, as note below, with slight weakening of the 'one-to-one' correspondence with A chaotic behaviors are observable; we can choose the set A in **22.10** observable (e.g., measure positive). That is, in contradistinction to Li-Yorke chaos whose 'core feature' = the scrambled set is never observable, our chaos is closely tied to observability, when chaos is observable. We can show that if the system is chaotic in our

Devaney's definition of chaos," Am. Math. Month. **99**, 332 (1992). In the latter X need not be bounded.

[257] [1968: Prague Spring, The Tet Offensive, Down to the Countryside Movement, Nuclear nonproliferation treaty, Assassinations of Rev. M. L. King and Senator R. F. Kennedy.]

[258] Y. Oono, Period $\neq 2^n$ implies chaos, Prot. Theor. Phys., 59, 1029 (1978). "In the present Letter, we give another definition of chaos which is not directly related to the nonperiodicity of the solution, and sketch the proof of the theorem asserting that period other than 2" implies chaos (Theorem 1). The assertion also holds even if the definition of chaos by Li, Yorke and Nathanson [J. Combinatorial Theor. (A) 22 61 (1977)] is adopted (Theorem 2)." However, the proof in this letter is not very elegant.

sense, the system is also Li-Yorke chaotic, BUT the converse is not generally true. Furthermore, even when chaos is observable, it is not on the scrambled set.

We will show the following theorem:

22.15 General theorem for chaos of C^0-endomorphisms of intervals

Theorem[259] Let I be a finite interval and $F : I \to I$ be a continuous endomorphism. Then, the following (1)-(4) are equivalent.[260]
(1) F exhibits chaos.
(2) F has a periodic orbit whose period is not equal to the power of 2.
(3) There is a positive integer m such that F^m has a mixing invariant measure.
(4) F has an invariant measure whose Kolmogorov-Sinai entropy is positive.

An intuitive explanation of the theorem is given here.

"Invariant measure" is a steady distribution.

"Mixing" implies that the system relaxes toward some steady state.

"Kolmogorov-Sinai entropy" is the required extra information to predict the next time step state as accurately as the current state. Its positivity implies that (since more information is needed to determine the future state) the system behavior in the future becomes increasingly difficult to predict as the future is further away.

Practically, the following Proposition equivalent to (1)-(4) in the theorem is useful:
(5) There are two closed intervals J_1 and J_2 in I that share at most one point such that $f^p(J_1) \cap f^q(J_2) \supset J_1 \cup J_2$ holds for some positive integers p and q.

That Ito's earthquake model in Section 2.1 exhibits chaos is immediately seen from the 'folded paper' model (Fig. 18.12). We may draw a periodic orbit whose period is not a power of 2, but to check (5) may be the easiest.

Is A observable? The set A constructed in the proof is not observable, since it is measure zero and not attractive. However, if we ignore the 1 to 1 correspondence on a measure zero set, we can make A to be an interval (very often, especially when chaos is observable).

22.16 (2) implies (5)

[259] M. Osikawa and Y. Oono, "Chaos in C^0-diffeomorphism of interval," Publ. RIMS **17**, 165 (1981); Y. Oono and M. Osikawa, "Chaos in nonlinear difference equations. I," Prog. Theor. Phys. **64**, 54 (1980) (There is no part II, since the author was 'expelled' from Japan in 1979). Also see: L. Block, "Mappings of the interval with finitely many periodic points have zero entropy," Proc. Amer. Math. Soc. **67**, 357 (1978); "Homoclinic points of mappings of the interval," *ibid.* **72**, 576 (1978).

[260] (1) ⇒ (2), (3), (4) is trivial. (4) ⇒ (1) is also almost trivial.

If F has a periodic orbit whose period is not equal to the power of 2, then there are two closed intervals I_0 and I_1 in I that share at most one point such that $F^p(I_0) \cap F^q(I_1) \supset I_0 \cup I_1$ holds for some positive integers p and q.

[Demo][261]

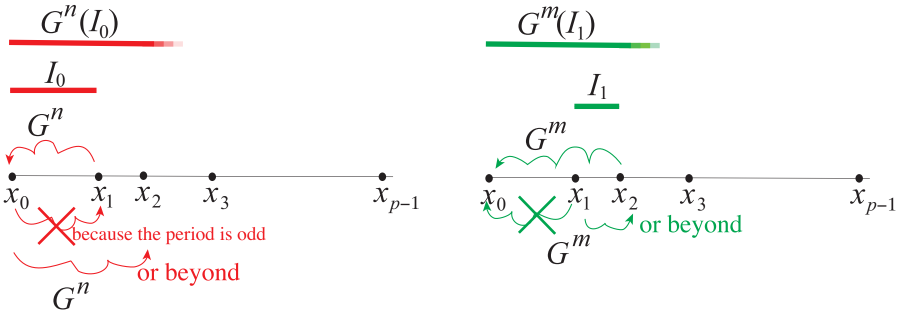


Figure 22.4: Illustrated proof of (2) ⇒ (5). (Due to possible folding, the open of kernels of the images of the intervals may well contain $I_0 \cup I_1$.)

$G \equiv F^{2^n}$ has a periodic orbit with odd periodicity p: $O_p = \{x_0, x_1, \cdots, x_{p-1}\}$, where $x_0 < x_1 < \cdots < x_{p-1}$ (Needless to say, the orbit is not necessarily chronologically in this order).

Let $I_0 = [x_0, x_1]$ and $I_1 = [x_1, x_2]$. There must be an integer n $(0 < n < p)$ such that $G^n(x_1) = x_0$. For this n $G^n(x_0) \neq x_1$; otherwise, $G^{2n}(x_1) = x_1$, that is, $p = 2n$.[262] Therefore, $G^n(x_0) \geq x_2$. That is, $G^n(I_0) \supset I_0 \cup I_1$.

There must be an integer m $(0 < m < p)$ such that $G^m(x_2) = x_0$. For this m obviously $G^m(x_1) \neq x_0$ nor $G^m(x_1) \neq x_1$, so $G^m(x_1) \geq x_2$. That is, $G^m(I_1) \supset I_0 \cup I_1$.

22.17 (5) implies (1)

If there are two closed intervals I_0 and I_1 in I that share at most one point such that $F^p(I_0) \cap F^q(I_1) \supset I_0 \cup I_1$ holds for some positive integers p and q, then F exhibits chaos.

[Demo]

The key to this assertion consists of two parts:

(i) If there are two disjoint closed intervals I_0 and I_1 in I such that $F^n(I_0) \cap F^n(I_1) \supset I_0 \cup I_1$ for some positive integer n, then F exhibits chaos.

(ii) (5) implies the existence of I_0 and I_1 required by (i).

Let us prove (i) first. Set $G = F^n$.

[261] Proof due to Hamachi.

[262] Note that $2n < 2p$, the smallest even number that is a multiple of p.

First we note the following elementary fact: (ia) Let I_0 be a closed interval, and $G(I) \supset I_0$ for a continuous function G. There there is a closed interval $Q \subset I$ such that $G(Q) = I_0$. It is essentially the intermediate value theorem.

Next, we note that (ib) if $G(I) \supset I_0, I_1$, which are disjoint closed intervals, then we can find $G(Q_i) = I_i$ ($i = 1$ or 2) such that $Q_i \subset I$ and $Q_1 \cap Q_2 = \emptyset$. The existence of Q_1 and Q_2 in I follows from the preceding elementary fact. If $Q_1 \cap Q_2$ were not empty and had x in it, then $f(x) \in I_0 \cap I_1$, contradicting the disjointness of I_0 and I_1

We can recursively construct a closed interval sequence $\{I_{a_1a_2\cdots a_n}\}_{n=1}^{\infty}$ where $a_i \in \{0, 1\}$ as

$$G(I_{a_1a_2\cdots a_n}) = I_{a_2\cdots a_n}, \tag{22.13}$$

where

$$I_{a_1a_2\cdots a_{n-1}a_n} \subset I_{a_1a_2\cdots a_{n-1}} \tag{22.14}$$

with the aid of (ia).

If s and s' ($\neq s$) are the length n 01 sequences. Then (ib) tells us $I_s \cap I_{s'} = \emptyset$.

Let $b \in [0, 1)$ and its binary expansion $b = 0.a_1a_2\cdots a_n\cdots$. We can define I_b as

$$I_b = \cap_{n=1}^{\infty} I_{a_1a_2\cdots a_n}, \tag{22.15}$$

which is closed nonempty set thanks to (22.14) and if $b \neq b'$, $I_b \cap I_{b'} = \emptyset$. Thanks to (22.13)

$$G(I_b) = I_{\{2b\}}, \tag{22.16}$$

where $\{x\}$ is the fractional portion of a real x (i.e., $2b = a_1.a_2\cdots a_n\cdots$, so $\{2b\} = 0.a_2a_2\cdots a_n\cdots$)

Define $A = \cup_{b\in[0,1)} I_b$. Introduce an equivalence relation $\sim$ on A as $x \sim x'$ iff $x, x' \in I_b$ for some $b \in [0, 1)$. Then, there is a standard surjection $\tau : A \to A/\sim$, and if we define $G^* = \tau \circ G \circ \tau^{-1} : A/\sim \to A/\sim$, it is isomorphic to the (one-sided) Bernoulli shift on $\{0, 1\}$ of the coin-tossing process.

We must prove (ii). Let us take an inelegant but the simplest path. If we take $m = pq$, obviously

$$F^n(I_0) \cap F^n(I_1) \supset I_0 \cup I_1. \tag{22.17}$$

Let $G = F^m$ and construct $I_{a_1a_2}$ as in the proof of (i). They must be on I, so we can always choose a pair $I_{a_10} = J_0$ and $I_{a_1'1} = J_1$, where $a_1, a_1' \in \{0, 1\}$m, such that they are not adjacent. Then, if we choose $G^2 = F^{2m}$ (i.e, $n = 2pq$)

$$G^2(J_0) \cap G^2(J_1) \supset J_0 \cup J_1. \tag{22.18}$$

Remark Actually, we can prove the odd periodicity lemma:
If G has a periodic orbit of odd periodicity, then there exists two disjoint closed intervals I_0 and I_1 such that $G^2(I_0) \cap G^2(J_1) \supset I_0 \cup J_1$.

Therefore, if F has a periodic orbit of period $2^n p$, $F^{2^{n+1}}$ has such intervals.

22.18 (1) (2) and (5) are equivalent

This is shown via (1) ⇒ (2). Then, 22.16 closes the demonstration with 22.17.

22.19 (1) and (3) are equivalent

There is an F^m invariant mixing measure μ for some positive integer k:

$$\lim_{n \to infty} \mu(A \cap F^{-nm} B) = \mu(A)\mu(B), \tag{22.19}$$

if and only if F is chaotic.[263]

22.20 (1) and (4) are equivalent

(4) ⇒ (1) is shown via (2). The converse is generally hard to prove (proved with H Takahashi's results with his help).

22.21 Time correlation function of chaos

Let x_n be the state at time step n. The, the time correlation function is defined as (here we assume the time average of x_n vanishes or $x_n - \langle x_n \rangle$ is considered)

$$c(n) = \frac{1}{N} \sum_{k=0}^{N-1} x_{n+k} x_k. \tag{22.20}$$

We are interested in this quantity when the system allows a steady state.
If the map (or dynamics) allows an ergodic invariant measure μ,[264],

$$c(n) = \int d\mu(x_0) x_n x_0 \quad \left[\text{or} = \int d\mu(\omega) x_n(\omega) x_0(\omega) \text{ is better}\right]. \tag{22.21}$$

$c(n) \le c(0)$, because the average of $(x_n - x_0)^2$ must be non-negative (use stationality). Therefore, we often normalize this with $c(0)$:

$$C(n) = \frac{\langle x_n x_0 \rangle}{\langle x_0^2 \rangle}. \tag{22.22}$$

[263] This is a special assertion generally true for C^0-endomorphism of an interval.

[264] We assume it is observable; thus, it is almost surely an absolutely continuous invariant measure.

If $C(n) \to 0$ in the large n limit, we say the system is mixing,[265] where $\langle\ \rangle$ implies time or invariant measure average.

We usually think that the faster the time correlation function decays, the more irregular/disordered/chaotic is the dynamical system. However, the situation is not that simple, because isomorphisms of dynamical systems can change the decay rate (or even its algebraic nature), but the KS entropy stays the same.[266]

For the standard tent map its time correlation decays to zero at time 1 (!), but for the logistic map, its decay is not that quick (though exponential). Both are mixing and the KS entropy is log 2, the same.

22.22 Power spectrum

Experimentally (and practically), the time correlation function is computed via its Fourier transform called the power spectrum:

$$\sigma(\nu) = \sum_n C(n) e^{-2\pi i n\nu}. \tag{22.23}$$

This strategy is advantageous, because the Fourier transform $\hat{x}(\nu)$ of a signal can be efficiently computed with the aid of the Fast Fourier Transform algorithm. The square (average) of the $|\hat{x}(\nu)|^2$ is the power spectrum, and its inverse Fourier transform is the time correlation function thanks to the Wiener-Khinchin theorem **22.23**.

22.23 Wiener-Khinchine theorem[267]

Let $x_n(\omega)$ be a sampled trajectory with summable stationary time correlation function $C(n) = \langle x_n x_0 \rangle$ (we normalize the signal). Let us compute its power spectrum $\sigma(\nu)$.

We compute the Fourier transform of the signal as

$$\hat{x}(\nu) = \sum_{-\infty}^{\infty} x_n e^{-2\pi i n\nu}. \tag{22.24}$$

Let us compute

$$\langle \hat{x}(\nu)\hat{x}(\nu') \rangle \;\; = \;\; \sum_n \sum_m C(n-m) e^{-2\pi i n\nu - 2\pi i m\nu'}, \tag{22.25}$$

[265] A more official definition of mixing will be given later, when we discuss ergodicity.

[266] The mixing property is isomorphism invariant, so the correlation is guaranteed to decay eventually to zero.

[267] The needed normalization is not carefully traced.

$$= \sum_m \sum_p C(p) e^{-2\pi i(\nu+\nu')m - 2\pi i \nu p}, \tag{22.26}$$

$$= \delta_{\nu,\nu'} \sum_p C(p) e^{-2\pi i \nu p} = \delta_{\nu,\nu'} \sigma(\nu). \tag{22.27}$$

That is,

$$C(n) = \sum \sigma(\nu) e^{2\pi i n \nu}. \tag{22.28}$$

This is called *Wiener-Khinchine's theorem.*

This is a practically very important theorem, because the equality (22.28) is the most convenient method to compute the correlation function: the power spectrum is easy to compute numerically from the original data thanks to the fast Fourier transformation (FFT). If you have never heard of the FFT, study the Cooly-Tukey algorithm.[268] 32B.12 of

https://www.dropbox.com/home/ApplMath?preview=AMII-32+FourierTransformation.pdf

explains the principle of FFT.

Note that the time correlation function is positive definite in the sense appearing in Bochner's theorem **22.25**). This implies that if the correlation function is continuous at the origin and if normalized as $C(0) = 1$, it is a characteristic function of a certain probability distribution.

We know that this 'certain function' is the corresponding power spectrum. Therefore, the power spectrum may be interpreted as a probability distribution function (of a certain quantity). Then, the information minimization technique may be used to infer or model the time correlation function.

22.24 Long correlation means sharp spectrum

As we have learned, the power spectrum is easy to observe,[269] so it is advantageous to have some 'feeling' about the relation between correlation functions and power spectra.

A summary statement is:

"A signal that has a short time correlation has a broad band."

This is intuitively obvious, because short time correlation means rapid changes that must have high frequency components.

There may be several mathematical statements that are related to the above assertion. Perhaps the most elementary is to look at

$$C(t) = e^{-\alpha|t|} \;\leftrightarrow\; \sigma(\omega) = \frac{2\alpha}{\omega^2 + \alpha^2}. \tag{22.29}$$

[268] As usual, Gauss used this method long before Cooly and Tukey.

[269] Simply turn on a spectrum analyzer.

That is, the half-width of the power spectrum is the reciprocal of the 'life-time' of the signal. (The power spectrum with the Cauchy distribution is often called the *Lorentzian spectrum*)

The result can be understood intuitively with the aid of dimensional analysis $\tau\omega \sim 1$.

Exercise 1. The *Riemann-Lebesgue lemma* tells us that $\lim_{|\omega|\to\infty} \sigma(\omega) = 0$ for time-continuous signals. Learn about this lemma. A related statement was used in the demo of the KAM theorem.□

Exercise 2. The Riemann-Lebesgue lemma also tells us that the decay rate of the power spectrum for large $|\omega|$ gives information about the smoothness of the signal.

Demonstrate that if the signal is m-times continuously differentiable with respect to time, $\sigma(\omega) = o(\omega^{-2m})$ for large ω.

22.25 Bochner's theorem.

A positive definite function φ on $\boldsymbol{R}^n$ that is continuous at the origin and $\varphi(0) = 1$ is a characteristic function of some probability measure on $\boldsymbol{R}^n$.□

Here, 'positive definite' means the following: for any $n \in \mathbb{N}$, for any $a_i \in \mathbb{C}$ $(i = 1, \cdots, n)$ and for any 'time points' $t_i \in \mathbb{R}$

$$\sum_{i,j\le n} a_i \overline{a_j} \varphi(t_i - t_j) \ge 0. \tag{22.30}$$

22.26 Sarkovskii's theorem[270]

Sharkovsky ordering of the set of natural numbers is given by the following ordering

$$\begin{aligned} &3 \prec 5 \prec 7 \prec 9 \prec \cdots \prec 3\cdot 2 \prec 5\cdot 2 \prec 7\cdot 2 \prec 9\cdot 2 \prec \\ &\quad \cdots \prec 3\cdot 2^2 \prec 5\cdot 2^2 \prec 7\cdot 2^2 \prec 9\cdot 2^2 \prec \cdots \prec 2^3 \prec 2^2 \prec 2 \prec 1. \end{aligned} \tag{22.31}$$

Let I be either the real line or an interval and $f : I \to I$ be a continuous map. Let us say $a \prec b$, if a precedes b in the Sharkovskii ordering. The three parts of the full Sharkovsky Theorem are:

[270] A. N. Sarkovskii, "Coexistence of cycles of a continuous map of a line into itself," Ukr. Mat. Z. 16, 61-71 (1964) [See P. Stefan, "A theorem of Sarkovski on the existence of periodic orbits of continuous endomorphisms of the real line," Commun. Math. Phys. 54, 237 (1977) as well]. M. Misiurevicz, "Remarks on Sharkovsky's theorem," Am. Math. Month. 104 864 (1997) is a good summary. Here, the summary follows the third paper.

Theorem 1. Let $f : I \to I$ be a continuous map. If f has a cycle of period n and if $n \prec k$, then f has a cycle of period k.
Theorem 2. For every k there exists a continuous map $f : I \to I$ that has a cycle of period k, but has no cycles of period n for any $n \prec k$.
Theorem 3. There exists a continuous map $f : I \to I$ that has a cycle of period 2^n for every n and has no cycles of any other periods (the existence of the critical map). [This part is actually included in Theorem 2.]

Sharkovskii's theorems are beautiful theorems, and tell us clearly the existence of the critical map. Also they tell us that having a periodic orbit of odd period is the essence of the complicated trajectories as Hamachi's proof of the key lemma **22.16** exploits. However, the theorems tell us virtually nothing about chaos, so I will not give a proof.

23 Lecture 23. Randomness

23.1 Summary so far

Chaos, according to my definition, is characterized (defined) by its 'close relation' to randomness. How 'good' is this characterization? Consistency with intuition, close connection to fundamental concepts, equivalence to definitions based on very different points of view, etc., may be signs of goodness. As long as we assume that "(apparent) random behavior is fundamentally important characteristic of chaos," consistency with intuition is captured. Randomness must be a fundamental concept, so the characterization has a close connection with a fundamental concept.[271]

What is then 'randomness'?

23.2 Random number table that is not so random

To have some 'feeling' about 'randomness', let us look at the random number table.

What is a random number table? Intuitively, it is a table on which numbers are arranged without any regularity. Whether there is regularity or not is checked actually by various statistical tests, so a random number table is a table tabulating a number sequence that is recognized to have no regularity, passing all the statistical tests.[272]

As an example, take the well-known Kitagawa random number table.[273] The Kitagawa table was constructed as follows:[274]

(i) On each page of the random number table of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver and Boyd, 1938),[275]

[271] It is also important to check the consistency and relations with other definitions of chaos to confirm the naturalness of the definition. The comparison is already summarized in **22.13**.

[272] For example, see A. L. Rukhin, "Testing randomness: a suite of statistical procedures," Theory Probab. Appl., **45**, 111 (2001). The author's math mentor Professor H. Watanabe once said,"Random number is like God. Its existence might be admissible, but, if you are shown 'this is Him,' it is quite doubtful."

[273] See T. Kitagawa, *Statistical Inference* (Suisoku-Tokei Gaku) (Iwanami, 1958).

[274] in O. Miyataka and T. Nakayama, *Monte Carlo Method* (Nikkan-Kogyo, 1960)

[275] The Fisher-Yates table was constructed as follows: from the 20 digit logarithm table (of A. J. Thompson, *Logarithmetica Britannica: Logarithms to 20 Decimal Places 10,000-100,000* (This work of Dr. Thompson's is an attempt to commemorate in a worthy manner the first great table of common logarithms, which was computed by Henry Briggs and published in London in 1624. It brings together the series of nine separate parts, issued between 1924 and 1952 from University College, London, in Karl Pearson's Tracts for Computers series; reprinted from Cambridge University Press, 2008), 15,000 digits were selected randomly, and then they were randomly arranged.

two consecutive numbers are paired, and
(ii) The resultant 25 × 50 pairs were scrambled.
(iii) Then, the columns on different pages of the resultant scrambled table were exchanged.
(iv) Then, the following statistical tests were performed to each page, and the best 4 pages were kept:

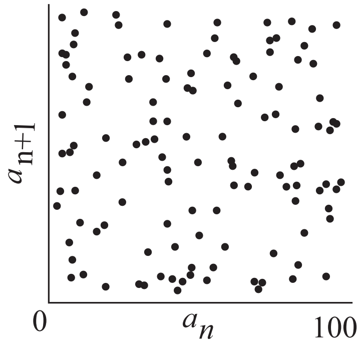
(1) Frequency test: using the χ^2-test, to check whether all the digits appear equally frequently,

(2) 'Pair' test: the table is considered as the table of two consecutive number pairs, and the frequencies of the pairs were tested just as in (1).

(3) Poker test: the frequencies of the various patterns of the consecutive, e.g., 5 digit blocks (say, *abcde, abacd, aabac*, etc.) were tested.

(4) Gap test: Reading the table column-wisely, the spacing between consecutive identical digits were tested.

Is the Kitagawa table random? As illustrated in Fig. 23.1 it fails a simple test: numbers tend to change oscillatingly:



3	0	2	1	0	2	0	2	2	1
3	0	2	1	0	1	2	2	3	0
2	1	1	1	3	1	1	1	0	0
1	1	0	3	0	0	1	2	0	3
1	1	1	2	1	1	0	0	1	2
2	1	1	2	1	0	3	1	2	1
3	0	1	1	0	0	2	1	3	2
1	2	0	1	2	1	0	0	1	1
2	0	0	1	1	1	0	1	1	0
2	1	1	2	3	2	3	1	1	1

Figure 23.1: Shown on the Left is the first 5 rows of the Kitagawa table converted into a sequence of two digit number sequence $a_1, a_2, \cdots$, and then a_{n+1} is plotted against a_n. The square is divided into 100 square boxes of 10 × 10, and the numbers found in each square is counted as shown in the Right. If the sequence is random, the points must be distributed uniformly on the square, so in the right figure, we expect 79 points inside the square and 45 on the periphery. Actually, there are 65 inside and 59 outside, failing the uniformity test with the P value less than 0.5 %. In short, $\{a_n\}$ HAS a tendency to change in an oscillatory fashion.

Incidentally, it is said that, when a person is asked to write down as random a number sequence as possible, the randomness of the produced number sequence and

However, the digit 6 appeared slightly more frequently than others, so 50 of them were randomly selected and replaced with other digits randomly.

the intelligence of its writer are positively correlated.

The best random number is supposedly the natural random number that can be made by counting the number of, e.g., β-decay. Thus, we might say that without quantum phenomena there is no truly random numbers.

23.3 Featurelessness = randomness?

Statistical tests check the non-existence of particular patterns in the number sequence. The Kitagawa table was disqualified, because it has a significant (or detectable) pattern. Therefore, it seems to be a good idea to declare that the sequence without any feature (lawlessness) is a random number sequence. However, within our usual logical system that admits the exclusion of middle (i.e., there are only two possibilities A or non-A), 'that there is no feature' becomes a respectable feature;[276] we fall into an impasse that the random number sequence is a sequence with a characteristic that it has no characteristics.

von Mises (1883-1953) wanted to systematize probability theory based on randomness, but it was difficult because formalizing featurelessness or lawlessness was difficult. However, if we could positively characterize the feature that there is no feature, in other words, if we can define 'being without features' by an explicitly specifiable property, then 'there is no characteristic feature' is no more the negation of 'there are characteristic features.' Still, the characteristics such as 'lawlessness' or 'featurelessness" are ambiguous, allowing various interpretations.

23.4 Featurelessness = incompressibility?

Suppose we can find a feature (regularity) in a number sequence, we could save the phone charge by exploiting the regularity when we wish to send it to the second person. For example, if we wish to send $10101010\cdots10101010$ that has one million 10's, it is far better to send the message, "repeat 10 1,000,000 times," than to send the raw sequence itself. The number of digits required to describe a number N is asymptotically proportional to $\log N$, so such a regular sequence may be sent with the cost proportional to the logarithm of the original message length.[277] This is, of course, far more money-saving than the raw message.

[276] Of course, we must understand what 'features' mean.

[277] A student suggested that to send the number $\log N$, we could take its logarithm to compress it further and save money. Is it a good idea, or what is wrong?

Let us consider another example:

```
0273900749 7297363549 6453328886 9844061196 4961627734 4951827369 5588220757 3551766515
8985519098 6665393549 4810688732 0685990754 0792342402 3009259007 0173196036 2254756478
9406475483 4664776041 1463233905 6513433068 4495397907 0903023460 4614709616 9688688501
4083470405 4607429586 9913829668 2468185710 3188790652 8703665083 2431974404 7718556789
3482308943 1068287027 2280973624 8093996270 6074726455 3992539944 2808113736 9433887294
0630792615 9599546262 4629707062 5948455690 3471197299 6409089418 0595343932 5123623550
```

This sequence may look random, but it is the 480 digits starting from the 10,501st digit of π. If we wish to send this sequence, we can send a message, "the $\underline{480}$ digits starting from the 10,501st digit of π," and it is already shorter than the original message.[278] In this case as well, the message we must send is asymptotically proportional to the length of the part specifying the length of the sequence (the underlined part).

If we can compress the message, it is obviously non-random. Therefore, can we characterize the randomness of a message by the fact that it cannot be compressed (made shorter for communication) however we may try?

23.5 Algorithmic randomness

To compress a given sequence of symbols we must use its regularity and meaning, so whether we can recognize them or not is the key issue of the randomness of the sequence.[279] However, by whom should the regularities be recognized?

The basic idea of the algorithmic randomness due to Solomonov (1926-), Kolmogorov and Chaitin (1947-) is that this recognition should be done by the most powerful computer. Then, basic questions arise such as 'What is the most powerful computer?" and, in the first place, "What is a computer?"

A computer is a machine to perform computation. For this statement to make sense, we must know what 'computation' is.[280] Computation is to process a number into another number. The process of computation is not haphazard, but is understood to obey strictly certain rules. Therefore, we may say intuitively that

[278] However, the receiver of the message must perform a considerable procedure to obtain the message actually. Compressed information is often costly to expand. An interesting topic related to this is D. Bailey, P. Borwein and S. Plouffe, "On the rapid computation of various polylogarithmic constants," Math. Computat. **66**, 903 (1997) (`http://www.cecm.sfu.ca/ pborwein/`) and V. Adamchik and S. Wagon, "π: a 2000-year search changes direction," Mathematica in Education and Research, **5**(1) 11 (1996), D. H. Bailey, J. M. Borwein, P. B. Borwein, and S. Plouffe, "The quest for Pi," Math. Intelligencer **19**(1) 50 (1997).

[279] However, we do not discuss the vague concept called 'meaning.'

[280] What is a 'machine'?

computation is to transform one finite number sequence into another, using finitely many definite procedures.

23.6 Church's characterization of computation in plain terms

If we use the expression everyone understands by now, Church(1903-1995)'s proposal is essentially as follows. A function whose program can be accepted by a digital computer is a "partial recursive function,"and if the computation specified by the program is guaranteed to be completed within a finite time, the function is a 'recursive function' = computable function. Computation is a process to obtain the value of a computable function.

23.7 Number-theoretical functions

We consider only computation of a number with finitely many digits without any roundoff errors. Consequently, we can understand computation as a map: $\mathbb{N}^k \to \mathbb{N}$ ($k \in \mathbb{N}^+$). Such a map is called a number-theoretic function or arithmetic function. We consider only the computation of number-theoretic functions.

23.8 Church's basic idea

First, take a few functions that everyone intuitively accepts to be obviously computable as the starting point (see **23.10**). The totality of computable functions is constructed from these starting functions with a finite number of applications of the procedures that everyone agrees to be executable (see **23.11**).

To obtain values of the computable functions is called computation.

Thus, we need a characterization of 'computable functions.'[281]

23.9 Partial and total functions

In the theory of computation in contrast to the ordinary analysis, when we speak of a function, it need not be a map but can be a partial function. That is, $f(x_1, \cdots, x_n)$, where $x_1, \cdots, x_n$ are nonnegative integers (i.e., $x_i \in \mathbb{N}$), need not be meaningful (need not be defined) for all the n-tuples $\{x_1, \cdots, x_n\}$ of nonnegative integers (that is, the

[281] A classic introduction to the topic is M. Davis, *Computability and Unsolvability* (Dover, 1982). Newer textbooks include D. S. Bridges, *Computability, a mathematical sketchbook* (Springer, Graduate Texts in Mathematics 146, 1994), for example.

domain is not specified beforehand). For those tuples for which f is not defined, f is not evaluated. If f is defined on the totality of $\{x_1, \cdots, x_n\} \in \mathbb{N}^n$, it is called a total function.

23.10 Obviously computable functions

As the functions to start with, which everyone must agree to be computable, the following three functions S, P and C are adopted:
(A) $S(x) = x + 1$,
(B) $P_i^n(x_1, \cdots, x_n) = x_i$,
(C) $C_m^n(x_1, \cdots, x_n) = m$.
S is a function to give the successor of x in $\mathbb{N}$. P_i^n is a 'projection operator' to read the i-th variable out of n variables. C_m^n is a constant function assigning a constant m to all the n-tuples $\{x_1, \cdots, x_n\}$.

23.11 Basic operations

As unambiguous 'procedures' (basic operations) that everyone should agree to be applicable to any function let us accept the following I$-$III:
I Composition: From functions $g_1, \cdots, g_m$ and h we can make another function

$$f(x_1, \cdots, x_n) = h(g_1(x_1, \cdots, x_n), \cdots, g_m(x_1, \cdots, x_n)), \tag{23.1}$$

where h is an m-variable function and g_i $(i = 1, \cdots, m)$ are n-variable functions.
II (Primitive) recursion: Starting with $f(x_1, \cdots, x_n, 0) = g(x_1, \cdots, x_n)$, we can construct $f(x_1, \cdots, x_n, m)$ recursively as follows:

$$f(x_1, \cdots, x_n, m) = h(x_1, \cdots, x_n, m-1, f(x_1, \cdots, x_n, m-1)), \tag{23.2}$$

where g and h are, respectively, n and $n + 2$ variable functions.
III Minimalization (or minimization) or unbounded search: Let $f(x_1, \cdots, x_n)$ be a total function. For each $\{x_1, \cdots, x_{n-1}\}$ we can determine the smallest x_n satisfying $f(x_1, \cdots, x_n) = 0$.[282]

23.12 Partial recursive function

A function that can be constructed from the basic functions (A)$-$(C) with a finite

[282] Here, it is crucial that f is a total function. Each step of the algorithm must be guaranteed to end within a finite number of steps, so f must be total.

number of applications of the basic procedures I$-$III is called a *partial recursive function*.

There is no problem with I being a computable procedure. II is the same. Perhaps it may be tedious, but applying finitely many steps patiently step by step can complete the procedure.[283].

However, Procedure III (minimalization) is tricky. Since f is a total function, for any $\{x_1, \cdots, x_{n-1}\}$ we can certainly evaluate $f(x_1, \cdots, x_{n-1}, m)$ for any m with a finitely many steps. Therefore, fixing $\{x_1, \cdots, x_{n-1}\}$, and putting m starting with 0 in the ascending order into f one by one, we can check whether $f(x_1, \cdots, x_{n-1}, m)$ is zero or not. If f becomes zero for the first time with $m = q$, then we define $h(x_1, \cdots, x_{n-1}) = q$. However, the existence of such a non-negative integer q is not known beforehand (in other words, we do not know beforehand whether $h(x_1, \cdots, x_{n-1})$ is a total function or not). Therefore, we cannot know beforehand whether the minimalization process even ends or not. Indeed, there is a way to check whether a given m is an answer or not however large it may be, but no one knows the upper bound of m such that if there is no answer up to the value there is really no answer.

Thus, a partial recursive function is a number-theoretical function which construction procedure can be described unambiguously (i.e., its algorithm is given). However, whether it can be actually computed (constructed) cannot be known beforehand (due to minimalization).[284]

Suppose we begin evaluating a partial recursive function for variable x. If we have not obtained the value after some computation, this may imply that the function is not defined for this x (because the minimalization step does not have a solution) or it is defined but we must be much more patient. We hesitate to declare such a function computable.

23.13 Recursive functions

The functions we can really compute must be such that not only its each computational step is explicitly and unambiguously specifiable, but also the whole computation is guaranteed to be completed with a finite number of steps. Such functions are called recursive functions. That is, total partial recursive functions are called

[283] Functions that can be constructed only with the aid of these two procedures are called primitive recursive functions. These were defined first by Kleene.

[284] Informally (but actually in a not very inaccurate way), a function whose procedure to compute can be programmed on the usual digital computer is a partial recursive function. There is no guarantee that the program actually completes the computation and produces its value for all the inputs.

recursive functions. A recursive function is a function with an algorithm that is guaranteed to be completed with a finite number of steps for any (admissible) inputs.[285]

Note that there are only countably many recursive functions. All the recursive functions can be numbered as $f_1, f_2. \cdots$.

23.14 Computation, the Church thesis

Church defined 'computation' as a procedure to evaluate a recursive function.

That is, Church identifies the set of computable functions and the set of recursive functions. This proposal is called the Church thesis.[286] Impeccably unambiguous computation is possible only when the computational procedures are given purely syntactically (that is, given just as symbol sequences that do not require any interpretation). This is a sort of ultimate reductionism.

The crucial points of this proposal are that the algorithm is explicitly given and that the whole process is completed with a finite number of steps.

23.15 Church's proposal was not easily accepted

When Church's thesis was proposed, it was not immediately and generally accepted that being a recursive function is a convincing characterization of any computable function. The reason was that there was no clear feel for constructive procedures that may be explicitly written down; aren't there not recursive (that is, not I-III above) completely new types of algorithms with which different class of functions may become computable? Isn't the above proposal under the restriction of the era (i.e., the level of mathematics of the day)? Furthermore, since such an intuitively appealing concept as continuity requires, to be defined clearly, the axioms of the topological space, it is possible that apparently intuitively obvious basic procedures

[285] Continuing the above informal expression, we can say that a recursive function is a function which can be programmed on the usual digital computer, and the program produces a number (with sufficient but finite computational time and memory) for any (admissible) input.

[286] The definition here is consistent with M. Davis, *Computability and Unsolvability* (Dover, 1982). However, names and definitions are different in different books. In M. Li and P. Vitànyi, *An Introduction to Kolmogorov Complexity and Its Applications* (Springer, 1993) and J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation* (Addison Wesley, 1979) Church's thesis is the proposal that partial recursive functions are computable. Bridges call them computable partial function. Davis call partial recursive functions as partially computable functions.

and basic functions may not be logically simple.[287]

Subsequently, various definitions of computability were proposed, but interestingly all the definitions were equivalent to Church's thesis. That is, it gained a certain signature of naturalness. Usually, this is the explanation of the relevant history, but actually, one of the reasons that Church confidently proposed his thesis was that various definitions were equivalent.[288] Still, as mentioned above, the oppositions could not be quenched.

However, the characterization of computability in terms of the Turing machine explained below (roughly speaking, a digital computer with indefinitely large memory capacity) silenced all the oppositions to Church's proposal. Church himself wrote that Turing computability did not require any preparation to relate constructibility and the existence of effective procedures in the ordinary sense. In short, for basic concepts clear consistency with our intuition is crucial, so the best characterization of basic concepts is often supplied by explicitly visualizable machineries.

23.16 Basic idea of Turing

If a computer is a machine that performs computation, to characterize computation, one should formalize/construct a machine that can do all the elementary steps 'mechanical calculation' can make. This is the basic idea of the Turing machine.[289]

> Turing chose the restrictions built in the basic mechanism/function of the Turing machine, taking account of the limits of our sensory organs and intelligence. Our sensory organs can distinguish only finitely many distinct signals, and the number of distinguishable states of our brain is also finite. The number of kinds and quantities of tasks our brain and effectors can perform are also finite. These conclusions may be reasonable, because even if the signals and states are continuous, they are meaningful only after 'quantized' in the actual noisy world, and because our brain is a finite object. In short, Turing conceived

[287] See R. Gandy, "The Confluence of Ideas in 1936," in *The Universal Turing Machine, a half-century survey* (Oxford University Press, 1988). Neither Gödel nor Post accepted the proposal. [1936 was full of disastrous events: remilitarization of the Rhineland, Spanish Civil War started, X'ian Incident; J. M. Keynes, *The general theory of employment, interest, and money*].

[288] The situation is explained in detail in W. Sieg, "Step by recursive step: Church's analysis of effective calculability," Bull. Symbolic Logic **3**, 154 (1997).

[289] C. Petzold, *The annotated Turing* (a guided tour through Alan Turing's historic paper on computability and the Turing machine) (Wiley, 2008) is strongly recommended. It includes historical comments on many related topics, many interesting anecdotes (and gossips). It is a serious book, but fun to read.

our brain as a finite automaton = an automaton that relies on finitely many symbols, finitely many rules and operations, and finitely many internal states.

Thus, he required an ‘artificial brain’ the following:
(i) Only finitely many kinds of symbols can appear on each cell (on the tape).
(ii) It can survey only a finitely many cells at once.
(iii) At each instance, it can rewrite only one cell.
(iv) Only finite range of the whole tape can be used (scanned).
(v) There are only finitely many states of the black box, and there are only finitely many instructions that can be performed.

The only idealization, from Turing’s point of view, is that there is no memory capacity limit. However, since computers and our brains can indefinitely increase their external memory capacity, it is a benign idealization.

23.17 Turing machine[290]

A Turing machine consists of an infinite tape, a read/write head, and a black box with a finite number of internal states (Fig. 23.2). The tape is divided into cells. The read/write head can scan only one cell on the tape at a time. The head position and what it should do is specified by the Turing program. See **23.18**.

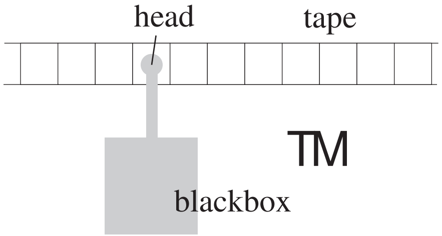


Figure 23.2: A Turing machine consists of an infinite tape that is divided into cells, a read/write head that can scan a single cell at a time, and a black box with a finite number of the internal states.

23.18 Turing program

A Turing program is a finite set of four tuples $(q, S, *, q')$, where
(1) q and q' are the internal states of the black box,

[290] There are many different versions, e.g., with many tapes or a tape that is only infinite in one direction, but the reader has only to be able to have a general notion of TM. See J. H. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison Wesley, 1979), for example. The exposition here uses the machine revised by Kleene and M. Davis (who was in Illinois Math department and is know for his contribution to Hilbert’s tenth problem).

(2) S is the symbol in the cell being scanned now (it may be 1 or blank B, or 0, 1, B, etc., depending on authors, but in any case there are only finitely many of them),
(3) $*$ denotes R, L or S': R (resp., L) implies that the head moves to the right (resp., to the left) in the next time step (or it may be better to say the tape is moved in the opposite direction); S' implies that the head does not move but rewrite the symbol in the cell being scanned as $S \to S'$.

The implication of $(q, S, *, q')$ is as follows:
If the internal state of the black box is q and if S is written in the tape cell being scanned by the head, then the head performs $*$, and the internal state of the black box is changed to q'.

A *Turing machine* may be identified with its Turing program.[291] That is, the machine is understood as a single task machine.

The input non-negative integer x is written on the tape according to a certain rule as, e.g., a 01 sequence (there are many different schemes, but here no details are important). The Turing machine (its black box) is initially in the *initial state* q_I, and the head is located at the leftmost non-blank cell on the tape (the portion where numbers are written is bounded).

The machine has a special final state called the *halting state* q_H. The sequence written on the tape when the machine state reaches q_H is interpreted as the computational result y of the Turing machine. Thus, the Turing machine defines a function $x \to y$. If the Turing machine halts for any input, it defines a total function. However, a Turing machine may not define a total function, because it may not halt (q_H may never be reached) for particular inputs.[292]

23.19 Turing computability

A function defined by a Turing machine that halts for any input is called a Turing computable function. The fundamental theorem is:

Theorem [Turing] Turing computable functions are computable functions in Church's sense and vice versa.

This theorem is proved through demonstrating that the three basic functions S, P and C and the fundamental operations I-III can be Turing-programmable. For

[291] To identify the program and the machine is not as simple as we think. The program lives in the world of symbols, which is distinct from the real world, so how to specify the correspondence between the symbols and the actual movements of the parts of the actual machine is nontrivial. We have already mentioned the concept called 'adaptors.' This will be discussed in Chapter 4.

[292] Will the Turing machine ever halt for input x? This is the famous *halting problem*. There is no algorithm to answer this question. This is a typical *decision problem* that cannot be decided.

example, Davis' textbook explains detailed construction of the programs.[293]

The equivalence of Church's computability and Turing's computability convinced many people the naturalness of Church's thesis.[294]

23.20 Universal Turing machine

Notice that a Turing machine can be encoded as a number sequence (i.e., we can make an integer that can be deciphered to give a Turing program[295], e.g., the Gödel numbering of the Turing program).[296] Now, it is possible to make a machine (= compiler) that spits the four-tuples of the Turing program upon reading the corresponding Gödel number. Then, if we subsequently feed the ordinary input to this master Turing machine, it can perform any computation that can be done by any Turing machine (it is a by-now common programmable digital computer).

A formal proof of this statement may not be simple, but we are not at all surprised thanks to our daily experience. This machine is called a *universal Turing machine*. We may understand it as an ideal digital computer without any memory capacity limit. Turing demonstrated that everything a machine with constraints (i)-(v) above

[293] Reading the programs to check their functions (to check that indeed they work) is a matter of patience; to write such a program is something like writing a program in a machine language (less efficient, actually), so for the ordinary scientists checking the demonstration is probably useless. Therefore, no further discussion will be given on the equivalence demonstration. Now that digital computers are everywhere, the theorem should be almost obvious.

[294] However, this does not mean the acceptance of Turing's basic idea, which seems to have motivated the formulation of Turing machines, that we are finite automata. Gödel (1906-1978) and Post (1897-1954) always believed that our mathematical intelligence was not mechanical. Especially, Gödel argued that our capability to manipulate abstract concepts is not restricted by the finiteness that Turing respected literally; the restrictions apply only when we manipulate (potentially) concrete objects such as symbol sequences; we had to take into account non-finitary creative thoughts to understand our mathematical capabilities.
⟪**Natural intelligence and finiteness constraints**⟫ Perhaps, Gödel may have wished to say that our natural intelligence is not restricted by the finiteness constraints. Our natural intelligence is 'embodied.' It is open to the external physical world through our body. It may be more sensible to idealize our brain as a nonfinitary system. Thus, after all, Gödel's intuition may well be correct. Indeed, abstract concepts are much more concretely supported than concrete concepts by the materialistic basis (or, we could even say, by the molecular biological basis) that is ancient phylogenetically. That is, abstract concepts are directly connected to our body. Abstract concepts are more embodied than concrete concepts.

[295] q, L, etc., appearing in the Turing program is encoded into numbers.

[296] ⟪**Gödel numbering**⟫ Assigning positive integers to symbols, we can convert any symbol sequence into a positive number sequence $n_1n_2\cdots n_k\cdots$. This is converted into $N(n_1n_2\cdots n_k\cdots) = 2^{n_1}3^{n_2}\cdots p_k^{n_k}\cdots$, where p_k is the k-th prime. We can decipher n_i form N uniquely.

can perform can be done by a universal Turing machine. Thus, we may regard a universal Turing machine as the most powerful computer. If we accept Turing's analysis of our intelligence, any intelligent task we can do can be done by a universal Turing machine (with a suitable input).

A universal Turing machine is not a parallel computer. However, whether a machine is parallel or not is not fundamental. It is still a finite automaton. Therefore, the universal Turing machine being not parallel is an unimportant restriction. In the theory of computation we totally ignore computational speed; what matters is whether a required computation can be performed within a finite time.[297]

23.21 All the universal TM are equivalent

Since there are many ways to formulate Turing machines, universal Turing machines cannot be unique. We wish to have the most powerful computer, so perhaps we have to look for the most powerful universal Turing machine. Actually, all the universal Turing machines are equally powerful, so we may choose any of them for our purpose:

Theorem [Solomonov-Kolmogorov] Let M and M' be two universal Turing machines, and $\ell_M(x)$ (resp., $\ell_{M'}(x)$) be the length of the shortest program for M (resp., M') to produce output x (measured in, say, bits). Then, we have

$$\ell_M(x) \preceq \ell_{M'}(x),\ \ell_{M'}(x) \preceq \ell_M(x), \tag{23.3}$$

where $A(x) \preceq B(x)$ implies that there is a positive constant c independent of x such that $A(x) \leq B(x) + c$; c may depend on A and B.

The key to prove this theorem is that M can emulate M' and vice versa. For example, the program for M to emulate M' must be finite, however long it may be. Therefore, if we disregard the length of the program for this overhead (that is, if we write this length as c in the definition of $\preceq$), the length of the needed program does not change whether x is computed directly on M' or computed on the emulated M' on M. The meaning of $\preceq$ is just the inequality disregarding an additive constant (corresponding to the overhead), so the theorem should hold.

Thus, when we wish to write the shortest program length $\ell_M(x)$ (i.e., when we wish to compress x), we will not explicitly specify the universal Turing machine M to use and write simply $\ell(x)$.

[297] Quantum computers may drastically change the required time, but even they cannot compute Turing noncomputable functions. Thus, when we ask the fundamental question what computers can ever do, quantum computers need not be considered.

23.22 Program length, a preliminary observation

Let $\omega[n]$ be the first n digits of a binary sequence ω. How does $\ell(\omega[n])$ (see **23.21**) behave generally?

For example, for the uninteresting $1111\cdots$, asymptotically the information required to specify the number of digits n dominates the program, so $\ell(11\cdots[n])$ behaves as $\log n$. For a number sequence ω with an obvious regularity the shortest program to specify $\omega[n]$ is independent of n except for the portion needed to specify n itself. There are only countably many such regular sequences.

On the other hand, in the $n \to \infty$ limit if $\ell(\omega[n])/\log n$ can be indefinitely large, the pattern in the sequence cannot be specified asymptotically by a finite length program, so it should not be simple. However, such sequences include many subtle sequences.[298]

If no regularity is discernible at all in ω, to specify $\omega[n]$ requires to specify almost all the n digits, so we expect $\ell(\omega[n]) \sim n$. In other words, a typical random sequence must be such ω that for infinitely many n $\ell(\omega[n]) \sim n$.

23.23 Algorithmic randomness

Let us call a sequence ω such that for infinitely many n $\ell(\omega[n]) \sim n$ an algorithmically random sequences.

Definition 2.11.1

$$K(\omega) \equiv \limsup_{n\to\infty} \ell(\omega[n])/n \tag{23.4}$$

is called the randomness of a binary sequence $\omega \in \{0,1\}^{\mathbb{N}}$. Here, $\omega[n]$ is the first n digits of ω and $\ell(\omega[n])$ is the length of the shortest program (written in 01) to produce $\omega[n]$.[299]

Thanks to Kolmogorov and Solomonov (**23.21**) the randomness does not depend on the choice of the universal Turing machine M (so it is not written already).

Kolmogorov used the word 'complexity' to describe the above quantity. in this lecture notes the word should be reserved for genuine complex systems, so we use the word 'randomness' instead.[300]

[298] This situation is exactly the same as the KS entropy zero dynamical systems.

[299] The reason why we must use lim sup is, for example, we cannot ignore the appearance of meaningful sequences occasionally. That is why it was said $\ell(\omega[n]) \sim n$ for infinitely many n instead of all n.

[300] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*

23.24 Undecidability of random sequence

Notice that no program can be written to compute $K(\omega)$ (there is no algorithm for it). This should be easily inferred from the appearance of the expression as 'the shortest program' in the definition. For an arbitrarily chosen number sequence, except for almost trivial cases, it is hard to show that it is not a random sequence, because we must write down a short program to produce it. Worse still, it is harder to claim that a given sequence is random, because we must show that however hard one tries one cannot write a short program. Therefore, for a given number sequence, generally we cannot compute its randomness K (except for some almost trivial cases with $K = 0$, it is very hard, if not impossible.)

However, we can say something meaningful about a set of number sequences. For example, for almost all (with respect to the Lebesgue measure) numbers in $[0, 1]$ their binary expansion sequences are algorithmically random. All the algebraic numbers[301] are not random. In the next section we will discuss the average randomness.[302]

23.25 Is our characterization of randomness satisfactory?

To make the concept 'random' mathematical, we identified it with the lack of any computable regularity. This identification does not seem to contradict our intuition. However, is it really true that all the regularities are all those detectable by computers? There can be the following fundamental question: why can we say that if computers cannot detect any pattern, Nature herself cannot, either? In this characterization don't we admit that the Turing computer or computation itself is more fundamental a concept than randomness? Is this consistent with our intuition?

