

Fermi National Accelerator Laboratory

FERMILAB-Conf-89/24

Phase Space Concepts*

Leo Michelotti
Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

January 1989

*To be published in the proceedings of the 1987 US Particle Accelerator Summer School, Fermi National Accelerator Laboratory, Batavia, Illinois, August 3-14, 1987.



Operated by Universities Research Association Inc. under contract with the United States Department of Energy

PHASE SPACE CONCEPTS

Leo Michelotti

Fermi National Accelerator Laboratory

If we want to know where Jupiter will be so as to plan properly the Jupiter shot, then we may proceed in one mathematical direction. If we are interested in whether the solar system is dynamically stable or unstable, we will have to proceed in another. In view of the inherent difficulties of the mathematics, the art of modelling is that of adopting the proper strategy.

— Philip J. Davis and Reuben Hersh
The Mathematical Experience

1 RESONANCES ON INVARIANT TORI.

Much has been written about a revolution taking place in the asymptotic analysis of dynamical systems, usually in association with beautiful, color photographs of objects like chaotic attractors, fractal basin boundaries, and other such Julia-Fatou sets. The first shots of this revolution were fired, however, not in the 1980's or 1970's but towards the close of the nineteenth century. By 1892, Henri Poincaré already had published his landmark work *Les Méthodes Nouvelles de la Mécanique Céleste* in which he advanced the theses that differential equations should be viewed as *geometric* objects, in particular, as vector fields on manifolds, and that questions concerning the long term stability of a dynamical system might be attacked by studying the topological properties of these objects. His work even led him to recognize the extraordinarily complicated behavior of orbits in the vicinity of a separatrix, what today we would call "chaos." Much like his predecessor Newton, Poincaré found that the ideas and language which he needed did not yet exist and that he had to create entirely new mathematics in order to progress. In time the seeds which he planted grew into branches of modern topology, with all its trappings of tangent and cotangent bundles, differential forms, exterior algebra and calculus, homology and cohomology — all of which are frequently associated with general relativity, string theories, or gauge theories, but are *almost never mentioned in connection with one of their sources, good old classical mechanics*. (This is a fate shared by the ideas of Sophus Lie, which generally are not introduced until the study of quantum angular momentum or $SU(3)$.)

In these lectures we shall look into this geometric approach to the study of Hamiltonian dynamical systems, especially in connection with the kinds of problems which arise in accelerator orbit theory. This is a vast subject, and we certainly shall not be able to treat it as fully or as carefully as it deserves. In recent years a number of books have been published on dynamics, and the reader who wants to learn more will find some of these titles included in the bibliography. Our own relatively modest goals will be to delineate the idea of invariant tori in phase space, to define and illustrate the importance of resonant orbits and their separatrices as structures for organizing dynamics, and to touch upon the meaning of chaotic orbits in nonintegrable systems. I hope, further, to lessen the impression

which some may have acquired during their formal education that classical mechanics is a dull, closed subject with no mysteries left to explore.

1.1 Manifolds, mappings, and vector fields.

A series of lectures with the title *Phase Space Concepts* should begin with a good, rigorous, explicit definition of “phase space.” Regrettably, doing this would require a full course in differential topology, so we shall settle for a bad, heuristic, implicit definition and shift our focus to the objects which live on phase spaces: dynamical systems.

A dynamical system is abstractly associated with two different but related mathematical objects: (1) a vector field and (2) a group of mappings. Both are defined on a manifold, \mathcal{M} , which is the “phase space” of the system. Let us deal with these three objects one at a time.

manifold: Everyone has an intuitive understanding of “manifolds” as spaces which look Euclidean on small scales, a formal generalization of surfaces, such as spheres, cylinders, or tori. A localized region of a manifold is thus representable as an open subset of R^n , called a “chart.” A collection of charts which covers the entire manifold, with some overlap between charts so we know how to patch them together smoothly, is called, appropriately enough, an “atlas.” Having said this, we must also include manifolds having exotic properties, such as nonorientability — the Mobius strip or the Klein bottle — or multiple connectivity — the torus. Although all are admissible in principle, no physically relevant system has been constructed on Mobius strips or Klein bottles; tori, however, are another matter. If two charts overlap, there must exist a “smooth” transformation connecting the coordinates associated with points that belong to both charts. If $\underline{z} \in R^n$ and $\underline{z}' \in R^n$ are the two n -tuple representatives of a point $\mathbf{p} \in \mathcal{M}$, then the transformation is a one-to-one, highly differentiable function, $T: \underline{z} \mapsto \underline{z}'$. The key idea which these elaborate constructions is meant to convey is this: \mathcal{M} is an object which can be represented, or coordinatized, in a large number of ways but which we want to think of as an entity independent of all these representations.

vector field: A good heuristic image of a “vector field defined over a manifold” is conveyed by picturing a surface with a tangent vector attached to each point. The principal tool required to make that image precise is the atlas of charts used to define the manifold. Since a manifold is specified by an atlas of charts, we can define a vector field over it by giving its representation, that is, its components, on each chart and making sure that everything is consistent and gets patched together smoothly as you jump from one chart to another. To insure that this is a geometric object requires a set of instructions for transforming components of vector fields when you change the coordinates representing points on the manifold. We shall go into this in a little more detail later. The set of all possible tangent vectors which can be attached to a point $\mathbf{p} \in \mathcal{M}$ is itself a vector space which is typically labelled $T\mathcal{M}_{\mathbf{p}}$, the “tangent space” at \mathbf{p} ; the union of all these vector spaces is the “tangent bundle” associated with the manifold, $T\mathcal{M} \equiv \bigcup_{\mathbf{p} \in \mathcal{M}} T\mathcal{M}_{\mathbf{p}}$.¹ (A particular vector field is called a “cross-section” of this bundle by mathematicians.)

group of mappings: A mapping, ϕ , is a function which takes the manifold into itself: $\phi: \mathcal{M} \rightarrow \mathcal{M}$. Of course, we tacitly assume that the mappings under consideration are “nicely” behaved: in particular, we avoid mappings which have discontinuities or are otherwise not sufficiently “smooth.” We want to interpret a special set of mappings in terms of the time-evolution of a dynamical system. This set is indexed by R^2 , and we interpret

¹This is oversimplifying a little: the tangent bundle also contains the underlying base manifold as a projection. My purpose here is to motivate without slowing for details which would be necessary in a scholarly presentation.

the particular map, $\phi_{t_2 t_1} : \mathcal{M} \rightarrow \mathcal{M}$, as a time-evolution operator which takes a system from its state at time t_1 to its state at time t_2 . In order for this to be the case, we must insist that the set obey a group-like property known as the Chapman-Kolmogorov equation:

$$\phi_{t_3 t_2} \circ \phi_{t_2 t_1} = \phi_{t_3 t_1} .$$

From this we see easily that, for all t , $\phi_t = \text{id}$. Notice that reversibility is also implicit (unless we restrict ourselves to $t_2 > t_1$) since we have that $\phi_{t_2 t_1}^{-1} = \phi_{t_1 t_2}$. This is not quite a group: concatenation requires at least one “time” in common, t_2 in the above expression. However, if the system is periodic, so that for some τ and all t_1, t_2 we have $\phi_{t_2+\tau t_1+\tau} = \phi_{t_2 t_1}$, then we can construct a discrete group of mappings by taking as generator a period advance map, $\Phi \equiv \phi_{t+\tau t}$, for some t . The Chapman-Kolmogorov equation implies that the subset $\{\Phi^n \mid n \in \mathbb{Z}\}$ forms a group. The different groups that we get by varying the base time, t , are all isomorphic to each other. If the system is autonomous or time-independent, so that $\phi_{t_2+\tau t_1+\tau} = \phi_{t_2 t_1}$ for all values of the index parameters, then we can build a continuous group of mappings by defining, $\phi_\tau \equiv \phi_{t+\tau t}$. The Chapman-Kolmogorov equation then implies the group properties,

$$\phi_0 = \text{id} , \quad \phi_\tau \circ \phi_{\tau'} = \phi_{\tau+\tau'} , \quad \text{and} \quad \phi_\tau^{-1} = \phi_{-\tau} . \quad (1)$$

Comment 1: We have assumed that the index, τ , is a real variable, but it is possible to derive some deep theorems by enlarging its domain to the complex plane. In fact, Eq.(1) requires only that τ be a member of an (additive) Abelian group: this could be the integers, a module, the reals, rationals, p-adics, or whatever. Eq.(1) then essentially describes a morphism between the index group and the group of mappings.

These objects are the mathematical building blocks used to develop a geometrical theory of dynamical systems. We shall not employ their full power here, but they would be necessary to prove rigorously the heuristically plausible assertions which will be made throughout this paper. At any rate, it is good to know that they exist and could be called upon if necessary.

1.2 Dynamics.

Dynamics is introduced when we interpret these geometric structures as containing information about the behavior of a physical system. A dynamical system is actually associated with more than one manifold: First, there is the configuration manifold which labels the instantaneous “position” of the system without information as to how it is changing. For example, the configuration manifold of a harmonic oscillator is (topologically equivalent to) R , the set of reals, while that of a pendulum would be (topologically equivalent to) a circle, S^1 . The tangent bundle of the configuration manifold contains both positional and velocity information about the system. It is itself a manifold: for the harmonic oscillator, a plane, $R \times R = R^2$; for the pendulum, a cylinder, $S^1 \times R$. If instead of using the vector spaces $T\mathcal{M}_p$ one uses their duals, labelled $T\mathcal{M}_p^*$, then one is working with the cotangent bundle of the configuration manifold, which is again itself a manifold (generally equivalent to the tangent bundle). Typically, the tangent bundle is associated with (*position, velocity*) information and Lagrangians, while the cotangent bundle is associated with (*position, momentum*) and Hamiltonians. At the level at which we shall work in this paper, such nice distinctions do not matter, and we shall ignore them. The important idea that we shall try to motivate is this: differential equations of motion are the components

of a representation of a *geometric* object, a vector field on either the tangent bundle or the cotangent bundle of a configuration manifold. (Put another way, a cross-section of the tangent bundle of either a tangent bundle or a cotangent bundle; it gets a little involved.)

Having touched upon these points, it is time to reflect on an important example.

1.2.1 Harmonic oscillators.

The archetypal dynamical system is the harmonic oscillator, the phase space of which is R^2 , with interpretation,

$$\underline{z} \in R^2, \quad \underline{z} \equiv \begin{pmatrix} z \\ p \end{pmatrix} .$$

With momentum defined as $p = m\dot{z}$, and including a damping term, Newton's law takes the familiar form,

$$\dot{p} = -\Gamma p - m\omega^2 z,$$

which we write as a vector field over phase space.

$$\dot{\underline{z}} = \begin{pmatrix} \dot{z} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} 0 & 1/m \\ -m\omega^2 & -\Gamma \end{pmatrix} \underline{z} \equiv \mathbf{A}\underline{z} \quad (2)$$

In order to analyze the flow of orbits implied by Eq.(2) it is easiest to do an eigenanalysis of \mathbf{A} .

$$\begin{aligned} \det[\mathbf{A} - \lambda \mathbf{1}] &= \lambda^2 + \Gamma\lambda + \omega^2 = 0 \\ \Rightarrow \lambda_{\pm} &= \frac{1}{2} \left[-\Gamma \pm \sqrt{\Gamma^2 - 4\omega^2} \right] \\ \text{Eigenvectors: } \begin{pmatrix} z \\ p \end{pmatrix} &\propto \begin{pmatrix} 1 \\ \lambda m \end{pmatrix} \end{aligned}$$

Consider the nature of the flow for various regions in the (ω, Γ) control space. For $\Gamma \gg |2\omega|$, we have $\lambda_+ \approx 0^-$ and $\lambda_- \approx -\Gamma$. There is one and only one zero-dimensional invariant subset of phase space: the origin itself is the only fixed point. The two eigenvectors of \mathbf{A} lie along invariant, one-dimensional submanifolds. Orbits on one of these, corresponding to λ_- , move rapidly (since Γ is large) toward the origin, while those on the other approach the origin more slowly. Any other orbit is, by linearity, a superposition of these two motions: it will be dominated by the "fast" direction until it approaches the vicinity of the "slow" invariant manifold. The two one-dimensional invariant submanifolds thus form a *separatrix*, partitioning phase space into the two regions seen in Figure 1. The orbits in one region overshoot the origin and must reverse themselves, while those in the other move monotonically toward the origin.