(Springer, 1993) is the standard textbook of the Kolmogorov complexity. The reader will realize that there are many kinds of definitions and concepts, but here, to be simple, the most basic definition is used. The explanation in this section roughly follows A. K. Zvonkin and L. A. Levine, "The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms," Russ. Math. Surveys **25**(6), 83 (1970).

[301] numbers that can be zeros of integer coefficient polynomials.

[302] There is an attempt to make a computable measure of randomness. One approach is to use much less powerful computers. However, such an approach appears to be fundamentally off the mark to characterize the concept of randomness, because 'randomness' may well be a transcendental concept. There can be a point of view that within our mathematics it is natural that we cannot tell whether a given sequence is random or not in general.

23.26 Axiomatic approach to randomness

If we consider randomness as a fundamental concept, as long as there is no other concepts that we can accept as more fundamental, we cannot define it. In the end, randomness would be formalized only as a primitive concept of an axiomatic system for randomness, just as points and lines in Euclidean geometry. As far as the author is aware, the most serious approach in this direction is due to van Lambalgen.[303] As we have seen, randomness is algorithmically characterized by incompressibility. We needed the theory of computation to define incompressibility unambiguously. The axiomatic approach may be roughly interpreted as an attempt to axiomatize the concept corresponding to 'incompressibility.'[304]

As we have seen in Chapter 1, randomness is really significant only in nonlinear systems. We have also seen that the standard axiomatic system of sets is a legitimate heir of Fourier analysis, so to speak. A wild and heretic guess is that even on the foundation of mathematics is a shadow of linear systems, so it is not very suitable to study nonlinear systems and the concept of randomness.

[303] M. van Lambalgen, "The axiomatization of randomness," J. Symbolic Logic **55**, 1143 (1990).

[304] This axiomatic system is not compatible with the standard axiomatic system of sets (ZFC). The author does not understand how serious this incompatibility is. See M. van Lambalgen, "Independence, randomness and the Axiom of Choice," J. Symbolic Logic **57**, 1274 (1992). See also "Logic: from foundations to applications, European logic colloquium" edited by W. Hodges, M. Hyland, and J. Truss (Clarendon Press, 1996) Chapter 12 "Independence structures in set theory."

24 Lecture 24. Characterization of chaos

24.1 Basic motivation

Chaos is characterized by ‘complicate’ or ‘apparently random’ trajectories. We have refined the concept of randomness in the preceding section. ‘Algorithmic randomness’ is the refined concept, but, as noted there, there is no general way (no algorithm) to judge whether a given sample sequence or an object (coded appropriately as a number sequence) is random or not. However, collectively we can determine whether a set consists of mostly random numbers or not. We know almost all the binary expansions of the numbers in $[0, 1]$ are algorithmically random (because there are only countably many non-random numbers). Therefore, the basic idea to connect chaos and randomness is to check whether there is a bunch of trajectories that are ‘almost surely’ random or not. Thus, we wish to claim that a dynamical system which has an invariant set on which most trajectories are algorithmically random (after appropriate discretization = coding).

24.2 Can we be quantitative?

The claim “a dynamical system which has an invariant set on which most trajectories are algorithmically random” is actually already built in into my definition of chaos **22.10** through the coin-tossing process, because most numbers in $[0, 1]$ is algorithmically random.

We must recall that the definition of algorithmic random numbers is in terms of complexity K. We wish to define a corresponding number within the theory of dynamical systems and compare it with K (**23.23**). Since K measures how much information we need to specify a symbol in the coding sequence, it is a very natural idea that it must be compared with the Kolmogorov-Sinai entropy, which was intuitively introduced as the information loss rate (or required information to predict the future; **17.18**). The KS entropy h will be discussed in detail later mathematically (-**32.9**), but here let us use its intuitive meaning, and actually show $K = h$, completing our characterization of chaos even quantitatively.

About coding: We have extensively used symbolic dynamics (or shift dynamics) isomorphic or homomorphic to the original dynamical system to analyze it. Some people bitterly criticize this strategy, saying that this approach does not respect how ‘random sequences’ are actually produced. We could produce the same ‘01 sequence’ from a black box containing a person with a coin. Therefore, if one observes only the coded results, one can never infer the content of the black box (even if it is driven by a tent map). Hence, characterizing the dynamical system in terms of the coded result is impossible. *A fortiori* characterizing chaos with the randomness of the trajectories is flawed. How do you respond?[305]

24.3 Randomness of a trajectory

Let $\mathcal{B} = \{B_1, \cdots, B_k\}$ be a generator (= the best coding scheme; see **32.10**) for a map T defined on M with an invariant measure (= stationary distribution) μ (denoted as (T, μ, M) in standard books). We can code a trajectory starting from x at time 0 as ω in terms of k symbols with the rule $\omega_n = a$ if the trajectory goes through B_a at time n ($T^n x \in B_a$). Define the randomness of the trajectory starting from x as

$$K(x, T) \equiv \limsup_{n\to\infty} \frac{1}{n}\ell(\omega[n]), \tag{24.1}$$

where $\omega[n]$ is, as before, the first n symbols of ω, and $\ell(z)$ is the code length of the minimal program for z in terms of k symbols as defined in the preceding lecture (there, k was 2).

24.4 Brudno's theorem[306]

Brudno's theorem adapted to the current situation: coding of (T, μ, M) with k symbols reads:

> For μ-almost all $x \in M$
>
> $$K(x, T) = h_\mu(T)/\log k.$$
>
> Here, h_μ is the Kolmogorov-Sinai entropy of (T, μ, M) defined (as usual) in terms of the natural logarithm, but it is divided by $\log k$, so the right-hand side gives the entropy defined in terms of the base k logarithm.

[305] Since this is a good discussion topic, no comment should be added, but note that we do not simply treat dynamical systems as black boxes. The correspondence between a shift and a dynamical system must be at least homomorphic. There cannot be any deterministic dynamical system homomorphic to the package of a person + a coin.

[306] A. A. Brudno, THE COMPLEXITY OF THE TRAJECTORIES OF A DYNAMICAL SYSTEM, Russ. Math. Surveys 33 197 (1978); Entropy and the complexity of the trajectories of a dynamical system. Trans. Moscow Math. Soc., 2 127 (1983). See also H. S. White, Algorithmic complexity of points in dynamical systems, Ergod. Th. & Dynam. Sys., 13 807 (1993).

The following exposition is taken from TNW Chapter 2, but probably more readable (reader-friendly).

What is claimed is, qualitatively, the equality between the amount of the extra information required for prediction of the state time t in the future and the amount of information to describe the trajectory for the time span t. It is quite a natural assertion. To predict a chaotic trajectory for a long time, we need a tremendous amount of information initially. Since such an amount of information is needed to single out a trajectory, the coding sequence needed to describe the trajectory cannot be simple and information compression is out of question. Thus, Brudno's theorem confirms the goodness of our definition of chaos.

24.5 Brudno's theorem for symbolic dynamics

The information needed to describe a trajectory may be considered in terms of the symbol sequence after coding. Therefore, the core of Theorem in **24.4** is the following fact about the shift dynamical system (isomorphic to the dynamical system under consideration):

Theorem. Let (σ, Ω) be a certain shift dynamical system with k symbols, and μ its ergodic invariant measure. Then, for μ-almost all $\omega \in \Omega$

$$K(\omega) = h_\mu(\sigma)/\log k,$$

where $h_\mu(\sigma)$ is the Kolmogorov-Sinai entropy of the measure theoretical dynamical system (σ, μ, Ω), and $K(\omega)$ is the randomness defined in (23.4) (as in **24.4**, it is defined for k-symbol sequences instead of binary sequences and h uses natural log; that is why $\log k$ appears).

The above theorem is proved by showing the following two statements:
(i) ω satisfying $K(\omega) < h_\mu(\sigma)/\log k$ is μ-measure zero.
(ii) μ-almost surely (= for μ-almost all ω) $K(\omega) \leq h_\mu(\sigma)/\log k$.

24.6 Demonstration of (i)

The number of $\omega[n]$ satisfying

$$K(\omega) \sim \ell(\omega[n])/n \leq s \tag{24.2}$$

is no more than k^{ns} (in our context ω is the k-symbol sequence). On the other hand, according to the Shannon-McMillan-Breiman theorem **32.13**, the measure of the cylinder set specified by $\omega[n]$ is estimated as $e^{-nh_\mu(\sigma)}$. Therefore, the measure of all ω satisfying (24.3)

$$K(\omega) < h_\mu(\sigma)/\log k \tag{24.3}$$

is bounded by $e^{n(s\log k - h_\mu(\sigma))}$. This exponent is negative ($s\log k < h_\mu(\sigma)$; do not forget that h_μ is defined with the natural logarithm), so the upper bound converges to zero in the large n limit. Therefore the possibility of (24.3) is almost surely ignored.

24.7 Demonstration of (ii)

Next, we wish to show that μ-almost surely (= for μ-almost all ω)

$$K(\omega) \le h_\mu(\sigma)/\log k. \tag{24.4}$$

If this is demonstrated, then, since we just showed that the cases with $K(\omega) < h_\mu(\sigma)/\log k$ may be ignored almost surely, only the equality remains.

Since we have only to estimate the upper limit of $K(\omega)$, let us estimate the upper limit of $\ell(\omega[n])$. $\omega[n]$ is decomposed as follows in terms of q length-m-symbol-sequences ω_i^m ($i = 1, \cdots, M$, where M is the total number of distinct length-m-symbol-sequences; $n = mq + r$, i.e., $q = [n/m]$ and r is the residue):

$$\omega[n] = \omega_0^r \omega_{i_1}^m \omega_{i_2}^m \cdots \omega_{i_q}^m. \tag{24.5}$$

Here, ω_i^m is the ith kind of length-m-symbol-sequence, which is assumed to appear s_i times. With this representation, $\omega[n]$ can be uniquely specified by r, m, $s_1, \cdots, s_M$, ω_0^r and the arrangement of q length-m-symbol sequences $\omega_{i_1}^m \omega_{i_2}^m \cdots \omega_{i_q}^m$. Therefore, the needed information to specify $\omega[n]$ is given by (or, the length of the shortest required program with k symbols, or the information measured with the base k logarithm is given by)

$$\ell(\omega[n]) \le \ell(r) + R + \ell(m) + \ell(q) + \sum_{j=1}^{M} \ell(s_j) + H(\omega_{i_1}^m \omega_{i_2}^m \cdots \omega_{i_q}^m), \tag{24.6}$$

where $H(\omega_{i_1}^m \omega_{i_2}^m \cdots \omega_{i_q}^m)$ is the information needed to specify the arrangement (ordering) of q length-m-symbol sequences $\omega_{i_1}^m \omega_{i_2}^m \cdots \omega_{i_q}^m$ and R is the information required to specify ω_0^r, which is bounded by a constant independent of n. That is, except for the last term, all the terms are $o[n]$ and unrelated to the randomness. Hence,

$$K(\omega) \le \limsup_{n\to\infty} H(\omega_{i_1}^m \omega_{i_2}^m \cdots \omega_{i_q}^m)/n. \tag{24.7}$$

$H(\omega_{i_1}^m \omega_{i_2}^m \cdots \omega_{i_q}^m)$ is bounded by the information (in terms of base k logarithm) carried by the possible sequences under the assumption that all such sequences appear with equal probability. Therefore, it cannot exceed the logarithm (base k) of the number of sequences that can appear as $\omega[n]$. That is, $K(\omega) \leq \lim_{n\to\infty}[\log_k N(n)]/n$. [Beyond this point you need rudiments of KS entropy: Lecture 31] $N(n)$ is equal to the number of non-empty elements $\vee_{k=0}^{n}\sigma^{-k}\mathcal{B}$ for a generator $\mathcal{B}$. According to the Shannon-McMillan-Breiman theorem, for the cylinder sets contributing to entropy $\mu(\omega[n])/e^{-nh_\mu(\sigma)}$ must not vanish in the $n \to \infty$ limit. Therefore, the number of cylinder sets we must count must be, since the sampling probability of each cylinder set is $e^{-nh_\mu(\sigma)}$, the order of its inverse: $N(n) \sim e^{nh_\mu(\sigma)}$. Thus, we can understand (24.4).

As can be seen from the explanation here, Brudno's theorem is based on very crude estimates, so it is a natural theorem. Such a theorem should have been discovered by theoretical physicists without any help of mathematicians. Most physicists in the US did not know this theorem well into the 1990s.

24.8 What is chaos, after all?

Out conclusion is that chaos is a deterministic dynamical system whose trajectories are algorithmically random. Actually, equivalently, we can say that a dynamical system with a positive KS entropy invariant measure is a chaotic dynamical system.

Is this a satisfactory outcome? That a trajectory is random is, with the quantification in terms of the Kolmogorov-Sinai entropy, invariant under the isomorphism of the dynamical systems. Isomorphism is a crude correspondence ignoring even the topology of the phase space, so the characterization we have pursued has nothing at all to do with how the correlation function decays or what the shape of the attractor is. Even the observability of chaos by computer experiments is not invariant under isomorphism.

Perhaps we should conclude that chaos is very common random phenomenon exhibited by deterministic dynamical systems and is far more basic than the exponential decay of the correlation function, or the invariant sets being fractal.

24.9 Does chaos exist in Nature?

Since we started this chapter with a simple realizable example, the question whether there is actually chaos may sound strange. However, it is difficult to tell whether the actual apparently chaotic phenomenon is really chaos or not due to the existing noise.[307] For example, it is easy to make an example that apparently exhibits observable chaos, even though there is no observable chaos without noise. It is difficult

[307] Here, 'noise' need not mean the effect of the unknown scale, but any unwanted external disturbance as usual.

to answer affirmatively the question whether there is really chaos without external noise in a system for which the existence of chaos is experimentally confirmed (this is in principle impossible[308]). Therefore, whether the concept of chaos is meaningful in natural science or not depends on whether it is useful as an ideal concept to understand the real world just as points and lines in elementary geometry. The relevance of chaos to the instabilities in some engineering systems or instabilities in numerical computation shows that it is a useful ideal concept.

For actual systems, what is important is its response to small perturbations. For a deterministic chaos, its practically important aspect is almost exhausted by the exponential separation of nearby trajectories. Whether the system is deterministic or not is unimportant. What is practically important is that the phase space is bounded and the trajectory itinerates irregularly various 'key' points in the phase space that are crucial to the system behavior. However, in order to model a system that easily exhibits such trajectories with small external perturbations, use of a chaotic system is at least metaphorically effective.[309]

[308] For example, for a one dimensional map, indefinitely small modification of the map can change it to have a stable fixed point. Such a system behaves just as before the modification, if a small noise is added. M. Cencini, M. Falcioni, E. Olbrich, H. Kantz, and A. Vulpiani, "Chaos or noise: difficulties of a distinction," Phys. Rev. E **62**, 427 (2000) recommend a more practical attitude toward chaos.

[309] See for chaotic itinerancy view of brain by I. Tsuda, "Hypotheses on the functional roles of chaotic transitory dynamics," Chaos **19**, 015113 (2009).

25 Lecture 25. Computable analysis

An accessible reference is: M. B. Pour-El and J. I. Richards, *Computability in Analysis and Physics* (Springer, Perspective in Mathematical Logic, 1987).

25.1 Decision problem

Given a set of problems (or a set of instances of a problem, e.g., whether a polynomial has 0 as its root or not), we ask whether there is an algorithm to answer all the problems in the set. This problem is called a *decision problem.* If there is such an algorithm, we say the set (or the problem) is *decidable.* If not, we say it is *undecidable.* The word 'algorithm' was unclear before Turing, but now we can clearly state that 'algorithm = existence of Turing program.'

25.2 Remark

If a set is finite, or the problem has only finite instances, then it is trivially decidable, because we can check all of them one by one blindly. The decision problem becomes nontrivial only if the problem has infinitely many instances like the one due to Diophantus in **25.3**.

25.3 Decision problem examples.

(1) Hilbert's 10th problem. Decide whether a polynomial $P(x_1, x_2, \cdots, x_n)$ with integer coefficients (Diophantine equations) has an integer root. This is decidable if $n = 2$,[310] but is undecidable for general n.[311]
(2) Is $\exists x_1 \exists x_2 \exists x_3 \forall y_1 \cdots \forall y_m \mathcal{U}$ true ($m \in \boldsymbol{N}$)? Here, $\mathcal{U}$ is any logical formula within the first order logic[312] without including $\exists$, $\forall$ and free object variables. This is undecidable.

25.4 Halting problem of Turing machine

Suppose we have a Turing machine T. We feed a number n to it, and ask whether T ever stops (that is, the solution is given within a finite time or not). Certainly, we

[310] A. Baker, Phil. Trans. Roy. Soc. London A 263 (1968).

[311] Ju. V. Matyasevich, 1970: The theorem is called the Matyasevich-Robinson-Davis-Putnam (MRDP) theorem. There is a moving short movie of Julia Robinson: "Julia Robinson and Hilbert's Tenth Problem" (A film by G Csicsery, Zala Films, 2008).

[312] I recommend H. D. Ebbinghaus, J. Flum and W. Thomas, *Mathematical Logic* (Springer Undergraduate Texts in Mathematics, 1984; there is a new edition).

can run the program on T, but that the machine has not yet stopped does not mean anything about its final result. Is there any algorithm to judge that T halts upon n? This is called the *halting problem*.

25.5 Halting problem is undecidable

We can identify a Turing machine X and its program = its Gödel number $n(X)$. Since $n(X)$ is a number, we can feed this into X. Let us collect all the Turing machines X such that $X[n(X)]$ (the result of the computation) is computed (that is, X halts on input $n(X)$) and make a set K of such Turing machines:

$$K = \{X \mid X[n(X)] \text{ is well defined.}\}. \tag{25.1}$$

If the halting problem is decidable, then, since K is a collection of special Turing machines, in particular, we must have an algorithm to judge whether a given Turing machine X is in K or not. Thus, we can make a Turing machine Y that halts for any input and gives the following output:

$$Y[n(X)] = \begin{cases} X[n(X)] + 1, & \text{if } X \in K, \\ 0, & \text{otherwise .} \end{cases} \tag{25.2}$$

Note that $Y \in K$, since it halts for any input.

This Y is defined for any input Gödel number, so, in particular, we may input $n(Y)$. The outcome is

$$Y[n(Y)] = Y[n(Y)] + 1, \tag{25.3}$$

a contradiction; we cannot make such Y. There is no algorithm to decide the halting problem.

25.6 Recursive set

A set whose characteristic function is a recursive function (**23.13**) is called a *recursive set*.

What this means is: if a set is a recursive set, then we have an algorithm to tell whether a given number is in the set or not. In this sense, we can tell the member of the set without referring to how to generate the set.

In other words, a set is a recursive set, if and only if we can construct a Turing machine (or a program for a universal Turing machine) such that it can print 1 if the element[313] is in the set and 0 otherwise with finite steps.

[313] Of course, this must be suitably encoded so that the machine can understand it.

25.7 Recursively enumerable set

A set A which is a range of a recursive function f is called a *recursively enumerable set*. We say f enumerates A.

Hence, if a set is a recursively enumerable set, we know how to produce the set (there is a computer program which generates the set). We simply feed the elements in $\mathbb{N}$ one by one to the recursive function, and collect its outcomes.[314]

25.8 Nondeterministic Turing machines

To produce a recursively enumerable set, we feed $\mathbb{N}$ one by one into a Turing machine. Instead, we could prepare infinitely many copies of the Turing machine and feed $\mathbb{N}$ at once. This is an ideal parallel computation, and the infinite bank of the Turing machine is called a nondeterministic Turing machine.

The output of a nondeterministic Turing machine is a recursively enumerable set.

The index of a recursively enumerable set is partially recursive **23.6**, since we can use its enumerating function (**25.7**) to make the required index function.

25.9 Existence of recursively enumerable non-recursive set

Theorem: There exists a recursively enumerable but not recursive (RENR) set. This is important, so a demonstration is given here. You will realize the following is very similar to the logic in **25.5**.

Let $\phi_x(y)$ be the output (if any) of the Turing machine whose Gödel number is x (remember that there are only countably many Turing machines), when its input is y. Here, both x and y are in $\mathbb{N}$. Make a set

$$K \equiv \{\, x : \phi_x(x) \text{ is defined}\}. \tag{25.4}$$

That is, K is the set of all the numbers x such that the corresponding Turing machine halts with the input x. Certainly, this is a recursively enumerable set, because we know how to perform each step needed to compute $\phi_x(x)$, although we do not know whether it actually gives a number or not. Now define a function f such that

$$f(x) \equiv \begin{cases} \phi_x(x)+1, & \text{if } \phi_x(x) \text{ is defined,} \\ \quad 0, & \text{otherwise.} \end{cases} \tag{25.5}$$

[314] In this case it is known that the enumeration can be done without repetition. See e.g., Zvonkin and Levine, *op. cit.* Theorem 0.4.

That is,

$$f(x) \equiv \begin{cases} \phi_x(x) + \chi_K(x), & \text{if } \phi_x(x) \text{ is defined,} \\ \chi_K(x), & \text{otherwise,} \end{cases} \tag{25.6}$$

where χ_K is the characteristic function of K. If K is recursive, then χ_K is recursive, so there must be a Turing machine which reproduces f. However, there cannot be such a Turing machine; if any, there must be an x such that $f(z) = \phi_x(z)$ for any $z \in \mathbb{N}$, but obviously this is untrue for $z = x$. Thus we cannot assume that χ_K is a recursive function.

Remark.
(1) I believe most recursive sets are RENR sets, but I do not find any such statement.
(2) Notice that whether RENR or not the totality of recursively enumerable sets is a countable set, since the number of Turing machines is countable. In contrast, the totality of algorithmic random numbers is uncountable.

25.10 Condition for recursiveness

Theorem. A set Q is recursive if and only if both Q and Q^c are recursively enumerable (**25.7**).

[Demo] We assume Q and Q^c are non-empty.
($\Rightarrow$) Let χ be the index function of Q. Then there is a TM that compute χ (or χ is a recursive function f). Then $1 - f$ is also recursive, which is the index of Q^c, so Q^c is recursive. A recursive set is recursively enumerable,[315] so Q and Q^c are both recursively enumerable.
($\Leftarrow$) Suppose f enumerate Q and f^c enumerates Q^c. Both are recursive functions. Therefore, we can make the index function for Q, since we know Q explicitly.

25.11 Computable rational sequence

We say a rational number sequence $\{r_k\}$ is a *computable rational number sequence*, if for any $k \in \mathbb{N}$ there are recursive functions (**23.13**) a, b and s ($b \neq 0$) such that

$$r_k = (-1)^{s(k)} \frac{a(k)}{b(k)}. \tag{25.7}$$

[315] If A is a recursive set, we have a Turing machine that accepts it; that is $T(a) = 1$ for any $a \in A$ and $T(b) = 0$ for any $b \notin A$. Make a TM based on T such that if $T(a) = 1$, spit a. For $b \notin A$, we could leave this TM not to halt, or to print one known element in A.

25.12 Effective convergence

Let $\{r_k\}$ be a computable rational sequence. We say it converges effectively to $x \in \mathbb{R}$, if there is a recursive function $e(N)$ such that

$$k \geq e(N) \Rightarrow |r_k - x| \leq 2^{-N}. \tag{25.8}$$

That is, if $\{r_k\}$ converges to x in the ordinary sense of this word and if there is an algorithm to estimate error, we say $\{r_k\}$ converges effectively to x.

25.13 Computable real number

x is a *computable real number*, if there is a computable rational number sequence effectively converging to x.

25.14 Remark: Effectiveness

We say we can do something effectively, if we have an algorithm. We say a concept is effective, if we can define it with an algorithm (for example, whether it is correct or not can be decided). An asymptotic object such as irrational numbers is said to be an effective object when its construction and the distance (error) from the asymptotic limit can be estimated effectively. Thus, 'effectiveness' is a precise formalization of 'constructibility.'

25.15 How to destroy effectiveness

Let $A = \{a(n)\}$ be a RENR set **25.9** without repetition (i.e., $a(n) \neq a(m)$, if $n \neq m$). We can compute each $a(n)$, but we cannot effectively tell whether, say, 10 appears in A or not. Hence, if we can construct a procedure whose error estimate is bounded by $2^{-a(m)}$, then effective estimation is destroyed.

25.16 Waiting lemma

Let $A = \{a(n)\}$ be a RENR set (**25.9**) without repetition (i.e., $a(n) \neq a(m)$ if $n \neq m$). Let

$$w(n) \equiv \max\{m \,|\, a(m) \leq n\}. \tag{25.9}$$

Then, there is no recursive function (**23.13**) $c(n)$ such that

$$w(n) \leq c(n). \tag{25.10}$$

That is, there is no algorithm to estimate the needed m so that $\{1, \cdots, n\} \subset \{a(1), \cdots, a(m)\}$.

[Demo]
If $c(n)$ were recursive, then we could tell whether $n \in A$ or not with a finite number of steps. First, compute $c(n) = m$, then check all $a(m')$ for m' up to m. If we could find n among the output, certainly $n \in A$; if we could not, then $n \neq A$. Hence, A would be a recursive set, a contradiction.

25.17 Existence of converging noneffectively converging series

Theorem. There is a bounded monotone increasing series consisting of computable rational numbers that does not converge effectively (that is, although its convergence is guaranteed in the ordinary mathematics, we have no means to compute its value for sure).
[Demo]
Take A in the above and construct

$$S = \sum_{n=0}^{\infty} 2^{-a(n)}. \tag{25.11}$$

This is a desired example of the series claimed in the theorem, because A is not repetitive; it is bounded from above and the partial sums are monotone increasing. Since, for example, we do not know whether 2 is in A or not effectively, we cannot estimate S (which must be less than 2) better than the error of 1/4.

S is not computable according to the expansion (25.11). There must be many other series converging to the same S, so you might think the non-computability of S may 'series expression'-dependent. This is not the case at least if there is a monotone converging noncomputable series.[316]

25.18 Computable function

We say a function from ℝ into itself is computable, if its values at computable reals

[316] Pour-El & Richards p20.

are computable reals. Pour-El and Richards impose further the following *effective uniform continuity.* There is a recursive function d such that for any $n \in \boldsymbol{N}$

$$|x-y| \le 1/d(n) \Rightarrow |f(x)-f(y)| < 2^{-n}. \tag{25.12}$$

25.19 'Ordinary functions' are computable

sin, cos, exp, J_n, etc., are computable. Behind this statement lies the following 'effective Weierstrass' theorem.'

If we can find a recursive function $D(n)$ such that

$$p_n(x) = \sum_{i=0}^{D(n)} r_{nj} x^j, \tag{25.13}$$

where r_{nj} are computable rationals, we say $\{p_n\}$ is a *computable sequence of rational polynomials.*

Effective Weierstrass. If we can find a recursive function $e(n)$ such that

$$m \ge e(N) \Rightarrow |f(x)-p_n(x)| < 2^{-N}, \tag{25.14}$$

then f is a computable function.[317]

25.20 Computable operations on functions

Composition $f \circ g$, sum $f \pm g$, multiplication fg, and many other elementary operations preserve computability. Integration also preserves computability. Hence, it is not hard to guess that the derivatives of computable analytic functions are again computable. However,

25.21 Myhill's theorem

Theorem [Myhill]. Even if f is a computable C^1 function, f' may not be computable.

[Demo]

[317] See Pour-El and Richards, Chapter 0, Section 5 and 7.

The following is the counterexample. Let

$$\varphi(x) = \begin{cases} \exp(-x^2/(1-x^2)) \text{ for } |x| < 1, \\ 0 \text{ otherwise,} \end{cases} \tag{25.15}$$

which is a C^∞ function. Let $A = \{a(n)\}$ be the RENR set mentioned before. Define

$$\varphi_n(x) = \varphi[2^{n+a(n)+2}(x - 2^{-a(n)})]. \tag{25.16}$$

Construct

$$f(x) = \sum_{k=0}^{\infty} 4^{-a(k)} \varphi_k(x). \tag{25.17}$$

This is computable, but

$$f'(2^{-m}) = 4^{-m}\chi_A(m), \tag{25.18}$$

where χ_A is the characteristic function of A, which cannot be computed.

25.22 PDE and computability

(1) Laplace and diffusion equations preserve the computability of the auxiliary conditions.
(2) In $d(\geq 2)$-space, the wave equation cannot preserve computability. More explicitly, even if the initial data is computable, the solution at time, say, $t = 1$ is not computable. It is not hard to understand this, if we notice that the Radon transformation formula ($d \geq 2$) involves differentiation. See **32D.9-10** (For the Radon transformation see **32D.2**) of
https://www.dropbox.com/home/ApplMath?preview=AMII-32+FourierTransformation.pdf.

25.23 Real physics implications?

There are uncountably many algorithmically random numbers, but the recursive and recursively enumerable sets are both only countably many, because all the number of Turing machines or Turing programs ("Gödel numbers") is countable.

There is no way to construct a non-computable functions nor RENR sets. However, it is expected that most recursively enumerable sets are not recursive. Therefore, if we 'sample' a series 'randomly', we will hit unpleasant examples easily. [TBC].

26 Lecture 26. Symbolic dynamics

26.1 Shift

Let us consider a finite set S whose elements we call symbols (thus S may be understood as a set of alphabets). A finite sequence of symbols is called a word. Make a set Σ consisting of some infinite sequences (both-sided or one-sided) $\omega = (\omega_n)_n$ ($\omega_n \in S$). Then we can make a shifted sequence $\omega' = \sigma\omega$ from ω as $\omega' = (\omega_{n+1})_n$:

$$\omega = \cdots \omega_{-k}\omega_{-k+1}\cdots\omega_{-1} \quad \omega_0 \quad \omega_1\omega_2\cdots\omega_k\omega_{k+1}\cdots, \tag{26.1}$$

$$\sigma\omega = \omega' = \cdots \omega_{-k+1}\omega_{-k+2}\cdots\omega_0 \quad \omega_1 \quad \omega_2\omega_3\cdots\omega_{k+1}\omega_{k+2}\cdots. \tag{26.2}$$

or in the one-sided case

$$\omega = \quad \omega_0 \quad \omega_1\omega_2\cdots\omega_k\omega_{k+1}\cdots, \tag{26.3}$$

$$\sigma\omega = \omega' = \quad \omega_1 \quad \omega_2\omega_3\cdots\omega_{k+1}\omega_{k+2}\cdots. \tag{26.4}$$

The operator σ is called a shift operator or shift.

26.2 Shift dynamical systems

$\Sigma \equiv S^{\mathbb{Z}}$ (resp. $\Sigma^+ = S^{\mathbb{N}}$) is the totality of both-side (resp. one-side) infinite symbol sequences on S. We can define shift on Σ and is called, as a dynamical system, the full shift on S.

A subset M of Σ is an invariant set if $\sigma M = M$. Then, we may define a restriction of σ to M (sometimes, it is written as σ_M), which is called a subshift.

For example, we can consider all the sequences $M \subset \{0, 1\}^{\mathbb{Z}}$ that never contains 11. Needless to say, the shift never changes the sequence structures, so we can define a shift dynamical system (σ, M). This is an example of a Markov subshift.

A finite sequence of symbols is called a word. Thus M above may be characterized as a sequence without word 11.

26.3 Topology or metric in symbol sequence space

We wish to compare different sequences. As can be seen from the illustration 22.4 if the length n words close to 0 are identical we should regard the sequences close. Thus, we introduce the following metric in Σ:

$$d(\omega, \omega') = \sum_n 2^{-|n|}\delta_{\omega_n\omega'_n}. \tag{26.5}$$

The full shift $(\sigma, \Sigma^{\mathbb{Z}})$ (or $(\sigma, \Sigma^{\mathbb{N}})$) has dense periodic orbits and topologically mixing 26.4.

26.4 Topological mixing

Let $f : M \to M$ be a topological dynamical system (= C^0-endomorphism of a (smooth) manifold M). For any $U, V \subset M$ if $f^n(U) \cup V \neq \emptyset$ for any sufficiently large $n \in \mathbb{N}$ (i.e., there is a positive integer $N(U, V)$ and for all $n > N(U, V)$), we say this dynamical system is topologically mixing.

26.5 Cylinder set

A subset of sequence space M with a particular word at a particular position is called a cylinder set: for example

$$\{\omega \,|\, \omega_0 = \alpha_1, \omega_1 = \alpha_2, \cdots \omega_{10} = \alpha_{11}, \omega \in M\} \tag{26.6}$$

is a length (or rank) 11 cylinder set consisting of the totality of sequences in M such that

$$\cdots \omega_{-2}\omega_{-1}\alpha_1\alpha_2 \cdots \alpha_{10}\alpha_{11}\omega_{11}\omega_{12} \cdots, \tag{26.7}$$

where ω_k can be anything as long as $\omega \in M$.

26.6 Markov subshift

Let $S^\circ = n$ (the size of the alphabet). Let A be a $n \times n$ matrix whose elements are 0 or 1. Define

$$\Sigma_A = \{\omega \in \Sigma \,|\, A_{\omega_n, \omega_{n+1}} = 1, n \in \mathbb{Z}\}. \tag{26.8}$$

Σ_A^+ for one-sided systems may be analogously defined by replacing $\mathbb{Z}$ with $\mathbb{N}$. (σ, Σ_A) is called a Markov subshift whose structure matrix is A.

26.7 Fixed points of Markov subshift

If there is an orbit starting from symbol x that returns to itself after p shift, $(A^p)_{xx} > 0$. This number is actually the number of ways to go from x to itself in p steps.

The fixed points of σ^p has the following structure:[318]

$$(i_1 i_2 \cdots i_p i_1 i_2 \cdots i_p \cdots). \tag{26.9}$$

Therefore, the total number $N_{\text{Fix}}(\sigma^p)$ of fixed points of σ^p is $\mathrm{Tr}\, A^p$:

$$N_{\text{Fix}}(\sigma^p) = \mathrm{Tr}\, A^p. \tag{26.10}$$

Notice that this is closely related to the number of periodic orbits:

$$N_{\text{Fix}}(\sigma^n) = \sum_{(k,\tau):k\tau=n} \tau, \tag{26.11}$$

where τ is the period of a periodic orbit τ, and the summation is over all the pairs (k, τ).

26.8 Zeta function of dynamical system

The zeta function ζ_f for a dynamical system $f \in C^r(M)$ is generally defined as

$$\zeta_f(s) = \exp\left[\sum_{n=1}^{\infty} \frac{N_{\text{Fix}}(f^n)}{n} s^n\right]. \tag{26.12}$$

Using (26.11), we get

$$= \exp\left[\sum_{n=1}^{\infty} \frac{1}{n} \sum_{(k,\tau):k\tau=n} \tau s^n\right]. \tag{26.13}$$

Here, all the combinations (k, τ) appear once and only once; the totality of τ (if orbits are distinct, even if they have the same τ they must be counted separately) is system dependent, but if the system has one τ, $k\tau$ for all $k \in \mathbb{N}^+$ appear. Thus, for each τ the summation over k runs from 1 to ∞:

$$\begin{aligned} \zeta_f(s) &= \exp\left[\sum_{(k,\tau)} \frac{1}{k\tau} \tau s^{k\tau}\right] && (26.14)\\ &= \exp\left[\sum_{\tau}\sum_{k=1}^{\infty} \frac{1}{k} s^{k\tau}\right] = \exp\left[-\sum_{\tau} \log(1-s^\tau)\right] && (26.15)\\ &= \prod_{\tau} (1-s^\tau). && (26.16) \end{aligned}$$

[318] Notice that (26.9) and its 'cyclic permutation', e.g., $(i_2 i_3 \cdots i_p i_1 i_2 i_3 \cdots i_p i_1 \cdots)$ are distinct points.

Here the summation (or product) over τ is over all the periods allowed to the system. This is why it is called the zeta function.

26.9 Zeta function for Markov subshift

Let us compute the zeta function for (σ, Σ_A) with the structure matrix A. We have an explicit formula for $N_{\mathrm{Fix}}(\sigma^n)$, so[319]

$$\zeta_\sigma(s) = \exp\left[\sum_{n=1}^{\infty} \frac{\mathrm{Tr}\, A^n}{n} s^n\right] = \exp[-\mathrm{Tr}\, \log(1 - sA)] = 1/\det(1 - sA). \tag{26.17}$$

Thus, the inverse of the eigenvalue of A with the largest modulus gives the convergence radius ρ_A for the zeta function.

As we will see, $-\log \rho_A$ is the topological entropy of the dynamical system (the sup of the KS entropy). The invariant measure that realizes this sup value (thus, actually the max value) is the invariant measure of the Markov chain defined by the transition matrix A.

26.10 Topologically transitive Markov subshift

If there is $m \in \mathbb{N}$ such that $A^n > 0$ (all the elements positive), we say the resultant Markov subshift is topologically transitive: there is an orbit starting from any symbol to reach any symbols with a finite number of steps.

A topologically transitive Markov subshift is topologically mixing and periodic orbits are dense in Σ_A.[320]

The Perron-Frobenius theorem **26.11** tells us that $\rho_A = 1/\lambda_{\mathrm{PF}}$, where λ_{PF} is the Perron-Frobenius eigenvalue of A.

26.11 Perron-Frobenius theorem

Let A be a square matrix whose elements are all non-negative, and there is a positive integer n such that all the elements of A^n are positive. Then, there is a nondegenerate real positive eigenvalue λ such that
(i) $|\lambda_i| < \lambda$, where λ_i are eigenvalues of A other than λ,[321]

[319] Recall $\log\det(A) = \log(\prod \text{eigenvalue of } A) = \sum \log(\text{eigenvalue of } A) = \mathrm{Tr}\, \log A$.

[320] P1.9.9 KH p51

[321] That is, λ gives the spectral radius of A.

(ii) the elements of the eigenvector belonging to λ may be chosen all positive.
This special real eigenvalue giving the spectral radius is called the Perron-Frobenius eigenvalue.

For the 'shortest demonstration' see **26.13**.[322]

26.12 Preview of thermodynamic formalism

As you realize, thus dynamical systems are closely related to 1D spin systems. Periodic states correspond to ordered states in spin systems; thus, periodic dynamical systems correspond to spin systems with long-range interactions. Chaotic states corresponds to high-temperature states. Critical phenomena in spin systems correspond to Feigenbaum critical phenomena **8.7**.

Sinai introduced the thermodynamic formalism to understand dynamical systems, which systematize the ideas above. Invariant measures correspond to Gibbs measures. Observability of chaos may be characterized by a special temperature.

26.13 Proof of the Perron-Frobenius theorem[323]

Let us introduce the vectorial inequality notation for n-column vectors): $\boldsymbol{x} > 0$ (≥ 0) implies that all the components of $\boldsymbol{x}$ are positive (non-negative). Also let us write $\boldsymbol{x} \geq (>)$ $\boldsymbol{y}$, if $\boldsymbol{x} - \boldsymbol{y} \geq (>)\ 0$.

We use analogous symbols for $n \times n$ matrices A, $B, \cdots$ as well. $A > 0$: (≥ 0) implies that all the components of A are positive (non-negative). Also let us write $A \geq (>)$ B, if $A - B \geq (>)\ 0$.

We consider $n \times n$ matrices $A \geq 0$ for which there is a positive integer m such that $A^m > 0$.

Let $\boldsymbol{x}$ be a vector such that $|\boldsymbol{x}| = 1$ and $\boldsymbol{x} \geq 0$. The largest ρ satisfying

$$A\boldsymbol{x} \geq \rho \boldsymbol{x} \tag{26.18}$$

is denoted by $\Lambda(\boldsymbol{x})$: $\Lambda(\boldsymbol{x}) = \max\{\rho \,|\, A\boldsymbol{x} \geq \rho\boldsymbol{x}\}$.

(o) Note that

$$\Lambda(\boldsymbol{x}) = \min_{i \text{ s.t. } x_i \neq 0} \frac{(A\boldsymbol{x})_i}{x_i}, \tag{26.19}$$

so it is continuous for $\boldsymbol{x} > 0$.

(o′) $A^m > 0$ implies $A^m\boldsymbol{x} > 0$ for $\boldsymbol{x} \geq 0$ $(\neq 0)$.

[322] The newest version, taking account of H Tasaki's critical comments (April, 2018).

[323] A standard reference may be E. Seneta, *Non-negative matrices and Markov chains* (Springer, 1980). The proof here is an eclectic version due to many sources, including N. Iwahori, *Graphs and Stochastic Matrices* (Sangyo-tosho, 1974) and S. Sternberg: `http://www.math.harvard.edu/library/sternberg/slides/1180912pf.pdf`..

(i) Define a compact[324] set $U = \{\boldsymbol{x} \,|\, \boldsymbol{x} \geq 0, |\boldsymbol{x}| = 1\}$ and a continuous map Q on it as

$$Q\boldsymbol{x} \equiv \frac{A^m\boldsymbol{x}}{|A^m\boldsymbol{x}|} > 0. \tag{26.20}$$

Then, $\Lambda(Q\boldsymbol{x}) \geq \Lambda(\boldsymbol{x})$ on U.
[Demo] Let $A\boldsymbol{x} \geq \lambda\boldsymbol{x}$ for $\boldsymbol{x} \in U$. Then,

$$A^m A\boldsymbol{x} = AA^m\boldsymbol{x} \geq \lambda A^m\boldsymbol{x} \;\Rightarrow AQ\boldsymbol{x} \geq \lambda Q\boldsymbol{x}. \tag{26.21}$$

Therefore, $\Lambda(Q\boldsymbol{x}) \geq \Lambda(\boldsymbol{x})$ for any $\boldsymbol{x} \in U$.

(ii) There is a vector $\boldsymbol{z} \in U$ that maximizes $\Lambda(\boldsymbol{x})$.
[Demo] Let $C = Q(U)$. Notice that

$$C \subset \{\boldsymbol{x} \,|\, \boldsymbol{x} > 0, |\boldsymbol{x}| = 1\}, \tag{26.22}$$

because for any $\boldsymbol{x} \in U$ $\tilde{\boldsymbol{x}} = Q\boldsymbol{x}$ satisfies $|\tilde{\boldsymbol{x}}| = 1$ and $\tilde{\boldsymbol{x}} > 0$. Thus. $C \subset U$. Since Q is continuous on U, C is compact. $\boldsymbol{x} > 0$ for any $\boldsymbol{x} \in C$, so $\Lambda(\boldsymbol{x})$ is continuous and has a max value on C. For any $\boldsymbol{x} \in U$ (i) says $\Lambda(Q\boldsymbol{x}) \geq \Lambda(\boldsymbol{x})$, so Λ defined on U is maximum on C, since $C \subset U$.

(iii) Let us write $\lambda(A) = \Lambda(\boldsymbol{z})$. That is, $\lambda(B) = \max_{\boldsymbol{x} \in U}\{\Lambda(\boldsymbol{x}) \,|\, B\boldsymbol{x} \geq \Lambda(\boldsymbol{x})\boldsymbol{x}\}$ for any $n \times n$ matrix with some positive m such that $B^m > 0$. $\lambda(A)$ is an eigenvalue of A, and $\boldsymbol{z}$ belongs to its eigenspace: $A\boldsymbol{z} = \lambda(A)\boldsymbol{z}$.
[Demo] Even if not, we have $\boldsymbol{w} = A\boldsymbol{z} - \lambda(A)\boldsymbol{z} \geq 0$ (not equal to zero) by definition of $\lambda(A)$. Notice that for any vector $\boldsymbol{x} \geq 0$ $A^m\boldsymbol{x} > 0$, so unless $\boldsymbol{w} = 0$

$$A^m\boldsymbol{w} = AA^m\boldsymbol{z} - \lambda(A)A^m\boldsymbol{z} > 0. \tag{26.23}$$

This implies $\Lambda(A^m\boldsymbol{z}) > \lambda(A)$, but $\lambda(A)$ is the maximum of Λ, so this is a contradiction. Therefore, $\boldsymbol{w} = 0$. That is, $\boldsymbol{z}$ is an eigenvector belonging to λ.

(iv) $\boldsymbol{z} > 0$ because max of Λ is on C.

(v) λ is the spectral radius of A.
[Demo] Suppose $A\boldsymbol{y} = \lambda'\boldsymbol{y}$. Let $\boldsymbol{q}$ be the vector whose components are absolute values of $\boldsymbol{y}$: $q_i = |y_i|$. Then, $A\boldsymbol{q} \geq |\lambda'|\boldsymbol{q}$ (as seen from (26.19)). Therefore, $|\lambda'| \leq \lambda(A)$.

(vi) The absolute values of other eigenvalues are smaller than $\lambda(A)$. That is, no eigenvalues other than $\lambda(A)$ is on the spectral circle.
[Demo] Suppose λ' is an eigenvalue on the spectral circle but is not real positive. Let $\boldsymbol{q}$ be the vector whose components are absolute values of an eigenvector belonging to λ'. Since $A\boldsymbol{q} \geq |\lambda'|\boldsymbol{q} = \lambda(A)\boldsymbol{q}$, actually we must have $A\boldsymbol{q} = \lambda(A)\boldsymbol{q}$. Therefore, the absolute value of each component of the vector $A^m\boldsymbol{y} = \lambda'^m\boldsymbol{y}$ coincides with the corresponding component of $A^m\boldsymbol{q}$. This implies

$$\left|\sum_j (A^m)_{ij}y_j\right| = \sum_j (A^m)_{ij}|y_j| = \sum_j |(A^m)_{ij}y_j|. \tag{26.24}$$

[324] Since we deal with finite-dimensional vector spaces, 'closed' can always replace 'compact.'

All the components of A^m are real positive. Therefore, all the arguments of y_j are identical,[325] so $\boldsymbol{y}$ is parallel to $\boldsymbol{q}$. Hence, $\lambda' = \lambda(A)$.

(vii) $\lambda(A)$ is non-degenerate.
We show that $\lambda(A)$ is a non-degenerate root of the characteristic equation $\det(\lambda I - A)$. We use
(viia) For any matrix $B \geq 0$ ($\neq 0$) $\det(\lambda I - B) > 0$ for real $\lambda > \lambda(B)$ (= the spectral radius).
(viib) Let $A_{(i)}$ be the matrix with the ith row and ith column removed from A. Then,[326]

$$\frac{d}{d\lambda}\det(\lambda I - A) = \sum_i \det(\lambda I - A_{(i)}). \tag{26.25}$$

Let $A_{[i]}$ be the matrix with all the elements in ith row and ith column of A being replaced with 0. Suppose $A_{[i]}\boldsymbol{z} = \sigma\boldsymbol{z}$ ($\boldsymbol{z} \neq 0$). Then, for all i

$$A|\boldsymbol{z}| \geq A_{[i]}|\boldsymbol{z}| \geq |\sigma||\boldsymbol{z}|. \tag{26.26}$$

That is $\lambda(A) \geq |\sigma|$. If $\lambda(A) = |\sigma|$, then $A|\boldsymbol{z}| = |\sigma||\boldsymbol{z}|$, so $|\boldsymbol{z}| > 0$ and $(A - A_{[i]})|\boldsymbol{z}| = 0$, but this is impossible. Therefore, the spectral radius of $A_{[i]}$ is smaller than $\lambda(A)$ for any i. Since $A_{[i]}$ and $A_{(i)}$ have the same spectral circle, (viia) implies all the terms in (26.25) are positive for $\lambda = \lambda(A)$, so (viib) implies $\lambda(A)$ is non-degenerate.

[325] $a, b \neq 0$ and $|a + b| = |a| + |b|$ imply the real positivity of a/b. This means inductively that $|\sum a_i| = \sum |a_i|$ implies all a_i must have the same argument.

[326] Let $X = n \times n$ diagonal matrix with the diagonal $x_1, \cdots, x_n$. Then, expanding the determinant with respect to the ith row, obviously,

$$\frac{d}{dx_i}\det(X - A) = \det(X_{(i)} - A_{(i)}),$$

where $X_{(i)} = (n-1) \times (n-1)$ diagonal matrix with the diagonal $x_1, \cdots, (x_i), \cdots, x_n$ (x_i omitted). The rest is due to the chain rule.

27 Lecture 27. Baker's transformation

27.1 Baker's transformtion

The baker's transformation is a one-to-one map $T : M = [0,1)^2 \to [0,1)^2$ defined as follows:

$$T(x,y) = \begin{cases} (2x, y/2) & x \in [0,1/2) \\ (2x-1, (y+1)/2 & x \in [1/2,1) \end{cases} . \tag{27.1}$$

The transformation and its inverse are illustrated in Fig. 27.1

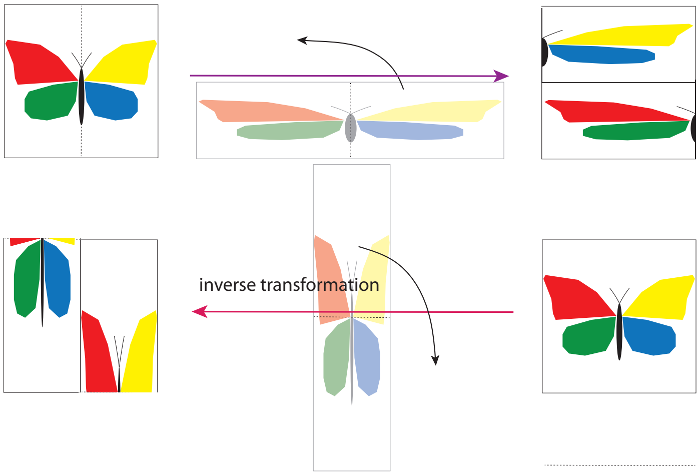


Figure 27.1: Baker's transformation and its inverse

27.2 Stable and unstable manifolds of baker's transformation

M may be decomposed into vertical sets and horizontal sets. The vertical set containing (x,y) is

$$\gamma_-(x,y) = \{(x,y_1)\,|\,y_1 \in [0,1)\} \tag{27.2}$$

and the horizontal set containing (x,y) is

$$\gamma_+(x,y) = \{(x_1,y)\,|\,x_1 \in [0,1)\}. \tag{27.3}$$

Since T expands $\gamma_+(x,y)$, it is an (local) unstable manifold of (x,y); we see $\gamma_-(x,y)$ is a (local) stable manifold of (x,y).[327]

Notice that this decomposition defines an equivalence relation: If $x \in \gamma_\pm(y) \Rightarrow$

[327] 'local' in general, because they may not be continuous.

$\gamma_\pm(x) = \gamma_\pm(y)$.

27.3 Invariant measure of baker's transformation

Obviously the area is preserved, so the usual Lebesgue measure is an invariant measure.