If we now reduce the control parameter Γ , a bifurcation occurs when $\Gamma = |2\omega|$. At this critical point \mathbf{A} has only one eigenvalue and one eigenvector. The middle region of monotonic flow toward the origin has vanished; all orbits, except the origin itself, overshoot. Under these conditions \mathbf{A} cannot be diagonalized by a similarity transformation: at best, it is brought to its Jordan canonical form via the transformation,

$$\begin{aligned} \mathbf{V} &\equiv \begin{pmatrix} 1 & 0 \\ -m\omega & m \end{pmatrix} \\ \mathbf{V}^{-1}\mathbf{A}\mathbf{V} &= \begin{pmatrix} -\Gamma/2 & 1 \\ 0 & -\Gamma/2 \end{pmatrix} . \end{aligned}$$

If we continue to decrease Γ below 2ω , the eigenvalues and eigenvectors of \mathbf{A} will be complex, and orbits spiral in towards the origin. The spirals get tighter as $\Gamma \rightarrow 0$ until

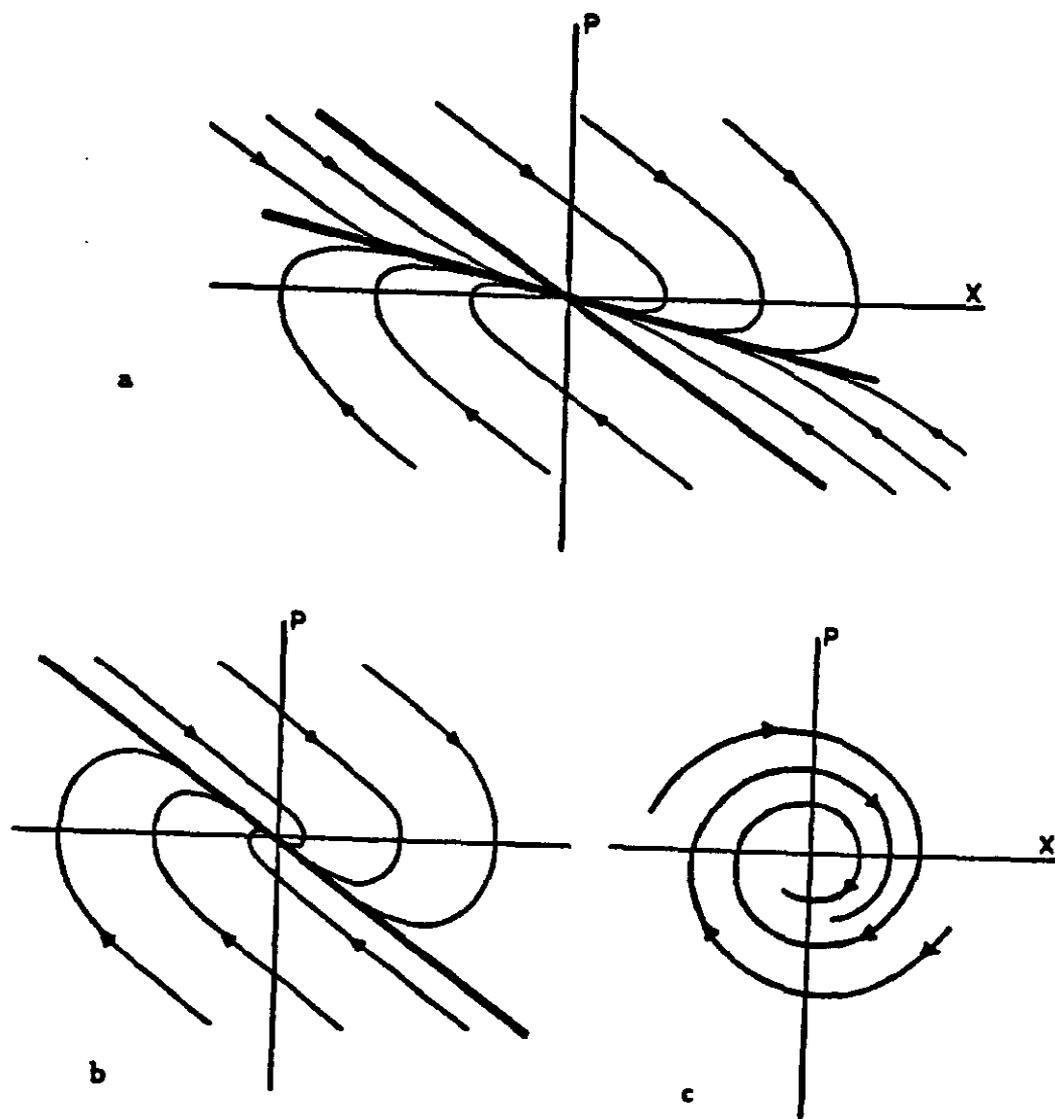


Figure 1: Phase space of a damped harmonic oscillator. (a) Overdamped, (b) critically damped, (c) underdamped.