We can show that the measure is a mixing (of course ergodic) measure.

27.4 Symbolic dynamical expression of baker's transformation

Let us define the partition $\mathcal{A} = \{M_0, M_1\}$ (see **32.2** for a precise definition): $M = M_0 \cup M_1$, where $M_0 = [0, 1/2) \times [0, 1)$ and $M_1 = [1/2, 1) \times [0, 1)$.

How this partition is transformed according to T or T^{-1} may be understood easily from the figure 27.2.

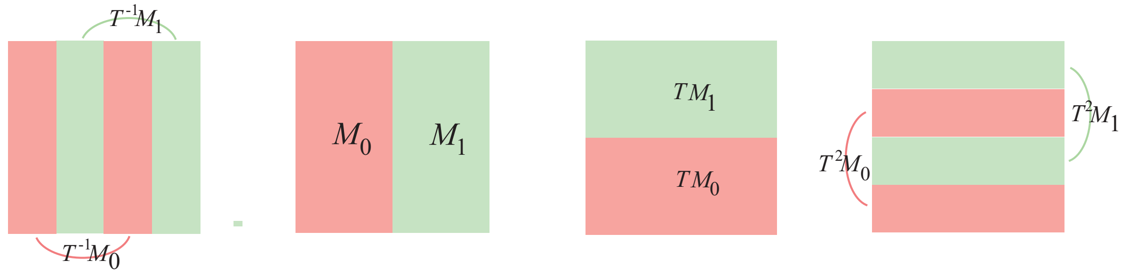


Figure 27.2: The fate of partition $M_0 \vee M_1$

As is clear from the figure $\vee_{n=0}^{\infty} T^{-n}\mathcal{A}$ consists of the totality of γ_+, and $\vee_{n=0}^{\infty} T^{n}\mathcal{A}$ consists of the totality of γ_-. Thus, an element of $\vee_{-\infty}^{\infty} T^{n}\mathcal{A}$ specifies a point in M. This allows us to assign a 01 sequence to each point in M such that TM corresponds to the shift on $\{0, 1\}^{\mathbb{Z}}$. The rule may be more explicitly stated as follows: For $x \in M$ if $T^{-n}x \in M_\alpha$ ($\alpha = 0$ or 1) $\omega(x)_n = \alpha$.

The correspondence is not one to one. $M \to \{0, 1\}^{\mathbb{Z}}$ is injective. However, for binary rational numbers its binary expansion is not unique. However, these points are measure zero, so as a probabilistic system (= measure-theoretical dynamical system) we may totally ignore them and identify baker's transformation (with the Lebesgue measure) and the Bernoulli system $B(1/2, 1/2)$.

28 Lecture 28 Horseshoe

28.1 Horseshoe dynamical system

The horseshoe dynamical system is constructed as follows:
First a map $g : D \to D$ in Fig. 28.1 is constructed that maps the (yellow) square ABCD to a horseshoe overlapping the original square (which we will call R):

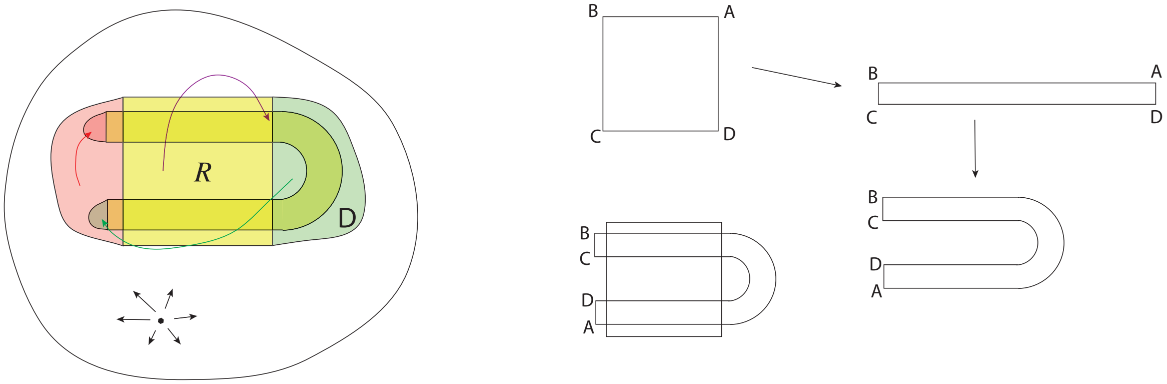


Figure 28.1: Construction of horseshoe

Then, prepare one more disk and complete S^2. Thus as an example of $C^r(S^2, S^2)$ the horseshoe dynamical system has been constructed. Let us write this (η, S^2).

There is a unique sink in the red semidisk in Fig. 28.1.

The inverse map η^{-1} is also a horseshoe map: $\eta^{-1}(R)$ (R = yellow square in Fig. 28.1) has again a horseshoe shape. Can you illustrate this?

28.2 Where do horseshoes appear naturally?

The construction of the horseshoe system might look rather artificial, but horseshoes appear 'everywhere' when we have homoclinic points as illustrated in Fig. 28.2.

28.3 All the nontrivial nonwandering points are in R

By construction, all the points outside R is attracted to the sink of η. If we reverse time, all the points outside R is attracted to the source in Fig. 28.1. Therefore, all

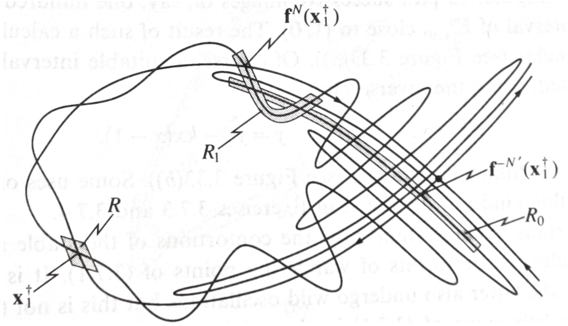


Figure 28.2: Horseshoes can appear naturally when there is a homoclinic point, [Fig. 3.32 of AP]

the other nonwandering points must be in R.

Therefore, all other nonwandering points must be in

$$\Lambda = \cap_{n\in\mathbb{Z}}\eta^n(R). \tag{28.1}$$

28.4 Nontrivial nonwandering set of horseshoe

To understand the structure of Λ let us consider how R is successively mapped onto R. We can understand the preimage of the blue rectangles as follows (Fig. 28.3).

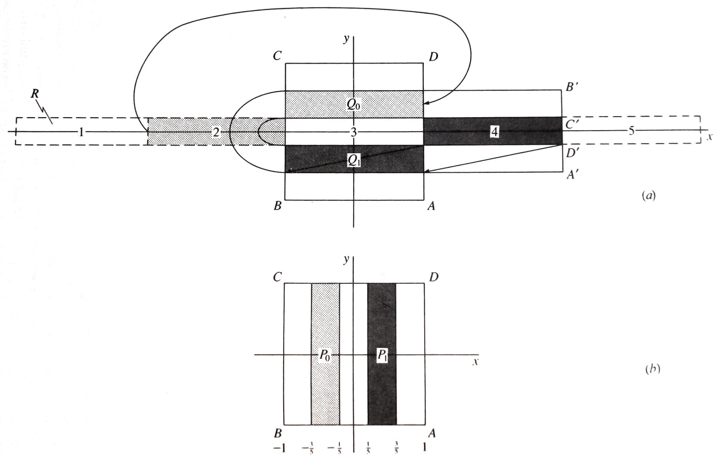


Figure 28.3: η on R. P_a is mapped to Q_a [Fig. 3.13 of AP]

If we repeat mapping we have Fig. 28.4.

Thus, $\vee_{n\in\mathbb{N}}Q^{(n)}$ consists of local unstable manifolds of points in Λ (See Fig. 28.4).

We know η^{-1} is also a horseshoe (90° rotated). Therefore, we can repeat the above argument in the reverse time direction to construct (local) stable manifolds of points

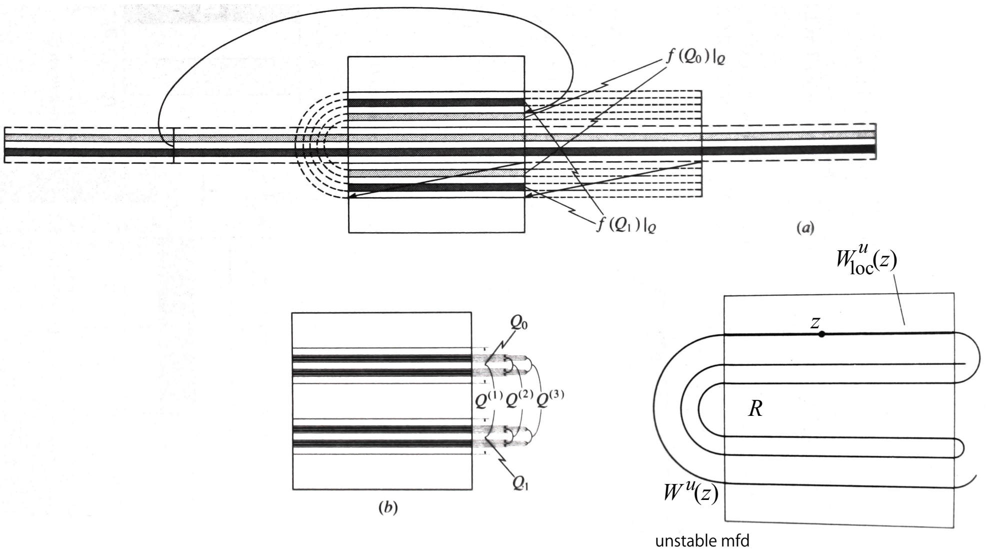


Figure 28.4: η, η^2 and η^3 on R. Here, $Q_0 \cup Q_1$ is $Q^{(1)}$, $\eta(Q_0 \cup Q_1) \cap R$ is $Q^{(2)}$ and $\eta^2(Q_0 \cup Q_1) \cap R$ is $Q^{(3)}$. [Fig. 3.14 of AP]

in Λ; $Q^{(-1)}, Q^{(-2)}, \cdots$, instead (Fig. 28.5)

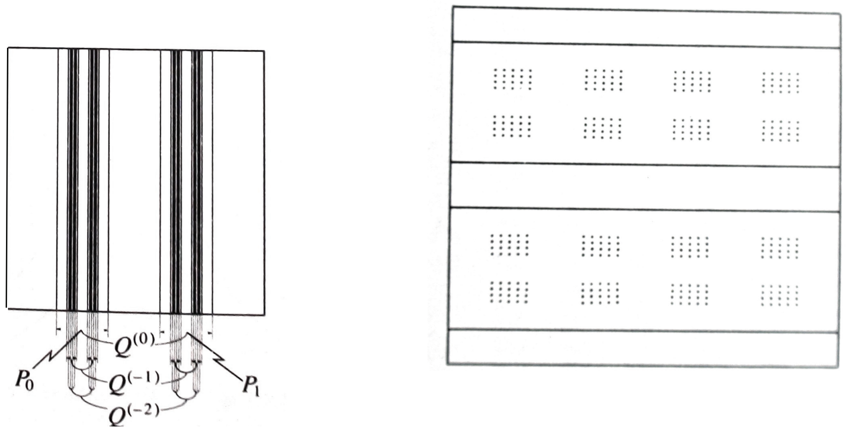


Figure 28.5: Left: η^{-1}, η^{-2} ... on R. As Fig. 28.4 we can make $Q^{(-1)}, Q^{(-2)}, \cdots$. Right: Λ[Fig. 3.15c of AP, 8.5 of Yano]

The invariant set of the horseshoe consists of a source, a sink (in the red region in Fig. 28.1) and a Cantor set Λ.
https://www.youtube.com/watch?v=ItZLb5xI_1U&frags=pl%2Cwn illustrates the horseshoe invariant set (perhaps you feel the pace of explanation a bit too slow).

The horseshoe is structurally stable as can be guessed from the following illustration (Fig. 28.6)

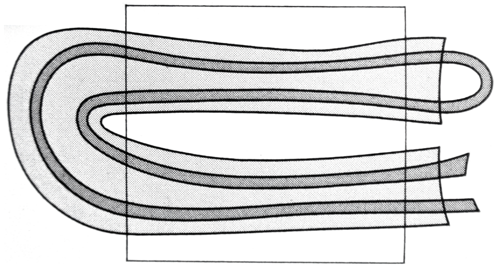


Figure 28.6: The structure of the horseshoe is quite stable [Fig. 8.6 of Yano]

https://www.youtube.com/watch?v=2aeFG5YN_mk&frags=pl%2Cwn Smale talks about his dynamical system study

28.5 Symbolic dynamics of horseshoe

As can be clear from the structure of Λ and its construction, each point in Λ must be coded uniquely in terms of an element in $\{0,1\}^{\mathbb{Z}}$. Thus $\eta|_\Lambda$ is (as a measure-theoretical dynamical system) isomorphic to $B(1/2,1/2)$. Thus the horseshoe system is not Morse-Smale.

Actually $\eta|\Lambda$ is an Anosov system, so it is Ω-stable.

29 Lecture 29. Measure-theoretical dynamical systems

29.1 Why measure-theoretical dynamical systems?

We have realized that even for a single dynamical system (that is defined by a definite law or rule) its behavior depends on initial conditions. Thus, even if we are told that the system exhibit periodic movements or apparently random (that is, chaotic) behaviors, we may not be lucky enough to observe them. How lucky can we be? That is a question of probability, and probability is a 'measure' of our confidence.

For example, for a continuous time dynamical system $\mathcal{X}^r(M)$ defined on M, a trajectory starting from $x_0 \in M$ at $t = 0$ $\varphi_t(x_0)$ depends on x_0. Suppose the system exhibits an unstable periodic orbit, you can observe it only when x_0 is on the orbit. What is your chance to observe it? If we throw a dart 'randomly' on M, you would expect that you would never hit the orbit. However, if there is some device channeling the darts on to the orbit, with probability one, you will observe it. If x_0 is in the basin of attraction of a stable periodic orbit and the basin is of positive Riemann volume defined on M, then we may have a finite (not infinitesimal) chance to observe the periodic motion. Even in this case if there is a fixed point to whose basin of attraction all the initial conditions are channelled, there will be no chance to observe the stable periodic orbit.

These trivial considerations clearly tell us that it is very important (especially for physicists) to specify how we can sample the initial condition. This is the question of the sampling measure.

Once the system settles down to a steady state, that is, the system is in a certain attractor, then we may be interested in its average behavior. For example, in the case of a chaotic attractor, we could ask how nearby orbits leave each other on the average (that is, we ask 'how chaotic' the system is). Inevitably we need a measure to average the behavior.

Thus, if we ask about observability and about the average behavior, we need (probability) measures.

29.2 Measure

What is a measure? See Appendix. You should be able to answer the question: What is the area?

29.3 Remark on family of measurable sets

If you read a respectable book on probability you will see a setup of probability space $(M, \mathcal{F}, \mu)$, where μ is a (probability) measure defined on M and $\mathcal{F}$ is a set of all the measurable sets on which μ is defined. Not all the subsets of M can be assumed to have 'probability.' This restriction is a must under the usual axioms of mathematics.

Therefore, when we consider a measure-theoretical dynamical system on M with an invariant measure the 'dynamical rule' (say, the map defining a discrete time dynamical system) must be compatible with (preserved) $\mathcal{F}$. When we say two dynamical systems are isomorphic (= equivalent), then there must be a correspondence (modulo μ-measure zero sets) between the families of measurable sets for these dynamical systems.

However, in this exposition we will not mention about $\mathcal{F}$.

29.4 Invariant measure: introduction

Prepare numerous clones of the dynamical system under consideration, and plot their states in the same phase space. Then, we would see a cloud of points that describe individual clones in the phase space (Fig. 29.1).

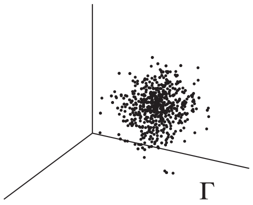


Figure 29.1: Distribution or ensemble. The space is the phase space Γ. Each point represents an instantaneous state of each system. An invariant measure corresponds to a steady state for which the cloud as a whole becomes time-independent, although individual points may keep wandering around.

The distribution described by the cloud, if the total mass is normalized to be unity, may be understood as a probability measure on the phase space. Following the time evolution of the system, the cloud may change its shape.

After a sufficiently long time, often the cloud representing an ensemble ceases to change its shape. Then, we say that the ensemble has reached its steady state. Individual points corresponding to the members of the ensemble may still keep wandering around in the cloud, but the cloud as a whole is balanced and the distribution on the phase space becomes invariant. The (probability) distribution corresponding to this invariant cloud is called an *invariant measure* of the dynamical system.

29.5 Invariant measure of dynamical system: discrete time

Discrete time case: Measure μ on the phase space M is an invariant measure of a dynamical system $T \in C^r(M)$, if

$$\mu = \mu \circ T^{-1} \tag{29.1}$$

holds. That is, for an arbitrary (μ-measurable) subset $A \subset M$

$$\mu(A) = \mu(T^{-1}A) \tag{29.2}$$

holds, where $T^{-1}A$ is the totality of points that comes to A after a unit time step (Fig. 29.2).

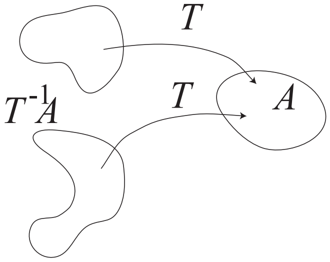


Figure 29.2: The preimage $T^{-1}A$ of A by the time evolution operator T need not be connected (in this figure it consists of two connected components). Since $T(T^{-1}A) = A$, $T^{-1}A$ is the totality of the points that come to A after one time step. As you see from the figure, unless T is invertible, $\mu = \mu \circ T$ does not guarantee the invariance of μ.

For an arbitrary measurable set A, if we know $T^{-1}A$ at present, we know everything we can discuss probabilistically that will occur after one time step. (29.2) expresses the condition that the evaluation of the weights of A and $T^{-1}A$ by μ is consistent with the conservation law of th members of the ensemble.

29.6 Invariant measure of dynamical system: continuous time

Continuous time case: Measure μ on the phase space M is an invariant measure of a dynamical system φ_t defined by a vector field $\mathscr{X}(M)$, if

$$\mu = \mu \circ \varphi_t^{-1} \tag{29.3}$$

holds for all t. That is, for an arbitrary (μ-measurable) subset $A \subset M$

$$\mu(A) = \mu(\varphi_t^{-1}A) \tag{29.4}$$

holds (quite analogous to the discrete time case).

29.7 Invariant measures are not unique for a given dynamical system

Generally speaking, an invariant measure for a given dynamical systems is not unique (as already suggested in **29.1**). For example, $Tx = \{2x\}$ discussed previously has actually uncountably many distinct invariant measures.

For a Hamiltonian system we know the phase volume is invariant, but there are many other invariant measures. Furthermore, not all the invariant measures can give 'natural averages' that agree with averages obtained by continuous even observations.[328]

29.8 Measure-theoretical dynamical system

Let μ be an $T \in C^r(M)$-invariant measure on M. The triplet (T, μ, M) is called a *measure theoretical dynamical system*. This may be interpreted as a mathematical expression of a steady state allowed to the dynamical system $T \in C^r(M)$; given a dynamical system $T \in C^r(M)$, for each invariant measure μ, a distinct measure theoretical dynamical system is constructed.

For a continuous time dynamical system defined by $X \in \mathcal{X}^r(M)$, (X, μ, M) may be understood as a measure-theoretical dynamical system.

29.9 Absolutely continuous invariant measure

If a measure μ satisfies $\mu(A) = 0$ for any Lebesgue-measure zero set[329] A, μ is called an absolutely continuous measure. For a dynamical system with an absolutely continuous invariant measure, chaos may often be observed by numerical experiments.

If a measure is absolutely continuous, its probability density g may be defined as $d\mu = g d\lambda$, where λ is the Lebesgue measure.[330]

Absolutely continuous invariant measures are observable (see **22.3**) (e.g., computationally).

[328] As we will see later, invariance does not mean ergodicity.

[329] More precisely, we should say 'for any set whose Riemann volume is zero.' Riemann volume is the volume based on the Riemann metric = the usual length. It seems that the volume based on the length is the most natural measure for us.

[330] That is, g is the Radon-Nikodym derivative $d\mu/d\lambda$. See, e.g., Kolmogorov and Fomin cited already.

29.10 Ergodic measure

For any invariant set $A \subset M$ (that is, any set satisfying $A = T^{-1}A$[331]), if $\mu(A) = 0$ or 1, the measure theoretical dynamical system is said to be *ergodic*.

Note that this is neither the property of the measure nor the property of the dynamicarule T.

Roughly speaking, if a dynamical system is ergodic, the trajectory starting from any initial point[332] can eventually go into any μ-measure positive set.

In fact, if there is a positive measure set C to which any trajectory starting from a certain positive measure set D cannot reach, then there must be an invariant set B with $0 < \mu(B) < 1$ such that $B \supset C$ and $B^c \supset D$.

Conversely, if there is an invariant set B such that $0 < \mu(B) < 1$. Then, we can choose a measure positive set C with $B \cap C = \emptyset$ such that any trajectory starting from C never visits B [if possible, $\mu(\cup_{k=1}^{\infty} T^{-k}B \cap C) > 0$ (note that $T^{-k}B \cap C$ means the points now in C and will be in B after k), but $T^{-1}B = B$, so $B \cap C \neq \emptyset$, a contradiction].

Let Γ be a unit circle, and f be a rigid rotation around the center by an angle that is irrational multiple of 2π. Since the uniform distribution μ_U on the circle is rotation-invariant, we can make a measure-theoretical dynamical system (f, μ_U, Γ), which is ergodic. This not so interesting example illustrates why ergodicity is totally insufficient for modeling irreversible phenomena such as relaxation exhibited by usual many-body systems.

29.11 Mixing measure

For any sets[333] A and B (both $\subset M$), if

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B) \tag{29.5}$$

holds, the measure-theoretical dynamical system is said to be *mixing*, where T^{-n} implies to apply T^{-1} n-times. $T^{-n}A$ is the set that agrees with A after n time steps. $T^{-n}A \cap B$ is the totality of the points that is at present in B and will be in A after n time steps. Intuitively, the cloud of points starting from B spread over the phase space evenly if n is sufficiently large, so the probability for the cloud to overlap with A is proportional to the 'statistical weight' of A. It is obvious that initial conditions

[331] This condition must be, precisely speaking, the equality ignoring the μ-measure zero difference.

[332] Precisely speaking, starting from μ-almost all initial points. 'μ-almost all' implies that all except for μ-measure zero sets.

[333] Precisely speaking, any μ-measurable sets. From now on, such statements will not be added.

cannot be used to predict the future. Physically, irreversible processes such as relaxation phenomena occur. In particular, the time-correlation function eventually decays to zero.

Appendix: What is measure[334]

The concept of measure does not appear in elementary calculus, but it is a fundamental and important concept. It is not very difficult to understand, since it is important. Besides, the introduction of the Lebesgue measure by Lebesgue is a good example of conceptual analysis, so let us look at its elementary part. A good introductory book for this topic is the already quoted Kolmogorov-Fomin. It is desirable that those who wish to study fundamental aspects of statistical mechanics and dynamical systems have proper understanding of the subject.

29.12 What is the volume?

For simplicity, let us confine ourselves to the two dimension. Thus, the question is: what is the area? Extension to higher dimensions should not be hard. If the shape of a figure is complicated, whether it has an area could be a problem,[335] so let us begin with an apparently trivial case.

"The area of the rectangle $[0, a] \times [0, b]$ is ab."

Is this really so? If so, why is this true? Isn't it strange that we can ask such a question before defining 'area'? Then, if we wish to be logically conscientious, we must accept the following definition:

Definition. The area of a figure congruent to the rectangle $\langle 0, a\rangle \times \langle 0, b\rangle$ (here, '$\langle$' implies '[' or '(', '$\rangle$' is ']' or ')', that is, we do not care whether the boundary is included or not) is defined as ab.

Notice that the area of a rectangle does not depend on whether its boundary is included or not. This is already included in the definition.

29.13 The area of a fundamental set

A figure made as the direct sum (that is, join without overlap except at edges and vertices) of a finite number of rectangles (whose edges are parallel to the coordinate axes and whose boundaries may or may not be included) is called a fundamental set (Fig. 29.3). It should be obvious that the join and the product (common set) of two fundamental sets are both fundamental sets. The area of a fundamental set is defined as the total sum of the areas of the constituent rectangles.

29.14 How to define the area of more complicated figures; a strategy

For a more complicated figure, a good strategy must be to approximate it by a sequence of fundamental sets allowing increasingly smaller rectangles. Therefore, following Archimedes, we approximate the figure from inside and from outside (that is, the figure is approximated by a sequence of fundamental sets enclosed by the figure and by a sequence of fundamental sets enclosing the figure). If the areas of the inside and the outside approximate sequences

[334] Taken from TNW Chapter 2 Appendix

[335] (Under the usual axioms of mathematics) we encounter figures without areas.

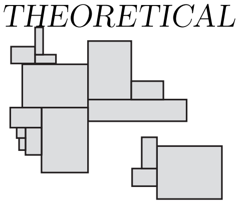


Figure 29.3: Fundamental set: it is a figure made of a finite number of rectangles whose edges are parallel to a certain Cartesian coordinate axes and the rectangles do not overlap except at edges and vertices. Its area is the total sum of the areas of the constituent rectangles.

agree in the limit, it is rational to define the area of the figure by the limit.

Let us start from outside.

29.15 Outer measure

Let A be a given bounded set (that is, a set that may be enclosed in a sufficiently large disk). Using a finite number of (or countably many) rectangles P_k ($k = 1, 2, \cdots$), we cover A, where the boundaries of the rectangles may or may not be included, appropriately. If $P_i \cap P_j = \emptyset$ ($i \neq j$) and $\cup P_k \supset A$, $P = \{P_k\}$ is called a finite (or countable) cover of A by rectangles (Fig. 29.4O). Let the area of the rectangle P_k be $m(P_k)$. We define the outer measure $m^*(A)$ of A as follows:

$$m^*(A) \equiv \inf \sum_k m(P_k). \tag{29.6}$$

Here, inf[336] is taken over all the possible finite or countable covers by rectangles.

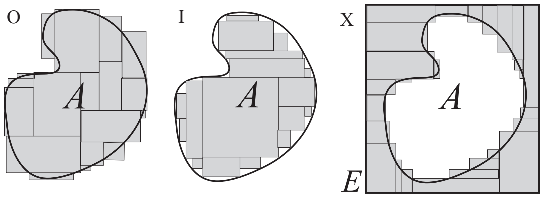


Figure 29.4: Let A be the set enclosed by a closed curve. O denotes a finite cover by rectangles. If there is an area of A, it is smaller than the sum of the areas of these rectangles. The outer measure is defined by approximating the area from outside. In contrast, the inner measure is computed by the approximation shown in I by the rectangles included in the figure A. In the text, by using a large rectangle E containing A, $E \setminus A$ is made and its outer measure is computed with the aid of finite covers; the situation is illustrated in X. The relation between I and X is just the relation between negative and positive films. If the approximation O from outside and the approximation I from inside agree in the limit of refinement, we may say that A has an area. In this case, we say A is measurable, and the agreed area is called the area of A.

[336] The infimum of a set of numbers is the largest number among all the numbers that are not larger than any number in the set. For example, the infimum of positive numbers is 0. As is illustrated by this example, the infimum of a set need not be an element of the set. When the infimum is included in the set, it is called the minimum of the set. The above example tells us that the minimum need not exist for a given set.

29.16 Inner measure

For simplicity, let us assume that A is a bounded set. Take a sufficiently large rectangle E that can enclose A. Of course, we know the area of E is $m(E)$. The inner measure of A is defined as[337]

$$m_*(A) = m(E) - m^*(E \setminus A). \tag{29.7}$$

It is easy to see that this is equivalent to the approximation from inside (Fig. 29.4I). Clearly, for any bounded set A $m^*(A) \geq m_*(A)$ holds.

29.17 Area of figure, Lebesgue measure

Let A be a bounded set. If $m^*(A) = m_*(A)$, A is said to be a measurable set (in the present case, a set for which its area is definable) and $\mu(A) = m^*(A)$ is called its area (2-dimensional Lebesgue measure).

At last the area is defined. The properties of a fundamental set we have used are the following two:
(i) It is written as a (countable) direct sum of the sets whose areas are defined.
(ii) The family of fundamental sets is closed under $\cap$, $\cup$ and $\setminus$ (we say that the family of the fundamental sets makes a set ring.[338])
An important property of the area is its additivity: If P_i are mutually non-overlapping rectangles, $\mu(\cup P_i) = \sum \mu(P_i)$. Furthermore, the σ-additivity for countably many summands also holds.[339]

Notice that such a summary as that the area is a translationally symmetric σ-additive set-theoretical function which is normalized to give unity for a unit square does not work, because this does not tell us on what family of sets this set-theoretical function is defined.[340] The above summary does not state the operational detail about how to measure the areas of various shapes, so no means to judge is explicitly given what figures can be measurable. Lebesgue's definition of the area outlined above explicitly designates how to obtain the area of a given figure.

29.18 General measure (abstract Lebesgue measure)

The essence of characterization of the area is that there is a family of sets closed under certain 'combination rules' and that there is a σ-additive set-theoretical function on it. Therefore, we start with a σ-additive family $\mathcal{M}$ consisting of subsets of a set X: A family of sets satisfying the following conditions is called a σ-additive family:
(s1) $X, \emptyset \in \mathcal{M}$,
(s2) If $A \in \mathcal{M}$, then $X \setminus A \in \mathcal{M}$,

[337] $A \setminus B$ in the following formula denotes the set of points in A but not in B, that is, $A \cap B^c$.

[338] More precisely, that a family $\mathcal{S}$ of sets makes a ring implies the following two:
(i) $\mathcal{S}$ includes $\emptyset$,
(ii) if $A, B \in \mathcal{S}$, then both $A \cap B$ and $A \cup B$ are included in $\mathcal{S}$.

[339] Indeed, if $A = \cup_{n=1}^{\infty} A_n$ and A_n are mutually exclusive (i.e., for $n \neq m$ $A_n \cap A_m = \emptyset$), for an arbitrary positive integer N $A \supset \cup_{n=1}^{N} A_n$, so $\mu(A) \geq \sum_{n=1}^{N} \mu(A_n)$. Taking the limit $N \to \infty$, we obtain $\mu(A) \geq \sum_{n=1}^{\infty} \mu(A_n)$. On the other hand, for the external measure $m^*(A) \leq \sum_{n=1}^{\infty} m^*(A_n)$, so $\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

[340] If we assume that every set has an area, under the usual axiomatic system of mathematics, we are in trouble. See Banach-Tarski's theorem (Discussion 2.4A.1).

(s3) If $A_n \in \mathcal{M}$ $(n = 1, 2, \cdots)$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{M}$.
$(X, \mathcal{M})$ is called a measurable space. A nonnegative and σ-additive set-theoretical function m defined on a measurable set that assigns zero to an empty set is called a measure, and $(X, \mathcal{M}, m)$ is called a measure space. Starting with this measure m, we can define the outer measure on a general set $A \subset X$, mimicking the procedure already discussed above. The inner measure can also be constructed. When these two agree, we can define a set-theoretical function μ as $\mu(A) = m^*(A)$, and we say A is μ-measurable. Thus, we can define μ that corresponds to the Lebesgue measure explained above in the context of the area. μ is called the Lebesgue extension of m (this is called an abstract Lebesgue measure, but often this is also called a Lebesgue measure). This construction of μ is called the completion of m. In summary, if $(X, \mathcal{M}, m)$ is a measure space, we define a new family of subsets of X based on $\mathcal{M}$ as

$$\overline{\mathcal{M}} = \{A \subset X : \exists B_1, B_2 \in \mathcal{M} \text{ where } B_1 \subset A \subset B_2, m(B_2 \setminus B_1) = 0\}, \tag{29.8}$$

and if μ is defined as $\mu(A) \equiv m(B_2)$ for $A \in \overline{\mathcal{M}}$, $(X, \overline{\mathcal{M}}, \mu)$ is a measure space, and is called the completion of $(X, \mathcal{M}, m)$.[341]

The final answer to the question, "What is the area?" is: the area is the completion of the Borel[342] measure, where the Borel measure is the σ-additive translation-symmetric measure that gives unity for a unit square and is defined on the Borel family of sets which is the smallest σ-additive family of sets including all the rectangles. Generally speaking, a measure is something like a weighted volume. However, there is no guarantee that every set has a measure (μ-measurable). It is instructive that a quite important part of the characterization of a concept is allocated to an 'operationally' explicit description (e.g., how to measure, how to compute). Recall that Riemann's definition of the integral was based on this operational spirit, so it can immediately be used to compute integrals numerically.

[341] The completion is unique. In a complete measure space, if A is measure zero ($\mu(A) = 0$), then its subsets are all measure zero. Generally, a measure with this property is called a complete measure. Completion of $(X, \mathcal{M}, m)$ may be understood as the extension of the definition of measure m on the σ-additive family generated by all the sets in $\mathcal{M}$ + all the measure zero set with respect to m.

[342] E. Borel (1871-1956)

30 Lecture 30 Ergodic theorems

This section is based on I Kubo *Dynamical systems I* (Iwanami 1997) Section 2.2.

30.1 Measure preserving transformation

Let $(M, \mathcal{F}, \mu)$ be a probability space (but we will 'ignore $\mathcal{F}$.' See **29.3**). A map $T : M \to M$ preserves μ if $\mu = \mu \circ T^{-1}$. In this case μ is said to be T-invariant. For any μ-integrable function f

$$\int_M f(Tx)d\mu(x) = \int_M f(x)d\mu(x). \tag{30.1}$$

This can be shown easily if we note[343] for any $A \subset M$[344]

$$\int_M \chi_A(Tx)d\mu(x) = \int_M \chi_A(x)d\mu(x), \tag{30.2}$$

because

$$\int_M \chi_A(Tx)d\mu(x) = \int_M \chi_A(y)d\mu(T^{-1}y) = \int_M \chi_A(y)d\mu(y). \tag{30.3}$$

30.2 Poincare's recurrence theorem

Consider a measure-theoretical dynamical system (T, μ, M).

Theorem. Let $B \subset M$ with $\mu(B) > 0$. Then, μ-almost all $x \in B$ (i.e., μ-$\tilde{\forall} x \in B$) $T^k x \in B$ for infinitely many $k \in \mathbb{N}$.

[Demo] We show that the totality of the points $x \in B$ that do not return to B infinitely many times is μ-measure zero. Since x does not return to B infinitely many times, there must be $n \in \mathbb{N}$ such that $T^k x \in B^c$ (i.e., $x \in T^{-k}B^c$) for $\forall k > n$.

Let $E_n = \cap_{k \geq n} T^{-k} B^c$ (the totality of points that stays in B_c after time n; i.e., points that never return to B after time $n-1$). $T^{-1}E_n$ consists of points going into E_n by one time step, that is, points that do not return after time $n-1$ + one. Thus, $T^{-1}E_n = E_{n+1} \supset E_n$. $E = \cup_{n=0}^{\infty} E_n$ is the totality of the points never returns to B

[343] Integrals are all Lebesgue integrals, so f is considered as a limit of simple functions (= piecewise constant functions).

[344] $A \in \mathcal{F}$, precisely speaking, but as announced, I will not mention such a thing again.

or never in B (that is E_0). Suppose for some n $\mu(E_{n+1} \setminus E_n) > 0$. Then, this holds for all n, because μ is T-invariant:

$$\mu(E_{n+1} \setminus E_n) = \mu(T^{-1}[E_{n+1} \setminus E_n]) = \mu(T^{-1}E_{n+1} \setminus T^{-1}E_n) = \mu(E_{n+2} \setminus E_{N+1}). \quad (30.4)$$

Therefore,

$$\mu(E) = \mu[\cup_{n=1}^{\infty}(E_{n+1} \setminus E_n) + E_0] = \sum_{n=1}^{\infty} \mu(E_{n+1} \setminus E_n) + \mu(E_0) \nearrow \infty. \quad (30.5)$$

Thus, $\mu(E \setminus E_0) = 0$, but E_0 is the set of points never touching B, so $E_0 \cap B = \emptyset$. Therefore,

$$0 = \mu(B \cap (E \setminus E_0)) = \mu(E \cap B). \quad (30.6)$$

There is almost no points in B that cannot return to B infinitely often.

Remark: Note that $Q_n = E_{n+1} \setminus E_n$ is the totality of points in B that returns to B at time n and stays in B ever since. Thus, $T^{-1}Q_n = Q_{n+1}$. All have the same measure and disjoint and in $B^c = M \setminus B$. All Q_n must stay in B^c without overlap, so $\mu(Q_n)$ must be zero. As is clear from this what matters is that μ is normalizable. Whether M is bounded or not (as a metric space) is irrelevant. Thus, this theorem can be used to destroy Boltzmann's logic.

30.3 Zermelo's Wiederkehreinwand[345,346]

This theorem played an important role when Zermelo[347] pointed out Boltzmann's logical error. Boltzmann claimed that (after responding to the criticism Umkehreinwand[348] by Loschmidt just after the Boltzmann equation paper) his approximate dynamical system explains irreversibility: due to approximation irreversibility ensues. Zermelo (1895) pointed out still reversibility occurs indefinitely accurately due to this theorem (published in 1890); recurrence is not due to any property of the dynamics, but merely due to the finiteness of the total mass of the relevant invariant measure. He was confident because of the correct math and the moral support of his boss M. Planck.

[345] 'Wiederkehr' = recurrence, 'Einwand' = objection.

[346] H.-D. Ebbinghaus, *Ernst Zermelo, an approach to his life and work* (Springer, 2007) p15-26 1.4 Boltzmann Controversy.

[347] A road in Freiburg was named in his honor in 2017 (Wikipedia).

[348] 'Umkehr' = turning back.

30.4 Birkhoff's individual ergodic theorem

Theorem. Under the same condition as **30.5**
(i) The time average exists for μ-$\tilde{\forall} x \in M$

$$\overline{f}(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x). \tag{30.7}$$

(ii) This convergence is also in L_1.
(iii) For any invariant set B

$$\int_B \overline{f}(x) d\mu(x) = \int_B f(x) d\mu(x). \tag{30.8}$$

Remark: Notice that for this theorem to hold whether μ is ergodic or not does not matter. We need ergodicity to reduce (iii) to our familiar formula (see **30.8**):

$$\text{Time average of } f = \int_M f(x) d\mu(x). \tag{30.9}$$

We demonstrate these below **30.6**-**30.7**. The maximal ergodic theorem **30.5** drastically simplifies the original almost unreadable proof.

30.5 Maximal ergodic theorem

Let $T : M \to M$ preserves a measure μ. For μ-integral function f define the following functiions for $n \in \mathbb{N}$

$$S_n(f; x) = \sum_{k=0}^{n-1} f(T^k x),\ \ S_0(f; x) = 0. \tag{30.10}$$

Notice that

$$S_n(f; Tx) = S_{n+1}(f; x) - f(x). \tag{30.11}$$

Theorem

$$\int_{\{x \,|\, \sup_{n\geq 0} S_n(f;x)>0\}} f(x) d\mu(x) \geq 0. \tag{30.12}$$

[Demo][349]
We make a 'finite approximation' of $B = \{x \,|\, \sup_{n\geq 0} S_n(f; x) > 0\}$ as $B_n = \{x \,|\, M_n(x) > 0\}$:

$$M_n(x) = \max\{0, S_1(f; x), \cdots, S_n(f; x)\}. \tag{30.13}$$

[349] All the detailed are filled in, so you should be able to follow the proof without pencil.

Note $S_0(f;x)=0$. Since (30.11)

$$M_n(Tx)=\max\{0,S_2(f;x)-f(x),\cdots,S_{n+1}(f;x)-f(x)\}. \tag{30.14}$$

We also introduce

$$M_n^*(x)=\max\{S_1(f;x),\cdots,S_n(f;x)\}. \tag{30.15}$$

This means (notice that $S_1(f;x)=f(x)$ from (30.10))

$$M_{n+1}^*(x)=\max\{S_1(f;x),\cdots,S_{n+1}(f;x)\}=\max\{f(x),S_2(f;x),\cdots,S_{n+1}(f;x)\}. \tag{30.16}$$

Since $\max\{x_n+c\}=\max\{x_n\}+c$,

$$M_n(Tx)+f(x)=M_{n+1}^*(x). \tag{30.17}$$

That is, (note that $M_n^*(x)$ is an increasing sequence in n)

$$f(x)=M_{n+1}^*(x)-M_n(Tx)\geq M_n^*(x)-M_n(Tx). \tag{30.18}$$

We make an approximation of the integral (recall $B_n=\{x\,|\,M_n(x)>0\}$)

$$\int_{B_n}f(x)d\mu(x)\geq\int_{B_n}[M_n^*(x)-M_n(Tx)]d\mu(x). \tag{30.19}$$

If $x\in B_n$, then you can ignore 0 in the definition of $M_n(x)$, so $M_n^*(x)=M_n(x)$. Therefore,

$$\int_{B_n}M_n^*(x)d\mu(x)=\int_{B_n}M_n(x)d\mu(x). \tag{30.20}$$

Therefore, (30.19) reads

$$\int_{B_n}f(x)d\mu(x)\geq\int_{B_n}M_n(x)d\mu(x)-\int_{B_n}M_n(Tx)d\mu(x). \tag{30.21}$$

Since $M_n=0$ oputside B_n, we have

$$\int_{B_n}f(x)d\mu(x)\geq\int_{M}M_n(x)d\mu(x)-\int_{B_n}M_n(Tx)d\mu(x). \tag{30.22}$$

Since $M_n(Tx)\geq 0$,

$$\int_{B_n}M_n(Tx)d\mu(x)\leq\int_{M}M_n(Tx)d\mu(x). \tag{30.23}$$

Therefore, we have obtained

$$\int_{B_n} f(x)d\mu(x) \geq \int_M M_n(x)d\mu(x) - \int_M M_n(Tx)d\mu(x). \tag{30.24}$$

Now, take the $n \to \infty$ limit, and the right-hand side vanishes to reach (30.12).

30.6 Existence of time average

Notice that (i) in **??** is a generalization of the (weak) law of large numbers.

We wish to show that $S_n(f;x)/x$ converges almost μ-surely. If this does not converge lim sup and lim inf must be different. Therefore, there must be reals $a < b$ such that the following set has a positive measure:

$$B(a,b) = \left\{ x \,\middle|\, \liminf \frac{1}{n} S_n < a < b < \limsup \frac{1}{n} S_n \right\}. \tag{30.25}$$

Since $B(a,b)$ is an invariant set with a positive measure, let us confine ourselves to $B(a,b)$ and apply the maximal ergodic theorem to $a - f(x)$ and $f(x) - b$. For example, for $g(x) = a - f(x)$ the maximal ergodic theorem reads:

$$S_n(g;x) = na - \sum_{k=0}^{n-1} f(T^k x). \tag{30.26}$$

and

$$\int_{\{x \,|\, \sup_{n\geq 0} S_n(g;x)>0\}\cap B(a,b)} g(x)d\mu(x) = \int_{\{x \,|\, a>\sup_{n\geq 0} S_n(f;x)/n\}\cap B(a,b)} (a-f(x))d\mu(x) \geq 0. \tag{30.27}$$

However, $\{x \,|\, \sup_{n\geq 0} S_n(h;x) > 0\} \supset B(a,b)$, so this means

$$a\mu(B(a,b)) - \int_{B(a,b)} f(x)d\mu(x) \geq 0. \tag{30.28}$$

Analogously, for $h(x) = f(x) - b$

$$0 \leq \int_{B(a,b)} h(x)d\mu(x) = \int_{B(a,b)} f(x)d\mu(x) - b\mu(B(a,b)). \tag{30.29}$$

Hence,

$$a\mu(B(a,b)) \geq \int_{B(a,b)} f(x)d\mu(x) \geq b\mu(B(a,b)), \tag{30.30}$$

a contradiction.

30.7 L^1 convergence and (iii)

First we show (ii) to guarantee the commutativity of the integration and limit in (30.31).

If f is bounded as $|f(x)| \le K$, $|S_n/n| \le K$, so $|\overline{f}(x)| \le K$. Therefore, we may apply Lebesgue's bounded convergence theorem, implying the L^1-convergence of the time average. For general f we L^1-approximate f with a sequence of bounded functions to complete the proof. Thus (ii) holds.

If g is integrable on M, then (ii) allows

$$\int_M \overline{g}(x)d\mu(x) = \lim_{N\to\infty} \int_M \frac{1}{N}\sum_{K=0}^{N-1} g(T^k x)d\mu(x) = \int_M g(x)d\mu(x), \tag{30.31}$$

Now let $g(x) = \chi_B(x)f(x)$. Since B is invariant, $\overline{g}(x) = \chi_B(x)\overline{f}(x)$, which concludes the demonstration.

Notice that the above theorem holds for any invariant measure.

30.8 Time average and phase average

If μ is an ergodic invariant measure, then the time average agrees with the phase average (or the ensemble average):

$$\int_M f(x)d\mu = \lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} f(T^k x). \tag{30.32}$$

Notice that $\overline{f}(x)$ is time independent: $\overline{f}(Tx) = \overline{f}(x)$. Then, $\overline{f}(x)$ must be μ-almost surely constant: Let $A(c) = \{x \,|\, \overline{f}(x) = c\}$. It is an invariant set, so its measure is 0 or 1. (iii) in **??** tells us $c =$ the phase average value.

30.9 von Neumann's mean ergodic theorem

Before Birkhoff von Neumann proved the L^2 convergence

$$\lim_{N\to\infty} \int_M \left| \frac{1}{N}\sum_{k=0}^{N-1} f(T^k x) - \overline{f}(x)\right|^2 d\mu = 0. \tag{30.33}$$

No proof will be given, but if f is bounded, then L^1 and L^2 convergence on M is equivalent.

30.10 Weyl's equidistribution theorem

Historically, the first ergodic theorem is the following:

Theorem. For an integrable function f defined on a unit circle

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1} f(2\pi k\gamma\dot{+}\theta) = \int f(x)d\mu(x), \tag{30.34}$$

where $\dot{+}$ is the sum mod 2π, γ an irrational number and μ the uniform probability measure on the unit circle.

μ is an ergodic invariant measure for rotation. A more direct proof (e.g., using Fourier expansion) may be a good exercise.

30.11 Tragicomical history of ergodicity[350]

The word 'ergode' was used to specify the microcanonical ensemble. Maxwell in his "On Boltzmann's theorem on the average distribution of energy in a system of material points,"[351] he clarified the logic of Boltzmann's attempt (his second paper written at age 22) to derive equipartition of energy and the second law from mechanics, saying "The only assumption which is necessary for the direct proof is that the system, if left to itself in its actual state of motion, will, sooner or later, pass through every phase which is consistent with the equation of energy." Maxwell then considered an ensemble of systems having this property to develop a statistical theory. Actually Boltzmann adopted the idea of 'ensemble' and introduced his 'ergodic ensemble.'

Lord Kelvin on April 27, 1900 gave a talk on the clouds over the dynamical theory of heat and light[352] The second cloud he mentioned was the anomaly of the specific heat ratio $\gamma = C_P/C_V$. "Spectrum analysis showing vast numbers of lines for each gas makes it certain that the numbers of freedoms of the constituents of each molecule is enormously greater than those which we have been counting," γ should

[350] Based, in part, on I. Kubo's article on the history of ergodic theory (1982 September-December

[351] Cambridge Phil. Soc. Trans. 7 547 (1979). This is his last paper.

[352] "Nineteenth century clouds over the dynamical theory of heat and light," Phil. Mag. Series 6, 2: 7, 1-40 (1901). 'Cloud II' starts at Section 12. The outline I give here should be regarded as a hasty one by a very impatient theoretical physicist.

be very close to 1 in contrast to empirical results. Since Maxwell and Boltzmann premised “that the mean kinetic energies with which the Boltzmann-Maxwell doctrine (=equipartition of energy) is concerned are time integrals of energies divided by totals of the times,” their hypothesis (ergodic hypothesis) should be questioned. The, he goes on to the study of ergodicity of simple dynamical systems—billiards! He studies an elliptic billiard and says, “It seems not improbable that if the figure deviates by ever so little from being exactly ellipsoidal, Maxwell’s condition might be fulfilled.” “the meaning of the doctrine is that a single geodesic drawn long enough will not only fulfil Maxwell’s condition of passing infinitely near to every point of the surface in all directions.” “I have made many efforts to test it for the case in which the closed surface is reduced to a plane with other boundaries than an exact ellipse.’ He studied (not convex) polygons and ‘experimentally’ (by the assistant Mr Anderson with a straight rule) studied the equipartition: after 600 collisions $\langle K_x \rangle = \langle K_y \rangle$ was not fulfilled with an error of 7.5 %. He also studied a converging billiard as well (see Fig. 30.1; even he studied the ice cream cone!).

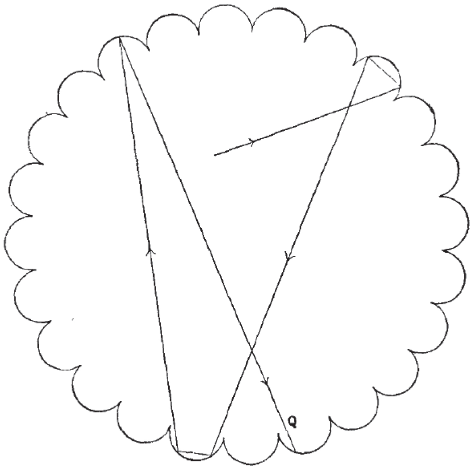


Figure 30.1: Kelvin’s ‘simulation’ for this failed to demonstrate ergodicity

Thus Lord Kelvin extremely seriously cast doubt on ergodicity.

A and T Ehrenfest clearly formulated ‘ergodic hypothesis’ in their encyclopedia article:[353] according to their definition, ‘ergodic’ means, as Maxwell said, a trajectory

[353] P. Ehrenfest & T. Ehrenfest (1911) Begriffliche Grundlagen der statistischen Auffassung in der Mechanik, in: Enzyklopdie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen. Band IV, 2. Teil (F. Klein and C. Mller (eds.). Leipzig: Teubner, pp. 390. Translated as *The conceptual Foundations of the Statistical Approach in Mechanics*, New York: Cornell University Press, 1959. ISBN 0-486-49504-3.

passes every point energetically allowed and 'quasiergodic' means, as formulated by Lord Kelvin, a trajectory passes any neighborhood of any point energetically allowed. Then, 'ergodic hypothesis' means that the time average along an ergodic trajectory is identical to the microcanonical average.

This hypothesis was immediately shot down by Plancherel and by Rosenthal (independently in 1913): there cannot be such an orbit; simply topologically impossible. In those days, physicists realized that Gibbs' statistical mechanics is practically usable, so the interest in ergodicity faded, and the topic came to be considered as a pure mathematical question.

Then, in 1914 Weyl proved his theorem **30.10**, showing that quasiergodicity may be enough. His theorem is about T^n. von Neumann extended this to a general domain[354]:[355]

Theorem [Mean ergodic theorem] If f is $L^2(X)$, then a function $\overline{f}(x)$ exists such that

$$\lim_{t\to\infty}\int_X \left|\frac{1}{t}\int_0^t f(T_t x) - \overline{f}(x)\right|^2 d\mu(x) = 0. \tag{30.35}$$

This means

Corollary. μ-almost every x there is an increasing time sequence $\{\tau_n\}$ such that

$$\overline{f}(x) = \lim_{n\to\infty}\frac{1}{\tau_n}\int_0^{\tau_n} f(T_t x)dt. \tag{30.36}$$

This is basically what physicists 'needed.'

On Oct 22, 1931, Birkhoff received von Neumann's letter on the proof of the mean ergodic theorem, and started his study vehemently.[356] He declared that the quasiergodic hypothesis was now replaced by its modern version: measure-theoretical transivity.

Has Birkhoff unraveled or resolved the secret of statistical mechanics? Simply, the question becomes whether the Liouville measure (i.e., the Lebesgue measure on the phase space) is ergodic or not. For almost all systems of interest, this has never been proved.

We now clearly recognize that ergodicity cannot found statistical mechanics; ergodicity has been a big red herring.

[354] 1932

[355] This possibility was suggested to him by Koopman (in 1930) and by A Weil (in 1931).

[356] You must learn a great lesson from this episode; you should not tell even what you are studying.

31 Lecture 31 Information loss rate

31.1 Observation with finite resolution

When a dynamical system is given, we wish to describe its state at each instant with a constant precision. Suppose the data with information[357] H is required to describe the current state with a prescribed precision. If we could predict the state with the same precision after one unit time with the aid of this initial data, all the states in any future can be predicted with the same precision in terms of the information H contained in the initial data. However, for chaotic systems the cloud of the ensemble is scrambled and the precise locations of the individual points in the phase space are lost. If we wish to locate the state of the system after one time unit as precisely as we can locate it now, we have to augment H with more information to counter the 'information-dissipating' tendency of chaotic dynamical systems. Thus:

(1) We need extra information ΔH to describe time one future as accurately as the present. That is, we need more information than H required at present to maintain the precision of the description.

(2) To describe a state at any given time in the steady state we need the same amount H to have 'equi-precision' description. Therefore, ΔH means the loss of information due to time evolution.

Let us try to quantify the information loss per time.

31.2 Interval map information loss rate

Consider a discrete measure-theoretical dynamical system (F, μ, I) on the interval I with a piecewise continuous and differentiable map[358] F with an absolutely continuous (**29.9**) invariant measure μ. The average amount $h_\mu(F)$ of the extra information required for the equi-precise description of the states is given by the following formula:

$$h_\mu(F) = \int_I \mu(dx)\, \log|F'(x)|. \tag{31.1}$$

Henceforth, we will often write $\mu(dx) = d\mu(x)$.[359] The generalized form that is also

[357] What is information? We only ask how we quantify it. An explanation is given in terms of 'surprisal' in **??** and **31.9**.

[358] 'piecewise continuous' implies that a map consists of a several continuous pieces. 'Piecewise continuously differentiable' implies that each piece is differentiable inside, and, at its ends, one-sided derivatives from inside are well-defined.

[359] Here, a formal calculation is explained as given in Y. Oono, "Kolmogorov-Sinai entropy as disorder parameter for chaos," Prog. Theor. Phys. **60**, 1944 (1978). The formula had been obtained

correct for non-absolutely continuous invariant measures is (31.3).[360]

31.3 Derivation of Rokhlin's formula

For simplicity, let us assume that F is unimodal (its graph has only one peak and no valley; Fig. 31.1).

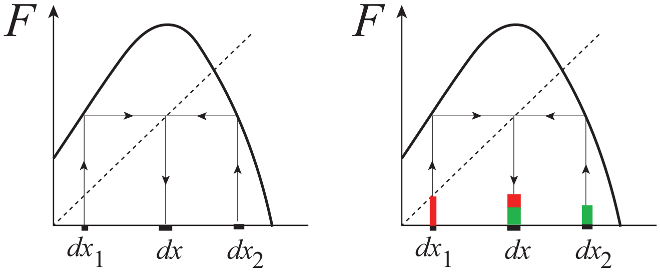


Figure 31.1: An explanation of (31.2). The points in a small interval dx was in dx_1 or in dx_2 one time step ago. These intervals are stretched and superposed onto dx. The right figure illustrates (31.2).

The invariance condition (29.2) for the measure μ reads, in this case (see Fig. 31.1),

$$\mu(dx_1) + \mu(dx_2) = \mu(dx). \tag{31.2}$$

Here, dx denotes a small interval and its preimage (the set mapped to dx by F) is written as $dx_1 \cup dx_2$ ($F^{-1}(dx) = dx_1 \cup dx_2$). For example, if F is applied once to dx_1, it is stretched and mapped onto dx, so without any doubt the precision to specify the location is reduced. That is, with the equal probability $\mu(dx_1)/\mu(dx) = \mu(dx_1)/\mu(F(dx_1))$ they are spread over dx. Consequently, the information was lost by $-\log(\mu(dx_1)/\mu(dx))$. We must also do an analogous calculation for dx_2.

Therefore, the expectation value of information deficit required to predict one time step later with the same precision as the current state must be obtained by

by Rohlin, "Lectures on the entropy theory of transformations with an invariant measure," Russ. Math. Surveys **22**(5), 1 (1967), so let us call it Rohlin's formula. Note that most important theorems of measure theoretical dynamical systems had been obtained by the Russian mathematicians long before chaos became popular among the US physicists in the 1980s.

[360] F. Ledrappier, "Some properties of absolutely continuous invariant measures on an interval," Ergodic Theor. Dynam. Syst. **1**, 77 (1981) proved the following theorem for C^2-maps (actually, for C^1-maps with some conditions):

Theorem Let f be a C^2-endomorphism on an interval. A necessary and sufficient condition for an ergodic invariant measure to be absolutely continuous is that the Kolmogorov-Sinai entropy is given by Rohlin's formula (31.4).

averaging this over x_1 with the aid of the (probability) measure μ. Consequently, the rate of information deficit is written as

$$h_\mu(F) = -\int_I \mu(dx) \log \frac{\mu(dx)}{\mu(F(dx))}. \tag{31.3}$$

This loss must be compensated with the extra information (ΔH in the above) for equi-precise description of the system.

If μ is absolutely continuous, we can introduce the invariant density g as $g(x)\lambda(dx) = \mu(dx)$, where λ is the Lebesgue measure. Hence,

$$\frac{\mu(dx_i)}{\mu(dx)} = \frac{\mu(dx_i)}{\lambda(dx_i)} \frac{\lambda(dx)}{\mu(dx)} \frac{\lambda(dx_i)}{\lambda(dx)} = \frac{g(x_i)}{g(x)|F'(x_i)|}. \tag{31.4}$$

Noting that $F(x_i) = x$ with the aid of (31.4), we may rewrite (31.3) as

$$h_\mu(F) = \int_I \lambda(dx) g(x) \log \frac{g(F(x))|F'(x)|}{g(x)}. \tag{31.5}$$

Here, the integral with respect to the Lebesgue measure is explicitly written as $\int \lambda(dx)$, but it is simply $\int dx$ with the usual notation.

If we apply the Perron-Frobenius equation **31.4**, we see

$$\int_I dy\, g(y) \log g(F(y)) = \int_I dy \int_I dz\, g(y)\delta(z - F(y)) \log g(z) = \int_I dz\, g(z) \log g(z), \tag{31.6}$$

so indeed (31.3) is nothing but (31.1) under the absolute continuity assumption.

31.4 Perron-Frobenius equation

If μ is absolutely continuous, we may introduce the density $g = d\mu/dx \equiv \mu(dx)/\lambda(dx)$. Then, the invariance condition (31.2) for measure μ reads

$$g(x_1)dx_1 + g(x_2)dx_2 = g(x)dx. \tag{31.7}$$

Here, $F^{-1}(x) = \{x_1, x_2\}$ and $F(dx_i) = dx$ $(i = 1, 2)$ (see Fig. 31.1). This means $|F'(x_i)|dx_i = dx$. Therefore, (31.7) reads

$$g(x_1)\frac{dx}{|F'(x_1)|} + g(x_2)\frac{dx}{|F'(x_2)|} = g(x)dx, \tag{31.8}$$

or

$$\frac{g(x_1)}{|F'(x_1)|} + \frac{g(x_2)}{|F'(x_2)|} = g(x). \tag{31.9}$$

With the aid of the elementary property of the δ-function,[361] regarding y as the independent variable, we have for F in **31.3**

$$\delta(x - F(y)) = \frac{1}{|F'(x_1)|}\delta(y - x_1) + \frac{1}{|F'(x_2)|}\delta(y - x_2), \tag{31.10}$$

where $F(x_1) = F(x_2) = x$. Thus can be rewritten as

$$g(x) = \int_I dy\, g(y)\delta(x - F(y)), \tag{31.11}$$

which is called the Perron-Frobenius equation. It is a general formula for the invariant density if a 1D-endomorphism (piecewise differentiable).

31.5 Can we solve Perron-Frobenius equation?[362]

Generally, no, but if we assume that the system (defined on [0,1]) allows an absolutely continuous invariant measure, then its density g may be amenable to analytic approach. We start from the Perron-Frobenius equation (31.11). Using Heaviside's step function[363]

$$\Theta(x) = \begin{cases} 0 & x \le 0, \\ 1 & x > 0. \end{cases}, \tag{31.12}$$

the delta function may be written as

$$\frac{d}{dx}\Theta(x - y) = \delta(x - y). \tag{31.13}$$

[361] Let a be an isolated (real) zero point (i.e., $f(a) = 0$), and f be differentiable around it. In a sufficiently small neighborhood of a with the aid of the variable change: $y = f(x)$ we get for sufficiently small positive ϵ

$$\int_{a-\epsilon}^{a+\epsilon} \delta(f(x))\varphi(x)dx = \frac{1}{|f'(a)|}\varphi(a).$$

Here, φ is a sufficiently smooth test function. Therefore, if f has a several isolated real zeros a_i, we can add each contribution to get

$$\int_{-\infty}^{\infty} \delta(f(x))\varphi(x)dx = \sum_i \frac{1}{|f'(a_i)|}\varphi(a_i).$$

[362] YO talk in Aug 1979 at Res Inst Math Sci, Kyoto

[363] Its definition at $x = 0$ is subtle, but since we do not care for measure zero sets, we may choose it for our convenience.

Introducing this into the Perron-Frobenius equation, we have

$$g(x) = -\int_0^1 dy\, g(y) \frac{1}{F'(y)} \frac{d}{dy} \Theta(x - F(y)). \tag{31.14}$$

Performing an integration by parts, this yields

$$g(x) = \frac{g(0)}{F'(0)}\Theta(x-F(0)) - \frac{g(1)}{F'(1)}\Theta(x-F(1)) + \int_0^1 dy\,\Theta(x-F(y)) \frac{d}{dy}\frac{g(y)}{F'(y)}. \tag{31.15}$$

For a map with $F(1) = 0$, we know $g(0) = -g(1)/F'(1)$, this reads

$$g(x) = g(0) + \frac{g(0)}{F'(0)}\Theta(x - F(0)) + \int_0^1 dy\,\Theta(x - F(y)) \frac{d}{dy}\frac{g(y)}{F'(y)}. \tag{31.16}$$

Iterative substitution can solve the above equation for unimodal maps,[364] β-transformations, etc.

$$g(x) = g(0)\left[1 + \sum_{k=1}^{\infty} \frac{\Theta(x - F^k(0))}{[F^k(0)]'}\right]. \tag{31.17}$$

Here the derivative of $F^k(0)$ at the breaking point of F is computed as the left derivative: $[f(x) - f(x - \varepsilon)]/\varepsilon$ ($\varepsilon > 0$).[365]

31.6 KS entropy for billiards

17.18 explains the general strategy to compute the KS entropy (= information loss rate) and the idea is applied to the Sinai billiard in **17.19**. For these expansive dynamical systems, we have only to pay attention to the stretching rate of the unstable manifolds. We will see this more systematically, when we study Axiom A systems later.

31.7 KS entropy of Markov chain

For an ergodic Markov chain[366] the extra information required for equi-precision may

[364] Ito et al. ibid.

[365] to be consistent with our choice of Θ.

[366] See, for example, Durrett, *ibid.*; Z. Brzeźniak and T. Zastawniak, *Basic Stochastic Processes, a course through exercises* (Springer, 1998) is a very kind measure-theoretical introduction to stochastic processes.

be computed by the same idea as explained above. If its transition matrix is given by $\Pi \equiv \{p_{i\to j}\}$, then

$$h(\Pi) = -\sum_{i,j} p_i p_{i\to j} \log p_{i\to j}, \tag{31.18}$$

where p_i is the invariant measure (stationary state) (i.e., $\sum_i p_i p_{i\to j} = p_j$). In particular, for a Bernoulli process $B(p_1, \cdots, p_n)$, we have

$$h(B(p_1, \cdots, p_n)) = -\sum_{i=1}^{n} p_i \log p_i. \tag{31.19}$$

It is easy to understand (31.18). Suppose we are in state i now. After one time step what do we know? With probability $p_{i\to j}$ we will be in state j. Thus, we lose on the average the following amount of information

$$-\sum_j p_{i\to j} \log p_{i\to j}. \tag{31.20}$$

We should average this over the probability of our being in state i.

31.8 Sanov's theorem, a large deviation of observable probabilities

Consider an uncorrelated (= statistically independent) symbol sequence consisting of n symbols. The probability to fine symbol k is given by $q_k > 0$: $\sum q_k = 1$. Suppose we know these probabilities. Observing N symbols, we can obtain the empirical distribution of symbols as

$$\pi_k = \frac{1}{N}\sum_{t=0}^{N-1} \delta_{x_t,k}, \tag{31.21}$$

where x_t is the t-th symbol we actually observe.

The law of large numbers for the symbol distribution tells us for any $\varepsilon > 0$

$$\lim_{N\to\infty} P\left(\left\| \frac{1}{N}\sum_{t=0}^{N-1} \delta_{x_t,k} - q_k \right\| > \varepsilon \right) = 0. \tag{31.22}$$

If N is finite, we inevitably observe fluctuations of $\{\pi_k\}$ around $\{q_k\}$. Using the multinomial distribution, we can estimate the probability to observe $\{\pi_k\}$.

If event i occurs n_i times ($\sum_{i=1}^m n_i = N$), the empirical probability is $p_i = n_i/N$.

Therefore, the probability to get this empirical distribution from N sampling is given by the following multinomial distribution:

$$P(\pi \simeq p) = \prod_{i=1}^{m} \frac{N!}{(Np_i)!} q_i^{Np_i}. \tag{31.23}$$

Taking its log and using Stirling's approximation, we obtain

$$\log P(\pi \simeq p) = \log N! + \sum_{i=1}^{m} \{Np_i \log q_i - Np_i \log(Np_i) + Np_i\}. \tag{31.24}$$

That is, we get a large deviation principle for the probability distribution.

$$P(\pi \simeq p) \approx e^{-N \sum_i p_i \log(p_i/q_i)}. \tag{31.25}$$

This relation is called *Sanov's theorem*.[367]

If we introduce the following quantity called the *Kullback-Leibler entropy*

$$K(p\|q) = \sum p_i \log \frac{p_i}{q_i}, \tag{31.26}$$

Sanov's theorem reads

$$P(\pi \simeq p) \approx e^{-NK(p\|q)}. \tag{31.27}$$

As to the nonnegative definite nature of the Kullback-Leibler entropy K, see **32.7**.

31.9 Information via Sanov's theorem

Since we know the true answer for the symbol distribution, we would not be surprised if $\pi \simeq q$. How should we quantify our surprise?

Suppose we actually observed an event whose probability is $p < 1$. What is the most rational measure of surprise. It is $-\log p$ (or $-\log_2 p$ bits). This is called the surprisal.

> The 'extent of surprise' $f(p)$ we get, spotting a symbol that occurs with probability p or knowing that an event actually happens whose expected probability is p, should be

[367] Sanov, I. N. (1957). On the probability of large deviations of random variables, *Mat. Sbornik*, **42**, 11-44. Ivan Nikolaevich Sanov (1919-1968); obituary in (1969). *Russ. Math. Surveys*, **24**, 159.

(1) A monotone decreasing function of p (smaller p should give us bigger surprise).
(2) Nonnegative.
(3) Additive: $f(pq) = f(p) + f(q)$.
Therefore, $f(p) = -c \log p$ ($c > 0$) is the only choice.[368]

Thus the average surprise per symbol we get is $K(p\|q)$. If we know that the true distribution is flat (even), then the surprisal may be measured by the Shannon formula:

$$H(p) = -\sum_k p_k \log p_k. \tag{31.28}$$

[368] We could invoke the Weber-Fechner law in psychology.

32 Lecture 32. Kolmogorov-Sinai entropy

32.1 Kolmogorov-Sinai entropy as information loss rate

The identity of the information deficiency rate = the extra information required for the equi-precise description and the Kolmogorov-Sinai entropy is generally expected, so the definition of the Kolmogorov-Sinai entropy for a measure-theoretical dynamical system (T, μ, M) is given. This section gives an elementary introduction to the concept and some basic theorems facilitating its intuitive understanding. Some preparation is needed.

32.2 Partition

A (finite) partition $\mathcal{A}$ of a set Γ (Fig. 32.1) is a family of subsets $\{A_1, \cdots, A_n\}$ of Γ satisfying the conditions:

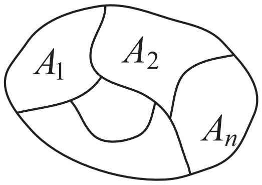


Figure 32.1: Partition $\mathcal{A} = \{A_1, \cdots, A_n\}$. A finite partition $\mathcal{A}$ of a set Γ is a family of finitely many subsets of Γ such that its members A_i do not have any overlap with each other and the total sum perfectly covers Γ.

(1) For any i and j $(\neq i)$ $A_i \cap A_j = \emptyset$, and
(2) $\cup_{i=1}^n A_i = \Gamma$

$\{A_i\}$ may be interpreted as the totality of the mutually exclusive observation results.[369] Our observation is always under finite precision. Therefore, if the phase space is continuous, we can never specify a particular point in it by observation. Therefore, it is reasonable to introduce such discrete (coarse-grained) observables.[370]

[369] Here, each set A_i may be a collection of discrete pieces or with holes (i.e., it need not be (singly) connected).

[370] We could not tell which A_i the phase point is actually in, so mustn't A_i be fuzzy? Here, we adopt an interpretation that the relation between the values of a macro-observable that can be observed with a finite precision and the actual microstates of the dynamical system is given by the system itself independent of our capability of observing each microstate. That is, an element of a partition $\mathcal{A}$ consists of a definite set of microstates as an intrinsic property of the system, although we cannot determine this with our finite precision observations.

32.3 Composition of partitions

The composition operation $\vee$ of two partitions of Γ, $\mathcal{A} \equiv \{A_1, \cdots, A_n\}$ and $\mathcal{B} \equiv \{B_1, \cdots, B_m\}$, is defined as follows (Fig. 32.2):

$$\mathcal{A} \vee \mathcal{B} \equiv \{A_1 \cap B_1, A_1 \cap B_2, \cdots, A_n \cap B_m\}. \tag{32.1}$$

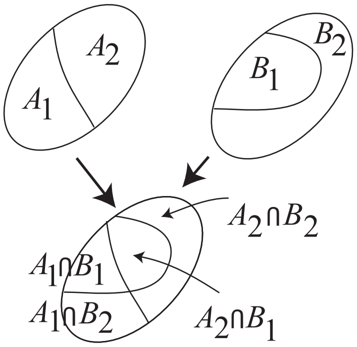


Figure 32.2: Composition of partitions: $\{A_1, A_2\} \vee \{B_1, B_2\} = \{A_1 \cap B_1, A_1 \cap B_2, A_2 \cap B_1, A_2 \cap B_2\}$

On the right-hand side appear all the nonempty pairs of A_i and B_j. By definition $\mathcal{A} \vee \mathcal{B} = \mathcal{B} \vee \mathcal{A}$. The elements of $\mathcal{A} \vee \mathcal{B}$ may be interpreted as the mutually exclusive outcomes obtained by performing two macroscopic observations corresponding to $\mathcal{A}$ and $\mathcal{B}$ simultaneously.

32.4 Information obtainable from observable $\mathcal{A}$

Let us return to the general measure-theoretical dynamical system (T, μ, Γ). The average information we can obtain about the system by a single observation of macroscopic observable $\mathcal{A}$ may be written with the aid of Shannon's formula (**31.9**) as[371]

$$H(\mathcal{A}) = -\sum_{A \in \mathcal{A}} \mu(A) \log \mu(A). \tag{32.2}$$

This information implies that unless we have on the average this amount of information about the system, we cannot infer in which A_i the observation result is in.

[371] The partition must be a measurable partition; $\mu(A_i)$ must be meaningful. Such a statement will not be written explicitly in the following.

32.5 Extra information needed for equi-precise description of observable $\mathcal{A}$

When a system evolves according to the dynamical law T, how much information do we need to infer its coarse-grained state (one of $\mathcal{A}$) at the next time (time $t+1$) under the condition that we know the coarse-grained state at present (time t)? To infer both the states at time t and time $t+1$ without any prior knowledge, we need information of $H(\mathcal{A} \vee T^{-1}\mathcal{A})$, where $T^{-1}\mathcal{A} = \{T^{-1}A_1, \cdots, T^{-1}A_n\}$. We can understand this as follows. An element of $T^{-1}\mathcal{A}$ coincides with some element of $\mathcal{A}$ after one time step. That is, to describe the state of the system after one time step in terms of the observable $\mathcal{A}$ is equivalent to knowing in which element of $T^{-1}\mathcal{A}$ the current state is. Therefore, to specify both the states at $t+1$ and at t is to specify a single element in $\mathcal{A} \vee T^{-1}\mathcal{A}$. Consequently, on the average without $H(\mathcal{A} \vee T^{-1}\mathcal{A})$ of information, we cannot infer in which A_i at time t and in which A_j at $t+1$ the system is in. If we already know the macrostate at time t, to predict the macrostate at $t+1$ we need $H(\mathcal{A} \vee T^{-1}\mathcal{A}) - H(\mathcal{A})$ extra information. This is the amount of extra information ΔH (appearing in **31.1**) required for equi-precision prediction of the one time step future.

32.6 Steady-state information loss

We need not stick to a particular time t and $t+1$ in the steady state. That is, to describe the extent of chaos for a measure-theoretical dynamical system we should consider the average deficiency for a long time:

$$\frac{1}{n}[H(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-n}\mathcal{A}) - H(\mathcal{A})] \to h_\mu(T, \mathcal{A}), \tag{32.3}$$

where the existence of this limit in $n \to \infty$ is guaranteed by the (intuitively plausible) subadditivity of H **32.7** and $H(T^{-n}\mathcal{A}) = H(\mathcal{A})$ ($n = 1, 2, \cdots$; this can be seen from the invariance of the measure (29.2)).

32.7 Subadditivity of entropy

The following inequality shows the subadditivity of information:

$$H(\mathcal{A} \vee \mathcal{B}) \le H(\mathcal{A}) + H(\mathcal{B}). \tag{32.4}$$

The inequality must be intuitively natural, because in order to describe two observables $\mathcal{A}$ and $\mathcal{B}$ it is better to use information about the relation between these two as well than to use information separately from each of the observables. Algebraically, we proceed as follows:

$$\begin{aligned} H(\mathcal{A} \vee \mathcal{B}) &= -\sum_{i,j} \mu(A_i \cap B_j) \log \mu(A_i \cap B_j), \qquad (32.5)\\ &= -\sum_{i,j} \mu(A_i \cap B_j) \left\{ \log \left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)\mu(B_j)} \right) + \log \mu(A_i)\mu(B_j) \right\}, \qquad (32.6)\\ &= -\sum_{i,j} \mu(A_i \cap B_j) \log \left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)\mu(B_j)} \right) + H(\mathcal{A}) + H(\mathcal{B}). \qquad (32.7) \end{aligned}$$

If the first term on the right-hand side of (32.7) is non-positive, the proof of $H(\mathcal{A} \vee \mathcal{B}) \leq H(\mathcal{A}) + H(\mathcal{B})$ is over. This is the following important inequality: Let p and q be probabilities ($p_i \geq 0$ and $\sum_i p_i = 1$, etc., hold)

$$\sum_i p_i \log \frac{p_i}{q_i} \geq 0. \qquad (32.8)$$

To show this we use the inequality $x \log x \geq x - 1$ for $x \geq 0$.[372] Introduce $x = p_i/q_i$ into this and sum over i after multiplying q_i:

$$\sum_i q_i \left(\frac{p_i}{q_i} \log \frac{p_i}{q_i} \right) \geq \sum_i p_i - \sum_i q_i = 0. \qquad (32.9)$$

32.8 Fekete's lemma

If $\{f(n)\}$ is subadditive (i.e., $f(n+m) \leq f(m) + f(n)$ holds for any positive integers, n, m), $\lim_{n\to\infty} f(n)/n = \inf_m f(m)/m$.[373]

[Demo] Obviously, $\liminf f(n)/n \geq \inf f(m)/m$. Writing $n = s + km$ ($m > 0$, $s \geq 0$ are integers), we get

$$\frac{f(n)}{n} = \frac{f(s+km)}{s+km} \leq \frac{f(s) + kf(m)}{s+km} \to \frac{f(m)}{m}. \qquad (32.10)$$

[372] The minimum value of $f(x) = x \log x - x + 1$ for $x \geq 0$ is zero as can be seen from the graph.

[373] ⟪**Infimum limit (lim inf), supremum limit (lim sup)**⟫ $\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_n, x_{n+1}, \cdots\}$. That is, we make the lower bound y_n of the sequence beyond x_n and then take its limit $n \to \infty$. Since $\{y_n\}$ is monotone increasing, the limit is well-defined (may not be bounded); similarly, $\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_n, x_{n+1}, \cdots\}$.

Therefore, $\limsup f(n)/n \le \inf f(m)/m$. Hence, the infimum and supremum limits agree, and the limit exists.

32.9 The best observable: definition of Kolmogorov-Sinai entropy

We should look for the 'best' observable to observe the system. Here, 'the best observable' should imply the observable that allows us to observe the system maximally in detail. Such an observable must be sensitive to the time evolution of the system, so the increasing rate of the information deficiency for this observable must be the largest. With this idea the Kolmogorov-Sinai entropy (or measure-theoretical entropy) is defined as follows:

$$h_\mu(T) \equiv \sup_{\mathcal{A}} h_\mu(T, \mathcal{A}), \tag{32.11}$$

where the supremum is taken over all the finite partitions of Γ (roughly speaking, we try all the finite-resolving power observations).

32.10 Generator

If all the future data of a certain macro-observable determines the future history (trajectory) uniquely, we do not need any more detailed observations. Take an arbitrary history ω. Let us write the element of $\mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-n+1}\mathcal{A}$ containing ω as $A^n(\omega)$ (which is an example of a cylinder set; see **26.5**).[374] Since $\omega \in A^n(\omega)$ for any $n = 0, 1, 2, \cdots$, $\omega \in \cap_{n=0}^{\infty} A^n(\omega)$. If this common set does not contain any history other than ω, in other words, if $\cap_{n=0}^{\infty} A^n(\omega) = \{\omega\}$, any further detailed observation is superfluous. If such a relation holds for μ-almost all ω (i.e., except for μ-measure zero set), the partition $\mathcal{A}$ is called a generator. If $\mathcal{A}$ is a generator, as expected,[375]

$$h_\mu(T) = h_\mu(T, \mathcal{A}). \tag{32.12}$$

For example, for $Tx = \{2x\}$ in Appendix 2.1A $\{[0, 1/2], (1/2, 1]\}$ is a generator.

If a measurable partition $\mathcal{A}$ separates all points (precisely speaking, for μ-almost

[374] An element of $\mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-n+1}\mathcal{A}$ has the form: $A_i \cap T^{-1}A_j \cap \cdots \cap T^{-n+1}A_k$, which is the totality $\{x : x \in A_i, Tx \in A_j, \cdots, T^{n-1}x \in A_k\}$. Therefore, around here in the text, $x \in \Gamma$ and a history $\omega = \{x, Tx, \cdots\}$ are identified.

[375] P. Walters, *An Introduction to Ergodic Theory* (Springer, 1982) is an excellent textbook of the Kolmogorov-Sinai entropy (but students outside mathematics may not be able to read it with ease). P. Billingsley, *Ergodic Theory and Information* (Wiley, 1960) is also excellent, but is slightly dated.

all points), in other words, for any two points $x \neq y \in \Gamma$, there is a positive integer n and an element $A \in \mathcal{A}$ such that $T^n(x) \in A$ and $T^n(y) \notin A$, $\mathcal{A}$ is a generator. Actually, if $\mathcal{A}$ were not a generator, there are two histories $T^n(x)$ and $T^n(y)$ that are always contained in a single element of $\mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-n+1}\mathcal{A}$ for any $n = 0, 1, 2, \cdots$. This contradicts the requirement that each point is separated.

32.11 Chaos and information loss/gain rate

A chaotic dynamical system is a dynamical system for which the extra information required for equi-precise description increases linearly in time.[376]

For chaos, if we wish to predict its state after N time steps with sufficient precision, we need a tremendous amount of information at present (we need $\sim e^{Nh}$ times as much precision as required to describe the present state, so N times as much information as required to know the state at present; see just below), so there is no wonder that sooner or later the behavior of the system becomes unpredictable.

32.12 Krieger's theorem on generator[377]

Let (T, μ, M) be an ergodic dynamical system. If its Kolmogorov-Sinai entropy satisfies $h_\mu(T) < \log k$ for some integer $k > 1$, there is a generator with k elements.

The theorem roughly tells us that if the Kolmogorov-Sinai entropy of the system is $\log k$, then we can encode its dynamics using k symbols without losing any information.

[376] In contrast to the above explanation, there were people who wished to introduce the Kolmogorov-Sinai entropy as the rate of generation of information by the dynamical system. Chaotic dynamical systems often exhibit us details of initial conditions later because of exponential separation of nearby trajectories. Consequently, (if noise is completely ignored) continuous observation of trajectories would tell us increasingly detailed information about the initial condition (in retrospect). In this sense, the dynamical system looks as if it is generating information. This observation itself is correct, but whether this generating rate can be measured by the Kolmogorov-Sinai entropy is another question. For chaos that may be observed numerically (observable chaos), this is correct, but, for example, for unimodal endomorphisms of intervals, except for at most one invariant measure, the assertion does not hold for any of (uncountably many) invariant measures; generally speaking, the generating rate of information just considered is larger than the Kolmogorov-Sinai entropy. This fact is captured by Ruelle's inequality (33.10) we will encounter later. Therefore, it is quite dangerous to interpret the Kolmogorov-Sinai entropy as the information generating rate.

[377] W. Krieger, "On entropy and generators of measure-preserving transformations," Trans. Amer. Math. Soc. **149**, 453 (1970); corrections *ibid.* **168**, 519 (1972).

32.13 Shannon-McMillan-Breiman's theorem[378]

Let (T, μ, Γ) be an ergodic dynamical system and $\mathcal{A}$ a finite partition of Γ. Let $A^n(x)$ be an element of $\mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-n+1}\mathcal{A}$ (a cylinder set of length n) containing $x \in \Gamma$. For μ-almost all x

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log\mu(A^n(x))\right] = h_\mu(T, \mathcal{A}). \tag{32.13}$$

In the above formula $A^n(x)$ is, as before, the bundle of histories (i.e., cylinder set) that cannot be distinguished from the history with the initial condition x for n time steps with a coarse-grained observation corresponding to the partition $\mathcal{A}$. $\mu(A^n(x))$ is the volume of the initial conditions for the history satisfying the condition (we could interpret it as the volume of the cylinder set). If the observation time span n is increased, the condition becomes increasingly stringent, so the volume decreases exponentially. (32.13) measures how fast the volume of the cylinder set decreases. The more chaotic the system is, the harder trajectories to be close to a particular one starting from x, so this value must become larger.

It may be more intuitive to rewrite (32.13) as

$$\mu(A^n(x)) \sim e^{-nh_\mu(T,\mathcal{A})}. \tag{32.14}$$

32.14 Asymptotic equipartition theorem

A special case of the Shannon-McMillan-Breiman theorem is the asymptotic equipartition (AEP) theorem of information theory:[379]

Let $\{X_i\}$ be a sequence of independently and identically distributed stochastic variables. Let us write the probability for a (consecutive) n samples, $X_1, X_2, \cdots, X_n$, as $p(X_1, \cdots, X_n)$. Then,

$$-\frac{1}{n}\log p(X_1, \cdots, X_n) \to H(X), \tag{32.15}$$

where $H(X)$ is the entropy of the individual stochastic variables. Notice that this is nothing but the weak law of large numbers (footnote 28 in Section 1.2) for the

[378] For a proof of the Shannon-McMillan-Breiman theorem, see, for example, W. Parry, *Entropy and Generators in Ergodic Theory* (Benjamin, New York 1969).

[379] It is quite important in information theory. See T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, 1991) p50-

logarithm of probability. $X_1, X_2, \cdots, X_n, \cdots$ may be interpreted as a history, so the Shannon-McMillan-Breiman theorem is the extension of the asymptotic equipartition theorem to histories with correlated events.

32.15 Brin-Katok's theorem

The Brin-Katok theorem explicitly counts the number of histories in the ε-neighborhood of the trajectory starting from x.[380]

Let the totality of the initial conditions that stay in the ε-neighborhood of the trajectory starting from $x \in \Gamma$ for N time steps be (the theorem holds for continuous dynamical systems as well)

$$B_N(x, \varepsilon) = \{y \in M : d(T^n x, T^n y) \le \varepsilon, 0 \le n \le N\}. \tag{32.16}$$

Theorem [Brin-Katok] Let (T, μ, Γ) be an ergodic dynamical system. For μ-almost all $x \in \Gamma$,

$$h_\mu(T) = -\lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{1}{N} \log \mu(B_N(x, \varepsilon)). \tag{32.17}$$

32.16 (some) Chaos can be predictable

Chaos may be predictable for a certain time span thanks to its deterministic nature, in contradistinction to noise, but not for a long time. We can estimate during how many steps n we can predict dynamics with the aid of the Kolmogorov-Sinai entropy h. According to the Brin-Katok theorem $n \sim (\log \delta x)/h$, where δx is our resolving power of the initial condition. However, there are chaotic dynamical systems with very small h, for which accurate prediction of considerable future is possible.[381]

32.17 Chaos under noise

As was stated around the Shannon-McMillan-Breiman theorem above, if a system is chaotic, the bundle of histories close to a given history thins quickly, so it is often the case that the external noise makes a chaotic system 'more chaotic.' This is

[380] M. Brin and A. Katok, "On local entropy" Lecture Notes Math. **1007**, 30 (1983).

[381] For asteroid 522 Helga the Lyapunov characteristic time (the time needed to magnify the initial error by e) is 6,900 years, so accurate prediction is possible. See A. Milani and A. M. Nobili, "An example of stable chaos in the Solar System," Nature, **357**, 569 (1992). J. J. Lissauer, "Chaotic motion in the Solar System," Rev. Mod. Phys. **71**, 835 (1999) is a review.

because noise can induce jumps between the elements of a generator that does not happen with a single time step by the intrinsic dynamics. However, with our crude description of dynamical systems so far given, we cannot claim anything general as to the noise response of a chaotic system because the 'effectiveness' of noise strongly depends on its details such as which trajectories actually come close in the phase space. We cannot conclude that a system with a larger Kolmogorov-Sinai entropy is more sensitive to noise. For example, due to noise transition into a particular element $A_i \in \mathcal{A}$ can become disproportionately frequent. If this element is embedded by the intrinsic dynamics into a (small portion) of another element, the Kolmogorov-Sinai entropy could become smaller due to noise. This is indeed the essence of the noise-induced order discovered by Tsuda.[382] More extremely, we could imagine a set D outside the range of the intrinsic steady dynamics B such that the trajectories perturbed into D from B by noise are sent back into a very small subset of B. Thus, we see that the noise effect is very sensitively dependent on the details of the system (Fig. 32.3).

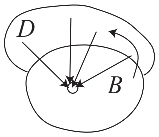


Figure 32.3: A dynamical system that can order by noise. If a phase point goes out of B to D, it is sent back to a very small portion of B. In such a case moderate noise could suppress chaos that is supported on B.

32.18 Open cover

A collection of open sets $\mathcal{O} = \{O_\lambda \subset M, \lambda \in \Lambda\}$ is called an open cover of M, if $M = \cup_{\lambda \in \Lambda} O_\lambda$.

We define the join $\mathcal{O} \vee \mathcal{P}$ of two open covers $\mathcal{O} = \{A_i\}$ and $\mathcal{P} = \{B_j\}$ just as the case of the partitions as (removing empty sets)

$$\mathcal{O} \vee \mathcal{P} = \{A_i \cap B_j\}. \tag{32.18}$$

32.19 Topological entropy

Consider a continuous dynamical system (T, M). Then , if $\mathcal{O}$ is an open cover of M,

[382] K. Matsumoto and I. Tsuda, "Noise-induced order," J. Stat. Mech. **31**, 87 (1983).

$T^{-1}\mathcal{O} \equiv \{T^{-1}A_i\}$ is again an ope cover. Therefore, we can define $\mathcal{O} \vee T^{-1}\mathcal{O}$.

The following limit is called the topological entropy of the open cover $\mathcal{O}$ of (T, M):

$$h_{\text{top}}(T, \mathcal{O}) = \lim_{n\to\infty} \frac{1}{n} \log(\mathcal{O} \vee T^{-1}\mathcal{O} \vee \cdots \vee T^{-n+1}\mathcal{O})^\circ, \tag{32.19}$$

where A° denotes the number of elements in the collection A.

Note that the limit exists just as in the case of te KS entropy, because $(A \vee B)^\circ \le A^\circ B^\circ$.

The topological entropy of (T, M) is defined as

$$h_{\text{top}}(T) = \sup_{\mathcal{O}} h_{\text{top}}(T, \mathcal{O}). \tag{32.20}$$

Just as in the case of te KS entropy, honestly computing the supremum is hard; there is a counterpart of generators, if M is metrizable.

32.20 Top ent is the lowest upper bound of KS entropies
For a dynamical system (T, M), its topological entropy is given by

$$h_{\text{top}}(T) = \sup_{\mu} h_\mu(T), \tag{32.21}$$

where μ is an invariant measure of T.

32.21 Number of fixed points and topological entropy If (T, M) is topological mixing, and has the pseudo orbit tracing property, then

$$h_{\text{top}}(T) = \limsup_{n\to\infty} \frac{1}{n} \log N(T,\ \text{fixed points}). \tag{32.22}$$

This implies that the convergence radius of the ζ-function is given by $e^{-h_{\text{top}}(T)}$.

For a endomorphism of an interval $F : I \to I$, if topologically mixing, we have only to count the number of peaks of F^n:

$$h_{\text{top}}(F) = \limsup_{n\to\infty} \frac{1}{n} \log N(F,\ \text{fixed points}). \tag{32.23}$$

33 Lecture 33. Lyapunov characteristic number

33.1 Lyapunov indices

As seen in examples above and from the Brin-Katok theorem, exponential separation of nearby trajectories may be regarded as a characteristic feature of chaos. Consider two trajectories starting from x and a nearby point $x + \varepsilon v$, where ε is a small positive number and v the directional vector. They tend to separate exponentially as time increases. At time t let us write the separation distance between these two trajectories as $\exp(t\lambda(x, v))$. The exponent $\lambda(x, v)$ is called the Lyapunov exponent (or Lyapunov characteristic exponent; A. M. Lyapunov 1857-1918) for the vector v at $x \in \Gamma$. Its precise definition is

$$\lambda(x, v) = \limsup_{n\to\infty} \frac{1}{n} \log \|D\phi^n(x)v\|, \tag{33.1}$$

where D is differentiation with respect to x[383] and $\| \ \|$ is the norm in the tangent vector space of Γ. The map T defining the dynamical system is written as ϕ to avoid confusion in the present note. At each x $\lambda(x, v)$ as a function of v takes q ($\leq$ the dimension of Γ) distinct values:

$$\lambda^{(1)}(x) > \cdots > \lambda^{(q)}(x). \tag{33.2}$$

Its existence is guaranteed by the following theorem:

33.2 Oseledec's theorem[384]

At each $x \in \Gamma$ the tangent vector space $T_x\Gamma$ ($\simeq \mathbb{R}^n$) of Γ may be decomposed into a direct sum of the form

$$T_x\Gamma = \bigoplus_{i=1}^{q(x)} H_i(x) \tag{33.3}$$

and for $v \in H_j(x)$

$$\limsup_{n\to\infty} \frac{1}{n} \log \|D\phi^n(x)v\| = \lambda_j(x). \tag{33.4}$$

[383] This gives a Jacobi matrix in general.

[384] For a proof, see, for example, A. Katok and B. Hasselblat, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge University Press, 1996) p665. The original theorem is about the general cocycle of a dynamical system and is called Oseledec's multiplicative ergodic theorem, but here it is quoted only in the form directly relevant to our topic.

If the dynamical system is ergodic, this does not depend on x.

33.3 Numerical calculation of Lyapunov spectrum[385]

If we choose an arbitrary vector v to compute (33.1), then almost surely we get λ_1 as can be clear from the illustration Fig. 33.1.

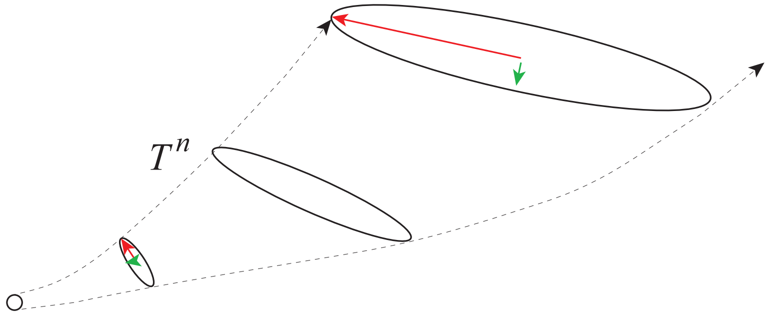


Figure 33.1: Fastest-growing direction wins. In this example, the green direction is also exponentially growing, but the ratio of the lengths of the red and green arrows increase exponentially.

If we could consider the dynamics in the tangent space perpendicular to the fastest-growing direction, then we should obtain the second fastest λ_2. However, the fastest-growing direction rotates as illustrated in 33.1. A natural approach is as follows. Let us assume M to be an n-manifold. Start from a basis $\langle e_1, e_2, \cdots, e_n\rangle$ (of $T_{x_0}M$). For example, the time evolution of the vector looks like the following succession of linear transformations

$$D\phi^n(x_0)e_1 = \phi'(\phi^{n-1}(x_0))\cdots\phi'(\phi(x_0))\phi'(x_0)e_1. \tag{33.5}$$

The expansion rate for this time span for this initial vector is given by

$$\frac{1}{n}\log\|D\phi^n(x_0)e_1\|. \tag{33.6}$$

After several (n) time steps, we may order the length of these vectors, and order them according to the length as $e^{i_1}, \cdots, e_{i_n}$. Then, apply the Gram-Schmidt orthonormalization to make a new ON basis $\langle e_1^{(1)}, e_2^{(1)}, \cdots, e_n^{(1)}\rangle$. Then, repeat the procedure and

[385] The original idea is due to Ippei SHIMADA and Tomomasa NAGASHIMA, "A Numerical Approach to Ergodic Problem of Dissipative Dynamical Systems," Prog. Theor. Phys. 61 1605 (1979). The exposition here is not for actual calculations. For an actual calculation see for example, C. Skokos, The Lyapunov characteristic exponents and their computation, arXiv:0882v2 (Jan 26, (2009).

compute the expansion rate as follows:

$$\frac{1}{N}\left[\frac{1}{n}\log\|D\phi^n(x_0)e_k\| + \frac{1}{n}\log\|D\phi^n(\phi^n(x_0))e_k^{(1)}\| + \cdots + \frac{1}{n}\log\|D\phi^n(\phi^{((N-1)n)}(x_0))e_k^{(N)}\|\right]. \tag{33.7}$$

33.4 Volume expansion rate calculation

In **33.3** each vector direction is calculated separately. Suppose in Fig. 33.1 we calculate the area spanned by the red and green vectors. Its area exponentially grows as $e^{N(\lambda_1+\lambda_2)}$. This means that calculating the volume of the fastest growing k-(rectangular) parallelepiped, we can compute $\lambda_1+\lambda_2+\cdots+\lambda_k$ (the sum up to the kth largest LCN).

33.5 Some notable properties of LCN

Suppose a continuous dynamical system has a bounded phase space (i.e., bounded M) and there is no fixed point, then there must be 0 LCN. This is obvious, because the direction tangent to the orbit cannot grow exponentially and its speed is bounded from below.

For a Hamiltonian dynamical system, the phase volume is conserved, so **33.4** implies that the total sum of LCN must vanish.

Moreover, if lam (>0) is a Lyapunov number, then so is $-\lambda$. This can be shown with the aid of the canonical invariance of

$$\int dq^r dp_r dq^s dp_s \cdots dq^t dp_t. \tag{33.8}$$

33.6 Pesin's equality[386]

For $T \in \text{Diff}^2(M)$ a ergodic measure theoretical dynamical system (T,μ,M) (if any). Then, the following illustration tells us the following equality:

$$h_\mu(T) = \sum_+ \lambda, \tag{33.9}$$

[386] Ya. B. Pesin, "Characteristic Lyapunov exponents and smooth ergodic theory," Russ. Math Surveys 32Z(4) 55 (1977). Section 5. This proves $h \geq \sum_+ \lambda$; the proof is technical and not enlightening; besides, the inequality $\leq$ uses Mather's paper.

where $\sum_+$ means the sum over all the positive LCN.

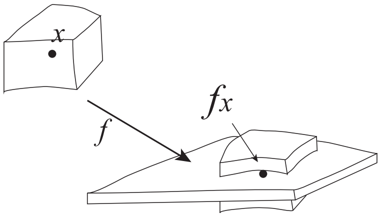


Figure 33.2: Pesin illustrated: the cube edges describe the local unstable and stable directions. Note that these directions are determined by the derivatives of the field. Since the map is C^2, they are smooth so locally expanding directions nicely agree with the original 'blocks' (elements of partitions). Although we may not be allowed to make Markov partitions, but still the partitions are geometrically nice and we can determine the expansion rate in the unstable directions.

33.7 Ruelle's inequality

Generally, if the dynamical system is less smooth, then in Fig. 33.2 nice overlap is not guaranteed, so the expansion direction of the mapped 'cube (the slab in the figure) may no more overlap nicely but only obliquely. Then, the expansion may not as effectively 'dilute' information. Therefore, generally,

$$h_\mu(\phi) \le \sum_+ \lambda \tag{33.10}$$

which is called Ruelle's inequality.[387]

33.8 Kingman's subadditive ergodic theorem[388]

Suppose $X_{m,n}$ $(0 \le m < n)$ satisfy:
(i) $X_{0,n} \le X_{0,m} + X_{m,n}$.
(ii) For each k, $\{X_{nk,(n+1)k}, n \ge 1\}$ is a stationary sequence.
(iii) The distribution of $\{X_{m,m+k}, k \ge 1\}$ does not depend on m.
(iv) $EX^+_{0,1} < \infty$ and for each n, $EX_{0,n} \ge \gamma_0 n$, where $\gamma_0 > -\infty$.
Then,

[387] D. Ruelle, "An inequality for the entropy of differentiable maps," Bol. Soc. Brasil. Mat. **9**, 83 (1978).

[388] based on R. Durrett, *Probability: Theory and examples* (Wadsworth & Brooks/Cole 1991). This is the best advanced introduction to probability.

(a) $\lim_{n\to\infty} EX_{0,n}/n = \inf_m EX_{0,m}/m \equiv \gamma$.
(b) $X = \lim_{n\to\infty} X_{0,n}/n$ exists a.s. and in L^1, so $EX = \gamma$.
(c) If all the stationary sequences in (ii) are ergodic, then $X = \gamma$ a.s.

(a) is clear from Fekete's inequality.

Remark. If we set $X_{n,m} = f(T^{m+1}) + \cdots f(T^n x)$, then (i)-(iv) hold. The outcome is the usual ergodic theorem, but we use this theorem to prove the subadditive version.

33.9 Estimate of $\limsup X_{0,n}/n$

(i) implies, for $n = km + l$ ($l \in [0, m)$),

$$X_{0,n} \le X_{0,km} + X_{km,n}, \quad X_{0,km} \le X_{0,(k-1)m} + X_{(k-1)m,km}, \tag{33.11}$$

so

$$X_{0,n} \le X_{0,(k-1)m} + X_{(k-1)m,km} + X_{km,n}. \tag{33.12}$$

Since

$$X_{0,(k-1)m} \le X_{0,(k-2)m} + X_{(k-2)m,(k-1)m}, \tag{33.13}$$

etc., repeated use of such inequalities to (33.12) gives

$$X_{0,n} \le X_{0,m} + X_{m,2m} + \cdots + X_{(k-1)m,km} + X_{km,n}. \tag{33.14}$$

Therefore,

$$\frac{X_{0,n}}{n} \le \left(\frac{k}{km+l}\right) \frac{X_{0,m} + X_{m,2m} + \cdots + X_{(k-1)m,km}}{k} + \frac{X_{km,n}}{n}. \tag{33.15}$$

Now, we use the ergodic theorem: there must be a limit

$$\frac{X_{0,m} + X_{m,2m} + \cdots + X_{(k-1)m,km}}{k} \to A_m. \tag{33.16}$$

Thus, assuming nice behaviors of the system,

$$\overline{X} \equiv \limsup_{n\to\infty} \frac{X_{0,n}}{n} \le \frac{A_m}{m}. \tag{33.17}$$

If we may assume the system to be ergodic, then (a) implies

$$\int d\mu \, \overline{X} \le \gamma. \tag{33.18}$$

33.10 Almost sure convergence of lim $X_{0,n}/n$

Let

$$\underline{X} \equiv \liminf_{n\to\infty} \frac{X_{0,n}}{n} \tag{33.19}$$

We wish to show

$$\int d\mu\, \underline{X} \geq \gamma. \tag{33.20}$$

This with (33.18) implies

$$\int d\mu\, (\underline{X} - \overline{X}) \geq 0. \tag{33.21}$$

That is, $\underline{X} = \overline{X}$ a.s., the a.s. convergence of $X_{0,n}/n$.