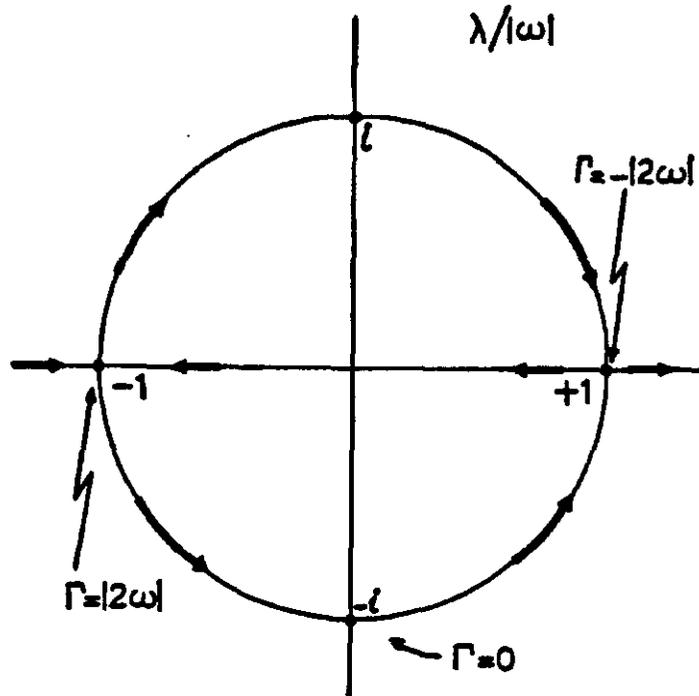


Figure 2: Trajectory of eigenvalues of the damped harmonic oscillator as Γ decreases from $+\infty$ to $-\infty$. Two bifurcations occur: (a) vanishing of the separatrix at $\Gamma = \pm|2\omega|$, and (b) a Hopf-like bifurcation, but without limit cycle, at $\Gamma = 0$.

finally at $\Gamma = 0$ all of phase space splits into a collection of invariant circles, the flow of the undamped harmonic oscillator. In principle we could continue this process and consider $\Gamma < 0$; Figure 2 illustrates the full set of all possible eigenvalues. Orbits would then flow away from the origin rather than toward it. (From Eq.(2) we see that the complete symmetry is: $\Gamma \rightarrow -\Gamma$, $t \rightarrow -t$, $p \rightarrow -p$.)

Notice that this qualitative analysis of the flow, in which we have learned how all the orbits are arranged in phase space and how this organizational structure varies with changes in the control parameters, has taken place *without actually solving the equations of motion*. The objective of this sort of analysis is to seek out special invariant submanifolds which partition phase space and thereby organize the flow of orbits. From them we get a qualitative and semi-quantitative understanding of the dynamics *as a whole*, the kind of information best suited for studying stability or developing statistics on an ensemble.

Of course, in this case obtaining exact, analytic solutions is easy. Since the matrix \mathbf{A} does not depend on t , the integration of Eq.(2) is immediate,

$$\underline{z}(t) = e^{t\mathbf{A}} \underline{z}(0) , \quad (3)$$

and we can identify the time-evolution map with the exponential: $\phi_\tau \leftrightarrow e^{\tau\mathbf{A}}$.

For the undamped case, $\Gamma = 0$, it is easy to confirm that

$$\left(\frac{1}{\omega} \mathbf{A}\right)^2 = -\mathbf{1} ,$$

so that we can write explicitly,

$$e^{t\mathbf{A}} = e^{\omega t(\mathbf{A}/\omega)} = \mathbf{1} \cdot \cos \omega t + \left(\frac{1}{\omega} \mathbf{A}\right) \cdot \sin \omega t \quad (4)$$

Substituting Eq.(4) into Eq.(3) and writing out the components yields the usual freshman physics result:

$$\begin{aligned} z(t) &= z(0) \cos \omega t + \frac{p(0)}{m\omega} \sin \omega t \\ p(t) &= p(0) \cos \omega t - m\omega z(0) \sin \omega t . \end{aligned} \quad (5)$$

Comment 2: The bifurcation which takes place as the eigenvalues of \mathbf{A} cross the imaginary axis and the origin changes from being an attractor to being a repeller is a variant of the Hopf bifurcation. In this particular case it does nothing interesting; linearity makes the entire phase space change simultaneously. In the presence of nonlinearities, as with the Van der Pol oscillator, it can lead to the creation of a limit cycle. For an easily readable discussion of this, see Gilmore.[8]

Comment 3: Eq.(4) is easily generalized. By the Cayley-Hamilton theorem, any square matrix satisfies its own characteristic equation. It follows immediately that for a $n \times n$ square matrix, \mathbf{A} , $\exp t\mathbf{A}$ is a $(n-1)$ -th degree polynomial in \mathbf{A} . In particular, for any 2×2 matrix, \mathbf{A} we can write

$$e^{t\mathbf{A}} = F(t)\mathbf{1} + G(t)\mathbf{A} ,$$

where F and G obey the following composition rules.

$$\begin{aligned} F(t+t') &= F(t)F(t') - |\mathbf{A}|G(t)G(t') \\ G(t+t') &= G(t)F(t') + F(t)G(t') + \text{Tr}[\mathbf{A}]G(t)G(t') \end{aligned}$$

Comment 4: The reader who considers linear systems too easy for serious study might like to consider the following statement.

“Suppose all the eigenvalues ... of the linear equation $\dot{\underline{x}} = \mathbf{A}\underline{x}$, $\underline{x} \in R^n$, $\mathbf{A} : R^n \rightarrow R^n$, are purely imaginary. Then under what conditions are two such equations topologically equivalent? The answer to this question is not known, and evidently the problem cannot be solved by presently available mathematical methods.” [2]

Wonderful! One can establish a life's work and a rewarding career merely by studying harmonic oscillators, a strategy not entirely unknown to some scientists.

1.2.2 Hamiltonian dynamics.

In addition to linearity, the (undamped) harmonic oscillator possesses a second important property: it is a Hamiltonian system. That is, its vector field can be derived from a real-valued function, H , via Hamilton's equations. In particular, the Hamiltonian for a collection of independent harmonic oscillators and the components of the corresponding vector field can be written in the familiar form,

$$H = \sum_k \left(p_k^2 / 2m_k + \frac{1}{2} m_k \omega_k^2 x_k^2 \right) \quad (6)$$

$$\begin{aligned}\dot{x}_k &= \frac{\partial H}{\partial p_k} = p_k/m_k \\ \dot{p}_k &= -\frac{\partial H}{\partial x_k} = -m_k \omega_k^2 x_k\end{aligned}$$

Hamilton's equations can be written more economically by using the $2N$ -tuple of chart coordinates directly, as displayed below in block form.