I cannot make the proof of (**??**) in Durrett understandable easily.

34 Lecture 34 KS entropy as isomorphism invariant

34.1 Kolmogorov-Sinai entropy as isomorphism invariant

The Kolmogorov-Sinai entropy was originally proposed as an invariant under isomorphism of measure theoretical dynamical systems. Suppose there are two measure theoretical dynamical systems (T,μ,Γ) and (T',μ',Γ'). Let $\phi:\Gamma\to\Gamma'$ be a one-to-one (except for measure zero sets) correspondent and measure-preserving: $\mu=\mu'\circ\phi$ (that is, the measure of any measurable set measured by the measure μ on Γ and measured by the measure μ' on Γ' after mapping from Γ by ϕ are identical).[389] If the following diagram:

$$\begin{tikzcd}
\Gamma \arrow[r, "T"] \arrow[d, "\phi"] & \Gamma \arrow[d, "\phi"] \\
\Gamma' \arrow[r, "T'"] & \Gamma'
\end{tikzcd}$$

is commutative (except on measure zero sets of Γ and Γ'), that is, if $T=\phi^{-1}\circ T'\circ\phi$, these two measure-theoretical dynamical systems are said to be isomorphic.

It is almost obvious that the Kolmogorov-Sinai entropies of isomorphic measure-theoretical dynamical systems are identical, because for a partition $\mathcal{A}$ of Γ $H(\mathcal{A})=H(\phi\mathcal{A})$ and $\phi\circ T=T'\circ\phi$. We say that the Kolmogorov-Sinai entropy is an isomorphism invariant. If two dynamical systems have different Kolmogorov-Sinai entropies, then they cannot be isomorphic. For example, as can be seen from (34.2) $B(1/2,1/2)$ and $B(1/3,2/3)$ cannot be isomorphic.

An isomorphism invariant that takes the same value if and only if dynamical systems are isomorphic is called a complete isomorphism invariant. If we find such an invariant, the classification of dynamical systems according to isomorphism is reduced to the computation of the invariant. Is the Kolmogorov-Sinai entropy such an invariant? When Meshalkin[390] demonstrated the isomorphism between $B(1/4,1/4,1/4,1/4)$ and $B(1/2,1/8,1/8,1/8,1/8)$ (both have the Kolmogorov-Sinai entropy $2\log 2$), the affirmative answer was expected (also there was a crucial contribution of Sinai: the weak isomorphism theorem for Bernoulli processes). In 1970[391] Ornstein proved that for Bernoulli processes the Kolmogorov-Sinai entropy is a complete invariant. Fur-

[389] Precisely, we must assume not only that $\phi:\Gamma\to\Gamma'$ is one to one as a map between measurable spaces, but also that measurable subsets of Γ are mapped on those of Γ' by ϕ, and vice versa by ϕ^{-1}.

[390] L. D. Meshalkin,"A case of Bernoulli scheme isomorphism," Dokl. Acad. Sci. USSR, **128**(1) 41 (1959).

[391] [1970: Russel died (1872-), Aswan High Dam, Allende became the President of Chile.]

thermore, completeness was proved for any finite mixing Markov chain with finite Kolmogorov-Sinai entropy.[392],[393]

34.2 Ornstein-Weiss' theorem

Ornstein and Weiss proved the following theorem:[394]

Theorem 2.7A.1 Not completely predictable systems (= systems with positive Kolmogorov-Sinai entropy) have Bernoulli flows (continuous dynamical systems whose periodically sampled sequences become Bernoulli processes) as their factor dynamical systems.

Here, "to have A as a factor dynamical system" implies that the original dynamical system behaves as dynamical system A if it is reduced to a certain space (more precisely, there is a homomorphism from the system onto A). They simply equate the chaotic system and the system with positive Kolmogorov-Sinai entropy in the quoted review article.

34.3 Bernoulli system

Let $M = \mathbb{Z}_n^{\mathbb{Z}}$, where $\mathbb{Z}_n = \{0, 1, \cdots, n-1\}$. We can define a shift dynamical system (actually it is a full shift). We introduce a measure μ on M through the measures of the cylinder sets as[395]

$$\mu([\omega_{k+1} = \alpha_1, \cdots, \omega_{k+q} = \alpha_q]) = \prod_{m=1}^{q} p_{\alpha_m}, \tag{34.1}$$

where $\alpha_m \in \mathbb{Z}_n$ and $p_1, \cdots, p_n$ are $p_m \in [0,1]$ with $\sum_{m=1}^{n} p_m = 1$ (i.e., p_i is the probability for a letter in the alphabet $\mathbb{Z}_n$.

(σ, M, μ) is a measure-theoretical dynamical system called the Bernoulli system and is denoted as $B(p_1, \cdots, p_n)$.

$B(1/2, 1/2)$ corresponds to the coin-tossing process, and is isomorphic to baker's transformation (see **27.4**). It is also isomorphic to the horseshoe dynamical system restricted on its nontrivial invariant set (see **28.5**).

[392] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory*, (Springer, 1982); D. S. Ornstein, *Ergodic Theory, Randomness, and Dynamical Systems* (Yale University Press, 1974).

[393] However, the cases with zero entropy are different; there are numerous non-isomorphic dynamical systems with zero entropy.

[394] D. S. Ornstein and B. Weiss, "Statistical Properties of Chaotic Systems," Bull. Amer. Math. Soc. **24**, 11 (1991), Theorem 1.4.3.

[395] This can actually uniquely specify μ on M according to Kolmogorov's extension theorem.

34.4 Kolmogorov-Sinai entropy of Bernoulli system

The Kolmogorov-Sinai entropy of $B(p_1, \cdots, p_n)$ is given by

$$h = -\sum_{i=1}^{n} p_i \log p_i. \tag{34.2}$$

We can compute the loss of information by σ easily (see **32.1**). It is indeed given by this formula.

34.5 Central statements of theory of Bernoulli processes[396]

Two central statements of the theory are the Sinai theorem and the Ornstein theorem:
(I) [Sinai's weak isomorphism theorem] Every ergodic process with positive entropy h has any Bernoulli shift of entropy smaller than or equal to h as a measure-theoretic factor.[397]
(II) [Ornstein's isomorphism theorem] Bernoulli shifts with equal entropies are isomorphic.

The Ornstein theorem fails for unilateral shifts (but the Sinai theorem holds, which is less known).

[396] T. DOWNAROWICZ and J. SERAFIN, A short proof of the Ornstein theorem, Ergod. Th. & Dynam. Sys. 32 587 (2012). The authors of this note managed to simplify the Burton, Keane and Serafin method. It is the residual Sinai theorem which is central in our method, and the number of applications of the Baire theorem is reduced to one. We continue to invoke the elementary combinatorial 'marriage lemma.' I give up to illustrate this paper.

[397] ⟪**Factor**⟫ A coarse-grained dynamical system is called a factor of the original dynamical system. Formally, (S, ν, N) is a factor of (T, μ, M), if there is a measurable 'homomorphism' $\varphi : M \to N$ such that
(i) $S \circ \varphi = \varphi \circ T$,
(ii) $\nu(A) = T(\varphi^{-1}(A)0$ for measurable set in N

35 Large deviation approach to dynamical systems

35.1 Level 1 Large deviation: review

We have already discussed large deviation theory in **31.8** for Bernoulli samples, but here we wish to be a bit more systematic.

Our starting point is the law of large numbers. Consider a chaotic endomorphism f. If it is chaotic enough (e.g., topologically mixing), the law of large numbers holds for the time average:

$$\lim_{N\to\infty} P\left(\frac{1}{N}\sum_{k=0}^{N-1} f^k(x) \sim \langle f\rangle\right) = 1, \tag{35.1}$$

where $\langle f\rangle$ is the average. Here, $f^k(x)$ may be denoted as $x(k,\omega)$ as well (i.e., the trajectory (specified by) ω observed at time k). The long-time average $\langle f\rangle$ is the true expectation value (ergodicity).

Now, what happens if time is not long enough? The average should deviate (fluctuate) from the true average. How? This is expressed as the large-deviation principle

$$P\left(\frac{1}{N}\sum_{k=0}^{N-1} f^k(x) \sim x\right) \approx e^{-NI(x)}, \tag{35.2}$$

where $I(x)$ is called the rate function or the large deviation function. It is a convex function with the unique minimum at $x = \langle f\rangle$. Thus, (35.2) includes the law of large numbers.

35.2 Level 2 Large deviation: introduction

Sanov's theorem **31.8** we already discussed for simple cases is actually a 'level 2' theory.

We can study the invariant measure empirically, taking statistics, so there must be a corresponding law of large numbers

$$\lim_{N\to\infty} P\left(\frac{1}{N}\sum_{k=0}^{N-1} \delta(x(k,\omega)-\cdot) \sim \frac{d\mu(\cdot)}{dm(\cdot)}\right) = 1. \tag{35.3}$$

Here, m is the basic sampling measure (for the real space, it is usually the Lebesgue measure λ, so m and λ may be used interchangeably).

There must be a corresponding large deviation principle. The LD for expectation values is called the level 1 LD, and this one the level 2.

$$P\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta(x(k,\omega)-\cdot)\sim\frac{d\mu(\cdot)}{dm(\cdot)}\right)\approx e^{-NI^{(2)}(\mu)}. \tag{35.4}$$

Here $I^{(2)}$ is the rate function(al) which is a convex functional of the measures with the unique minimum at the invariant measure compatible with the sampling measure. As is clear from the definition, μ must be absolutely continuous with respect to m.

35.3 Gärtner-Ellis theorem: level 1

We make a 'partition function' (generator)

$$Z(t)=\left\langle\exp\left(t\sum_{k=0}^{N-1}x(k,\omega)\right)\right\rangle, \tag{35.5}$$

where $\langle\ \rangle$ implies the average over the sampling measure m. Let us use (35.2) to compute this average:

$$Z(t)=\int dm(\omega)\int dx\,\delta\left(\frac{1}{N}\sum_{k=0}^{N-1}x(k,\omega)\sim x\right)e^{Ntx}=\int dx\,e^{N[tx-I(x)]}=e^{Nq(t)}, \tag{35.6}$$

where $q(t)$ is the $(-)$ free energy. Since N is very large, the integral is dominated by the peak value of the integrand, i.e., 'max'$_x[tx-I(x)]$. Therefore, we obtain

$$q(t)=\sup_x[tx-I(x)]. \tag{35.7}$$

Since $I(x)$ is convex, this is a proper Legendre transformation, and $q(t)$ is also a convex function. Thus,

$$I(x)=\sup_t[tx-q(t)]. \tag{35.8}$$

That is, 'entropy' (i.e., $-\log$ 'probability') may be obtained from 'free energy' through a Legendre transformation: q is much easier to compute than I, but I dictates what you can observe.

35.4 Gärtner-Ellis theorem: level 2

There must be a level 2 version of 'statistical mechanics.' We simply mimic the level 1 **35.3**. Make the partition function(al):

$$Z(\phi) = \int dm(\omega) \int \delta\mu\, \delta\left(\frac{1}{N}\sum_{k=0}^{N-1} \delta(x(k,\omega) - \cdot) \sim \frac{d\mu(\cdot)}{dm(\cdot)}\right) e^{N\int d\mu(\cdot)\,\phi(\cdot)} \quad (35.9)$$

$$= \int \delta\mu\, e^{N[\int d\mu(\cdot)\,\phi(\cdot) - I^{(2)}(\mu)]} = e^{Nq(\phi)}, \quad (35.10)$$

where $q(\phi)$ is the $(-)$ free energy functional, and $\int \delta\mu$ is a functional integral. Since N is very large, the integral is dominated by the peak value of the integrand. Therefore, we obtain

$$q(\phi) = \sup_{\mu}\left[\int d\mu(\cdot)\,\phi(\cdot) - I^{(2)}(\mu)\right]. \quad (35.11)$$

Since $I^{(2)}(\mu)$ is convex, this is a proper Legendre transformation, and $q(\phi)$ is also a convex function(al). Thus,

$$I^{(2)}(\mu) = \sup_{\phi}\left[\int d\mu(\cdot)\,\phi(\cdot) - q(\phi)\right]. \quad (35.12)$$

That is, 'entropy' may be obtained from 'free energy' through a Legendre transformation: q is much easier to compute than I, but I dictates what you can observe just as at level 1.

35.5 Sanov's theorem revisited

Let us compute $I^{(2)}(\mu)$ when the sampling measure is m. Notice that the partition function (35.9) can be rewritten as

$$Z(\phi) = \int dm(\omega)\, e^{\sum_{k=0}^{N-1}\phi(x(k,\omega))}, \quad (35.13)$$

because μ is an empirical measure obtained from $\{x(k,\omega)\}_{k=0}^{N-1}$. Let us assume that $x(k,\omega)$ are statistically independent for different k (i.e., let us assume the dynamics is Bernoulli). Then,

$$Z(\phi) = \int dm \prod_{k=0}^{N-1} e^{\phi(x(k,\omega))} = z^N, \quad (35.14)$$

where

$$z(\phi) = \int dm\, e^{\phi(x)}, \tag{35.15}$$

Therefore, $q(\phi) = \log z(\phi)$ or

$$I^{(2)}(\mu) = \sup_{\phi} \left[\int d\mu(\cdot)\, \phi(\cdot) - \log z(\phi) \right]. \tag{35.16}$$

Let us compute ϕ that 'maximizes' [ ]. The functional derivative gives[398]:

$$\int d\mu(x)\, \delta(x-y) - \frac{1}{z(\phi)} \frac{\delta z}{\delta \phi(y)} = 0, \tag{35.17}$$

where

$$\frac{\delta z}{\delta \phi(y)} = \int dm\, \delta(x-y) e^{\phi(x)} \tag{35.18}$$

Therefore, we get

$$d\mu(y) = \frac{dm(y) e^{\phi(y)}}{z(\phi)} \quad \Rightarrow \quad \frac{d\mu}{dm} = \frac{e^{\phi(y)}}{z(\phi)}. \tag{35.19}$$

That is,

$$\phi(y) = \log \frac{d\mu}{dm} + \log z. \tag{35.20}$$

Introducing this into (35.16), we obtain

$$I^{(2)}(\mu) = \int d\mu(\cdot)\, \log \frac{d\mu}{dm}. \tag{35.21}$$

This is called Sanov's theorem, justifying the use of information theory to search for the most probable results.

Notice, however, the result is only for Bernoulli processes. We must relax this constraint.

35.6 Extension of Sanov-type formula for general level 2

Warning: I identify the actual trajectories in M and their counterpart symbol sequences. In short I identify as measurable spaces the actual dynamical system and its isomorphic symbolic dynamics.

The easiest way is to regard a chunk of a trajectory (consecutive n time points) as a single state. Thus,

$$I^{(2)}(\mu) = \int d\mu(\cdot)\, \log \frac{d\mu}{dm}. \tag{35.22}$$

[398] The basic measure is m. That is δ-function is with respect to measure m

reads (recall the notation $[x] = x_1 \cdots x_n$)

$$I^{(2)}(\mu) = \int d\mu([x]_n) \log \frac{d\mu}{dm}([x]_n). \tag{35.23}$$

The Radon-Nikodym derivative can be rewritten as follows, assuming n is sufficiently large:

$$\frac{d\mu}{dm}([x]_n) = \frac{\mu([x]_n)}{m([x]_n)}. \tag{35.24}$$

Using the conditional probability

$$m([x]_n) = m(x_1 \,|\, x_2 \cdots x_n) m(x_2 \cdots x_n) \tag{35.25}$$

therefore, in the large n limit we can write

$$\begin{aligned} m([x_1 \cdots x_n]) &= \frac{m([x_1 x_2 \cdots x_n])}{m([x_2 \cdots x_n])} \frac{m([x_2 x_3 \cdots x_n])}{m([x_3 \cdots x_n])} \cdots & (35.26)\\ &= \frac{m([x_1 x_2 \cdots x_n])}{m(f[x_1 x_2 \cdots x_n])} \frac{m([x_2 \cdots x_n])}{m(f[x_2 \cdots x_n])} \cdots . & (35.27) \end{aligned}$$

Therefore,

$$\frac{1}{n} \log m([x_1 \cdots x_n]) = \frac{1}{n} \sum_{j=1}^{n-1} \log \frac{m([x_j \cdots x_n])}{m(f[x_j \cdots x_n])}. \tag{35.28}$$

Analogously we get

$$\frac{1}{n} \log \mu([x_1 \cdots x_n]) = \frac{1}{n} \sum_{j=1}^{n-1} \log \frac{\mu([x_j \cdots x_n])}{\mu(f[x_j \cdots x_n])}. \tag{35.29}$$

Thus, if we consider $(1/n)I^{(2)} \to I^{(3)}$ in the large n limit, we get

$$I^{(3)} = \int d\mu(x) \left[\log \frac{d\mu}{d\mu \circ f} - \log \frac{dm}{dm \circ f} \right]. \tag{35.30}$$

Recall (31.3). The first term is $(-)$KS-entropy. Notice that this is the large deviation function for trajectories (so it is called the level 3 rate function).

Although I said I would identify the real space and the symbol space, and although the identification is almost perfect (it is perfect for 1D endomorphism, so this immediately gives Rohlin's formula), the formula must be carefully interpreted. For a sequence $\{x_1, x_2, x_3, \cdots\}$, $f(\{x_1, x_2, x_3, \cdots\}) = \{x_2, x_3, \cdots\}$ (cf. the shift). Notice that one symbol disappears. This shortening of the cylinder set implies expansion. That is why the first term is (−)KS-entropy. The second term is the expansion rate; if we have a nice coding such as Markov partitions, then it is the sum of positive LCN. That is, (35.30) reads

$$I^{(3)} = \sum_{+} \langle \chi \rangle_\mu - h_\mu \geq 0 \tag{35.31}$$

Thus, Pesin's equality holds for the most probable state = equilibrium state.

35.7 Energy function

To do study statistical mechanic on the lattice we need an energy function $\varphi : \Sigma_n \to \mathbb{R}$. It is the 1/2 of the total interaction and the self energy of a spin. we impose the following constraint on ϕ. Let

$$\mathrm{var}_k \phi = \sup\{|\phi(x) - \phi(y)| \, : x_i = y_i \text{ for } |i| \leq k\}. \tag{35.32}$$

Our constraint is

$$\mathrm{var}_k \varphi \leq b\alpha^k, \tag{35.33}$$

where $b > 0$ and $\alpha \in (0, 1)$. Let us explicitly write

$$\mathcal{F}_A = \{\phi \,|\, \Sigma_A \to \mathbb{R}, \mathrm{var}_k \varphi \leq b\alpha^k, \forall k \in \mathbb{N}\}. \tag{35.34}$$

Here Σ_A is a Markov subshift with matrix A.

35.8 Gibbs measure

For any Hamiltonian $\phi \in \mathcal{F}_A$, there is a unique shift invariant measure μ_ϕ satisfying the following inequality for some positive constants C_1, C_2 and P (= free energy per spin)

$$C_1 \leq \frac{\mu\{y \,: y_i = x_i, \forall i \in \{0, 1, \cdots, n\}}{\exp(-Pm + \sum_{k=0}^{n-1} \phi(\sigma^k x))} \leq C_2, \tag{35.35}$$

35.9 Transfer operator

Consider the totality of the states on the right-half lattice S_A^+.[399] and $\phi \in C(\Sigma_A^+$.

[399] If you read the original math paper, there is a log discussion about how to justify considering of Σ^+ instead of S.

Define the transfer operator T_ϕ as

$$[T_\phi f](x) = \sum_{y \in \sigma^{-1}x} e^\phi(y) f(y). \tag{35.36}$$

Notice that $y = x_* x_1 x_2 \cdots$.

35.10 Ruelle-Perron-Frobenius theorem

Let Σ_A be mixing and $\phi \in \mathcal{F}_A \cap C(\Sigma_A^+)$. There is a unique positive eigenvalue λ_ϕ of T_ϕ, and the Gibbs measure is obtained from the partition func tion and the normalization obtained from λ_ϕ

$$\mu([x]_n) \simeq \frac{1}{\lambda_\phi^n} \exp\left(\sum_{i=1}^n \phi(x_i) \right), \tag{35.37}$$

where $[x]_n = x_1 \cdots x_n$.

35.11 Variational principle for Gibbs measure

The Gibbs measure μ_ϕ is the unique measure satisfying the following variational principle:

$$s(\mu) + \int \phi d\mu = P(\phi). \tag{35.38}$$

where s is the entropy per spin.

The T-invariant measure satisfying the variational principle is called an equilibrium state wrt to T and ϕ.[400]

[400] Th 2.7

36 Thermodynamic formalism

36.1-36.8 summarize the Gibbs measure theory on 1D lattice (for disordered systems). Mathematically, this measure and the observable measure for chaotic dynamical systems are closely related. Here, we wish to maximally abuse functional analysis and to respect theoretical physics aesthetics.

36.1 Preparatory comments on measures

Let $M(X)$ be the set of all the Borel measures on X.
Let T be an automorphism on X. Then, T induces a map T^* on $M(X)$: for $\mu \in M(X)$

$$\mu(T^{-1}E) = (T^*\mu)(E) = T^*\mu(E). \tag{36.1}$$

Since $T^{-1}(E)$ the totality of the points coming to E after one time step, $T^*\mu = \mu$ means the preservation of measure.

We may introduce a weak topology on $M(X)$. Actually,
Proposition: $M(X)$ is a compact convex metrizable space.

36.2 Invariant measure set is nonempty and closed

Proposition: Let $M_T(X)$ be the set of all the invariant measures. Then, it is nonempty and is closed in $M(X)$.

[Demo] Let

$$\mu_n = \frac{1}{n}(\mu_0 + T\mu_0 + \cdots + T^n\mu_0). \tag{36.2}$$

Since $M(X)$ is compact, we can always take a converging subsequence. The result is invariant.

$\mu \in M_T(X)$ means

$$\int (f \circ T) d\mu = \int f d\mu. \tag{36.3}$$

36.3 Lattice states

Let $x_i \in \{0, 1, \cdots, n-1\}$ be a state at a single lattice point. A lattice state is described by $\{x_k\}_{k=-\infty}^{\infty}$. The totality of these sequences is Σ_n.

36.4 Energy function

To do study statistical mechanic on the lattice we need an energy function $\varphi : \Sigma_n \to \mathbb{R}$. It is the 1/2 of the total interaction and the self energy of a spin. We impose the following constraint on ϕ. Let

$$\mathrm{var}_k\phi = \sup\{|\phi(x) - \phi(y)| \, : x_i = y_i \text{ for } |i| \leq k\}. \tag{36.4}$$

Our constraint is

$$\mathrm{var}_k\varphi \leq b\alpha^k, \tag{36.5}$$

where $b > 0$ and $\alpha \in (0, 1)$. Let us explicitly write

$$\mathcal{F}_A = \{\phi \,|\, \Sigma_A \to \mathbb{R}, \mathrm{var}_k\varphi \leq b\alpha^k, \forall k \in \mathbb{N}\}. \tag{36.6}$$

Here Σ_A is a Markov subshift with matrix A.

36.5 Gibbs measure

For any Hamiltonian $\phi \in \mathcal{F}_A$, there is a unique shift invariant measure μ_ϕ satisfying the following inequality for some positive constants C_1, C_2 and A (= free energy per spin)

$$C_1 \leq \frac{\mu\{y \, : y_i = x_i, \forall i \in \{0, 1, \cdots, n\}}{\exp(-An + \sum_{k=0}^{n-1} \phi(\sigma^k x))} \leq C_2, \tag{36.7}$$

36.6 Transfer operator

Consider the totality of the states on the right-half lattice S_A^+.[401] and $\phi \in C(\Sigma_A^+$. Define the transfer operator T_ϕ as

$$[T_\phi f](x) = \sum_{y \in \sigma^{-1}x} e^{\phi(y)} f(y). \tag{36.8}$$

Notice that $y = x_* x_1 x_2 \cdots$.

[401] If you read the original math paper, there is a long discussion about how to justify considering of Σ^+ instead of Σ_A.

36.7 Ruelle-Perron-Frobenius theorem

Let Σ_A be mixing and $\phi \in \mathcal{F}_A \cap C(\Sigma_A^+)$. There is a unique positive eigenvalue λ_ϕ of T_ϕ, and the Gibbs measure is obtained from the partition function and the normalization obtained from λ_ϕ

$$\mu([x]_n) \simeq \frac{1}{\lambda_\phi^n} \exp\left(\sum_{i=1}^{n} \phi(x_i)\right), \tag{36.9}$$

where $[x]_n = x_1 \cdots x_n$ is a cylinder set.

36.8 Variational principle for Gibbs measure

The Gibbs measure μ_ϕ is the unique measure satisfying the following variational principle:

$$s(\mu) + \int \phi d\mu = \tilde{A}(\phi). \tag{36.10}$$

where s is the entropy per spin, and $\tilde{A} = (1/N)\log Z = \log \lambda_\phi$ (that is, $-\beta A/N$ in the standard statistical thermodynamic notation).

The T-invariant measure satisfying the variational principle is called an equilibrium state wrt to T and ϕ.[402]

36.9 Large deviation and thermodynamics

In the following we assume H is bounded from above.

The partition function

$$Z(\beta) = \sum e^{(-\beta)\sum h} \tag{36.11}$$

may be interpreted as the generating function for energy with respect to the Liouville measure (the uniform or equal probability) measure. We ask the energy fluctuation with respect to this measure (for a portion of the system containing N subsystems, which are assumed to be more or less statistically independent)(. We assume the LD principle:

$$P\left(\frac{1}{N}\sum h \sim e\right) \approx e^{-NI(e)}. \tag{36.12}$$

[402] Th 2.7

Its generating function may be written as

$$\frac{Z(\beta)}{Z(0)} = \left\langle e^{(-\beta)\sum h}\right\rangle_0, \tag{36.13}$$

where $\langle\ \rangle_0$ denotes the expectation value with respect to the uniform measure.

(36.13) may be rewritten as

$$\frac{Z(\beta)}{Z(0)} = \left\langle e^{(-\beta)\sum h}\right\rangle_0 = \int dy \left\langle \delta\left(y - \frac{1}{N}\sum h\right)\right\rangle_0 e^{(-\beta)Ny} \tag{36.14}$$

$$= \int dy\, P\left(\frac{1}{N}\sum h \sim y\right) e^{(-\beta)Ny} = \int dy e^{-NI(y)} e^{(-\beta)Ny}. \tag{36.15}$$

Using the Laplace-type approximation, we have

$$\frac{1}{N}\log(Z(\beta)/Z(0)) = \sup_y[(-\beta)y - I(y)]. \tag{36.16}$$

Introduce

$$[-a](-\beta) = \frac{1}{N}\log(Z(\beta)). \tag{36.17}$$

(36.16) reads

$$[-a](-\beta) - [-a](0) = \sup_y[(-\beta)y - I(y)]. \tag{36.18}$$

In terms of the usual thermodynamic quantity $[-a](-\beta) = -\beta A/N$, $y = E/N = e$. Also so $[-a](0) = -\lim_{\beta\to 0}\beta A/N = -\lim_{T\to\infty} A/T = \lim_{T\to\infty}(S - E/T)$, but if we assume H to be bounded (as magnets), this converges to $S(\infty)$, the log of the phase volume, essentially. Therefore, (36.18) reads

$$-\beta A/N = \sup_e[(-\beta)e - I(e)] + S(\infty), \tag{36.19}$$

so $I(e) = \beta(A - E)/N - S_\infty = S_\infty - S$. The most probable e is the internal energy at $\beta = 0$. Thus, $I(e)$ is the information required to characterize the deviated energy distribution from uniform distribution: That is, the following KL entropy is

$$I(e) = \frac{1}{N}\sum_i f_i \log(Z_0 f_i) = S_\infty - S. \tag{36.20}$$

By the way, this can be obtained from Sanov's theorem + the contraction principle:

$$I(e) = \inf_{\nu:\int d\nu\, h = e} I^{(2)}(\nu). \tag{36.21}$$

Here $I^{(2)}$ is the level 2 large deviation functional with respect to the equal probability measure on the phase space. This relation is what Jaynes misunderstood as the principle to found statistical mechanics.

36.10 Perron-Frobenius eigenvalue problem

The Perron-Frobenius equation **31.4** reads

$$\varphi(x) = (\mathcal{L}_F \varphi)(x) = \int m(dy)\delta(x - F(y))\varphi(y) = \sum_{y \in F^{-1}(x)} \frac{\varphi(y)}{|F'(y)|}. \tag{36.22}$$

For reasonably chaotic systems ('(non-uniformly) hyperbolic systems'), the counterpart should read

$$\varphi(x) = (\mathcal{L}_F \varphi)(x) = \int m_+(dy)\delta(x - F(y)) = \sum_{y \in F^{-1}(x)} \frac{\varphi(y)}{L_+(y)}, \tag{36.23}$$

where m_+ is the Lebesgue measure (better, the Riemann volume of unstable manifolds; red curves in Lecture 39), L_+ is the expansion rate of the unstable manifold (the sum of all the positive Lyapunov characteristic numbers; cf. **33.6**), and δ must be defined wrt m_+.

Note the following formula:

$$(\mathcal{L}_{F^n} \varphi)(x) = \sum_{y \in F^{-n}(x)} \frac{\varphi(y)}{|(F^n)'(y)|} = (\mathcal{L}_F^n \varphi)(x). \tag{36.24}$$

Following the Fredholm integral equation, the corresponding eigenvalue problem reads

$$\varphi(x) = \lambda(\mathcal{L}_F \varphi)(x). \tag{36.25}$$

Note that $\mathcal{L}_F$ is a special case of the transfer operator **36.6** with $\phi = -\log |F'|$. The eigenvalue problem used in **35.10** is the reciprocal of λ in (36.25). The conventions in linear algebra and in integral equations are different.

36.11 Fredholm determinant for Perron-Frobenius operator[403]

(36.25) may be rewritten as

$$(1 - \lambda \mathcal{L}_F)\varphi = 0. \tag{36.26}$$

[403] Y Oono and Y Takahashi, "Chaos, external noise and Fredholm theory," Prog Theor Phys 63 1804 (1980).

Therefore, the eigenvalues should be the zeros of the 'determinant' $D(\lambda, F)$ of $1-\lambda\mathcal{L}_F$. To define this we use an identity $\det B = \exp(\mathrm{Tr}\,\log B)$:

$$\det(1-\lambda A) = \exp\left[\mathrm{Tr}\,\log(1-\lambda A)\right] = \exp\left[-\mathrm{Tr}\sum_{n=1}^{\infty}\frac{\lambda^n}{n}A^n\right]. \tag{36.27}$$

This implies that we have to define the trace of $\mathcal{L}_F^n$. Looking at (36.24) we define

$$\mathrm{Tr}\,\mathcal{L}_F^n = \sum_{z\in F^{-n}(z)}\frac{1}{|(F^n)'(z)|} \equiv Q_n. \tag{36.28}$$

Thus, we define

$$D(\lambda, F) \equiv \det(1-\lambda\mathcal{L}_F) = \exp\left[-\sum_{n=1}^{\infty}\frac{\lambda^n}{n}Q^n\right]. \tag{36.29}$$

Recall that the ζ-function (Artin-Mazur-Ruelle ζ-function) **26.9** is just the reciprocal of this quantity.

36.12 Significance of $D(\lambda, F)$

(1) The expansion of $D(\lambda, F)$ around $\lambda = 0$ is intimately related to the symbolic realization of the dynamical system.
(2) If $D(1, F) = 0$, then the system has an observable measure.[404]
(3) At criticality, $D(\omega, F) = 0$ for any ω such that $\omega^n = 1$. That is, the natural boundary of $D(\lambda, F)$ is a unit circle iff F is critical (under the condition that F' is bounded).

36.13 Free energy

Look at

$$-\log D(\lambda, F) = \sum_{n=1}^{\infty}\frac{\lambda^n}{n}Q^n. \tag{36.30}$$

The structure of Q_n is

$$Q_n = \mathrm{Tr}\,\mathcal{L}_F^n = \sum_{\text{fixed points of } F^n} e^{-\log|(F^n)'(y)|}. \tag{36.31}$$

[404] Precisely speaking, this is our conjecture.

Notice that the cain rule means

$$(F^n)'(y) = F'(F^{n-1}(x_0))F'(F^{n-2}(x_0)) \cdots F'(x_0) = \prod_i F'(x_i), \tag{36.32}$$

where $\{x_0, x_1, \cdots, x_{n-1}\}$ is a periodic orbit. $1/n$ in front of Q_n in (36.30) means that the actual sum is over the periodic trajectories. You could think that $\log |F'|$ is the energy (Hamiltonian) and that the trajectory positions correspond to 'microstates' of a statistical equilibrium system (see **35.10**), although there is no temperature (yet) β. Thus, (36.30) may be interpreted as the grand canonical partition function by setting the fugacity $\lambda = e^{\beta\mu}$. Q_n is the canonical partition function, so we may introduce the free energy A (per time = lattice point):[405]

$$A = -\limsup_{n\to\infty} \frac{1}{n} \log Q_n. \tag{36.33}$$

Notice that $\rho = e^A$ is the convergence radius of the zeta function = inverse of Perron-Frobenius eigenvalue.

If the nonwandering set is a stable periodic orbit, then $A < 0$.
If $A = 0$, there is an observable invariant measure.

36.14 Chaotic system under noise

Instead of $x \to F(x)$, let us add an additive noise

$$x_{n+1} = F(x_n) + \nu_n, \tag{36.34}$$

where we assume ν_n is a noise (at time n) which has a density distribution g. Then, the Perron-Frobenius equation is converted to

$$\varphi_\nu(x) = \int m(dy) g(x - F(y))\varphi_\nu(y) \equiv (\hat{\mathcal{L}}_F \varphi_\nu)(x). \tag{36.35}$$

We expect that in the $\nu \to 0$ limit (that is, the $g \to \delta$ limit), the system, if stable, should recover the noiseless system. Notice that, simply following the computational rule of the δ-function, we must conclude

$$\lim_{g\to\delta} \hat{\mathcal{L}}_F^n = \int m(dy)\delta(y - F_n(y)) = \sum_{z\in F^{-n}(z)} \frac{1}{|(F^n)'(z) - 1|} \equiv \hat{Q}_n. \tag{36.36}$$

[405] Mathematicians do not like '−' in front of the following definition, and from the convex analytical point of view mathematicians' convention is more rational than statistical-physicists', but here we stick to the physics tradition.

The difference between Q_n and $\hat{Q}_n$ is insignificant, if the system is sufficiently chaotic.

The measure stable under noise is the physical measure defined by Kolmogorov.[406]

Thus, the statistical properties of a chaotic system is stable against noise. Thus, we may say that the macroscopic stability is warranted by microscopic instability. This is very suggestive of the stability of tropical rain forest systems.

36.15 Let us introduce temperature[407]

Instead of (36.31) let us introduce the canonical partition function

$$Q_n(\beta) = \sum_{\text{fixed points of } F^n} e^{-\beta \log |(F^n)'(y)|}. \tag{36.37}$$

Accordingly, we may introduce the temperature-dependent free energy

$$A(F, \beta) = -\beta^{-1} \limsup_{n\to\infty} \log Q_n(\beta). \tag{36.38}$$

36.16 Thermodynamics

The internal energy reads (use the Gibbs-Helmholtz relation)

$$E(\beta) = \frac{\partial \beta A}{\partial \beta} = \langle \log |F'| \rangle_\beta. \tag{36.39}$$

Entropy is

$$S(\beta) = \beta^2 \frac{\partial A}{\partial \beta} = \beta(E(\beta) - A(F, \beta)). \tag{36.40}$$

Actually, we know this is a Legendre transformation. If we compare this with the LD formalism, this entropy is just the Kolmogorov-Sinai entropy.

From **36.12** (2), if F allows an observable chaos, $A(1, F) = 0$, so we recover Rohlin's formula:

$$S(1) = \langle \log |F'| \rangle. \tag{36.41}$$

[406] Eckmann-Ruelle says: by Kolmogorov (we are not aware of a published reference) a long time ago.

[407] Y Takahashi and Y Oono, "Towards he statistical mechanics of chaos," Prog Theor Phys 71 851 (1984).

Note that $S(0)$ is given by

$$S(0) = \limsup_{n\to\infty} \frac{1}{n}[\#\mathrm{Fix}(F^n)] \tag{36.42}$$

which is the topological entropy.[408] Entropy should be an increasing function of temperature: thus the KS entropy is bounded by topological entropy:

$$S(0) \geq S(1). \tag{36.43}$$

This is Dinaburg's theorem.[409]

36.17 What is $\boldsymbol{\beta}$?

The conjecture is:
The Hausdorff dimension of the support of the observable measure if β_c such that $A(\beta_c, F) = 0$.

36.18 What is the outstanding conjecture theoretical physicists can make?

This is a summary of what I was pursuing.

A necessary and sufficient condition for a dynamical system to have an observable measure defined by the stability against noises (i.e., a physical measure in Kolmogorov's sense) is that the Fredholm determinant of the Perron-Frobenius (+-version) satisfies $D(1, F_+) = 0$.

$D(\lambda, F_+)$ may be factorized into holomorphic factors each of which has nondegenerate $\lambda = 1$ as its smallest zero and corresponds to the unique observable measure supported on an invariant set (= basic set). The observable measure can be described as a Gibbs measure with $|\log F_+|$ as the energy function.

[408] R. Bowen, TAMS 184 125 (1973).
[409] E. I. Dinaburg, Math USSR Izv 5 337 (1971).

37 Takens embedding theorem

37.1 Original statement of Takens embedding theorem

Let $y : M \to \mathbb{R}$ be a smooth function (observable), where M is a smooth m-manifold. What can you know about a smooth dynamics φ_t on M from y?

Theorem. It is a generic property that the map $\Phi : M \to \mathbb{R}^{2m+1}$:

$$\Phi(x) = (y(x), y(\varphi(x)), \cdots, y(\varphi^{2m+1}(x))) \tag{37.1}$$

is an embedding at least C^2.
Technical conditions further required are:
(i) If x is periodic, its period is less than $2m + 1$, all eigenvalues of $d\varphi^k$ are different and distinct from 1.
(ii) No two fixed points of φ gives the same y.
(iii) For Φ to be an immersion near a fixed point x the covectors $dy, d(y\varphi), \cdots d(y\varphi^{2m})$ span $T_x^*(M)$.[410]

Some extension: J. P. Huke and D. S. Broomhead. Embedding theorems for non-uniformly sampled dynamical systems. Nonlinearity, 20(September):2205, 2007.

https://www.youtube.com/watch?v=6i57udsPKms&frags=pl%2Cwn good

Taken's paper

37.2 Differential topological rudiments

M and N are manifolds and $\dim M = m < n = \dim N$. If $F : M \to N$ is C^1, then $N \setminus f(M)$ is dense in N.
M and N are manifolds and $\dim M = m > n = \dim N$. If $F : M \to N$ is C^1, and submersive[411] at every point, then $F^{-1}(p)$ is a submanifold in M with dimension $m - n$.

37.3 Takens' embedding theorem

[410] The condition on dy means that $dy = dx_i \frac{\partial y}{\partial x_i}$

[411] $DF : T_pM \to T_{F(p)}N$ is surjective.

Let M be a compact manifold of dimension m. Let $\phi \in \mathrm{Diff}^2(M)$ (dynamics) and $y : M \to \mathbb{R}$ be C^2 (observable). Then, $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$ defined as

$$\Phi_{(\phi,y)} = (y(x), y(\phi(x)), \cdots, y(\phi^m(x)). \tag{37.2}$$

is, generically,[412] an embedding.

A version for a given diffeo ϕ.
Let $\phi \in$Diff(M) with
(i) only finitely many periodic points of period less than or equal to $2m$
(ii) At periodic points of period k thje eigenvalues of ϕ^k are all distinct.
Then, $\Phi_{(\phi,y)}$ is embedding.

37.4 Obvious continuity properties

$\mathcal{F} : y \to \Phi_{(\phi,y)}$ is continuous.

37.5 Obvious openness

Let $K \subset M$ be compact. The set of y such that $\Phi_{(\phi,y)}$ is immersive[413] (or injectively so) on K is open in $C^2(M,\mathbb{R})$).

37.6 Denseness proof

Given $y \in C^2(M,\mathbb{R})$, in its any nbh is y' such that $\Phi_{(\phi,y')}$ is an embedding of M.

Note first that **37.5** implies that if y is immersive, then y' in its sufficiently small nbh are all immersive. That is, (injective) immersiveness is not destroyed by perturbation.

(i) Periodic orbits of small periods make $\Phi_{(\phi,y)}$ degenerate. However, if there is only finitely many undesirable periodic orbits we can kill them with arbitrarily small perturbation of y. This is not enough; we must maintain y to be immersive at all the periodic points. $D(y(\phi^s)) = DyD(\phi^s)$ must have rank = dim M. This is possible only if $D(\phi^s)$ has this property.

(ii) Make Φ immersive. M is separated into a finite number of parts, so that the map is basically $\mathbb{R}^{2m}$ into $\mathbb{R}^{2m+1}$. (chart to chart). Perturb y on each patch.

(iii) Make Φ embedding on orbital segments. Let us call $o = \{x, \phi x, \cdots, \phi^{2m}x\}$ the

[412] open dense

[413] $DF : T_pM \to T_{F(p)}N$ is injective (F need not be). That is, rank$DF = \dim M$.

orbital segment of x. If there is another orbital segment o' of x' that overlaps with o, it could make period $4m$ orbits. Thus, $o_2 = \{x, \phi x, \cdots, \phi^{4m}x\}$ and make all the points of the orbital segments are separated by a nbh X_x of x such that $\phi^j(X_x)$ $(j = 0, 1, \cdots 4m)$ are disjoint. Next, Since M is compact, choose a finite cover that can separate all the orbital segments. (iv) injective immersion on M.

38 Peixoto's theorem

We consider compact 2-differentiable mfd M. We introduce a C^1-topology in $\mathcal{X}(M)$. As you read from the Wikipedia article of Peixoto[414] Peixoto's theorem was a springboard for the study of dynamical systems, Smale was interested in Peixoto's work, and defined the Morse-Smale system. Then to cover more generic dynamics, he introduced the horseshoe dynamical system and then the Axiom A system.

In this section, we taste the original proof of Peixoto's theorem.

38.1 ε-homeomorphism

A homeomorphism that does not move points more than ε is called an ε-homeomorphism.

[414] Usually, M M Peixoto, "Structural stability on two-dimensional manifolds" Topology, 1 101 (1962) is cited, but a large chunk of the proof is in M C Peixoto and M M Peixoto, "Structural stability in the plane with enlarged boundary conditions" Ann Acad Bras Sci 31 135 (1959) [MCP (1921-1960, the first Brazilian woman to receive a doctorate in mathematics) is his wife]. Read Wikipedia M M Peixoto (1920-):
"Once, while talking with his mentor, Solomon Lefschetz, Mauricio Peixoto commented that no one cared about structural stability of dynamical systems and that was the main problem in working with it. But to Peixoto's surprise Lefschetz's answer was no less than "No Mauricio, this is no trouble, this is your luck. Try to work as hard and as fast as you can on this subject because the day will come when you will not understand a single word of what they will be saying about structural stability; this happened to me in topology." Lefschetz's support was very important to Peixoto at the time. In 1957, Peixoto went to research the subject with Lefschetz at the Princeton University, where he spent uncountable hours talking to the Russian professor about Mathematics and other subjects. Despite of the great age difference (Peixoto was 36 years old and Lefschetz 73), they became good friends.

With Lefschetz incentive, Peixoto wrote his first paper on structural stability, that would be later published on the Annals of Mathematics, of which Lefschetz was editor. In 1958, they went to the International Mathematical Congress, in Edinburgh, Scotland, where Lefschetz introduced Peixoto to the Russian mathematician Lev Pontryagin, whose work on dynamical systems was used by Peixoto as a basis for his studies. Pontryagin, though, showed no interest whatsoever in Peixoto's work.

Back to Princeton, Peixoto met Steve Smale, the mathematician that would later become a reference in dynamical systems. Smale was interested in Peixoto's work and realized he could extend his own based on it. Their contact intensified and, when Peixoto came back to Brazil, the American mathematician spent six months at the Instituto Nacional de Matemática Pura e Aplicada (Institute of Pure and Applied Mathematics or IMPA) at Rio de Janeiro. Through Smale, Peixoto would meet the French mathematician René Thom, who would help Peixoto to formulate his theorem, that was finalized during Thom's visit to IMPA."

38.2 Structural stable vector field

$X \in \mathcal{X}(M)$ is structurally stable, if there is a nbh U of X such that any $T \in U$ is homeomorphic to X.

The original definition by Pontryagin and Andronov required ε-homeomorphy instead of the simple homeomorphy. Peixoto demonstrated that homeomorphy implies ε-homeomorphy at least for $\mathcal{X}(M)$ with M being 2-mfd.

The 'smallness of perturbation is usually in C^1-topology, since C^0-small perturbations are "too strong and may destroy any singularity or closed orbit."

38.3 Peixoto's theorem

The *Topology* paper contains two theorems:
THEOREM 1. In order that the vector field X (at least C^1) be structurally stable on M it is necessary and sufficient that the following conditions be satisfied:
(1) there is only a finite number of singularities, all hyperbolic;
(2) the α and ω-limit sets of every trajectory can only be singularities or closed orbits;
(3) no trajectory connects saddle points;
(4) there is only a finite number of closed orbits, all hyperbolic.

THEOREM 2. The set Σ of all structurally stable systems is open and dense in $\mathcal{X}(M)$.

Here, the openness can be shown in any dimension, so the 2D specialty is the denseness.

38.4 How about higher dimensions?

Does something like **38.3** hold for $d \geq 3$? The Morse-Smale system **38.5** was introduced to consider this problem.
Theorem[415] Morse-Smale systems are dense in Diff(M) for any d.
Theorem[416] For Diff(M) for compact M the Morse-Smale systems is structurally stable.

However, there are structurally stable non-MS systems, the converse is not true.

Incidentally,
Theorem[417] In Diff$^r(M)$ structurally stable systems are C^0-dense.

[415] Palis, Topology 8 385 (1969).
[416] Palis-Smale
[417] Shub

38.5 Morse-Smale systems[418]

A dynamical system is called a Morse-Smale system, if
(MS1) Ω is finite (so Ω consists of periodic orbits).[419]
(MS2) Periodic points are all hyperbolic.
(MS3) If $p, q \in \Omega$, then $W^s(p) \pitchfork W^u(q)$.[420]

Gradient dynamical systems are Morse-Smale.[421]

With this terminology
Peixoto's theorem A compact 2-mfd C^1-vector field is structural stable iff it is Morse-Smale.

38.6 Outline of the proof of 38.3

$\Leftarrow$
In this proof actually an ε-homeomorphism is constructed to relate perturbed systems. Thus, Theorem 1 implies that structural stability implies ε-structural stability.

To show vector fields allowing the flow satisfying (1)-(4) are structurally stable, we must construct homeomorphisms between the original and the perturbed fields. This is rather tedious and is shown in the AABS paper quoted above.

$\Rightarrow$
Starting from any vector field X, with a series of approximation lemmas it is shown that X is approximated by Y (i.e., there is a vector field Y in any nbh of X) that satisfies (1)-(4).

Thus, if X is structurally stable, then it must be homeomorphic to its perturbed version, but a perturbed version must satisfy (1)-(4), so X itself must have satisfied (1)-(4), because these properties are homeomorphism invariant.

Theorem 2 is shown almost simultaneously: Structural stable vector fields make an open set (in any dimension, actually), but the proof of $\Rightarrow$ implies such vector fields are dense (exists in any neighborhood of any vector field on a compact 2-manifold).

[418] M Shub: http://www.scholarpedia.org/article/Morse-Smale_systems.

[419] Thus, we may say $\cup_j W^s(P_j) = M$, $\cup_j W^u(P_j) = M$.

[420] Because of (MS1) transversality in this case means: no saddle connection nor homoclinic tangency. That is, $W^s(p) \pitchfork W^u(q)$ means they are not in contact except at separatrices.

[421] See also K. R. Meyer Energy Functions for Morse Smale Systems, Am J Math 90 1031 (1968).

38.7 Outline of the proof of $\Leftarrow$[422]

To explain this several definitions must be introduced:
Definition. A region on which the vector field is homeomorphic to one of the following (i)-(iii) in Fig. 38.1 is called 'parallel region.' Note that parallel regions are arcwisely connected.

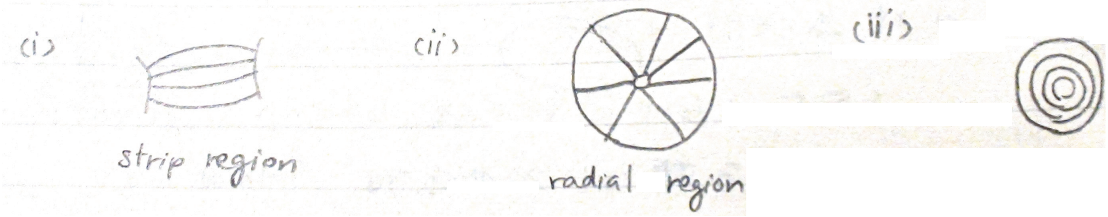


Figure 38.1: Parallel regions

Definition. A trajectory is called an ordinary trajectory, if it is in (i) the strip region. Other trajectories are called separatrices; limit cycles, homo and heteroclinic orbits, etc., are the examples.
Definition. A region $M \setminus \{\text{separatrices}\}$ is called the canonical region. Perhaps it is better to define separatrix set as the complement of the totality of parallel regions in M.