$$\begin{aligned}\underline{z} &= \begin{pmatrix} \underline{x} \\ \underline{p} \end{pmatrix} \\ \dot{\underline{z}} &= \mathbf{J} \cdot \partial H / \partial \underline{z} \\ \mathbf{J} &\equiv \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix}\end{aligned}\tag{7}$$

Now, let f be any observable, that is, any real-valued function defined over phase space, $f: \mathcal{M} \rightarrow R$. The "time"-derivative of f as observed by a system on an orbit of H is called a **Lie-derivative** and is evaluated using the usual chain rule.

$$\begin{aligned}\dot{f} &= \frac{\partial f}{\partial \underline{z}} \cdot \dot{\underline{z}} + \frac{\partial f}{\partial t} \\ &= \frac{\partial f}{\partial \underline{z}} \cdot \mathbf{J} \cdot \frac{\partial H}{\partial \underline{z}} + \frac{\partial f}{\partial t} \\ &\equiv [f, H] + \frac{\partial f}{\partial t}\end{aligned}$$

The object $[f, H]$ is, of course, the **Poisson bracket**, sometimes called the **Lie bracket**, of f with H . By interpreting H as *any* observable, this definition is generalized to form a binary operation on the space of observables whose fundamental properties are:

bilinearity: $[f, g + h] = [f, g] + [f, h]$ and $[f + g, h] = [f, h] + [g, h]$;

antisymmetry: $[f, g] = -[g, f]$;

Jacobi identity: $[f, [g, h]] + [g, [h, f]] + [h, [f, g]] = 0$.

Another fruitful construction is the **adjoint operator algebra**. With each observable f we associate a unary operator \hat{f} which acts on the space of observables as follows:

$$\hat{f}: g \mapsto [f, g]$$

Because of the bilinear property, \hat{f} is a linear operator.

Comment 5: Any binary operator that is bilinear, antisymmetric, and obeys a Jacobi identity belongs, by definition, to a **Lie algebra**. Exponentiation, as in Eq.(8) below, leads to a **Lie group**. It is in this way that Lie groups and Lie algebras were brought into the world. Because they are usually presented to students in connection with quantum theories — angular momentum, $SU(3)$, and so forth — their initial connection with dynamical systems, and especially with Hamiltonian dynamics, tends to be forgotten. *The appropriate place to introduce Lie groups is in a course on classical mechanics.*

Comment 6: It is an easy exercise to show that, because of the Jacobi identity, the commutator algebra of the adjoint operators is a homomorphism of the original Lie algebra.

In this way, any Lie algebra is homomorphic to a commutator algebra.

Comment 7: Dirac's quantization procedure establishes a morphism between the adjoint algebra of a classical dynamical system and the operator algebra of its quantum analog, i.e., between Poisson brackets and commutators.[7] What is frequently not mentioned in introductory quantum mechanics courses is that this procedure fails for any system more complicated than harmonic oscillators.² Loosely speaking, this is the content of a theorem, due to Groenwald and Van Hove, which should be much more widely known; a very readable discussion of it (albeit with some errors) can be found in Guillemin and Sternberg.[10]

If we now interpret the z_k and p_k themselves as *observables* rather than real numbers, then Hamilton's equations of motion can be written,

$$\dot{z} = [z, H] = -\hat{H}z$$

The solution is *formally* straightforward when H does not depend explicitly on t , although the detailed evaluation may be formidable, or even impossible:

$$z(t) = e^{-t\hat{H}}z(0) . \quad (8)$$

This provides us with the formal identification, $\phi_\tau \leftrightarrow e^{-\tau\hat{H}}$.

Comment 8: Meditate on the following specious argument, remaining ever on the alert for plausible sounding nonsense, and resolve its implied paradox. Since \hat{H} is a linear operator, for any two orbits $z_1(t)$ and $z_2(t)$ we have,

$$\begin{aligned} \frac{d}{dt}(z_1(t) + z_2(t)) &= \dot{z}_1(t) + \dot{z}_2(t) \\ &= \hat{H}z_1(t) + \hat{H}z_2(t) \\ &= \hat{H}(z_1(t) + z_2(t)) . \end{aligned}$$

Therefore $z_1 + z_2$ is also an orbit (!!!): it obeys the same Hamilton's equations of motion. We appear to have proved that *all* Hamiltonian systems are linear, a result of enormous benefit to humanity. Why, in fact, is this *not* a proof?

Example: Constant force field. Let us see how Eq.(8) might work in practice. Consider the Hamiltonian

$$H = p^2/2m + mgx$$

describing a non-relativistic particle moving under the influence of a constant gravitational field. Let us construct an orbit using Eq.(8) .

$$\begin{aligned} \hat{H}z &= [H, z] = [p^2/2m, z] \\ &= -p/m \\ \hat{H}^2z &= \hat{H}(-p/m) \\ &= [mgx, -p/m] \\ &= -g \end{aligned}$$

²More precisely, it is impossible to extend it to observables that are cubic or higher order polynomials in z and p .

$$\begin{aligned}\hat{H}^3 \mathbf{z} &= 0 \\ \therefore \mathbf{z}(t) &= e^{-t\hat{H}} \mathbf{z}(0) = \left(1 - t\hat{H} + \frac{1}{2}t^2\hat{H}^2\right) \mathbf{z}(0) \\ &= \mathbf{z}(0) + (p(0)/m)t - \frac{1}{2}gt^2\end{aligned}$$

Example: Harmonic oscillator. For the harmonic oscillator, we have

$$\begin{aligned}\hat{H}\mathbf{z} &= [H, \mathbf{z}] = [p^2/2m, \mathbf{z}] \\ &= -p/m \\ \hat{H}^2\mathbf{z} &= -\hat{H}p/m = -[\frac{1}{2}m\omega^2 x^2, p/m] \\ &= -\omega^2 \mathbf{z}\end{aligned}$$

By iterating these two equations, we find

$$\begin{aligned}(-t\hat{H})^{2n}\mathbf{z} &= (-1)^n(\omega t)^{2n}\mathbf{z}, \\ (-t\hat{H})^{2n+1}\mathbf{z} &= (-1)^n(\omega t)^{2n+1}p/m\omega.\end{aligned}$$