(i) Canonical regions are classified into five types (I)-(V).
(ii) On each canonical region, flows are structurally stable. This is shown by constructing homeomorphisms.
(iii) From these homeos a homeo for M is constructed (which is actually ε-homeo).

38.8 There are five types of canonical regions

Trajectories in a single canonical region R share α and ω limits (Fig. 38.2). Note that ∂R consists of separatrices, so when a band of parallel trajectories is extended none of the trajectories inside the band cannot reach saddles. Therefore, all the members of the bands must share the same destination.

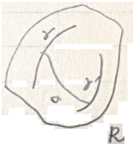


Figure 38.2: Connecting two trajectories on R. The connecting curve σ must always be in R, so it is singly connected.

Thus, R has one α and one ω limit set on its boundary. Each boundary

[422] In MCP+MMP

connecting these limit points cannot have more than one saddle because of **38.3**(3), so we have only the following 5 types of Rs (Fig. 38.3):

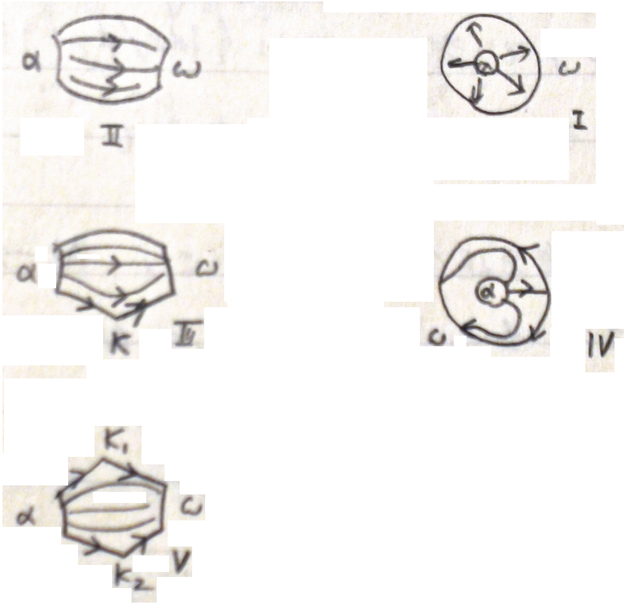


Figure 38.3: Five types of canonical regions

Their stability against perturbation is obvious.

38.9 Structural stability of separatrix set

Let X (separatrix set Λ) and $\tilde{X}$ (separatrix set $\tilde{\Lambda}$) are vector fields within δ in the C^1-norm. Then, Λ and $\tilde{\Lambda}$ are one-to-one correspondent:
(i) Corresponding separatrices are of the same type (homeomorphic),
(ii) If a subset of Λ is the boundary of a canonical region R, then for the corresponding canonical region $\tilde{R}$ its boundary is a homeomorphic subset of $\tilde{\Lambda}$.

This follows from **38.3**(1)-(4) and the usual discussion of the stability of hyperbolic structures (+, if you wish, degree-theoretical arguments).

38.10 Strategy to prove the structural stability of *X*

To show the structural stability we must construct a homeomorphism between X and $\tilde{X}$. The strategy is as follows. Remove small discs D_i containing fixed points and tubular neighborhoods S_i of limit cycles from M to make N.
(i) For each canonical region R on $R^* = R \cap N$ a homeomorphism is constructed.
(ii) Homeomorphisms are constructed on D_i and S_i.

(iii) Connect all these homeo to make a homeo for X.

Notice that both X and $\tilde{X}$ live on the same manifold M and All these local sets, Rs, Ds and Ss for X are perturbed (deformed a bit) to become the counterparts of $\tilde{X}$. Since D and S are built around hyperbolic limit sets, they are stable under perturbation and, e.g., for a D, there is its counterpart $\tilde{D}$. Thus, you can imagine that the structures due to X and their perturbed counterparts are overlapping on M; deformations and displacements are small.

Therefore, first, local homeomorphisms are constructed on these local sets. This is **38.11** and **38.12**.

38.11 Construction of homeo on canonical regions

We wish to show (i) in **38.10**.

For type V region R^* (Fig. 38.3), we wish to do the following (Fig. 38.4). We make a ε-homeomorphism for R^* and $\tilde{R}^*$ by constructing φ and φ' to a square.[423] The homeo we need is just $\varphi \circ \varphi'^{-1}$.

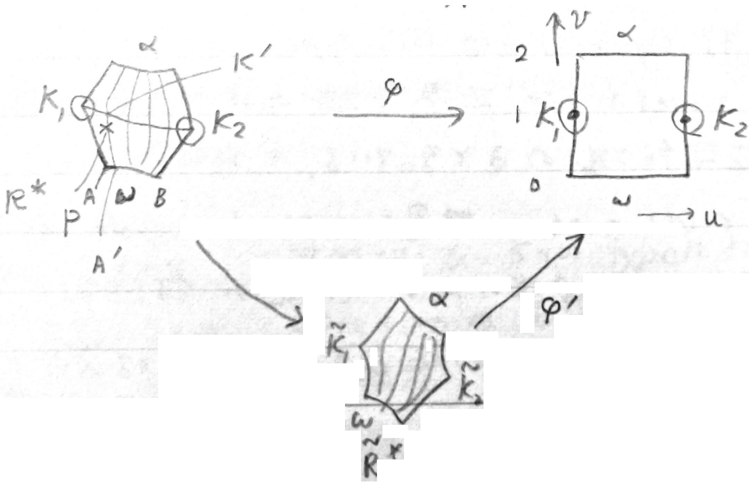


Figure 38.4: Construction of homeo on type V region R^*

In this case, the smooth parallel portion has no problem, since X and $\tilde{X}$ are close. The only (slightly) unclear situations are around the saddle points K_1 and K_2. This is clear if we note that if we take sufficiently small disk D_i around K_i, then the lengths L of the trajectories of $\tilde{X}$ in D_i can be as small as we

[423] Precisely speaking, between R and the square, homeo can be constructed, but not ε-homeo. Thus, we should say that we use the square to construct homeos, and then we must adjust them so that $\varphi \circ \varphi'^{-1}$ is an ε-homeo.

wish. Therefore, displacement of the trajectories due to perturbation must be smaller than the sum of $D - \tilde{D}$ displacement + L, so we can make this as small as we wish.

Types IV and III are considered as 'degenerate' version of type V as illustrated in Fig. 38.5.[424]

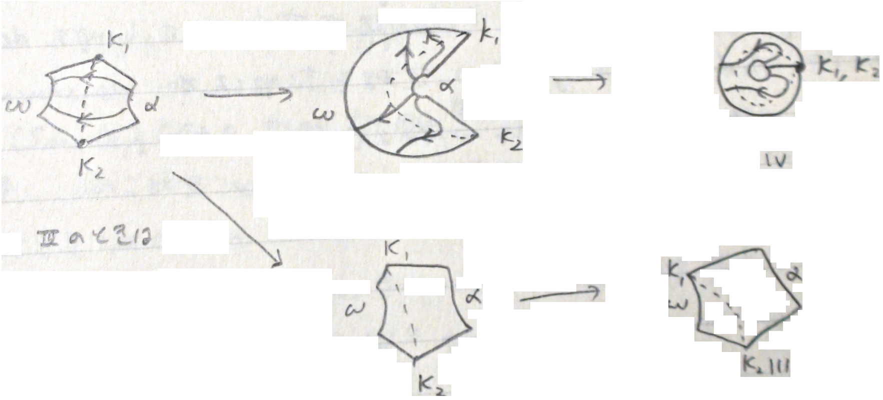


Figure 38.5: Construction of homeo on type IV region and III (bottom) is reduced to that of type V.

Type II should be understood as a portion of V.

For Type I, instead of a square in the type V case, we make maps to a circular annulus.

All these maps are ε-homeo.

38.12 Construction of homeo on nbh of separatrices

Since saddle points are not in the nonwandering set of the system in our case, we have only to consider fixed points for D_i. The situation is exactly the case of Type I.

For limit cycles we can introduce 'polar coordinates' specified by the crossing position on σ and the distance along the curve from σ in the tubular nbh as illustrated in Fig. 38.6. Using such coordinates, we can construct a homeo.

38.13 Completion of homeomorphism

We construct a global homeo by patching the homeos on R^*, D_i and S_i. To do so we must glue these along the boundaries, so we need appropriate local deformations. These deformations can be homeos, so we can adjust all the homeos

[424] If you wish to be precise, for IV you introduce a coordinate system with a periodic boundary condition along the dotted curve direction, and glue side edges (containing Ks) continuously.

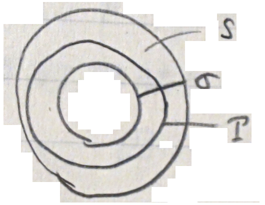


Figure 38.6: Coordinate system near a limit cycle

with these local deformation homeos and patch them continuously over M to construct a global homeo.

Historically, in this way, Peixoto's theorem restricted on a disk was proved first.[425]

38.14 Outline of the proof of ⇒

As is noted in **38.6**, ⇒ is proved by a chain of approximation lemmas.

Start from any X.

(A) X may be approximated by (= X has in its sufficiently small nbh) Y_1 that satisfies **38.3**(1), i.e., with finitely many hyperbolic fixed points. Note that we do not pay any attention to other members of the non-wandering set Ω of X. Thus, te resultant Y_1 may have a 'horrible' invariant set μ.[426]

Definition. A closed invariant set whose genuine subset is not a closed invariant set is called a minimal set. A minimal set that is neither a fixed point nor a limit cycle is called a nontrivial minimal set. [Note its closedness.]

(B) Y_1 may be approximated by Y_1' without nontrivial minimal sets. Here, minimal sets are converted to saddle connections and periodic orbits.

(C) Y_1' may be approximated by Y_2 with fixed points (sink, source and saddle), closed orbits and saddle connections as nonwandering sets.

(D) Y_2 may be approximated by Y_3 satisfying **38.3**(1)-(3): We can sever all the saddle connections, but perhaps new periodic trajectories are created. Note that periodic

[425] M M Peixoto, On structural stability, Ann Math 69, 199 (1959).

[426] You might say on 2-mfd not much complicated dynamics is possible. We know chaos requires three dimensions, and putting trajectories in 2-mfd is like putting noodles on a tray. Indeed on T^2 there is no flow with positive KS entropy. However, Denjoy constructed a highly non trivial invariant set whose 'cross section' is related to a Cantor set. If the genus g of 2-mfd increases (g is the number of holes: $g(S^2) = 0$, $g(T^2) = 1$, etc. Adding a handle increases g by one), then M is riddled with holes, so perhaps trajectories could come back through holes and mingle with the staying trajectories in a nontrivial fashion. Thus, it is safe to assume such μ exists in the flow due to X (and so in Y_1).

trajectories need not be limit cycles.
(E) Y_3 may be approximated by Y_4 satisfying **38.3**(1)-(4). This step uses Whitney's embedding theorem and Weierstrass' polynomial approximation of continuous functions.

38.15 Any vector field may be approximated by a vector field with finitely many hyperbolic fixed points

We can make all the singularities as simple by arbitrarily small perturbations. Simple singularities are isolated (and since M is compact, there are only finitely many of them), so we may handle them separately. A nonhyperbolic fixed point can be converted by sufficiently small deformation to a hyperbolic fixed point. Thus, X may be approximated by a vector field Y_1 satisfying **38.3**(1).

38.16 Perturbatively nontrivial minimal sets can be killed: outline

Suppose Y_1 has a nontrivial minimal set μ. Since there are only finitely many saddle points, we consider the following two cases (recall that usually dynamical systems are defined on the time range $(-\infty, +\infty)$)
(A) There is no trajectory connecting μ and a saddle.[427]
(B) There is a trajectory connecting μ and a saddle.

For case (A) we can show (**38.17**) that for any $p \in \mu$ there is a coordinate nbh such that all the trajectories leaving it return to it,[428] and the trajectory going through p can be converted into a periodic orbit. For case (B) such a trajectory can be converted into a saddle connection (**38.18**).

After these perturbations, if nontrivial minimal sets still remain, we repeat the procedure. It will be shown (**38.19**) that only finitely many repetition is required, so Y_1 can be perturbed with an arbitrarily small perturbation into Y_1' without nontrivial minimal set.

38.17 Case A: no saddle connection

For (A) in **38.16** Y_1 has a closed orbit passing through μ. This closed orbit does not bound any cell (see Fig. 38.7) [Closing lemma].

[427] As we will see, this happens only if $M = T^2$ pf a Klein bottle, so the system actually has no singularity at all.

[428] Since μ is non trivial, there is a trajectory returning to any nbh of p infinite times.

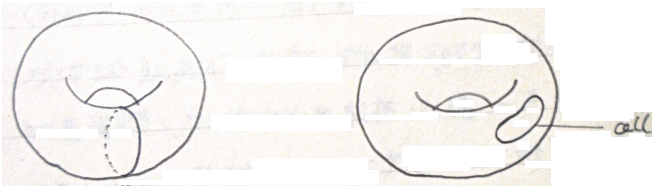


Figure 38.7: Two kinds of closed orbits on 2-mfd.

To show this take a local coordinate square 'abcd' around a point P in μ (Fig. 38.8).

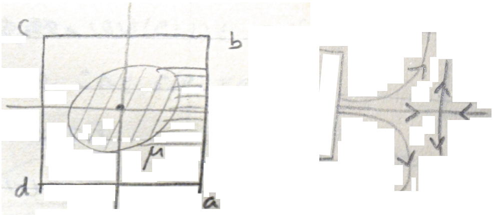


Figure 38.8: Local coordinates for μ; Right illustrates the impossible case for (A)

Since μ is minimal, at least one trajectory leaving from edge ab must come back to μ through edge cd.[429] The boundary trajectory between the trajectories coming back to μ and those not coming back must go to a saddle point (See Fig. 38.8 Right).[430] However, our assumption is that such μ-saddle connection does not exist. Therefore, all the trajectories leaving μ must come back to μ (infinitely many times to different points on edge cd), since μ is nontrivial.

Also these orbits cannot encircle a cell (see Fig. 38.7 Right); μ is a closed set, so its boundary cannot be a cycle; otherwise, since μ is a closed set, this boundary cycle belongs to μ, destroying its minimal nature.

Since trajectories do not cross each other, the ordering of the trajectories is preserved (or reversed). Consider the perturbation illustrated in Fig. 38.9. We can create a limit cycle going through P.

Needless to say, this may not totally erase μ; some potion(s) may survive as nontrivial minimal sets. Thus, we need to repeat the procedure.

[429] The local square around P is so chosen that the trajectories foliate the square. This is possible, because there must be a trajectory coming back to any nbh of P infinitely many times, and M is a 2-mfd, so near P the returning portion of the trajectory must be almost parallel.

[430] A formal proof is lengthy. Lemma 3 in the original.

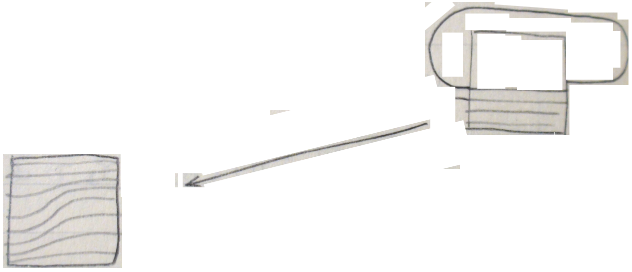


Figure 38.9: Closing lemma

38.18 Case B: when μ-saddle connection exists

For (B) in **38.16** the μ-saddle point connection can be converted to a saddle-saddle connection.

Actually, we have only the case of Fig. 38.10 Left. Suppose there is a μ-saddle connection in the future direction (the right-side edge situation in Fig. 38.10 Left). Then, above and below the connection, the trajectories have different fate. However, these trajectories come back to this coordinate square R. Then, reverse the time, we must say there must be a saddle on the left side just as illustrated in the figure. This saddle γ_1 cannot be the same saddle as γ_2., because if so the trajectories must cross the unstable manifold of γ_1.

We may assume that $\omega(\gamma_1) = \alpha(\gamma_2) = p$. If not, we have a situation like Fig. 38.10 Right, and the curly bracketed portion may be subjected to a surgery as performed in case (A).

If this surgery fails to connect γ_1 and γ_2, then the boundary between the curly bracketed portion and its outside must go to a saddle point as illustrated in Fig. 38.8 Right. Thus, we get the situation as Fig. 38.10 Left. When γ_1 and γ_2 are sufficiently close, we can short-circuit them with a small perturbation to create a saddle connection.

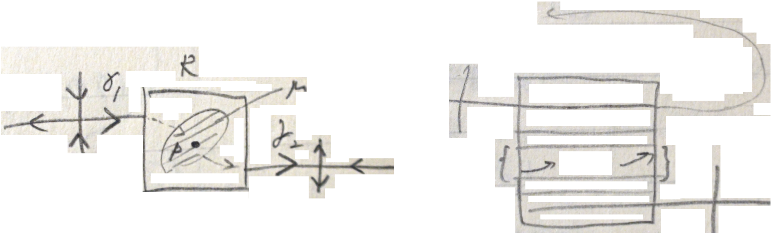


Figure 38.10: Saddle connection case

38.19 Finite repetition of surgeries totally kill nontrivial minimal sets

Since the total number of saddle point in Y_1 is finite, so the number of surgeries in (B) must be finite.

If (A) does not end with finitely many surgeries, we have very many closed orbits. Since M is a 2-mfd, with a finitely many cuts M can be expanded into a 2-disk.[431] Since trajectories cannot cross closed orbits, all the orbits must be on this disk. Suppose we still have some μ on this disk, since all the saddle points have been used up to make saddle connections, it produces an orbit encircling a cell, an impossibility as already discussed in **38.17**.

Thus, the procedure explained above ends with finite repetitions and an arbitrarily small perturbation can convert X into a vector field with finitely many hyperbolic fixed points without any nontrivial minimal set (but with too many periodic trajectories and possibly bands of periodic trajectories (as centers)).

38.20 Almost homoclinic orbits are converted to homoclinic orbits

(C) has almost been demonstrated, but there can be a situation where an almost homoclinic situation occurs for a saddle (Fig. 38.11), because the situation does not produce a nontrivial minimal set.

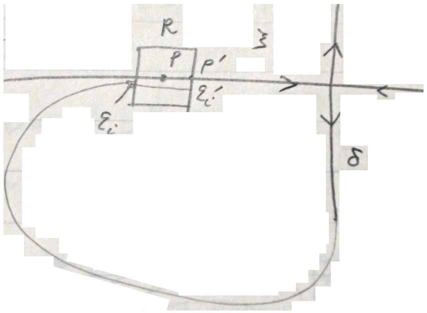


Figure 38.11: With a little perturbation a homoclinic orbit may be created.

If we look at R, since there is trajectory coming back to any nbh of P, so we can convert this into a homoclinic trajectory. Since the number of saddle points is finite, just as discussed in **38.19**, we have only to repeat such a procedure finite times and the perturbation to Y_2 is complete.

38.21 Perturbation can remove all the saddle connections

Graphically, we can exhibit the situation as in Fig. 38.12 [Actually such a diagram is called a graph].

[431] Genus g 2-mfd may be cut open to a disk (or a polygon) with $2g$ cuts, just as T^2 is converted to a square by 2 cuts.

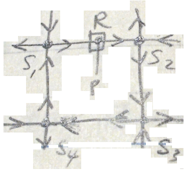


Figure 38.12: Graphic representation of saddle connections

There are two situations:
(1) S_1S_2 is not a part of a large graph.
(2) S_1S_2 is a part of a large graph

For (1) the following perturbation does not produce a saddle connection anew.

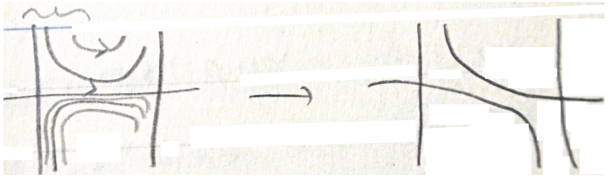


Figure 38.13: Surgery of saddle connections

However, in case (2) a new saddle connection might be formed (Fig. 38.14), but, in this case, if perturbed further n (pushed down further in the figure), the saddle connection would be reconnected to a periodic orbit. Since there are only finitely many saddle points, this can be done by arbitrarily small perturbation, so we may virtually ignore such cases.

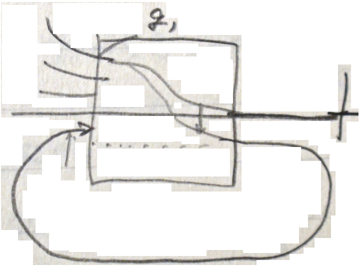


Figure 38.14: Possible emergence of a new saddle connection, but this can be killed by the small perturbation indicated by an arrow

In any case, we can repeat the above procedure finite times. All the saddle connections will be gone. Thus Y_3 has been produced from X. That is, except for the condition for periodic trajectories we are done.

38.22 Periodic orbits are made hyperbolic by perturbation

To make periodic orbits hyperbolic, we apply the perturbation illustrated in Fig. 38.15 to box R:

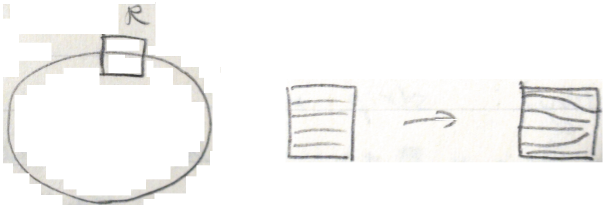


Figure 38.15: Make orbits hyperbolic

Also marginal periodic orbits can be understood as a degenerate version of two hyperbolic orbits. Thus, We can make all the separatrices hyperbolic (the stabilized Y_3). However, still there can be infinitely many closed orbits (like a center).

Peixoto embedded the stabilized Y_3 into $\mathbb{R}^5$ using Whitney's embedding theorem, and then made a polynomial approximation of the vector field (using Weierstrass' theorem). That is, he made an analytic vector field, which has only isolated periodic trajectories and periodic bands. Then, he constructed a vector field homeomorphic to this polynomial approximation on M. The resultant vector field is no more analytic, but the topological features are all preserved. Thus, the vector field satisfies (1) and (3) in **38.3** and isolated periodic trajectories/bands. However, nontrivial minimal sets μ may reappear by the approximation procedure.

If this further satisfies (2), that is, no new μ shows up, then there are two cases:
(A) All the orbits are closed.
(B) Otherwise.

For (A) if M is not a result of gluing a square (i.e., T^2 or Klein's bottle), then we have something like Fig. 38.16.

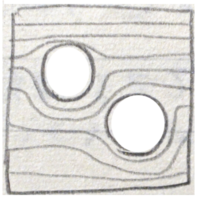


Figure 38.16: Expanded M should not have a hole; the holes mean these peripheries are glued to make a handle.

This implies there are saddle points, contradicting the assumption that there are only closed orbits. If T^2, for example, we can apply 'bunching' as exhibited in Fig. 38.15 to make a limit cycle.

For (B) even if we have a band of closed orbits (no returning orbits), the boundary of the band (black dots in Fig. 38.17) goes to a saddle (or comes from a saddle) just as in Fig. 38.8 Right, so this contradicts **38.3**(3), because Y_3 has already been constructed to satisfy this and polynomial approximation does not alter this.[432]

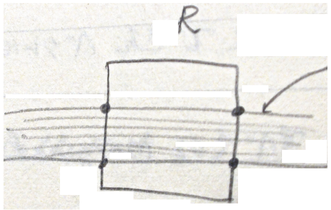


Figure 38.17: Killing the bands

If (2) is not satisfied, that is, there is nontrivial minimal set μ, then repeat the above argument (**38.16**). If, after this repetition, the number of orbits is finite, we are done. If not, we repeat the polynomial approximation, because polynomial fields allow only isolated or band of periodic orbits. Since the polynomial approximation maintains all the simple closed orbits, the newly added simple trajectories during the repetitions are maintained by each polynomial procedure step. However, as discussed in **38.19**, we can repeat the procedure only finitely many times, and μ will not show up eventually. We are done.

[432] need polynomial field check

39 Axiom A systems

39.1 Axiom A system

f is an axiom A system, iff $\Omega(f)$ is hyperbolic and

$$\Omega(f) = \overline{\{x : \text{ periodic point of } f\}} \tag{39.1}$$

Here, $\Omega(f)$ is te totality of the non-wandering set of f.

39.2 Hyperbolic set of f

Let X be a compact smooth manifold, $f : X \to X$ a diffeomorphism, and $Df : TM \to TM$ the differential of f. An f-invariant subset Λ of X (i.e., $f(\Lambda) = \Lambda$) is said to be hyperbolic, if the restriction to Λ of the tangent bundle of X (i.e., $T_\Lambda X$) admits a splitting into a Whitney sum of two Df-invariant subbundles, called the stable bundle and the unstable bundle (denoted E_s and E_u. With respect to some Riemannian metric on X, the restriction of Df to E_s must be a contraction and the restriction of Df to E_u must be an expansion.

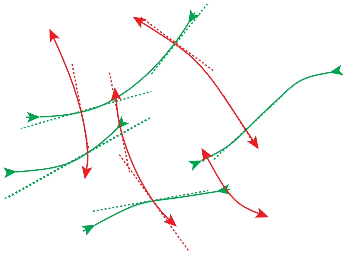


Figure 39.1: Hyperbolic set: red curves are unstable mfds and green curves are stable mfds. Dashed straight lines denote E^u and E^s.

39.3 How a small open set spreads

Let $f \in C^r(M, M)$, p its periodic point and $U \subset M$ is open.

(a)

$$U \cap W^s(p) \neq \emptyset \;\Rightarrow\; \overline{\cup_{m \in \mathbb{N}^+} f^m(U)} \supset W^u(p). \tag{39.2}$$

[If p is a sink $W^u(p) = \emptyset$, so the statement is trivially true.]

(b) Let $q \neq p$ be a periodic point, and x be a heteroclinic point of q and p: $x \in W^u(p) \cap W^s(q)$

$$U \cap W^s(p) \neq \emptyset,\; W^u(p) \,\overline{\pitchfork}_x W^s(q) \;\Rightarrow\; (\cup_{m \in \mathbb{N}^+} f^m(U)) \cap W^u(q) \neq \emptyset. \tag{39.3}$$

See the following Fig. 39.2.

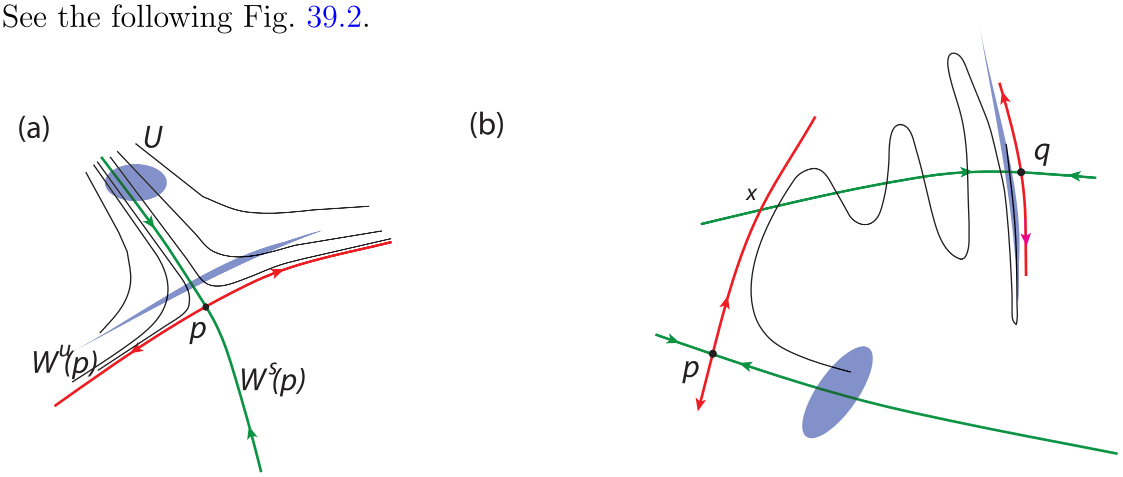


Figure 39.2: How small open sets on W^s spreads

39.4 Chain of periodic orbits

Let $f \in C^r(M, M)$, and p_i $(i = 0, \cdots, n)$ are hyperbolic periodic orbits $(p_0 = p_n)$ such that $x \in W^u(p_i) \cap W^s(p_{i+1})$ implies $W^u(p_i) \pitchfork_x W^s(p_{i+1})$. Then, x is non-wandering (i.e., $x \in \Omega$).

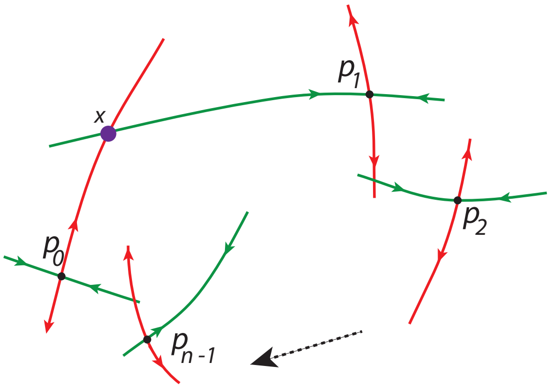


Figure 39.3: Chain of periodic orbits

39.5 Stability manifold theorem[433]

$f \in \mathrm{Diff}^r(X)$ and Λ is the hyperbolic set of f. Then, for small $\varepsilon > 0$

[433] Smale BAMS 73 747 (1967) Th(7.3) p781.

(a) Local stable and unstable mfd of x ($\in \Lambda$) is a C^r-disk in Λ and

$$T_xW^s_\varepsilon(x) = E^s_x,\ T_xW^u_\varepsilon(x) = E^u_x. \tag{39.4}$$

(b) On these local submanifolds f is just expanding or contracting as usual.
(c) $W^s_\varepsilon(x)$ and $W^u_\varepsilon(x)$ depend on x continuously.

39.6 Existence of canonical coordinate[434]

Let f satisfy Axiom A. Then, for any small ε (> 0) there is $\delta > 0$ such that for $x, y \in \Omega(f)$ and for $d(x, y) \le \delta$,

$$W^s_\varepsilon(x) \cap W^u_\varepsilon(y) = \{[x, y]\} \subset \Omega(f), \tag{39.5}$$

and the crossing point [,] depends on x and y continuously.

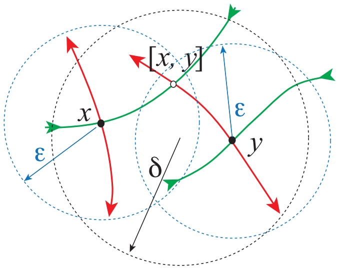


Figure 39.4: If we take $x, y \in \Lambda$ close enough (within δ).

The point is that for any ε we can find x and y in $\Omega(f)$ connected heteroclinically. Then, this is based on the following two lemmas (basically **39.3**):

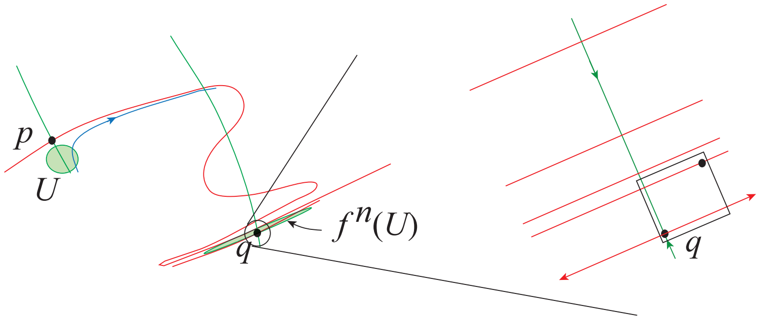


Figure 39.5: Lemma 1; Right: near q expanded, we can make a local cartesian coordinate system in the box.

[434] Bowen 3.3.

Lemma 1: Let p and q be periodic orbits and $W^u(p) \pitchfork W^s(q)$. Take $U \cap W^s(p) \neq \emptyset$. Then, $\cup_{m>0} f^m(U) \cap W^s(q) \neq \emptyset$. See Fig. 39.5.
Lemma 2: Let $p_0, p_1, \cdots, p_n$ ($p_0 = p_n$) be hyperbolic periodic points and $W^u(p_i) \pitchfork_{x_i} W^s(p_{i+1})$. Then, $x_i \in \Omega(f)$.

This can be shown by applying Lemma 1 to the successive periodic points. See Fig. 39.6:

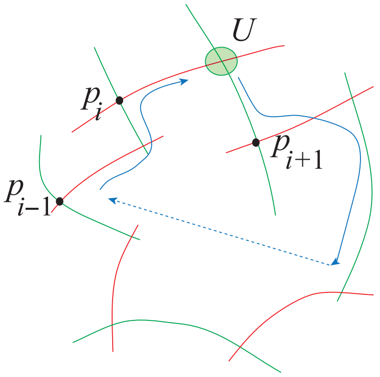


Figure 39.6: Lemma 2

The demonstration of the canonical coordinates goes as follows:
Periodic orbits are dense in $\Omega(f)$. If we take ε small enough, then $[x, y]$ becomes unique. If x and y are periodic, then Lemma 2 implies $[x, y] \in \Omega(f)$. Even if these are not periodic, they are dense in $\Omega(f)$, $[x, y]$ must be in $\Omega(f)$, because $\Omega(f)$ is a closed set.

39.7 Expansivity of hyperbolic set[435]

Let Λ be the hyperbolic set for f. Then there is $\varepsilon > 0$ such that $d(f^k(x), f^k(y)) > \varepsilon$ for some $k \in \mathbb{Z}$ for $x \in \Lambda$ and $y \in M$ $(x \neq y)$.

If there is no such ε, then $y \in W^s_\varepsilon(x) \cup W^u_\varepsilon(x)$, but this implies $x = y$, a contradiction.

39.8 Spectral decomposition theorem[436]

$\Omega(f)$ has the following structure:

[435] Bowen L3.4.
[436] Bowen Th 3.5

$$\Omega(f) = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_s \quad (\Omega_i \cap \Omega_j = \emptyset \text{ for } i \neq j), \tag{39.6}$$

where
(a) $f(\Omega_i) = \Omega_i$ and $f|_{\Omega_i}$ is topologically transitive.
(b) Each Ω_i (called a basic set) has the following partition: $\Omega_i = X_{i,1} \vee X_{i,2} \vee \cdots \vee X_{i,n_i}$ with $f(X_{i,k}) = f_{i,k+1(\text{mod } n_i)}$ and $f^{n_i}|_{X_{i,j}}$ is topologically mixing.

39.9 Strategy to show spectral decomposition theorem

Let p be a periodic point. Make

$$X_p = \overline{W^u(p) \cap \Omega}. \tag{39.7}$$

$f(X_p) = X_{f(p)}$. If n is the period of p, $f^n(X_p) = X_p$ and this must be topologically mixing on X_p. We collect X_p and its images $f^k(X_p)$ to make Ω_1.

Repeat this until this procedure exhaust Ω to realize (39.6).

39.10 Shadowing = tracing

We say the point sequence $\boldsymbol{x} = \{x_i\}$ is β-shadowed by a trajectory starting from x (by x, for short), if

$$d(f^n x, x_n) \leq \beta \tag{39.8}$$

for $i \in [a, b]$ $(a, b \in \mathbb{Z})$.

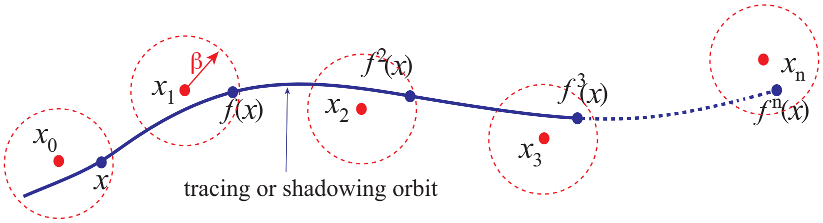


Figure 39.7: β-shadowing

39.11 α-pseudo orbit

$\boldsymbol{x}$ is called an α-pseudo orbit, if

$$d(fx_n, x_{n+1}) < \alpha. \tag{39.9}$$

39.12 Shadowing of pseudo orbits

For any $\beta > 0$, there is a $\alpha > 0$ such that any α-pseudo orbit $\boldsymbol{x}$ is β-shadowed by some point x.

If we choose α sufficiently small, then any α-pseudo orbit $\boldsymbol{y}$ satisfies

$$d(f^n y_0, y_n) < \delta/2, \tag{39.10}$$

where δ is the same δ as in **39.6**.

39.13 Strategy to show shadowability of pseudo orbits

(1) Note first that for a finite time span N an α-pseudo orbit $\boldsymbol{y} = \{y_i\}_{i=0}^M$, if α is sufficiently small, can satisfy, for all $i \in [0, M]$,

$$d(f^i y_0, y_i) < \delta/2. \tag{39.11}$$

(2) Let us make an α-pseudo orbit $\boldsymbol{x} = \{x_i\}_{i=0}^{rM}$, connecting such $\boldsymbol{y}$.
(3) The shadowing orbit $\boldsymbol{x}'$ is chosen as follows:
Starting from $x_0 = x'_0$, we recursively apply **39.6** as follows

$$x'_{(k+1)M} = [x_{(k+1)M}, f^M(x'_k)]. \tag{39.12}$$

See Fig. 39.8.

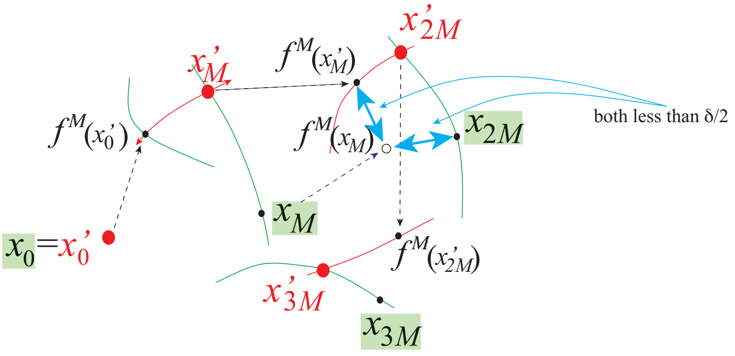


Figure 39.8: Constructing a orbit shadowing the α-pseudo orbit (pale green)

This way, we can find x_{rM}. The basic idea is to find the point we should aim at. That is x_{rM}.

Now, choose the initial condition such that $x = f^{-rM}(x_{rM})$, that is, from there we reach x_{rM}. Although we must carefully show that this orbit indeed δ-shadow th α-pseudo orbit, intuitively it should be clear by construction.

39.14 Closing $\boldsymbol{\alpha}$-pseudo periodic orbit

Let $x \in \Omega$ and $d(f^n(x), x) < \alpha$. For any $\beta > 0$ we can choose α so that there is a β-shadowing periodic orbit for this not-closed orbit such that $f^n(x') = x'$.

Concatenating $\{x, f(x), \cdots, f^{n-1}(x)\}$, we can make an infinite α-pseudo orbit which can be β-shadowed by $\boldsymbol{y}$ for any $\beta >$. Then,

$$\begin{aligned} d[f^i(y), f^i(f^n(y))] &\leq d[f^i(y), f^i(x)] + d[f^i(x), f^i(f^n(y))] \\ &\quad \leq d[f^i(y), f^i(x)] + d[f^i(f^n(x)), f^i(f^n(y))] + d[f^i(x), f^i(f^n(x))]. \end{aligned} \tag{39.13}$$

The first two terms are less than β due to shadowing, and the last term is zero by the periodic concatenation. Thus, for any $\beta > 0$ $d[f^i(y), f^i(f^n(y))] < 2\beta$ for any $i \in \mathbb{Z}$. However, since $x, y \in \Omega$ and since Ω is expansive, this means $y = f^n(y)$ (due to the contraposition of **39.7**).

39.15 Rectangle, proper rectangle

$R \subset \Omega_s$ is a rectangle, if and only if for $x, y \in R$ $[x, y] \in R$. R is called a proper rectangle if $R = [R^\circ]$ (i.e., R is identical to the closure of its open kernel).

39.16 Rectangles are 'registered to' stable and unstable manifolds

Let $\partial^s R$ (resp., $\partial^u R$) be the boundary of rectangle R parallel to W^s (resp., W^u). Then,

$$\partial R = \partial^s R \cup \partial^u R. \tag{39.14}$$

More precisely,

$$\partial^s R = \{x \in R, x \notin \text{ int } W^u(x, R) \equiv W^u_\varepsilon(x) \cap R\}, \tag{39.15}$$

$$\partial^u R = \{x \in R, x \notin \text{ int } W^s(x, R) \equiv W^s_\varepsilon(x) \cap R\}. \tag{39.16}$$

39.17 Markov partition

A partition consisting of proper rectangles $\mathcal{R} = \{R_1, \cdots, R_m\}$ satisfying the following conditions is called a Markov partition:
(a) int $R_i \cap$ int $R_j = \emptyset$ for $i \neq j$,
(b) If $x \in$ int R_i and $f(x) \in$ int R_j,

$$\begin{aligned} fW^u(x, R_i) &\supset W^u(fx, R_j), && (39.17)\\ fW^s(x, R_i) &\subset W^s(fx, R_j). && (39.18)\end{aligned}$$

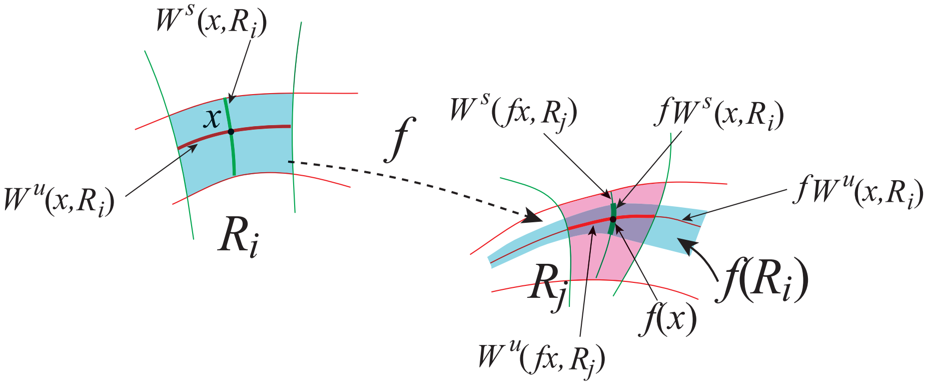


Figure 39.9: Markov partition

As seen later actually we may demand

$$\begin{aligned} f^{-1}(\partial^u R_i) &\subset \partial^u R_k, && (39.19)\\ f(\partial^s R_i) &\subset \partial^s R_l && (39.20)\end{aligned}$$

for some k and l.

39.18 Axiom A system has Markov partition

Let Ω_s be a basic set for an Axiom A diffeomorphism f. Then, Ω_s has Markov partitions $\mathcal{R}$ of arbitrarily small diameter.

An outline of the logic to demonstrate this key theorem is as follows:
(1) Cover Ω_s with a net whose mesh size is no more than γ, such that for any $x, y \in \Omega_s$ if $d(x, y) < \gamma$, $d(fx, fy) < \alpha/2$, where α is the α appearing in the pseudo orbits. Let $P = \{p_1, \cdots, p_n\}$ be the totality of the net vertices just constructed.

(2) Make the set $\Sigma(P)$ consisting of all the α-pseudo orbits using the points in P. Take $p_s \in P$ and collect all the α-pseudo orbits $\boldsymbol{q}$ starting from p_s (do not forget that we go to the negative time direction as well; the pseudo orbits are defined for all $j \in \mathbb{Z}$). For each such pseudo orbit we can choose a unique β-shadowing orbit: for any $\boldsymbol{q}$ there is a unique point $\theta(\boldsymbol{q})$ that β-shadows it. This is intuitively clear due to the expansivity of f.
(3) Let T_s be the totality of $\theta(\boldsymbol{q})$ with $q_0 = p_s \in P$. This is a rectangle (**39.15**).
(4) T_i and T_j may overlap, so choosing smaller γ to remove the overlaps. Thus obtained T_ss make $\mathcal{R}$.
(5) Finally, we demonstrate (b).

Let us try to understand why T_s's can make a nice Markov partition.

39.19 How to determine T_s

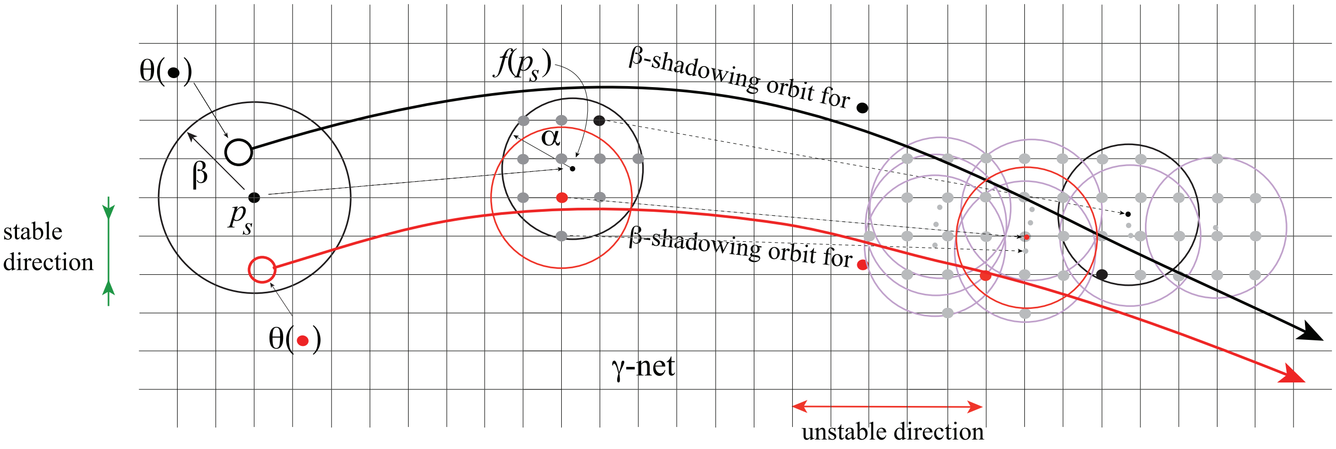


Figure 39.10: Construction of T_s. P consists of all the lattice points. Larger filled dots denote pseudo orbits; continuous curves denote shadowing curves; Open big dots give θ(pseudo orbits). Dotted lines denote f.

The illustration here assumes the forward time evolution, but as noted explicitly in **39.18** we must also consider the backward time evolution. The explanation below is for forward time evolution.

(1) Construct pseudo orbits. As may be guessed from Fig. 39.10 more and more points in P that can make pseudo orbits spread in the unstable direction (expanding direction) as gray dots indicate.

(2) The β-shadowing orbits for these pseudo orbits make a bundle whose width in the stable direction is basically determined by α; without this 'error' the width converges to zero. In the unstable direction the width increases exponentially. However, if the bundle is translated to its initial points (that is, in terms of $\theta(\boldsymbol{q})$), since backward in time there is a severe contraction along the unstable direction, the spread of $\theta(\boldsymbol{q})$ is again determined by α; again without this 'error' the width converges to zero.

39.20 Boundary of T_s

Let us determine the boundaries transversal to the unstable direction. To do so, we evolve the system backward in time. A caricature of what happens is in Fig. 39.11.

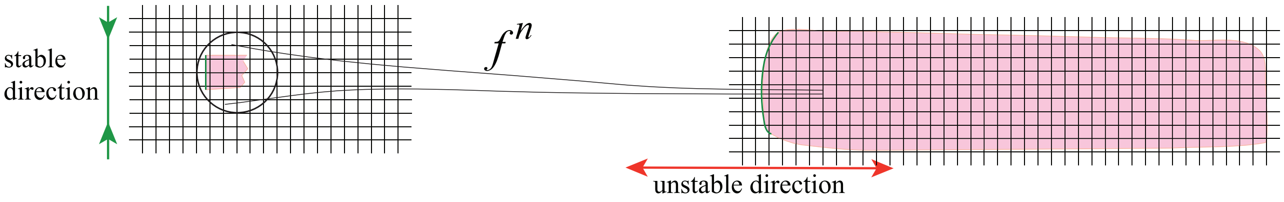


Figure 39.11: The green curve mapped backward by f^n $n \gg 1$ is (almost parallel to the stable manifold.

Imagine you come backward from the right situation to the initial conditions (i.e., the image of θ). There is a tremendous contraction along the red arrow and extremely expanded along the gree arrow, so however curvy the boundary transversal to the unstable direction is, as illustrated by the green line, the boundary becomes parallel to the stable mfd.

You can apply a parallel argument for f^{-n} to conclude that the to and the bottom boundaries are parallel to the unstable manifolds.

Therefore, T_s is bounded by stable and unstable manifolds and obviously it is a proper rectangle in the sense of **39.15**.

39.21 Construction of Markov partition

We have constructed T_ss, but they may have overlap.

As illustrated in Fig. 39.12, repartitioning the obtained T_ss into a set of smaller

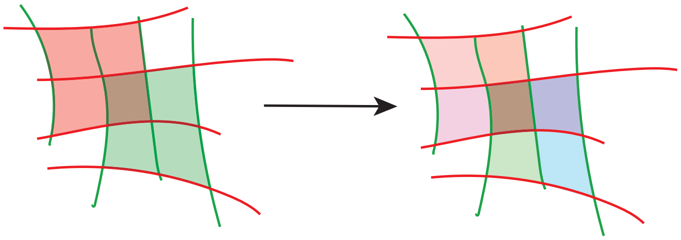


Figure 39.12: Constructing Markov components. Left: constructed T_ss. If there is an overlap we refine partition along the stable and unstable manifolds to mak smaller 'squares' with different colors.

rectangles so that there is no overlap among them, resultant refinement of the partition is a Markov partition. As can be seen from the construction we can make as fine Markov partition as we wish.

39.22 Markov subshift based on Markov partition

Let $\mathcal{R} = \{R_1, \cdots, R_n\}$ be a Markov partition. Basic properties of a Markov partition we use are:

(1) Let $\partial^s\mathcal{R} = \cup i\partial^s R_i$ and $\partial^u\mathcal{R} = \cup i\partial^u R_i$. Then

$$f(\partial^s\mathcal{R}) \subset \partial^s\mathcal{R}, \;\; f^{-1}(\partial^u\mathcal{R}) \subset \partial^u\mathcal{R}. \tag{39.21}$$

See Fig. 39.13.

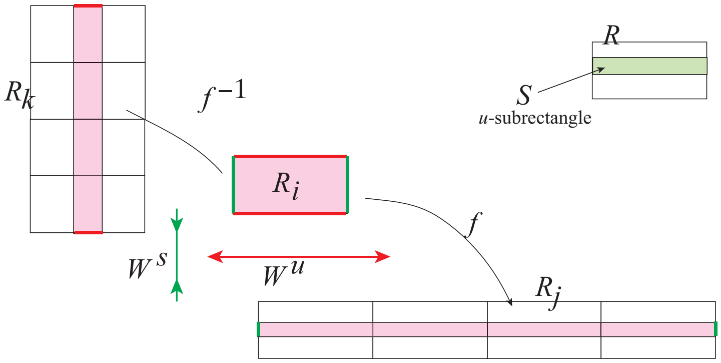


Figure 39.13: (1) illustrated: Boundaries are mapped onto boundaries; u-subrectangle is also illustrated for (2)

(2) A u-subrectangle S of $R \in \mathcal{R}$ is defined as a nonempty subset of R and $W^u(y, S) = W^u(y, R)$ (see Fig. 39.13).

If you can go from R_i to R_j and $S \subset R_i$ is a u-subrectangle, then $f(u) \cap R_j$ is a u-subrectangle of R_j.

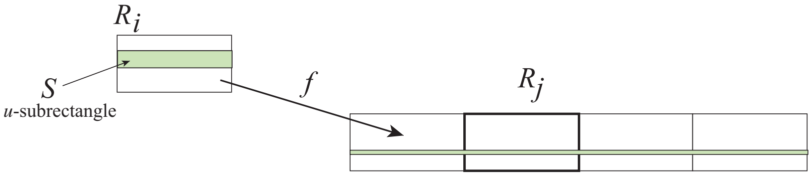


Figure 39.14: (2) illustrated

(3) The coding due to $\mathcal{R}$ is 'almost' one to one. More, precisely, (i) For any $a \in \Sigma_A$ $\pi(a) = \cap_{i\in\mathbb{Z}} f^{-j} R_j$ is a map from Σ_A to Ω_s which is continuous and onto.
(ii) $\pi \circ \sigma = f \circ \pi$ It is one to one on $Y = \Omega_s \setminus \cap_{j\in\mathbb{Z}} f^j(\partial^s\mathcal{R} \cap \partial^u\mathcal{R})$.
This should be clear from Fig. 39.15:

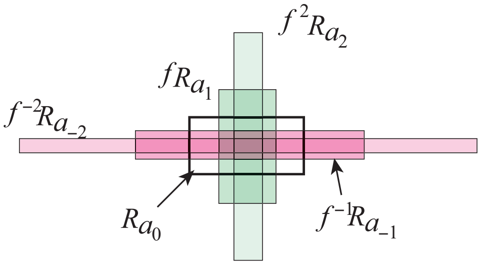


Figure 39.15: Markov coding rule: the points in $f^{-n} R_{a_n} \cap R_{a_0}$ have the code sequence $a_0 \cdots a_n$.

40 Anosov systems

40.1 Anosov system

$f \in C^r(M,M)$ is Anosov, if $\Omega(f) = M$ and hyperbolic.

In more detailed words:[437]

An Anosov diffeomorphism $f : M \to M$ is a diffeomorphism which satisfies the following:
(a) There is a continuous splitting of the tangent bundle $TM = E^s + E^u$ which is preserved by the derivative Df.
(b) There exist constants $C > 0$, $C' > 0$ and $\lambda \in (0,1)$ (i.e., hyperbolicity) and a Riemannian metric $\| \ \|$ on TM such that

$$\begin{aligned} \|Df^n(v)\| &\leq C\lambda^n \|v\| \text{ for } v \in E^s, && (40.1)\\ \|Df^n(v)\| &\geq C'\lambda^{-n} \|v\| \text{ for } v \in E^u. && (40.2) \end{aligned}$$

40.2 Some properties of Anosov systems

Per(f) is countable and dense in M.

Anosov systems are structurally stable.
If the Lebesgue measure may be introduced on M, Anosov systems are ergodic with respect to it.

Thus, clearly being Anosov excludes being Morse-Smale **38.5**. Obviously, MS diffeomorphisms are not dense for $d \geq 2$.

40.3 Toral diffeomorphism

Integer matrices L with $|\det L| = 1$ on T^n (constructed from the unit cube with periodic boundary conditions) is called toral diffeomorphisms.

For $TM = E^s \oplus E^u$, if $\dim E^u$ or $\dim E^s = 1$, the toral diffeomorphisms is said to be codimension one.

Theorem [Franks] An Anosov diffeomorphism f is homomorphic to a toral diffeomorphism, if f is codimension 1.

[437] Taken from J Franks, 'Anosov diffeomorphisms on tori," Trans AMS 145, 117 (1969).

40.4 Group automorphism on T^2

Let A be a regular 2×2 integer matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{40.3}$$

This defines a homomorphism $T_A : T^2 \to T^2$:

$$T_A(x, y) = (ax + by, cx + dy) \bmod 1. \tag{40.4}$$

If $\det T_A = \pm 1$, this is an automorphism (homeomorphism onto itself).

Notice that A need not be normal (i.e., $AA^* = A^*A$), so even if eigenvalues of A are distinct, it need not be diagonalizable with an orthogonal transformation. Thus, the two eigendirections may not be orthogonal. These are well illustrated by Thom's diffeomorphism **40.5**.

40.5 Thom diffeomorphism

A toral diffeomorphism $T_A : T^2 \to T^2$ with

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \tag{40.5}$$

is called th Thom diffeomorphism (physicists often call its 'Arnold's cat map, since they read easy-reading books only).

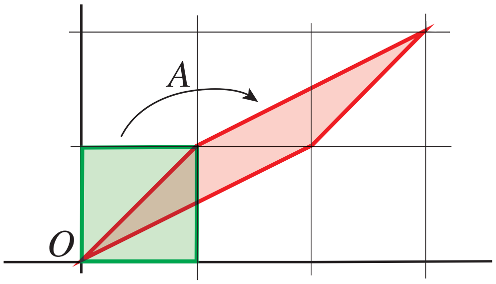


Figure 40.1: Thom's diffeomorphism is obtained from A by imposing a periodic boundary condition on the green square.

The origin corresponds to a hyperbolic fixed point p. The set of homoclinic points

$W^s(p) \cap W^u(p)$ is dense in T^2 (as you can guess from Fig. 40.4), so the periodic points are dense in T^2. This defines an Anosov system on T^2. It is called Arnold's cat map, because something like Fig. 40.2 was used by Arnold and Avez[438] what horrible things could happen for Anosov system.[439]

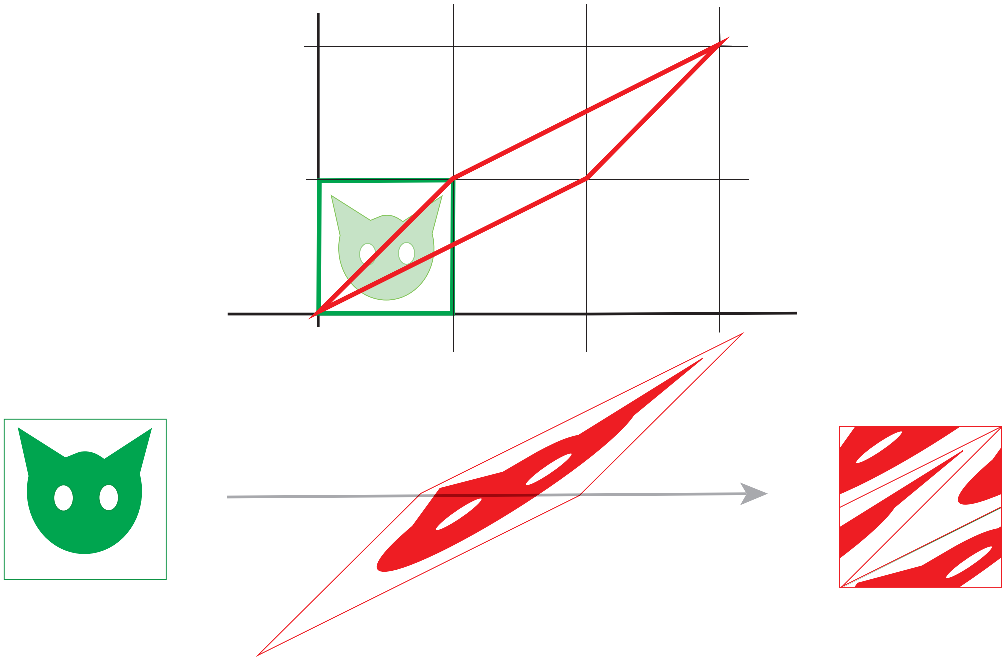


Figure 40.2: Thom's diffeomorphism

Note that $\det A = 1$. The eigenvalues and the corresponding eigenvectors of A are given by $(3 \pm \sqrt{5})/2$ (respectively) with $(1, (-1 \pm \sqrt{5})/2)^T$. We see that the Kolmogorov-Sinai entropy is $\log(3 + \sqrt{2})/2$.

[438] V I Arnold and A Avez, *Ergodic Problems of Classical Mechanics* (The Mathematical physics monograph series) (Benjamin 1968). This is a classic.

[439] A whole cat is kneaded here with T^2 illustrated: https://upload.wikimedia.org/wikipedia/commons/9/9e/Arnold%27s_cat_map.png.

40.6 Markov partition for Thom automorphism

Using the eigendirections in 40.5, we can make a Markov partition consisting of parallelograms, noting that stable and unstable manifolds must go through lattice points.

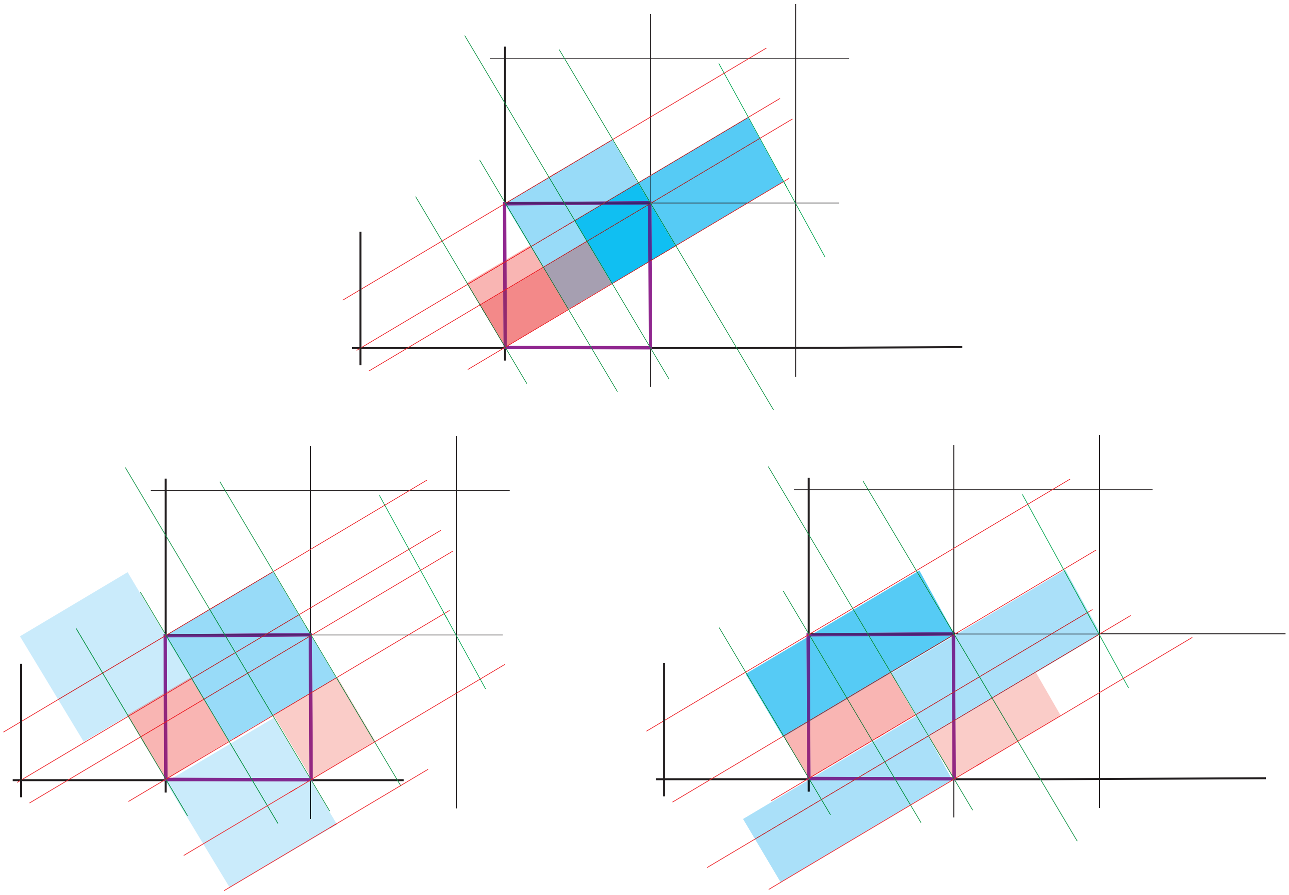


Figure 40.3: Markov partition; lower figures explain how to cover T^2 with the Markov partition above and its image. Redlines indicate W^u and the green W^s.

Needless to say, we can make many different Markov partitions, specifying the largest size of the piece.

Another example with nonnormal A is

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \tag{40.6}$$

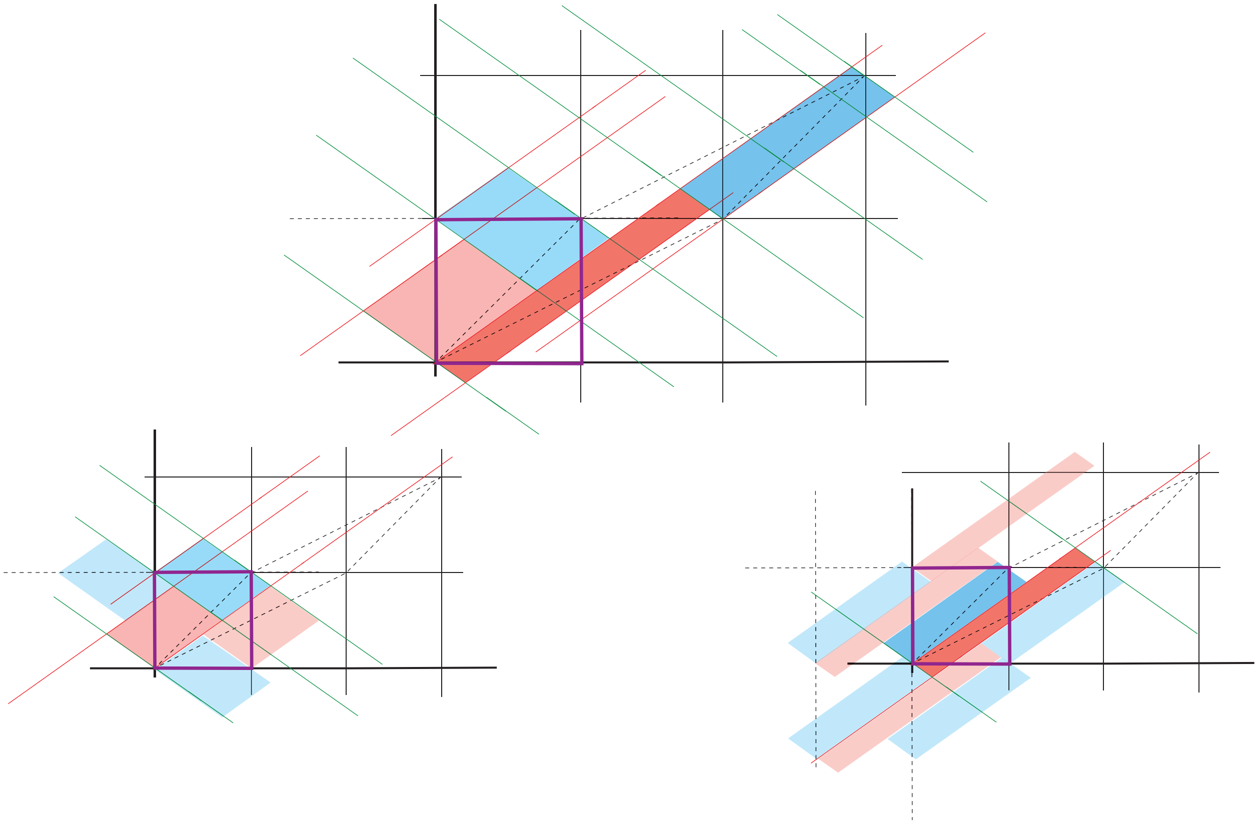


Figure 40.4: Markov partition; lower figures explain how to cover T^2 with the Markov partition above and its image. Redlines indicate W^u and the green W^s.

40.7 Pseudoorbit traceability

For a pseudoorbit $\{x_0, x_1, \cdots\}$, we can construct a true trajectory $T^k x$ always running close to it. This is the traceability of pseudoorbits.

For a system to have a traceability, necessary and sufficient condition (for C^1 systems) is that the system has a Markov partition (we have already seen this in Section 39). If a system has a Markov partition, the system is isomorphic to a symbolic dynamics called a Markovian subshift. Thus, Ornstein's theorem tells us that

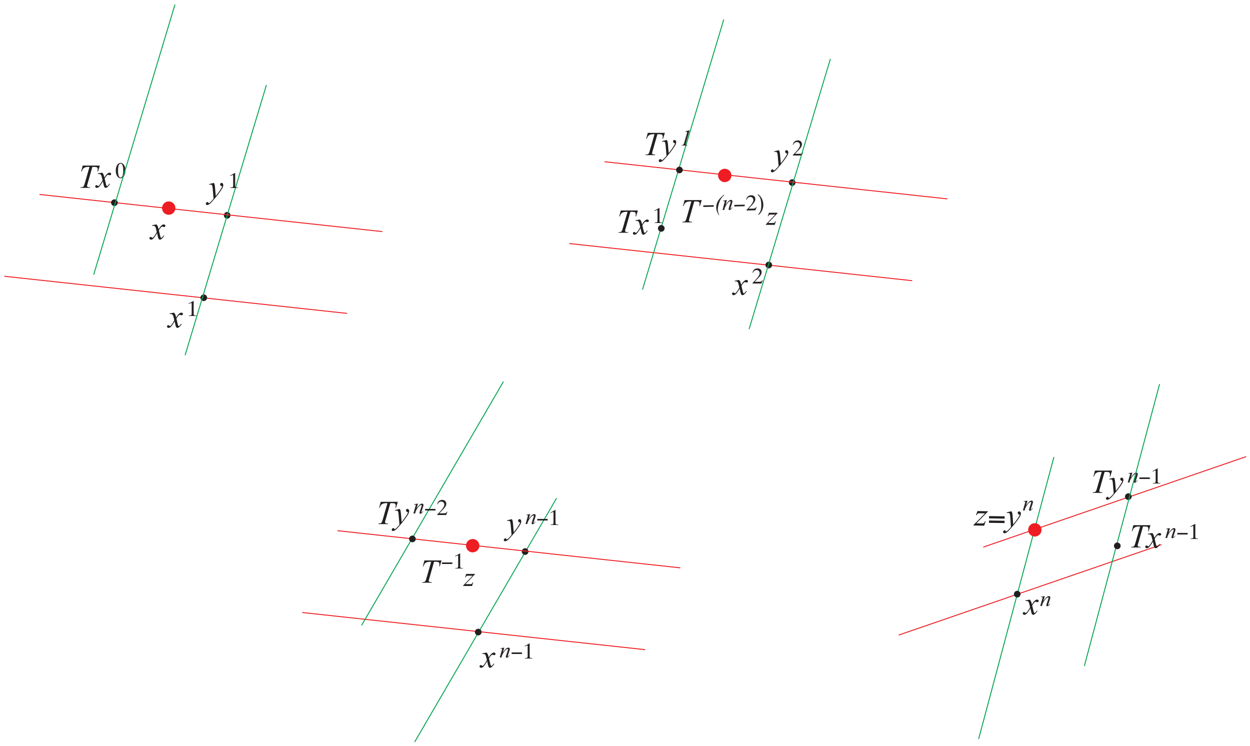


Figure 40.5: How to construct a shadowing orbit (red points)

the system is actually isomorphic to a Bernoulli system, a maximally chaotic system.

Thus ironically a numerically obtained trajectory can be a true trajectory only if the system is maximally chaotic.

41 Spectrum of almost periodic quantum systems

41.1 1D lattice Schrödinger equation

Consider a 1D lattice and a discrete version of the Schrödinger equation:

$$\psi_{n+1} + \psi_{n-1} + \lambda V(n\omega)\psi_n = E\psi_n \tag{41.1}$$

with the periodic potential

$$V(t+1) = V(t). \tag{41.2}$$

When $\omega \in \mathbb{Z}$, we have extended states and the usual energy band structure.

What happens if $\omega \notin \mathbb{Z}$? It is known to have complicated Cantor-set like spectrum. An explicitly provable case of self-similar energy band structure can be studied with an Axiom A system.

41.2 Related 2D dynamical system

We can write (41.1) may be rewritten as a 2D map problem:

$$\Psi_{n+1} = M(n\omega)\Psi_n, \tag{41.3}$$

where

$$M(t) = \begin{pmatrix} E - V(t) & -1 \\ 1 & 0 \end{pmatrix}, \quad \Psi_n = \begin{pmatrix} \psi_n \\ \psi_{n-1} \end{pmatrix}. \tag{41.4}$$

Define

$$M^k(t) = M(t + (k-1)\omega) \cdots M(t+\omega)M(t). \tag{41.5}$$

Then, we may write

$$M^k(t)\Psi_n = \Psi_{n+k} \tag{41.6}$$

and

$$M^{k+l}(t) = M^k(t + l\omega)M^l(t). \tag{41.7}$$

If we use the Fibonacci numbers F_m defined as[440]

$$F_{m+1} = F_m + F_{m-1}, \tag{41.8}$$

[440] Perhaps, https://www.math.ksu.edu/~cjbalm/Quest/Day7_slides.pdf is the best elementary page for the Fibonacci numbers.

with $F_0 = F_1 = 1$. Then,[441]

$$M^{F_{m+1}}(t) = M^{F_m}(t + F_{m-1}\omega)M^{F_{m-1}}(t). \tag{41.9}$$

41.3 Discontinuous potential case mappable to dynamical systems

If the periodic potential is two-valued:

$$V(t) = \begin{cases} -1 & \text{for } t \in (-\omega, -\omega^3], \\ +1 & \text{for } t \in (-\omega^3, \omega^2], \end{cases} \tag{41.10}$$

where $\omega = (\sqrt{5} - 1)/2 \simeq= 0.618$.[442] Note $\omega = \lim_m F_{m-1}/F_m$.

Then, (41.9) reads $M_m \equiv M^{F_m}(0)$:

$$M_{m+1} = M_{m-1}M_m \tag{41.11}$$

with $M_0 = M(0)$ and $M_1 = M(-\omega^3)$. Notice that $F_m\omega \simeq F_{m-1}$. This is OK, if, for all $m > 1$,

$$-\omega^3 < F_m\omega \bmod 1 \le \omega^2, \tag{41.12}$$

but $\omega F_m - F_{m-1} = (-\omega)^{m+1}$,[443] this is always true.

Thus, (41.11) may be used to study the spectrum. The 'extended state' implies $|M_m|$ to be bounded from 0 and from above.

41.4 Trace dynamics

Let

$$x_m = \frac{1}{2}\mathrm{Tr}\, M_m. \tag{41.13}$$

From (41.11) $M_m = M_{m-2}M_{m-1} \;\Rightarrow\; M_{m-2}^{-1} = M_{m-1}M_m^{-1}$

$$M_{m+1} + (M_{m-2})^{-1} = M_{m-1}M_m + M_{m-1}M_m^{-1}. \tag{41.14}$$

[441] M. Kohmoto, L. P. Kadanoff and C. Tang, Localization problem in one dimension: mapping and escape, PRL 50 1870 (1983).

[442] $\omega^2 = 0.382$, $\omega^3 = 0.236$; $\omega^2 + \omega = 1$.

[443] $-\omega = (-1/\omega) + 1$, $(-\omega)^2 = 1 + (-\omega)$, $(-\omega)^3 = 2(-\omega) - 1$, etc. gives this formula. Set $(-\omega)^n = A_n(-\omega) + B_n$. Then you see A_n and B_n are Fibonacci numbers F_n and F_{n-1}.

Since $\det M_m = 1$, the trace of the formula may be obtained after explicit matrix calculation as

$$x_{m+1} = 2x_m x_{m-1} - x_{m-2} \tag{41.15}$$

with

$$x_1 = \frac{1}{2}(E+\lambda), x_2 = \frac{1}{2}(E-\lambda), x_3 = 1. \tag{41.16}$$

41.5 Trace dynamics is (likely to be) Axiom A[444]

Thus, the trace of M obeys the 3D diffeo

$$T(x,y,z) = (2xy - z, x, y). \tag{41.17}$$

This has an invariant (with the initial condition (41.16) $I = \lambda^2$)

$$I = x^2 + y^2 + z^2 - 2xyz - 1 \tag{41.18}$$

which determines a 2-mfd (Fig.41.2)

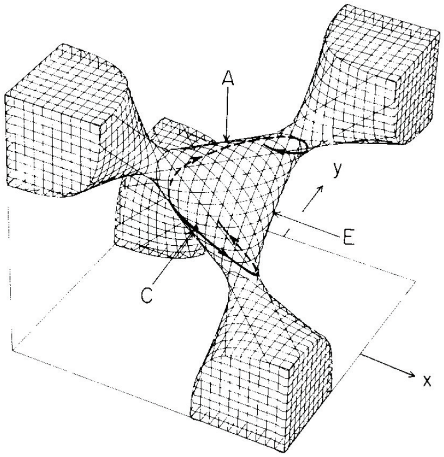


Figure 41.1: An example of the manifolds for $I = 0.2$.

Fig. 41.1 An example of the manifolds for $I = 0.2$. The four parts which are cut by a cube of size 6 actually extend to infinity. There are six saddle points, A, B, C, D, E, and F. The points B, D and F are located at the antipodal positions of E, A and C, respectively. A portion of the unstable manifold of C is drawn schematically. This crosses the stable manifold of A transversally near A.

[444] M Kohmoto and Y Oono, Cantor spectrum for an almost periodic Schrödinger equation and a dynamical map, PL 102A 145 (1984).

41.6 Determination of the spectrum

(41.17) defines a chaotic dynamics on a 2-mfd. It is Anosov if $\lambda = 0$.[445]

The most important periodic orbit of the map T to account for the spectra is the following six cycle:

$$A(0,0,a) \to B(-a,0,0) \to C(0,-a,0)n \to D(0,0,-a) \to E(a,0,0) \to F(0,a,0) \to A, \tag{41.19}$$

where $a = (1+I)^{1/2} = (1+\lambda^2)^{1/2}$. These six points are hyperbolic fixed points of T^6 whose eigendirections are tangent to the manifold:

$$\kappa_\pm = \{[1 + 4(1+\lambda^2)2]^{1/2} \pm 2(1+\lambda^2)\}^2. \tag{41.20}$$

The initial points are on $y = x - \lambda$, $z = 1$ near the fixed point A. The stable manifold of A crosses this straight line of the initial points. Most orbits starting from these crossings flow into the six cycle. The crossing point specified by the energy ε_0 which is closest to A along its stable manifold is clearly in the spectrum.

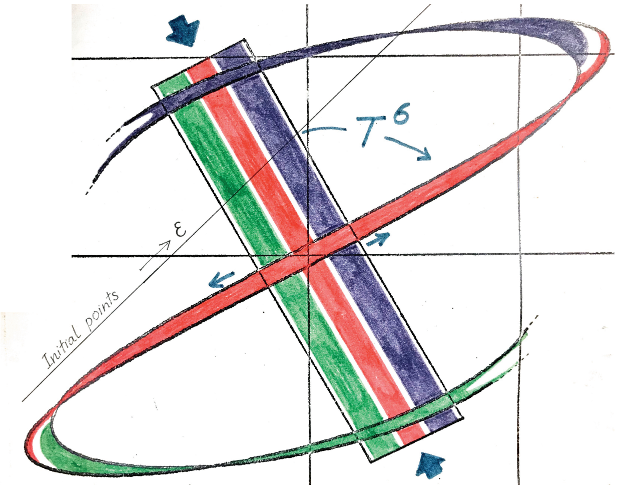


Figure 41.2: Horseshoe structure exhibited by the map T

[445] Pointed out by Y. Takahashi. This was the key observation; this statement immediately suggested a horseshoe, and a horseshoe hunt started.

41.7 Discrete cat map

The second order difference equation like discrete Schrödinger equation can always be rewritten as a first order 2D and vice versa. For example, Thom's map is

$$\begin{aligned} q_{t+1} &= 2q_t + p_t, & (41.21)\\ p_{t+1} &= q_t + p_t. & (41.22) \end{aligned}$$

Therefore,

$$q_{t+1} = 2q_t + (q_{t-1} + p_{t-1}) = 2q_t + q_{t-1} + (q_t - 2q_{t-1}), \tag{41.23}$$

so

$$q_{t+1} = 3q_t - q_{t-1}. \tag{41.24}$$

This may define a dynamics on $\mathbb{Z}$, but we could impose a mod N condition. Then this defines a dynamics on an integer ring $\mathbb{Z}/N$. Needless to say, all the orbits are periodic (with a period at longest N). This means the discrete version (41.22) mod N is also recursive as illustrated here or here.

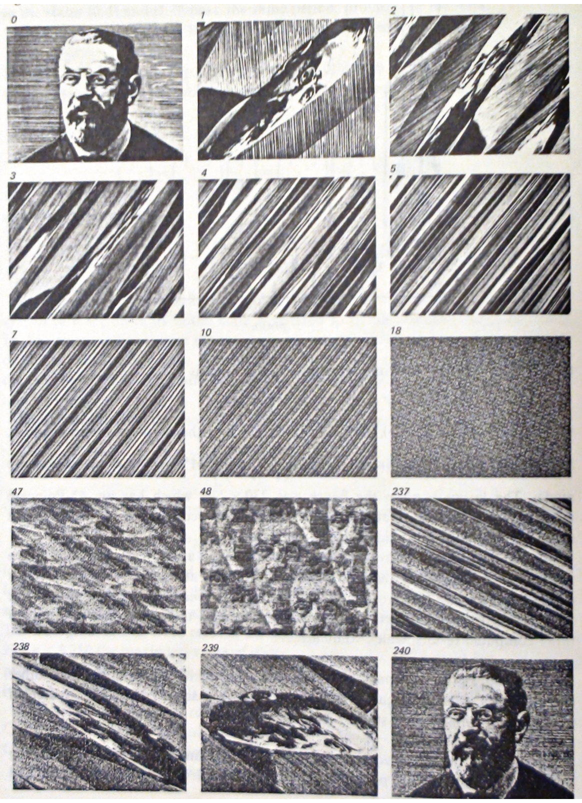


Figure 41.3: Discrete 'cat map' [Fig. 6.107 of Jackson

42 Newhouse phenomenon

42.1 Homoclinic point in modified linear map

The following an ex[licit example of a homoclinic point found in Palis-de Melo textbook.

Consider $\varphi(x, y) = (2x, y/2)$ and $\Psi(x, y) = (x - f(x + y), y + f(x + y))$, where f is continuous $f(x) = 0$ for $x < 1$ and $f(2) > 2$. Then $\Psi \circ \varphi$ has a homoclinic point on the y-axis between $y = 1$ and 2.

42.2 Homoclinic bifurcation

If a horseshoe map is slid as in Fig. 42.1, a non-transversal homoclinic orbit is formed. (b) has a homoclinic tangency.

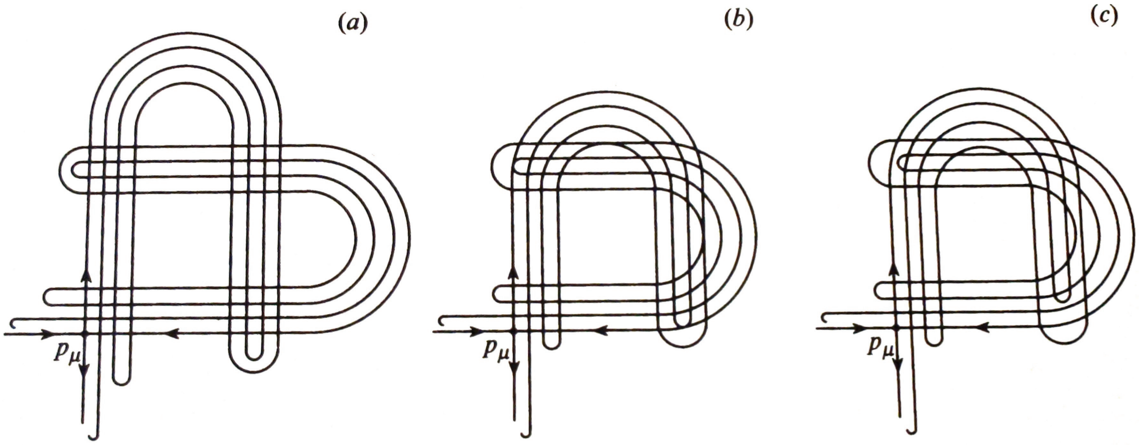


Figure 42.1: Homoclinic bifurcation ([Fig. 1.6 of Palis de Melo]

If we take a small rectangle R near p, then R has invariant foliations, and crossing points of these foliations make an invariant set that is a Cantor set. Poincare knew that transversal homoclinic points are accumulation points of other homoclinic points. Birkhoff showed that a transversal homoclinic point is an accumulation point of periodic orbits.

42.3 Cascade of homoclinic tangency

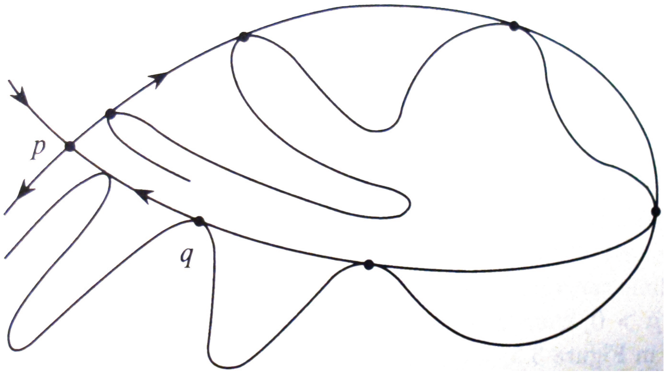


Figure 42.2: Homoclinic tangency [Palis-Takens Fig. 3.1]

Let φ_μ be a one-=parameter family of diffeomorphisms with a quadratic homoclinic tangency q at $\mu = 0$ associated to te fixed (periodic) saddle p and suppose it unfolds generically. Then, there is a sequence $\mu_n \to 0$ such that φ_{μ_n} has homoclinic tangencies $q_{\mu_n} \to q$ associated to $p_{\mu_n} \to p$.

42.4 Measure of homoclinic bifurcation set

We are interested in the measure of the set of μ in which φ_μ has a hyperbolic limit set. If the sum of the Hausdorff dimensions of W^s and W^u is less than 1, we can say this set has a relatively large Lebesgue measure.

Let B be the set of μ such that φ_μ is at a bifurcation point (homoclinic tangency). Then,[446]

$$\lim_{\mu_0 \to \mu} \frac{m(B \cap [0, \mu_0])}{\mu_0} = 0. \tag{42.1}$$

42.5 Newhouse phenomenon

Newhouse proves:

Let $\varphi \in \mathrm{Diff}^2(M)$, M a 2-manifold, be with a saddle point p whose stable and unstable manifolds have an orbit of tangency. Then,

(1) arbitrarily near φ there is an open set $U \subset \mathrm{Diff}^2(M)$ with persistent homoclinic tangencies.

(2) If moreover $|\det(d\varphi)_p| < 1$ (i.e., dissipative), then there is a residual set $R \subset U$ such that each member of which has infinitely many hyperbolic sinks.

This means that hyperbolic systems are not dense in $\mathrm{Diff}^2(M)$, M a 2-manifold.

[446] Theorem 2 of Palis-Takens p101.

However, for Diff$^i(M)$ nothing is known.

42.6 Persistent homoclinic tangency

Basic set: it is a maximal invariant set in its local) and canonical coordinate system may be taken if two points x y in it is close.
Persistent tangency: Let Λ_1 and Λ_2 are basic sets of φ. For any $\varphi \in U \subset$ Diff$^2(M)$, for $x_1 \in \Lambda_1$ $x_2 \in \Lambda_2$ there is a tangency between $W^s(x_1)$ and $W^u(x_2)$ or C^2-close φ'.
Thickness of a Cantor set: nongap length/gzp length $+ \tau$. This is not the Hausdorff dimension.
Overlaps of K_1 and K_2. If $\tau(K_1)\tau(K_2) > 1$ then $K_1 \cap K_2 \neq \phi$ because K's are closed.

Proposition 1
Let $\varphi \in$ Diff$^2(M)$ wutgh basic sets Λ_1 and Λ_2, both of saddle type, and let $p_i \in \Lambda_i$ be periodic points. Assume $\tau(L_1)\tau(L_2) > 1$[447] (this condition will be explained **42.9**) in and there is a orbit of tangency of $W^u(p_1)$ and $W^s(p_2)$. Then φ is in the closure of some $U \subset$ Diff$^2(M)$, where U has persistent tangencies involving $\Lambda_1(\tilde{\varphi})$ and $\Lambda_2(\tilde{\varphi})$ of Λ_1 and Λ_2 for $\varphi \in U$.

42.7 Persistence proof

Take $\tilde{\varphi}$ near φ, two basic sets $\Lambda_1(\tilde{\varphi})$ and $\Lambda 2(\tilde{\varphi})$ and periodic points $p_1(\tilde{\varphi})$ and $p_2(\tilde{\varphi})$. They depend on $\tilde{\varphi}$ continuously.

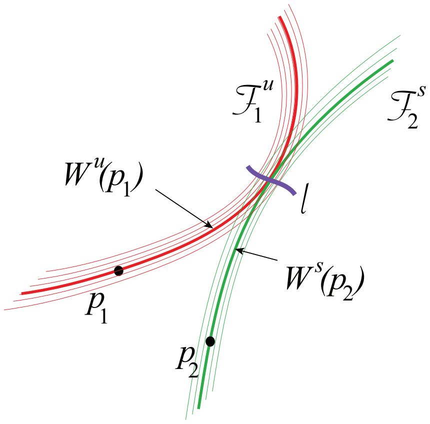


Figure 42.3:

We assume $W^u(p_1)$ and $W^s(p_2)$ are quadratically tangent (if needed, we can per-

[447] $\tau^u(\Lambda_1, p_1)\tau^s(\Lambda_2, p_2) > 1$ more preciesely in the original book.

turb the original system to enforce this).

We also have stable and unstable foliations $\mathcal{F}^u$ and $\mathcal{F}^s$ (recall a horseshoe). All these structures depend continuously on $\tilde{\varphi}$. If we change $\varphi(\varphi)$ the tangent point changes (C^1 change) its position transversally to the foliations. We can project leaves onto this trajectory ℓ of the tangent point.'

Now, take one-sided neighborhood boxes K_1 and K_2 as in Fig. 42.4. Then, we can consider their projections along the foliation onto ℓ. Generally, the projected images are Cantor sets.

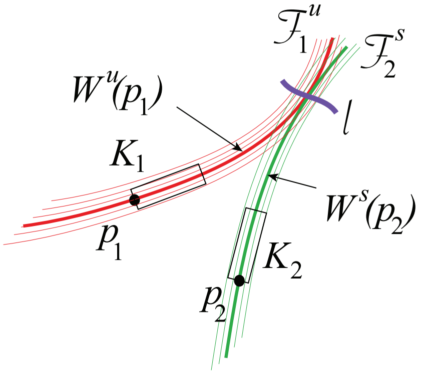


Figure 42.4:

These Cantor sets changes continuously with $\tilde{\varphi}$. We claim that under the condition "$\tau^u(\Lambda_1, p_1)\tau^s(\Lambda_2, p_2) > 1$", they persist to overlap ('really' as sets).

42.8 λ-lemma

If ℓ is a smooth curve intersecting $W^s(p)$ transversally, then its forward image $\ell^i = \varphi^i(\ell)$ contain compact arcs $m_i \subset \ell^i$ which approaches differentiably a compact arc m in $W^u(p)$ as in Fig. 42.5.

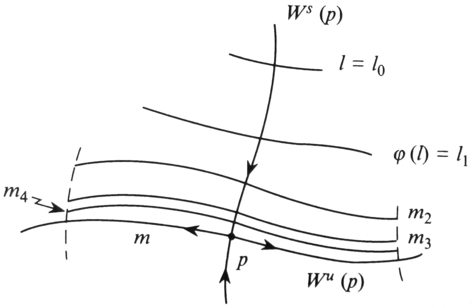


Figure 42.5: λ-lemma [Palis-Takens Fig. 1.7]

42.9 Linking Cantor sets

Suppose two Cantor sets 'generally' overlap (technical term = linked). Is there any actual overlap of points?

We define the thickness of a Cantor set K. A gap of K is a connected component of K^c. Let U be a bounded gap, and u its boundary point.

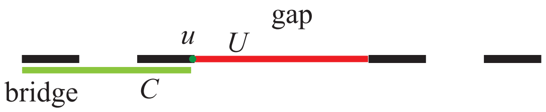


Figure 42.6: Gaps and bridges of a Cantor set

A bridge C of K at u is the maximal interval that does not contain any gap whose length is not equal or larger than U.

The thickness $\tau(K,u)$ of K at u is defined by ℓ is te length.

$$\tau(K,u) = \ell(C)/\ell(U). \tag{42.2}$$

Then, the thickness $\tau(K)$ of K is defined as

$$\tau(K) = \inf_u \tau(K,u). \tag{42.3}$$

Gap lemma:[448] If K_1 and K_2 are two Cantor sets. If

$$\tau(K_1)\tau(K_2) > 1, \tag{42.4}$$

then $K_1 \cap K_2 \neq \emptyset$, unless one of them is not engulfed in on of the other's gap.
[Explanation] (42.4) means (see Fig. 42.7)

$$\frac{\ell(C_1)}{\ell(U_1)}\frac{\ell(C_2)}{\ell(U_2)} > 1. \tag{42.5}$$

As seen in Fig. 42.7 if (42.4) holds, then we see overlaps may occur. Suppose there is certainly an overlap between some C and C' from the both Cantor sets. If there is not common point for $C \cap K$ and $C' \cap K'$, then all the points must be in U or U'. However, U and U' both shrink to zero...

[448] p63 of Palis and Takens.

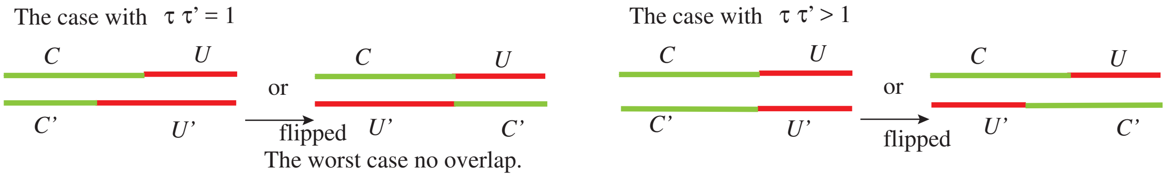


Figure 42.7: Meaning of (42.4).

42.10 Completion of 42.7

Let L_i be the projection of K_i in **42.7** onto ℓ along the foliations. These Cantor sets and the image of Ki with the dynamics (forward or backward) need not agree in general.

However, note that if the map is close to 'scaling' (i.e., the ratio of the max and min derivatives is close top unity), then the thickness is preserved. Therefore, we take K_i small enough and $\tilde{\varphi}$ is close enough to φ, this is realized.

Since L_1 and L_2 have a common set, and it consists of boundary points of Cantor sets (inevitably, since they do not have no internal point), so we have homoclinic tangency for $\tilde{\varphi}$.

42.11 Infinitely many hyperbolic sinks

Proposition 2.
Let $U \subset \text{Diff}^2(M)$ be an open set with persistent homoclinic tangencies, associated with a basic set $\Lambda(\varphi)$. Let $p(\varphi) \in \Lambda(\varphi)$ be a periodic point, say of period k and let $|\det{(d\varphi^k)_{p(\varphi)}}| < 1$. Then, there is a residual subset $R \subset U$ such that each $\varphi \in R$ has infinitely many hyperbolic periodic attractors (sinks). If $|\det{(d\varphi^k)_{p(\varphi)}}| > 1$, one gets infinitely may periodic repellers (sources).

The strategy to prove this proposition is to show that if φ has n hyperbolic sinks, then in its any neighborhood is a map with $n+1$ such sinks.

42.12 Demonstration of $n + 1$ sinks

Suppose φ has n hyperbolic sinks. They are stable against perturbations. $W^u(p)$ and $W^s(p)$ are both dense in $W^u(\Lambda)$ and $W^s(\Lambda)$ (resp.), and $W^u(\Lambda)$ and $W^s(\Lambda)$ have tangencies. Therefore, with (if needed) small perturbation we can make $W^u(p)$ and $W^s(p)$ in tangency. Let q be a point in this tangency orbit. This q is not among the already existing n periodic orbits, so we can take a neighborhood W of q that excludes n-periodic orbits.

Now, with an arbitrarily small perturbation, we can make a hyperbolic sink in W.

This can be understood from the horseshoe bifurcation **42.13**.

42.13 How does a horseshoe appear?[449]

As seen in Fig. 42.8 initially, there is no fixed point in R, but there is a horseshoe with all its hyperbolic fixed points at the 'right end.'

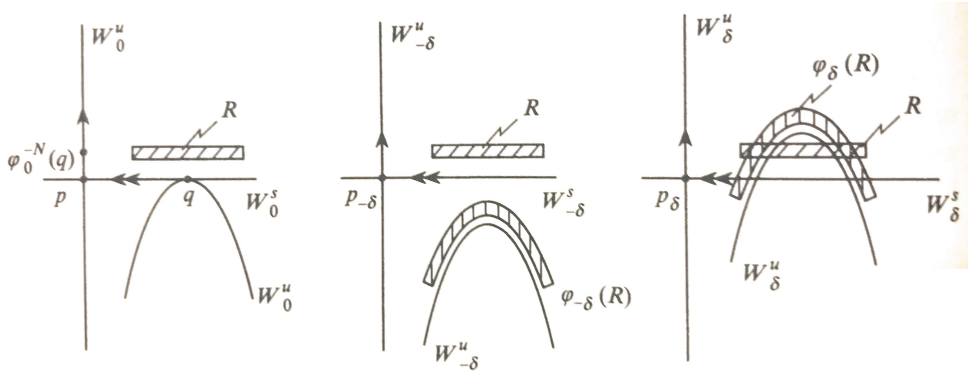


Figure 42.8: Horseshoe is closely related to the logistic map

What happens in between? It is the 'standard' story of bifurcations. We know there are three kinds of generic bifurcations in a one-parameter family of diffeomorphism; saddle-node, period doubling and Hopf. In the present context, the eigenvalues around fixed points are real, so we must consider the former two possibilities. Thus, sinks show up.

42.14 Hénon-like diffeomorphism

The Hénon map[450] $T : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$x_{n+1} = y_n + 1 - a x_n^2, \tag{42.6}$$

$$y_{n+1} = b x_n. \tag{42.7}$$

where $a = 1.4$ and $b = 0.3$ is the original choice of the parameters. This is written as the composition of the following three maps $T''' \circ T'' \circ T'$ (Fig. 42.9):

$$T' : \qquad x' = x, y' = y + 1 - a x^2, \tag{42.8}$$

[449] J. A. Yorke and K. T. Alligood, Cascades of period-doubling bifurcations: a prerequisite for horseshoe, Bull AMS 9 319 (1983); C. Robinson, Bifurcation to infinitely many sinks, Comm Math Phys 90 433 (1983) contain useful concrete examples.

[450] M. Hénon, A Two-dimensional Mapping with a Strange Attractor, CMP 50 69 (1976).

$$T'' : \qquad x'' = bx', y'' = y', \tag{42.9}$$
$$T''' : \qquad x''' = y'', y''' = x''. \tag{42.10}$$

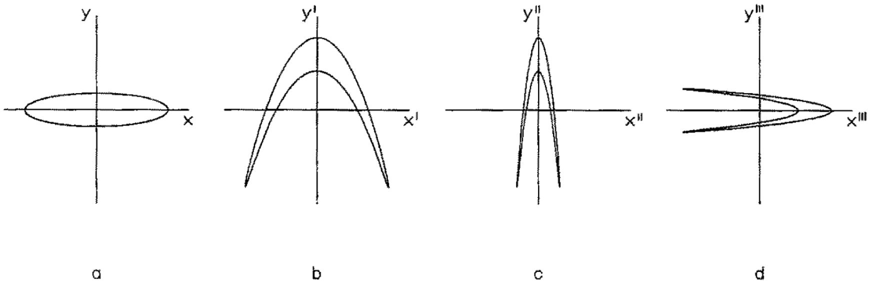


Figure 42.9: The Hénon map (d) is given by $T''' \circ T'' \circ T'$. [Fig. 1 of Hénon CMP 50 69 (1976)]

There are two fixed points:

$$x = \frac{1}{2a}\left[-1(1-b) \pm \sqrt{(1-b)^2 + 4a}\right], y = bx. \tag{42.11}$$

These points are real for $a > a_0 = (1-b)^2/4$. One is always a hyperbolic source, while the other is unstable for

$$a > a_1 = 3(1-b)^2/4. \tag{42.12}$$

Notice that not all the initial conditions give bounded orbits; they escape to infinity. The remaining set seems to be a Cantor set × smooth curves (locally). It is well-illustrated in

https://www.youtube.com/watch?v=42oeboRGqTo.

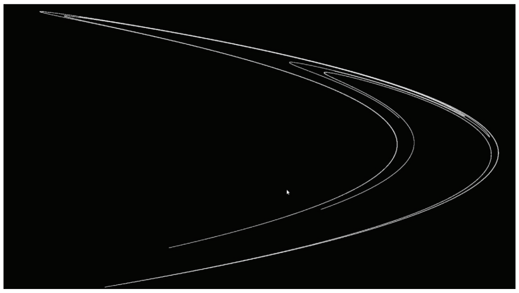


Figure 42.10: from the same YouTube.