We can now write the full solution, starting from Eq.(8).

$$\begin{aligned}\mathbf{z}(t) &= e^{-t\hat{H}} \mathbf{z}(0) = \sum_{n=0}^{\infty} \frac{1}{n!} (-t\hat{H})^n \mathbf{z}(0) \\ &= \sum_{n=0}^{\infty} \frac{1}{2n!} (-t\hat{H})^{2n} \mathbf{z}(0) + \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} (-t\hat{H})^{2n+1} \mathbf{z}(0) \\ &= \left(\sum_{n=0}^{\infty} (-1)^n \frac{(\omega t)^{2n}}{2n!}\right) \mathbf{z}(0) + \left(\sum_{n=0}^{\infty} (-1)^n \frac{(\omega t)^{2n+1}}{(2n+1)!}\right) p(0)/m\omega \\ \mathbf{z}(t) &= \mathbf{z}(0) \cos \omega t + \frac{p(0)}{m\omega} \sin \omega t\end{aligned}\tag{9}$$

Comment 9: This result is in complete agreement with and appears formally identical to Eq.(5). However, the interpretation is very different. In Eq.(5), the symbols $\mathbf{z}(0)$ and $p(0)$ represent real numbers; in Eq.(9) they represent observables, real-valued functions over \mathcal{M} . (The difference between these two points of view is analogous to the difference between the Schrödinger and Heisenberg pictures in quantum mechanics.) The *number* $\mathbf{z}(0)$ contains no information about the *number* $p(0)$, but the *function* $\mathbf{z}(0)$ does know that the *function* $p(0)$ is its canonical conjugate. This explains how we can start from $e^{-t\hat{H}} \mathbf{z}(0)$ and end with an answer which contains both $\mathbf{z}(0)$ and $p(0)$.

Equations of motion are the components of a vector field associated with a particular chart (coordinate system). Changing charts, say $T: \underline{z} \mapsto \underline{z}'$, alters these components, much as rotating the basis of a vector space alters the component representations of the vectors. A straightforward application of the chain rule of differentiation produces the new components.

$$\begin{aligned}\dot{\underline{z}}' &= (\partial \underline{z}' / \partial \underline{z}) \cdot \dot{\underline{z}} \\ &= (\partial \underline{z}' / \partial \underline{z}) \cdot \mathbf{J} \cdot (\partial H / \partial \underline{z}) \\ &= [(\partial \underline{z}' / \partial \underline{z}) \cdot \mathbf{J} \cdot (\partial \underline{z}' / \partial \underline{z})^T] (\partial H / \partial \underline{z}') \\ &\equiv \mathbf{J}'(\underline{z}') \cdot (\partial H / \partial \underline{z}')\end{aligned}\tag{10}$$

Poisson brackets transform in a similar way.

$$\begin{aligned}
[f, g] &= \partial f / \partial \underline{z} \cdot \mathbf{J} \cdot \partial g / \partial \underline{z} \\
&= (\partial f / \partial \underline{z}') \cdot [(\partial \underline{z}' / \partial \underline{z}) \cdot \mathbf{J} \cdot (\partial \underline{z}' / \partial \underline{z})^T] \cdot (\partial g / \partial \underline{z}') \\
&= (\partial f / \partial \underline{z}') \cdot \mathbf{J}'(\underline{z}') \cdot (\partial g / \partial \underline{z}')
\end{aligned} \tag{11}$$

These transformation equations attest to the fact that the bracket is a geometric invariant, a scalar; this will be discussed further in the next section.

If T is designed so that $\mathbf{J}'(\underline{z}') = \mathbf{J}$, that is, if the Jacobian matrix, $\mathbf{M} = \partial \underline{z}' / \partial \underline{z}$, satisfies

$$\mathbf{M} \mathbf{J} \mathbf{M}^T = \mathbf{J}, \tag{12}$$

then the forms for evaluating both the equations of motion and the brackets are the same on both charts. In such a case, T is called a canonical transformation, and \mathbf{M} is called a symplectic matrix. It is easy to show (do it!) that the set of all symplectic matrices is a group, the symplectic group.

EXAMPLE: Linear mappings on the configuration space. One of the simplest non-trivial examples consists of a linear transformation which does not mix the \underline{z} and \underline{p} coordinates. This is written in block form with $n \times n$ matrices \mathbf{m}_{11} and \mathbf{m}_{22} .

$$\begin{aligned}
\mathbf{M} &= \begin{pmatrix} \mathbf{m}_{11} & 0 \\ 0 & \mathbf{m}_{22} \end{pmatrix} \\
\mathbf{M} \mathbf{J} \mathbf{M}^T &= \begin{pmatrix} 0 & \mathbf{m}_{11} \mathbf{m}_{22}^T \\ -\mathbf{m}_{22} \mathbf{m}_{11}^T & 0 \end{pmatrix}
\end{aligned} \tag{13}$$

Therefore, a necessary and sufficient condition that this be a canonical transformation is that $\mathbf{m}_{11} \mathbf{m}_{22}^T = 1$. In particular, if this is just a scaling transformation, $\mathbf{m}_{11} = \lambda \mathbf{1}$, $\mathbf{m}_{22} = \mu \mathbf{1}$, then we require that $\lambda = 1/\mu$.

EXAMPLE: Scaled harmonic oscillator. A scaling transformation of this type puts the harmonic oscillator Hamiltonian, Eq.(8), into degenerate form. Take $(\mathbf{m}_{11})_{ij} = \delta_{ij} \sqrt{m_j \omega_j}$ and $(\mathbf{m}_{22})_{ij} = \delta_{ij} / \sqrt{m_j \omega_j}$, so that

$$\begin{aligned}
H &= \sum_k \left(p_k^2 / 2m_k + \frac{1}{2} m_k \omega_k^2 x_k^2 \right) \\
&= \sum_k \frac{1}{2} \omega_k \left((p_k / \sqrt{m_k \omega_k})^2 + (\sqrt{m_k \omega_k} x_k)^2 \right) \\
&= \sum_k \frac{1}{2} \omega_k \left(p_k'^2 + x_k'^2 \right)
\end{aligned} \tag{14}$$

EXAMPLE: Polar (action-angle) coordinates. An even more useful chart uses polar, or action-angle, coordinates.