See Michael Benedicks, Lai-Sang Young: Sina-Bowen-Ruelle measures for certain Hénon maps, Inventiones Mathematicae 112 541 (1993).

43 Heterodimensional cycles

43.1 Preliminary definitions[451]

The limit set $L(f)$ of f is

$$L(f) = \overline{\cup_{x \in M}(\alpha(x) \cap \omega(x))} \tag{43.1}$$

A point x is nonwandering if for every neighbourhood U of x there exists m, $m \neq 0$, such that $f^m(U) \cap U \neq \emptyset$. These points form the nonwandering set $\Omega(f)$. Obviously, $L(f) \subset \Omega(f)$.[452]

43.2 Axiom A related definitions and terminology

A diffeomorphism f satisfies Axiom A if $\Omega(f) = \overline{\mathrm{Per}(f)}$, and $\Omega(f)$ is hyperbolic. In this case $L(f) = \Omega(f)$.

There is a spectral decomposition: $\Omega(f) = \cup \Omega_i$, where Ω_i is f-invariant, transitive (= with a dense orbit), local maximal (i.e., there is a nbh U_i of Ω_i such that $\Omega_i = \cap_{\mathbb{Z}} f^n(U_i)$) and compact. Moreover $\Omega_i = \overline{H(P)}$, where $H(p)$ is the transversal homoclinic points related to P (i.e, $H(P) = W^s(P) \pitchfork W^u(P)$), where $P \in \Omega_i \cap \mathrm{Per}(f)$. Ω_i is called a basic set, and U_i isolating nbh of Ω_i.

The index of Ω_i is defined by $\dim W^s(P)$, where P is any periodic point in Ω_i. The index does not depend on the choice of P in Ω_i.

43.3 Local stability due to hyperbolicity

Hyperbolicity implies local stability: given a basic set Ω and its isolating neighbourhood U for any g C^r-close to f, $r > 1$, $\Omega(g)$ is hyperbolic and there is an homeomorphism $h : \Omega \to \Omega(g)$. $\Omega_g)$ is called continuation of Ω.

’43.4 Ω-stability

f is said to be Ω-stable if it has a conjugate continuation in its sufficiently small C^r nbh.

[451] Lan Wen *Differentiable Dynamical Systems* An Introduction to Structural Stability and Hyperbolicity.

[452] Recall Bowen’s non SBR counterexample.

43.5 We cannot ignore behaviors off Ω[453]

This is because we must worry about the mutual relations among basic sets. A simple examples are[454]

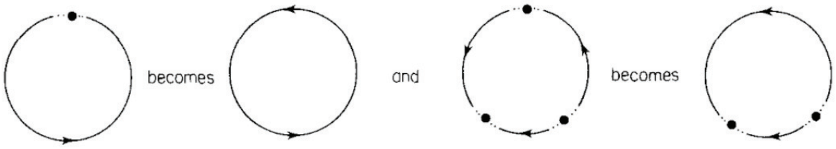


Figure 43.1: Simple Ω-explosion [Irwin Fig. 7.32]

Inn Fig. 43.1 Left, there is a one-way zero (non-hyperbolic zero). It is a the Ω set of the system. With a small perturbation we can remove it. The outcome is that the whole S^1 is Ω now, and example of the Ω-explosion. 43.1 Right, three fixed points reduce to two ('implosion').

43.6 Global explosion of Axiom A systems

Look at Smale's example (a diffeo in S^2): Its Ω consists of six hyperbolic fixed points Fig. 43.2.

We must see that the saddle connections are not in Ω, they are wandering. For example, take p. Its nbh eventually goes to sinks c or d.

Now, look at colored arrows in the figure. If we make a small surgery to cross $W^u(x)$ and $W^s(y)$, then p becomes non-wandering. Thus, $W^u(x) \cap W^s(y)$ is now non-wandering. However, $W^s(x) \cap W^u(y)$ is still wandering. Now, the Ω after perturbation consists of the previous fixed points + the new saddle connection.

43.7 Cycles

Let M be a closed C^∞-manifold and consider $\mathcal{X}^r(M)$ $(r \geq 1)$. We say that $X, Y \in \mathcal{X}^r(M)$ are Ω-conjugate if there is a homeomorphism $h : \Omega(X) \to \Omega(Y)$ sending trajectories of X into those of Y. $X \in \mathcal{X}(M)$ is Ω-stable if for any $\varepsilon > 0$ there is a

[453] Proc. Symp. Pure Math. Vol. 14, Amer. Math. Soc: Rhode Island, 1970.

[454] M C Irwin, *Smooth Dynamical Systems* (World Scientific, 2001) p185

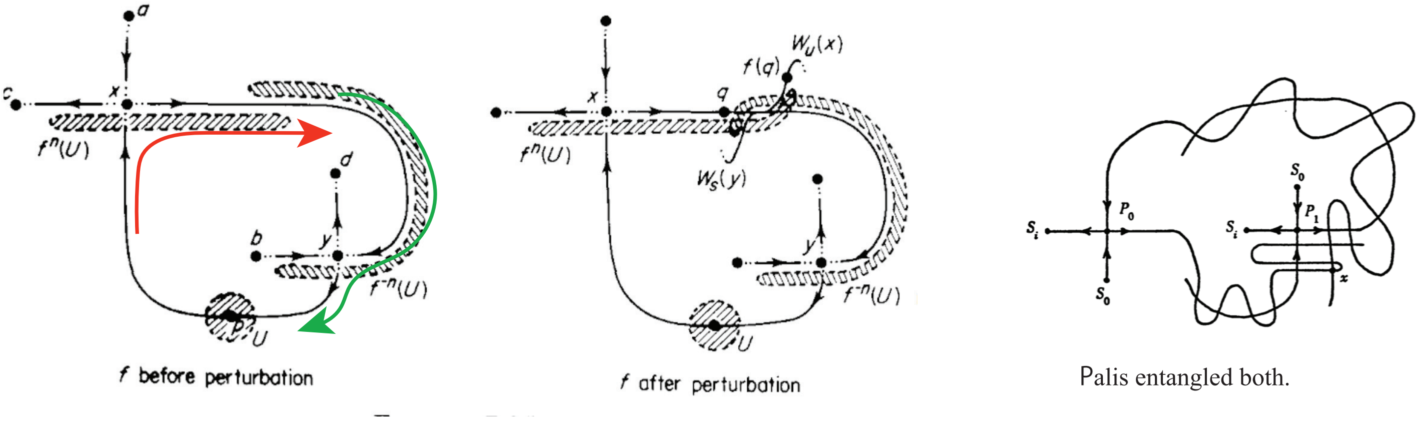


Figure 43.2: A is a global source and c a global sink. [7.35] Rightmost from Palis PAMS paper

neighborhood $N(X)$ in $\mathcal{X}^r(M)$ such that if $Y \in N(X)$ then X is Ω-conjugate to Y by a homeomorphism which is ε-C^0 close to the identity map in $\Omega(X)$.

For an Axiom A system for each basic set we can define its stable and unstable manifolds.

There is an n-cycle on Ω, if there is a sequence of basic sets $\Omega_0, \cdots, \Omega_{n-1}$ with $W_0 = W_n$, $\Omega_i \neq \Omega_j$if $i \neq j$ and

$$W^s(\Omega_i) \cap W^u(\Omega_{i+1}) = \emptyset. \tag{43.2}$$

A cycle is called equidimensional if index Ω_i making the cycle are identical and heterodimensional otherwise.

43.8 Ω-explosion[455]

For Smale's Axiom A' system:
(i) Ω is the disjoint union of the set of critical points F and the closure Λ of its periodic orbits,
(ii) each element of F is hyperbolic and Λ is a hyperbolic set.
Theorem: If X satisfies Axiom A' and there is a cycle on Ω, then X is not Ω-stable.

[455] J. Palis, Ω-explosion, Proc AMS 27 85 (1971). The diffeo version is J. Palis, A note on Ω-stability, Proc. Sympos. Pure Math., vol. 14, Amer. Math. Soc, Providence, R. I., 1970. In this paper Palis gives a sufficient condition for Ω-stability as well for special cases: If Ω is the finite union of hyperbolic critical points and closed orbits and has the no-cycle property, then X is Ω-stable.

That is, no-cycle condition is a necessary condition for Ω-stability.

Notice that so far we discussed the existence of explosive or dangerous cases

43.9 Stable non-Axiom A cycles[456]

Let M be a 3-mfd.

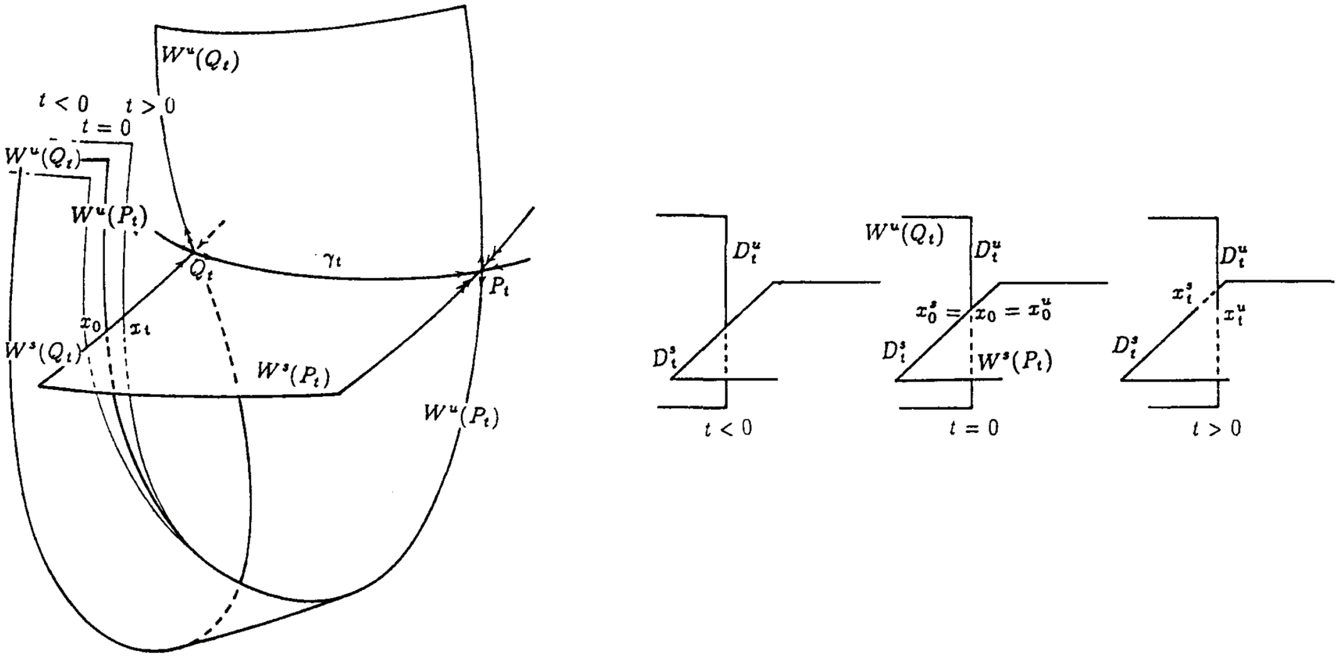


Figure 43.3: [Diaz Fig.1]

(I) Connected interesection
(1) $\dim W^s(P_0)) = 2$, $\dim W^u(Q_0) = 1$.
(2) There is an $_0$ invariant curve $\gamma_0 \subset W^s(P_0)) = 2 \pitchfork W^u(Q_0)$.
From now on $0 \to t$ indicates continuations.
(II) Creation and generic unfolding of the cycle
There are C^1 curves: $x_t \in W^s(Q_t)$ and $x_t \in W^s(Q_t)$. (III) Strong foliation condition.

[456] L J Diaz, Ribust nonhyperbolic dynamics and heterodimensional cycles, Ergod. Th. & Dynam. Sys. 15 291 (1995).

THEOREM 1. Let f_t satisfy the above conditions. Then for $t \in [0, t_0]$
(1) $\gamma_t \subset L(f_t)$, so $L(f_t)$ is not hyperbolic,
(2) f_t is Ω-stable. That is, f_t can be C^∞-approximated by a diffeo exhibiting a heterodimensional cycle.

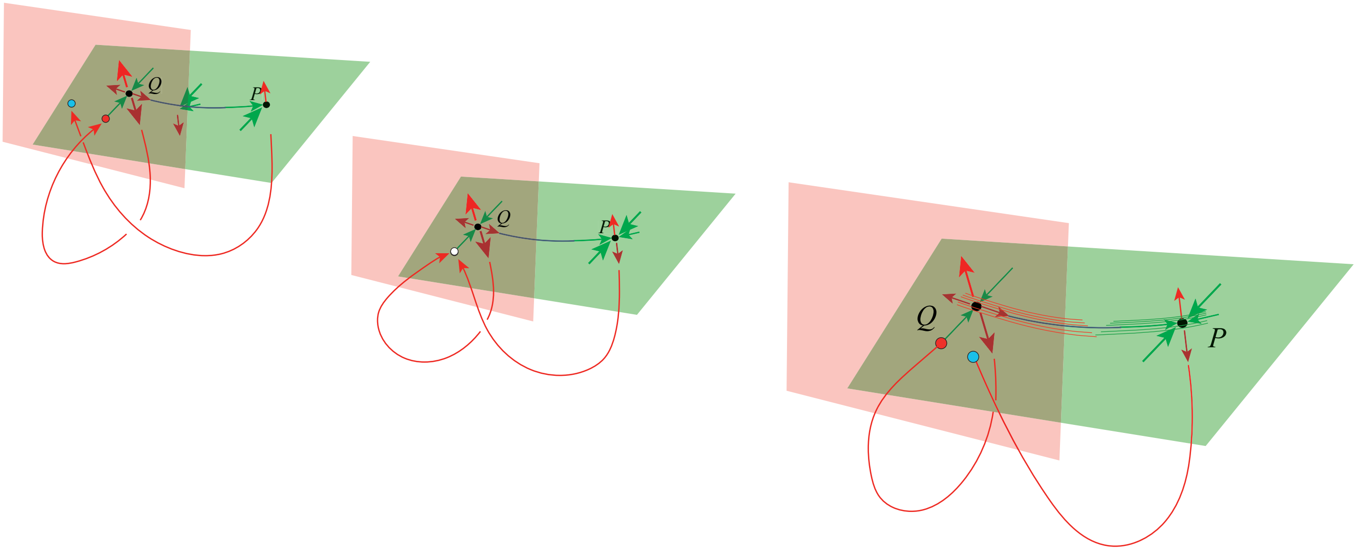


Figure 43.4: [Diaz Fig.1]

44 Palis conjecture

44.1 Stability conditions[457]

As we have seen Axiom A alone is not enough to guarantee the structural stability. Thus, Smale conjectured that:

Axiom A + no-cycle condition implies C^1-Ω-stability.[458]

This was proved by Smale (for diffeos) and by Pugh & Shub (for flows). Another conjecture is

Axiom A + strong transversality implies C^1-stability.

This was proved by Robbin and Robinson.

As in the case of Peixoto's theorem to show the necessity is much harder: For a system to be stable the above mentioned conditions are needed. Mañé proved the necessity for diffeomorphisms.[459] Hayashi showed that the C^1-dynamical system case: a C^1-dynamical system must be hyperbolic to be stable[460] A high point in Hayashi's work is his connecting lemma[461] creating homoclinic orbits by C^1-small perturbations of flow or diffeomorphism an unstable manifold accumulating on some stable one can be C^1-perturbed to make it intersect one another (the creation of homoclinic or heteroclinic orbits).

We usually claim that a phenomenon of a system relevant to natural science must be structurally stable, because science demands reproducibility. If we stick to this 'dictum,' we have only to study 'nice' Axiom A systems. However, we already know if the system dimension is not too low (1 for diffeo and 2 for flow as Peixoto showed and horseshoes cap.), we have something else as an open set. The members of such an open set is structurally unstable: we have 'stable structural instability.' There must be corresponding natural phenomena.

44.2 Palis' global conjecture on metric stability of attractors

I: There is a C^r-dense set D of dynamical systems such that any element of D has finitely many attractors whose union of basins of attraction has total probability.
II: The attractors of the elements in D support a physical measure.

[457] Hayashi 1997; J Palis, A GLOBAL VIEW OF DYNAMICS AND A CONJECTURE ON THE DENSENESS OF FINITUDE OF ATTRACTORS

[458] That is, the nonwandering sets are preserved.

[459] R Mañé, A proof of the C^1-stability conjecture, Publ Math IHES 66 161 (1988).

[460] S Hayashi Connecting invariant manifolds and the solution of the C^1 stability and Ω-stability conjectures for flows, Annals of Math 145 81 (1997).

[461] This is explained in his review

III: These properties are metrically stable (i.e., in D any k-parameter small C^r-perturbation preserves the dynamics).
IV: The attractors are stochastically stable. (Against noise)
V: For 1D maps the attractors are either sinks or abs cont measures.[462]

44.3 Palis conjecture 1993[463]

1. Every C^r-diffeo of a compact mfd M can be C^r-approximated by one of the following:
   (a) a hyperbolic system (Axiom A with strong transversality)
   (b) a system with heterodimensional cycle
   (c) a system exhibiting a homoclinic tangency.
2. If M is 2D then (a) or (c) occurs (in other words, Palis conjectured that avoiding homoclinic bifurcation, the generalized Peixotos picture can be recovered).

For $r = 1$ Pujas and Sabarino proved 2.[464], 1 is still open.

for $r \geq 2$ both are wide open. For $r \geq 2$ this is widely open. even for 2D.

[462] Lyubitch showed this for quadratic maps.

[463] A. Katok, B. Hasselblatt, Handbook of dynamical systems vol 1A(2005); E. R. Pujals From Peixoto's theorem to Palis's conjecture (2009).

[464] E. Pujals and M. Sambarino, Homoclinic tangencies and hyperbolicity for surface diffeomorphisms, Ann. of Math. 151 (2000), 961-1023.

Story Line

Here is a flow of story of Phys510 Fall 2018. Perhaps 6 semester materials are pushed into one semester, so I give here an overall outline of the flow, and summarize some important concepts and facts (theoretical) physicists should know. Key concepts are in boldfaces; you can look up the units relevant to the concepts in the main lecture notes by clicking the unit numbers (if you use the Story Line appended to the main body).
v

The course consists of four parts I - IV:
I. Introductory review of differential equations and maps (Lect 1-16).
II. Typical 'chaotic systems' and famous dynamical system (Lect 17-22).
III. Conceptual tools to understand dynamical systems (Lect 23-36).
IV. Outline of the modern theory: Peixoto to Palis (Lect 37-44).

Underlined statements are (my) conjectures perhaps theoretical physicists could look into.

Part Ia: non-conserved systems.

[**1**] Study of deterministic time evolution is the theory of dynamical systems (see **1.1**). Usually, we study **flows** defined by (sufficiently differentiable) **vector fields** on manifolds **2.5**, and **endomorphisms** (into-maps) defined on manifolds **2.2**. In many important cases, a **Poincaré map 6.3** and its **suspension 6.8** relate the continuous-time and its discrete-time descriptions.

[**2**] Theory of dynamical systems uses standard differential topological and geometrical terminologies freely. Therefore, to read the original math papers often demands some familiarity to differential topology. If you have no ambition to write math papers, intuitive understanding of the related concepts and theorems is sufficient (see **2.3**-**2.9**). However, the expression of tangent vectors in terms of ∂_i (**2.5**, **2.7**) is highly useful.

[**3**] As fundamental scientists we wish to have a general 'universal' or 'unified' understanding of many things, so we must properly characterize what we mean by 'general', '**generic**' (= residual **2.28**), etc. This is, however, not very simple, because simple characterizations are plagued with exceptions, and 'air-tight' characterizations tend to be cumbersome. Therefore, we confine ourselves to the study of systems stable against various perturbations (structurally stable systems; **2.13**). If the dimension

of the base manifold is not too large (2 or 3), then structural stable systems are quite numerous (generic or even sometimes open generic).

[**4**] Crudely put, **structurally stable systems** are characterized by **hyperbolicity** (**5.7**, **5.8**) and **transversality** (= relations are not 'tangential' or critical).

[**5**] Just as phase transitions, when qualitatively different stable features switch, structural stability is lost and systems become unstable against perturbations. This phenomenon is called **bifurcation** Lect 7, 8) **7.1**. Therefore, the theory of dynamical systems studies generic structural stable features and bifurcations exhibited by a collection of systems.

[**6**] There are two major ways to study dynamical systems, (i) **topological** and (ii) **measure-theoretical**. (i) is geometrical; we are interested in how trajectories go around (topologically). (ii) is statistical; we are interested in the average behaviors.

[**7**] Thus, our story line for 'Part I' goes as follows: In physics Hamiltonian dynamical systems are quite important. However, from the general dynamics point of view they are very special with the canonical structure. Therefore, first we discuss general ODE/Diffeo and their elementary bifurcations (Lect 3-9). Then, we go to Hamiltonian systems including elementary celestial mechanics (Lect 10-16).

[**8**] ODE (its origin: **3.26**) with continuous vector fields define flows **3.7**. Peano noted, however, that the uniqueness of the solution to initial value problems is not guaranteed as counterexamples show (**3.13**). Eyink points out such vector fields can be realized in fully developed turbulence as the flow velocity field (**3.14**). The spirit of the proof of **Peano's theorem** (**3.12**) with the aid of Arzela's compactness theorem (**3.11**) should be understood: we construct a sequence of approximate solutions, the totality of which makes a compact set, so we can find a limit which is a solution.

[**9**] The uniqueness of the flow defined by a vector field is guaranteed if the field is **Lipshitz** (Cauchy-Lipshitz theorem **3.18**). The reason may be intuitively understandable from the local **rectifiability** (**3.19**) of the field and its extension (**3.20**). The continuous dependence of the solution on its initial condition can be shown (**3.23**) almost constructively; **Gronwall's inequality** (**3.22**) is a standard tool.

[**10**] The uniqueness theorem breaks down at singularities where vector fields vanish. If the derivative at the singularity is non-singular, we say the singularity is simple

(**simply singularity 4.2**); simple singularities are isolated (**4.3**). We **linearize** the ODE around its isolated singular point; the solution to the linearized equation may be written in terms of a matrix A as (**4.5**)

$$\dot{x} = Ax.$$

[**11**] Its solution may be computed in terms of the evolution operator e^{tA} **4.6**. Constructing the (real or complex) Jordan form of A (Appendix 1 to Lect 4 for the general theory) is the standard way to compute this operator explicitly as a matrix; see a detailed example **4.17**.

[**12**] If the base manifold is a 2-manifold (2D manifold), then we can illustrate all the types of simple singularities: sink, source, saddle, focus, center **4.10**.

[**13**] On a given manifold not every vector field can live happily. There must be a topological consistency (**Poincaré-Hopf theorem 4.21**): the Euler characteristics of the manifold must be consistent with the sum of indices of the field (the **degree** of the field **4.20**).

[**14**] If A in [10] has no eigenvalue with vanishing real part, we say the isolated singularity x is **hyperbolic** (called a hyperbolic fixed point **5.1**). The stable (resp., unstable) eigenspace E^s (resp., E^u) is the subspace of T_xM (**2.5**) on which A has eigenvalues whose real parts are negative (resp., positive) (**5.7**). There is an invariant submanifold called **stable manifold** W_x^s (resp., **unstable manifold** W_x^u) of M on which the flow is contracting (resp., expanding) [**Stable manifold theorem 5.11**].

The renormalization-group flow near the critical point is a typical hyperbolic flow (**5.12**).

[**15**] Near a hyperbolic fixed point, the original dynamics and the linearized dynamics are homeomorphic (**Hartman's theorem 5.13**). Its proof is nontrivial, but it uses very standard functional-analytic tools; the basic idea of the proof is to construct the homeomorphism. The theorem justifies the **linear stability analysis** of a vector field around a hyperbolic fixed point. [See **5.20** and **5.21** for definitions of stability.]

[**16**] When bifurcation occurs, A is non-hyperbolic. Then, there is a neutral subspace on which eigenvalues of A have no real part. There is a manifold tangent to this subspace called the **center manifold** (not necessarily unique). The **center manifold theorem 5.19** allows us to make a reduced dynamics that is often with a smaller number of variables than the original (thus practically useful).

[**17**] For ODEs solution curves can be one of the following three: point, ring or line (**6.1**). The ring corresponds to a **periodic orbit**, for which we can construct a **Poincaré map 6.3**.

[**18**] To study the stability of a periodic orbit, we linearize its Poincaré map (**6.4**). The resultant matrix is periodic, so we may use **Floquet's theorem** to isolate the non-periodic components (Floquet multiplier or Lyapunov constant **6.5**) to study stability.

[**19**] Isolated periodic orbits are called **limit cycles 6.6**. For S^2 (or the domain embeddable in S^2) **Poincaré-Bendixson theorem 6.9** tells us the existence of periodic orbits. In practice, the use of null-clines (**6.17**) may also be useful.

[**20**] When singularities are not hyperbolic, **bifurcations 7.1** occur. To study systematically what can actually happen at or around bifurcation points, we make a 'standard form' (**normal form 7.6**) of the field at the bifurcation point, and then consider its most general deformation (unfolding). This is the **versal unfolding** (**7.3**) approach. Its first step is to make the lowest order nontrivial normal form using the cokernel technique (**7.7**, normal form theorem **7.9**). Look at **Hopf bifurcation** as an example (**7.14**).

[**21**] Normal form analogue can be constructed for maps as well. Accumulation of 2^n-periodic orbit for a continuous endomorphism (See Sarkovskii's Theorem 3 in **22.26**) is an accumulation of pitchfork bifurcations **Feigenbaum critical phenomena 8.7**). Feigenbaum constructed an RG theory (**8.7**-) to study this critical point. As you will see Sinai's thermodynamic formalism tells us the correspondence of this point and the critical phenomenon for 1D Ising model (with long-range interactions) (Lect 36 **36.1**-).

[**22**] The modifications appearing in versal unfoldings are perturbations that give qualitative changes to the system dynamics. That is, the perturbation series for such perturbations cannot converge. Such perturbations are called **singular perturbations 9.1**. However, there is a way to obtain the long-time behavior of the perturbed system systematically. They are collectively called singular perturbation theory, many of which may be unified as a renormalization group theory (**9.10**).

[**23**] The most important observation in the RG approach to singular perturbations is that the RG equation is the slow time equation that describes the long-time effect of singular perturbation. If the method is applied to PDE's, very often the lowest

nontrivial order RG equations are the 'named' equations, e.g., Boltzmann equation, Burgers equation, Swift-Hohenberg equation, etc. The reliability of RG equations is demonstrated by Chiba **9.18**. My conjecture is: if the original perturbed system is structurally stable, then the corresponding RG result is homeomorphic to it.[465]

Part Ib: conserved systems = Hamiltonian systems

[**24**] The **Newton-Laplace determinacy 10.1** and its compatibility with a **variational principle** (Veinberg's theorem **10.3**) imply that the equation of motion is a conserved second-order time-reversal symmetric equation. The **action principle** is locally a minimum principle **10.5**.

[**25**] A Legendre transformation of Lagrangian gives the Hamiltonian, and the action principle is rewritten as **Hamilton's principle 10.9**. In terms of **Poisson brackets 10.10** the Newton's equation of motion can be written symmetrically as the **canonical equation of motion 10.11**. Jacobi's identity may be demonstrated easily (**10.12**), if we introduce the infinitesimal canonical transformations (**13.5**).

[**26**] If a system with n-degrees of freedom[466] has a set of n independent invariants, then we say the system is **completely integrable 11.4**. Then, the phase space is foliated into nested invariant T^n (**Liouville-Arnold's theorem 11.6**; its demonstration is not so trivial as seen in **11.8**). Each torus is specified by the values of action variables, and the motion on it is described in terms of the angle variables **11.7**.

[**27**] Most (all?) completely integrable systems may be expressed in terms of a **Lax pair** A and L as (**12.1**)

$$\dot{L} = [A, L].$$

The eigenvalues of L are the invariants. The Toda lattice **12.3** is an example, which is related to the Kortweg-de Vries equation (**12.6**; a (not terribly) quick and dirty derivation **12.8**), which is famous for exhibiting solitons. Initially, the equation drew attention for its closeness to the Fermi-Ulam-Pasta problem **12.5** (but actually, not so close).

[**28**] In mechanics we consider only canonical transformations with **generators 13.1**. If the transformation is infinitesimal, it is called an **infinitesimal canonical trans-**

[465] The discrete time counterpart of this statement is a theorem.

[466] i.e., with n functionally independent variable pairs $\{q_i, p_i\}$

formation 13.4. Time evolution is an example, whose generator is the Hamiltonian (**13.6**). Noether's theorem may be understood with its aid (**13.7**).

[**29**] To show the invariance of Poisson brackets under canonical transformations, we introduce **Lagrange brackets 13.9**. This machinery allows us to prove **Liouville's theorem 13.15**. Also we can show that Poincaré maps preserve cross-sectional areas (**13.11**) for Hamiltonian systems.

[**30**] Note how special the Newtonian potential is (Bertrand's theorem **14.1**). It is almost impossible to show the stability (no collision, no escape) of $n(> 2)$-celestial body system theoretically (cf. **14.3**).

[**31**] Even the **restricted three body problem 14.6** is too complicated to study analytically. Poincaré showed that there is no integrable of motion functionally independent of the Hamiltonian that is analytic in the perturbation parameter (**14.8**). Thus almost all lost interest in solving the restricted problem.

[**32**] However, there are two stable fixed point solutions (Lagrangian points; from the SJ co-rotating coordinate system). These points describe Trojan asteroid group (**14.14** and more with respect to the earth and the moon **14.15**).

[**33**] Although Poincaré realized how complicated the three-body problem is (see **16.3** and the figure), Kolmogorov realized that still many invariant tori (esp highly non-resonating orbits **KAM tori** examples in Lect 16) guaranteed by the Liouville-Arnold theorem survive. There are two obstacles to prove the assertion. One is the **small denominator problem**, which was overcome by Siegel (Siegel's stability theorem **15.12**) with the so-called Diophantine approximation (**15.11**). The other is to prove the actual convergence of the perturbation series. The basic idea for the latter was furnished by Kolmogorov by 'partial linearization' (see **15.13**; **15.25**).

[**34**] What happens if the tori are deformed? This may be glimpsed from Poincaré-Birkhoff's theorem **16.8**. We have elliptic and hyperbolic periodic orbits, and the possible heteroclinic orbits produce chaos as shown in **16.9**. In the chaotic region the system can wander off far away from the original torus (especially if the system is high-dimensional; called **Arnold diffusion 16.13**)

[**35**] The FPU system does not thermalize due to the persistence of the KAM tori as noted in **16.12**. Motion of charged particles in electromagnetic fields is an important topic from the accelerator physics and plasma physics. A typical simple case is illustrated in **16.4** and may be understood in terms of the standard map **16.10**.

Part II: The Zoo

Here is a flow of story for Lectures 17-22. This is a showcase of representative examples, billiards, coupled relaxation oscillators, Lorenz system, Ruelle-Takens picture/strange attractors, interval endomorphisms + related concepts and theorems.

[**1**] Perhaps the simplest Hamiltonian system is a ballistically moving particle perfectly elastically colliding with boundaries/obstacles. Usually we discuss such systems defined on a 2-flat space. They are generally called **billiards**. Their overall dynamics may be understood from the mean free time and what happens at collisions (**Ambrose-Kakutani representation** **17.2**; **17.15**, mean-free time **17.16**; **Abramov formula** **17.17**).

[**2**] Noteworthy facts about billiards include:
(1) Even on polygons (triangles) a lot of things are not yet understood. See **17.4**.
(2) If the table is convex and if the boundary is sufficiently smooth, there is a **caustic**, so the system cannot be fully chaotic (even if chaotic) (Lazutkin **17.6**).
(3) **Sinai billiards** (dispersive billiards **17.7**) are 'maximally chaotic.'[467] Often they have, at least conceptually, related to geodesics on negative curvature surfaces **17.12**. These billiards are chaotic (intuitively), because information is lost upon collisions **17.18**. Some details about computing information loss rate (the Kolmogorov-Sinai entropy) is explained towards the end of Lecture 17.
(4) **Bunimovich billiards** (converging billiards) are also fully chaotic **17.13** [but the shapes are rather restricted (why? cf Lazutkin above)].

[**3**] Mutually hindering coupled relaxation oscillators are chaotic **18.1**. Methods to analyze such systems (converting to geometrical models and maps (e.g., **18.7**, **19.4**) are standard techniques). We can easily understand why the system is not predictable **18.5**.

Coupled relaxation oscillators can be related to the **Lorenz system** **19.3**; the motivation with its relation to the Rayleigh-Benard problem (Saltzman's equation **19.2**) is explained in **19.1**-; the reduction method is an example of the Galerkin method **20.8**.

[467] They are Bernoulli systems.

A very similar model is the Rikitake model of the earth dynamo **19.7**.

[**4**] Mathematical properties of the Lorenz model are highly nontrivial to study; even to establish the existence of the nontrivial attractor is not easy **20.3**. The existence of a physically observable invariant measure (introduction **29.4**) is hard to prove, although the binary Ising spin coding of dynamics on the Lorenz attractor by Shimada **20.7** was very suggestive.

Therefore, mathematically more transparent **geometrical models 20.2/templates 20.6** were studied. The latter are used to establish the existence of knotted orbits **20.1**.

The Lorenz system is not the usual chaotic system. For example, it is very likely to lack the tracing property **20.5**.

[**5**] The idea of the **strange attractor** (Definition, for example, **21.5**) was introduced by **Ruelle and Takens 21.2** to demonstrate that scenarios different from Landau's leading to turbulent flows exist. They constructed an example of a flow in T^4 and later in T^3 **21.3** (but actually observing them even numerically is almost impossible **21.4**).

[**6**] An endomorphism of an interval was used by Lorenz to show that his result is not due to simple numerical errors **19.4**. Also May pointed out such simple systems exhibit chaotic behaviors **22.1**. Li and Yorke published a paper, "Period three implies chaos' **22.2**. The simplest example of the interval map is illustrated in detail in **22.4** (you understand the essence of chaos if you understand this unit). Since I did not like Li-Yorke chaos, I introduced a more natural definition **22.10** equivalent to now popular definitions and showed necessary and sufficient conditions (e.g., "Period $\neq 2^n$ implies chaos") for a C^0-endomorphism to exhibit chaos **22.14**, **22.15**.

For periodic orbits of a C^0-endomorphism of an interval **Sarkovski's theorem 22.26** tells us the universal ordering of the appearance of periods.

[**7**] Other famous systems show up with more general discussion: baker's transformation **27.1**, horseshoe **28.1**, Bernoulli shift **34.3**, etc.

Part III: Conceptual tools

The portion is 23-36, which is more or less conceptual: symbolic dynamics, algorithmic randomness, Brudno's theorem, baker's transformation and horseshoes, ergodicity, entropy, Lyapunov indices, thermodynamic formalism, etc.

[**1**] What is the most natural characterization of chaotic dynamical systems (see **24.8**)? My intuition is: if (observable) orbits have natural relation to (say, after an appropriate coding) random number sequences, the system is chaotic.

To make this statement meaningful, we need precise mathematization (conceptual analysis) of 'randomness.' To this end algorithmic random numbers are introduced **23.23**; this requires clarification of algorithm and **computation**. Thus we have to go all the way back to **Church** **23.6**-**23.14** and **Turing** **23.16**-**23.19**.

[**2**] We use the most powerful machine **UTM** **23.20** and compress the number sequence. If you cannot compress it significantly, the sequence is random **23.23**. Roughly speaking, when we discretize a system along the time axis (say, with the aid of the Poincaré map **6.3**), the needed length of the shortest program to reproduce the code sequence divided by its duration time (the length of the sequence) is the **complexity of the trajectory**.

[**3**] If we can make faithful mapping (homomorphism) of a dynamical system to a shift (introduced in **22.6**; more formally **26.1**), we use the latter as a code sequence to analyze the trajectory. If we cannot information-compress it, then the sequence is random and the trajectory is chaotic.

One problem is that there is no way to judge whether a given sequence is random or not generally **23.24**, but collectively we can say, e.g., a set consists of mostly random numbers (for example, we can say that binary expansion of $\omega \in [0, 1]$ almost surely gives a random number).

[4] **Brudno's theorem** **24.4** tells us that the Kolmogorov-Sinai entropy (informally introduced in **17.18**, **17.19**) of a (measure-theoretical) dynamical system is identical to the (average) complexity of the trajectories.

This is probably the best characterization of chaos, or chaotic dynamical system at least for measure-theoretical systems (informally **1.2**; Lect 29, esp **29.1**). (If no measure is introduced, we can say a dynamical system is chaotic, if its topological entropy **32.19** is positive.)

[**5**] Whether a dynamical system itself is computable (e.g., can we compute the trajectory position at time 10?) is usually not discussed, but if a theory is a part of physics, its outcome must be compared with observations. If we demand some quantitative agreements, we must be able to compute the numerical outcomes of a theory. Thus the question whether the answer is numerically computable becomes a crucial question (**computable analysis** Lect 25). The prediction must be given in terms of computable reals (**25.13**, effective limits **25.12** of computable rational sequences **25.11**).

A noteworthy point is that even if a function is twice differentiable, its second derivative may not be numerically evaluated **25.21**. What is its implication in physics (say, Newton's equation of motion)?

[**6**] Although a time-discrete dynamical system always has a time continuous counterpart (constructed by suspension) whose Poincaré map can give the original system, a discretized time-continuous system may not be able to recover the original time-continuous system. However, for almost all natural systems we may go back and forth freely between the two descriptions (especially for statistical behaviors). Thus, study of symbol sequences or symbolic dynamics is quite important (as we have already seen in [III4]). Formally they are **shift dynamical systems** **26.2**. Its subclass called **Markov subshifts** **26.6** is quite important. Shift dynamical systems may be interpreted as 1D lattice equilibrium statistical mechanical models (entropy per spin = the KS entropy, for example) [Thermodynamic formalism]. Consequently the theory of Gibbs measures **36.5** becomes crucial. Since it is 1D the transfer matrix **36.6** is important, and the Perron-Frobenius theorem is a key (**26.11** or **35.10**).

[**7**] A typical use of symbolic dynamics is illustrated with the aid of **baker's transformation** **27.1** and **Smale's horseshoe** **28.1**. Horseshoes appear everywhere we see chaotic behavior; as we can see from Poincaré's celestial mechanical studies **16.3** homo- and heteroclinic crossings are everywhere (e.g., see **16.9**).

[**8**] For a given dynamical system usually there are infinitely (very often uncountably many) distinct **invariant measures** **29.5**, **29.6** (also a summary: **36.1**; For "What is measure?" see **29.12**-). For each invariant measure, that may be interpreted as a particular stationary state of the underlying dynamical system, we can make a measure-theoretical dynamical system **29.8**.

In physics observability is of superb importance. If an invariant measure is absolutely continuous **29.9**, it is very likely to be observable (numerically, or in actual experiments).

[**9**] **Ergodicity** **29.10** and **mixing property** **29.11** are properties of measure-theoretical dynamical systems, so invariant measures must be explicitly specified; topological transitivity and mixing **26.4** are topological counterparts of these concepts, but when we are interested in expectation values as in statistical mechanics relevant concepts are always measure-theoretical.

[**10**] Thus, when one says a system is ergodic in classical statistical mechanics, one means the Liouville measure is ergodic; this is almost never proved for any interesting systems. Although it is an irrelevant question for the foundation of statistical

thermodynamics, even it were relevant, we must note that an invariant measure is selected by the initial condition, so the invariant measure is subordinate to the sampling measure of the initial conditions. Everybody knows that the choice of the initial condition has nothing to do with the system dynamics. This clearly tells us the meaninglessness of the ergodicity question in statistical mechanics.

[**11**] The most important theorems related to the system ergodicity is **Birkhoff's ergodic theorem** **30.4** and **Poincaré's recurrence theorem** **30.2**. Zermelo used the latter to unravel Boltzmann's logical weak point in his second law argument **30.3** (see also its tragicomical history **30.11**). Note that Birkhoff's theorem has no direct relation to the ergodicity of the system (read the original theorem statements).

[**12**] We know already (see [III4]) the superb importance of information in understanding dynamical systems. An intuitive approach to the **Kolmogorov-Sinai entropy** **31.2**, Rokhlin's formula, etc., as well as a formal definition through partitions may be found in Lect 32 (e.g., **32.6**). We use a special partition called the generator **32.10**.

Around here why information is quantified by Shannon's formula is explained through Sanov's theorem **31.8**.

[**13**] Krieger's theorem **32.12** about coding of a trajectory anticipates Brudno's theorem. Practically, the most important theorem is the **Shannon-McMillan-Breiman theorem** **32.13** which states the relation between the size (measure) of the elements of the partition and the KS entropy. The theorem applied to Bernoulli systems is the **asymptotic equipartition theorem** **32.14** that justifies the principle of equal probability.

[**14**] Chaotic systems exhibit orbit instability: nearby orbits separate from each other exponentially. This extent (or the parting rate) is measured by the **Lyapunov characteristic number** (LCN) or indices **33.1** (related to the Lyapunov exponent for periodic orbits **6.5**). **Oseledec's multiplicative ergodic theorem** **33.2** guarantees their existence and initial-condition independence (if the system is ergodic); this theorem may be most conveniently proved with the aid of **Kingman's subadditive ergodic theorem** **33.8**.

[**15**] LCN is closely related to the KS entropy as expected from the Shannon-McMillan-Breiman theorem; if the system is sufficiently smooth, then the sum of positive LCN is equal to the KS entropy (**Pesin's theorem** **33.6**).

[**16**] Introduction of information into dynamical systems by Kolmogorov as an **iso-**

morphism invariant **34.1** led to an outstanding question: is it **complete**? This culminated in **Sinai's and Ornstein's theorems** **34.5**: For Bernoulli systems **34.3** the KS entropy is a complete invariant.

The original proof is very constructive (and long), but recent 'soft-proofs' are much shorter.

[**17**] We can study empirical time averages of observables as a function of the time span **35.1**, **35.2** (the **large deviation approach**). Extending Sanov's theorem to the current situation **35.6**, we can derive Rokhlin's theorem **31.2** and Pesin's theorem **33.6**.

[**18**] Since the code sequences and spin configurations on 1-lattices are one-to-one correspondent, Sinai introduced the **thermodynamic formalism**, in which entropy per spin = the KS entropy. This is closely related to the Fredholm theory of the Perron-Frobenius equation **36.11**. Even we could introduce temperature **36.16** that seems to be related to the Hausdorff dimension **2.25** of the support of the invariant measure.

In terms of the Fredholm determinant an outstanding conjecture may be **36.18**: the multiplicity of eigenvalue 1 is the number of distinct physical invariant measures **36.14** in the Kolmogorov sense (= stability against adding noises).

This approach tells us **36.13** that log of the expansion rate of the unstable manifold corresponds to the Hamiltonian.

[**19**] How can we observe strange attractors? This is answered by **Takens' embedding theorem** **37.1**. For example, plotting $(2n+1)$-vectors consisting of consecutively obtained $2n+1$ observed values of a scalar observable in $2n+1$-space, generically we can reconstruct the n-strange attractor of the system. The theorem is experimentally useful. Lect 37 is grossly incomplete: I believe a bit more restrictive but intuitively provable (i.e., that theoretical physicists can 'prove') statement is possible, but is not yet written up.

Part IV: from Peixoto to Palis

The last part is an outline of the modern theory of dynamical systems from Peixoto to Palis: Morse-Smale systems, Axiom A and Anosov systems and technical tools such as shadowing and Markov partition and SRB measures. Palis' conjecture about what we can find in the world conclude the lectures.

[**1**] For C^1 vector fields on 2-manifolds (differentiable), Peixoto asked a necessary

and sufficient condition for the structural stability **38.2** of the flows. Peixoto proved (with his wife) **Peixoto's theorem 38.3**: A vector field $X \in \mathcal{X}^1(M)$ is structurally stable if and only if: (i) there are only finitely many singularities (all hyperbolic), (ii) the limit set consists of fixed points or hyperbolic limit cycles (iii) without any saddle connections.

[2] The structural stability of X satisfying these condition (that is, the sufficiency part **38.7**) is proved explicitly classifying the possible 'patches of X and studying all of them one by one.

The sufficiency part **38.14** is proved through showing that any X may be converted to a field satisfying the itemized conditions above with arbitrarily small perturbations. If the original field X is structurally stable, it must have had the same features as after the perturbation.

[3] Peixoto's theorem was complete and clean, so it ignited interests of many good mathematicians including Smale. Smale wondered what happened on high-dimensional manifolds **38.4**. He defined the **Morse-Smale system 38.5**: (MS) Nonwandering sets consist of periodic points; (MS2) they are all hyperbolic; (MS3) the stable and unstable manifolds are always transversal. Peixoto's theorem may be stated: on 2-mfd flows are structurally stable iff MS.

[4] Since horseshoes can live on 2-mfd, it can be in a Poincaré map of a flow on 3-mfd, and since horseshoes are structurally stable (see Fig. 28.4), already on 3-mfd Peixoto's theorem does not hold.

[5] From Peixoto's theorem (and its proof) we see that hyperbolicity and transversality are crucial. Also horseshoes are structurally stable and appear 'everywhere,' Smale introduce another class of dynamical system **Axiom A 39.1**: a diffeo f satisfies Axiom A iff its nonwandering set $\Omega = \overline{\text{Periodic points of } f}$ and hyperbolic.

[6] The nonwandering set of an Axiom A diffeomorphism consists of invariant pieces (spectral decomposition theorem **39.8**). Intuitively (from the figures in the units) it is clear that we can introduce local 'canonical' coordinate systems consistent with local stable and unstable manifolds **39.6**. Using these coordinates, we can show the Axiom A systems have the **tracing property 39.13**. Using this and the local canonical coordinates, we can construct a **Markov partition 39.17**. In terms of this partition we can map M to a set of Markov subsequence, and map the original dynamical system isomorphically to a Markov subshift **39.22**.

[7] [This portion has not been explained.] Intuitively speaking the coding due to the constructed Markov partition is 'local' in the sense that the distance between x and y in the real space is monotonically reflected on the closeness of the corresponding code sequences. In the original space the 'Hamiltonian' $-\log L_+$ is a function of the position (i.e., totally localized in space without any interaction across space) but dynamics may be strongly correlated. After coding, a function of space spreads over symbols, but due to the local nature of the map explained above, the new Hamiltonian is still short-ranged (decaying exponentially). Thus, we can use the usual 1D thermodynamics to make the Gibbs state, which mapped back to the original space is an absolutely continuous invariant measure, which is the **Sinai-Ruelle-Bowen (SRB) measure** (mentioned in **2.29**), (thanks to Birkhoff's theorem **30.4**) because it is ergodic.

This is the standard approach, but as a theoretical physicist, I wish to go directly from the free energy expression to the canonical distribution **36.13**.

[**8**] If $\Omega = M$ Axiom A systems are called **Anosov systems** **40.1**. The cleanest example is the linear toral diffeomorphisms (group automorphisms) **40.4** including the famous Thom's map (called erroneously Arnold's cat map) **40.5**. In these cases Markov partitions are rather easy to construct **40.6**.

[**9**] As an example of applications of dynamical systems to 'standard physics problems' self-similar spectrum of almost periodic 1d-lattice discrete Schrödinger problem is discussed. Although the example uses a rather acrobatic relation between the original and the dynamical system descriptions, the Cantor structure is exhibited without any room for doubt by the presence of Horseshoes or related structure **41.6**.

To try to understand difference equations as a diffeo problem is often useful.

[**10**] Initially, Smale thought Morse-Smale systems are sufficiently general dynamical systems (generic, hopefully open dense, but at least dense) in the totality of dynamical systems. This was only true on a 2-mfd as Peixoto's theorem indicates; His own horseshoe, that can live in a flow of T^3 and that is structurally stable, destroyed the idea that Morse-Smale is dense or open in any dimension higher than 2. Thus. Smale proposed Axiom A, and expected such systems are at least Ω-stable (i.e., the nonwandering sets are stable, even if the whole system is not).

[**11**] However, again very soon he realized that if there is a cycle **43.7** connecting the basic sets (see **39.8**) of the system, Axiom A systems cannot even Ω-stable: there is an Ω-explosion **43.8**, although such examples are no more dangerous with arbitrarily small perturbations. Thus, generic picture is intact.

[**12**] Then, came a surprise: Newhouse phenomenon: if there is a homoclinic or

heteroclinic tangency **42.2**, then there is an open set of dynamical systems that have infinitely many sinks **42.5**.

[**13**] This is shown by demonstrating two things (1) homoclinic tangency can be stable under any small perturbations and (2) if s system has a homoclinic tangency with arbitrary small perturbations it can be converted to a system with infinitely many sinks,

(1) is shown by crossing of stable and unstable foliations packed close to the tangent point (**42.6**-**42.10**). (2) is shown by studying how horseshoe emerges **42.13**.

[**14**] As we will see up to 2-mfd for maps (3-mfd for flows) the Newhouse phenomenon is the 'worst.' Then, a multidimensional extension of Smale's Ω-explosion example **43.8** was discovered to be made stable: the heterodimensional cycles **43.9**. This time, there is a set of systems with stable saddle connections. Thus, Axiom A + no cycle condition is not dense not open.

[**15**] A separate question is the characterization of structural stability, the Peixoto's original question.

Palis and Smale conjectured: Axiom A + strong transversality iff structural stability. For C^1 dynamical systems (both maps and flows) this is now a theorem (Robbin+Robinson, Mañe, Hayashi) **44.1**.

This means there is a chance for physicists to encounter structurally unstable systems, since they make an open set.

[**16**] As to the genericity question or the question about the 'common' systems we encounter in the world, Palis formulated the following conjecture (**Palis conjecture**) **44.3**:

1. Every C^r-diffeo of a compact mfd M can be C^r-approximated by one of the following:

(a) a hyperbolic system (Axiom A with strong transversality)

(b) a system with heterodimensional cycle **43.9**

(c) a system exhibiting a homoclinic tangency **42.2**.

2. If M is 2D (for maps), then (a) or (c) occurs (in other words, Palis conjectured that avoiding homoclinic bifurcation, Peixotos picture can be recovered in one dimension higher space).

2 has been proved for $r = 1$.

[**17**] The YouTube movie Chaos illustrates another Palis conjecture **44.2**: There is a C^r-dense set of dynamical systems with finitely many attractors whose union of basins of attraction has total probability. These attractors support physical mea-

sures.