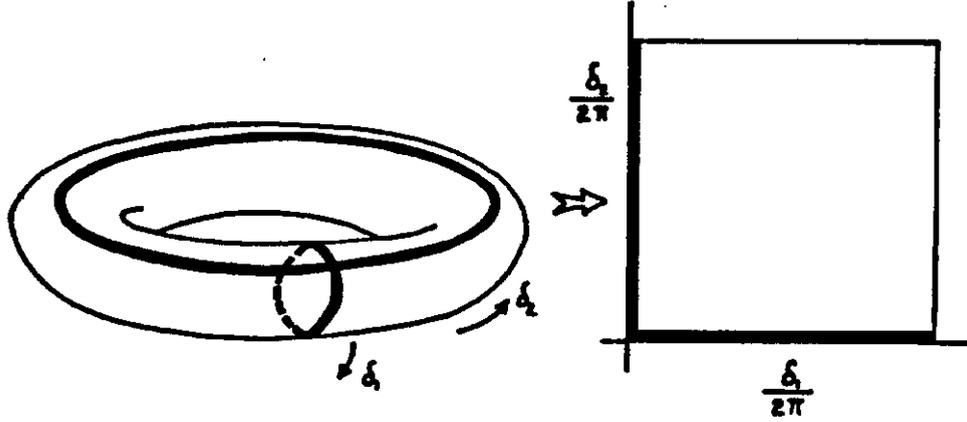


Figure 3: Angle coordinates on a torus.

$$z'_k = \sqrt{2I_k} \sin \delta_k \quad (15)$$

$$p'_k = \sqrt{2I_k} \cos \delta_k \quad (15)$$

$$M|_{k\text{th block}} = \partial(z'_k p'_k) / \partial(\delta_k I_k) \quad (16)$$

$$= \begin{pmatrix} p'_k & z'_k / 2I_k \\ -z'_k & p'_k / 2I_k \end{pmatrix}$$

It is now a trivial exercise to show that the symplectic condition Eq.(12) is indeed satisfied, so that the $\underline{\delta}$ and \underline{I} variables obey Hamilton's equations of motion. From Eq.(14), the Hamiltonian, expressed in these polar coordinates, is written,

$$H = \sum_k \omega_k I_k \equiv \underline{\omega} \cdot \underline{I} \quad (17)$$

Because the transformation is canonical, the coordinates of the vector field — also referred to as the equations of motion — are expressed according to the usual procedure.

$$\begin{aligned} \dot{\underline{I}} &= -\partial H / \partial \underline{\delta} = \underline{0}, \\ \dot{\underline{\delta}} &= \partial H / \partial \underline{I} = \underline{\omega} \end{aligned} \quad (18)$$

Eq's.(18) are parametric equations describing motion on an N -dimensional torus imbedded in a $2N$ -dimensional phase space. The "action" or "amplitude" variables, which are constants of the motion, serve to *label the torus*, and the "angle" or "phase" variables locate position on the torus. (See Figure 3) This angle chart is equivalent to slicing the torus along N independent directions and flattening it out to fit into a unit square in Euclidean N -space. This is, of course, just the identification of a torus as topologically equivalent to

R^N/Z^N . That orbits must lie on tori can also be seen from the following argument. The orbit of a single harmonic oscillator is topologically equivalent to a circle. Thus, the orbit of N independent oscillators lies on a surface that is the cross product of N copies of a circle, a torus.

EXAMPLE: Eigencoordinates. If we relax our implicit assumption that coordinates are real, then we can introduce a transformation,

$$a_k \equiv \frac{1}{\sqrt{2}}(p'_k + iz'_k) = \sqrt{I_k} e^{i\delta_k} \quad a_k^* \equiv \frac{1}{\sqrt{2}}(p'_k - iz'_k) = \sqrt{I_k} e^{-i\delta_k},$$

in terms of which the Hamiltonian in Eq.(14) or Eq.(17) is written,

$$H = \sum_k \omega_k a_k^* a_k .$$

The Jacobian matrix of this transformation is

$$\mathbf{M}|_{k\text{th block}} = \partial(a_k a_k^*) / \partial(z'_k p'_k) = \sqrt{1/2} \begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix}$$

from which we see immediately that $\mathbf{J}' = \mathbf{M}\mathbf{J}\mathbf{M}^T = i\mathbf{J}$. Eq.(10) then yields the equations of motion.

$$\dot{a}_k = i\partial H / \partial a_k^* = i\omega_k a_k, \quad \dot{a}_k^* = -i\partial H / \partial a_k = -i\omega_k a_k^*$$

These are the "eigencoordinates" of the harmonic oscillator problem; they could have been obtained by diagonalizing the matrix \mathbf{A} which appears in Eq.(2) (after zeroing Γ). Upon quantization they become the familiar annihilation and creation operators.

Comment 1: One of the most fundamental properties of Hamiltonian flow is that it induces an automorphism of the bracket algebra.

$$e^{-t\hat{H}} [f, g] = [e^{-t\hat{H}} f, e^{-t\hat{H}} g] \quad (19)$$

This means that the time-evolution map is canonical, and it is intimately connected with the fact that the Liouville measure over phase space,

$$d\mu = dx_1 \wedge dx_2 \wedge \dots \wedge dx_N \wedge dp_1 \wedge dp_2 \wedge \dots \wedge dp_N \quad (20)$$

is a dynamical invariant. Because of this, finding the interesting structure in a Hamiltonian dynamical system is much more difficult than in a dissipative system, where one can start practically anywhere in phase space and end up on attractors: Hamiltonian dynamical systems do not have attractors because Liouville measure must be conserved.

Comment 2: Physicists who learned classical mechanics from the first edition of Goldstein[9] are most familiar with canonical transformations in connection with generating functions rather than directly from the symplectic group. For example, the transformations in Eq.(13) arise from a generating function,

$$S(\underline{z}, \underline{p}') = \underline{p}' \cdot \mathbf{M} \cdot \underline{z} , \quad (21)$$

while those of Eq.(15) can be obtained from the generating function,

$$S(\underline{z}', \underline{\delta}) = \sum_k \frac{1}{2} z'_k{}^2 \cot \delta_k .$$

It is a straightforward but tedious exercise to confirm directly that generating functions do indeed produce canonical transformations.

1.3 Geometry.

“Linearity” is one of the most fundamental properties one can attribute to a dynamical system. A typical formulation goes something like this: A dynamical system is “linear” means that if $\underline{u}(t)$ and $\underline{v}(t)$ are orbits of the system, then $\underline{u}(t) + \underline{v}(t)$ is one also. Eq.(2) serves as the archetypal example of this. By long familiarity we are lulled into thinking this definition a simple, straightforward matter, but it is not. To begin with, it depends critically on representation; otherwise, what is meant in general by “adding” two orbits together? Consider, for example, a vector field, $\underline{v}(\underline{z}, t)$, given by

$$\underline{z} = \begin{pmatrix} x \\ p \end{pmatrix}, \quad \underline{v} = \dot{\underline{z}} = \begin{pmatrix} p \\ \bar{z} \end{pmatrix}$$

where \bar{z} is obtained from

$$t[\bar{z}(1-x^2) + 3x\dot{z}^2] - 2\dot{z}(1-x^2) + 9t^5x(1-x^2)^2 = 0. \quad (22)$$

This appears to be a complicated equation describing a system that is neither autonomous nor linear. However, the substitutions

$$t \leftarrow t^{1/3}, \quad x \leftarrow x \sqrt{\frac{1}{1+x^2}}$$

change Eq.(22) to

$$\bar{z} + x = 0.$$

The linearity of this system was disguised by a poor choice of (extended phase space) coordinates in which to represent it, that is, an inappropriate chart. Without knowledge of the transformation to the “correct” coordinates, how would one establish this property? Put another way, *is there a coordinate-free way of characterizing linearity?*

One can ask the same question about the Hamiltonian property. To treat this well would require a course in the topology of differential manifolds, but the “flavor” of the arguments can be appreciated without going into details.[13,1] To begin with, the way that analytic objects are made into geometric objects is to write their transformation laws under changes in chart coordinates and to recognize that these allow one to make statements — or write equations — which transform covariantly. For example, the inner product construction in an n-dimensional real vector space is defined, in the usual way,

$$(\underline{x}, \underline{y}) = \sum_i x_i y_i = \underline{x} \cdot \underline{y}. \quad (23)$$

If we confine ourselves to charts related by orthogonal transformations, then this defines a legitimate geometric object, an invariant of the group $O(n)$.

$$\forall \mathbf{M} \in O(n) : \underline{x}' = \mathbf{M}\underline{x} \quad \& \quad \underline{y}' = \mathbf{M}\underline{y} \implies \underline{x}' \cdot \underline{y}' = \underline{x} \cdot \mathbf{M}^T \mathbf{M} \cdot \underline{y} = \underline{x} \cdot \underline{y}$$

However, if we enlarge the allowed transformations and let \mathbf{M} be arbitrary, $\mathbf{M} \in GL(n)$, then Eq.(23) no longer defines a geometric scalar: it is not invariant under $GL(n)$. To correct it we must introduce a new object, the metric tensor, \mathbf{g} , with transformation laws appropriate to reestablish the invariance.

$$(\underline{x}, \underline{y}) \equiv \underline{x} \cdot \mathbf{g} \cdot \underline{y}$$

$$\begin{aligned}
(\underline{z}', \underline{y}') &= \underline{z}' \cdot \underline{g}' \cdot \underline{y}' \\
&= \underline{z} \cdot \mathbf{M}^T \underline{g}' \mathbf{M} \cdot \underline{y} \\
&= \underline{z} \cdot \underline{g} \cdot \underline{y} \\
&= (\underline{z}, \underline{y})
\end{aligned}$$

This reestablishes the quadratic form $(\underline{z}, \underline{y})$ as a geometric (scalar) object, a "two-form," at least under $GL(n)$. If it is symmetric and positive definite, so that \underline{g} is a symmetric, positive definite matrix, then matrix theory tells us that the eigenvalues of \underline{g} are positive reals and that \underline{g} is diagonalizable. That being the case, there will be some chart on which the form of the inner product reverts back to being the dot product of Eq.(23). That is, for some $\mathbf{M} \in GL(n)$ we shall have $\mathbf{M}^T \underline{g}' \mathbf{M} = \mathbf{1}$. This statement is a sort of "representation theorem": any symmetric, positive definite two-form (i.e., an inner product) must look like an ordinary dot product on some allowed chart.

The key issue, then, is the rule which governs the transformation of the components of a vector under a change of coordinates.³ Put another way, how should the basis vectors of $T\mathcal{M}_p$ change under a chart alteration? This can be answered in many ways, but the "natural" basis vectors attached to a point in a manifold will transform like the operator $\partial/\partial \underline{z}$. Thus, $T: \underline{z} \mapsto \underline{z}'$ induces the "natural" basis transformation $T: \{\bar{e}_k \mid k = 1 \dots N\} \mapsto \{\bar{e}_k' \mid k = 1 \dots N\}$, where, using Einstein's summation convention,

$$\bar{e}_k' \mapsto \partial/\partial z'^k = (\partial z^j / \partial z'^k) \partial/\partial z^j \mapsto (\partial z^j / \partial z'^k) \bar{e}_j \quad (24)$$

Despite this rather mysterious formulation the rule is heuristically pleasing; it is, in fact, nothing more than what we all learned as sophomores under another guise. That this is so is seen most easily by doing an

Example: Egg carton transformation. Consider the transformation illustrated in Figure 4.

$$\begin{aligned}
z_1 &\equiv z_1' + \cos \theta z_2' \\
z_2 &\equiv \sin \theta z_2'
\end{aligned}$$

From the matrix,

$$\partial \underline{z} / \partial \underline{z}' = \begin{pmatrix} 1 & \cos \theta \\ 0 & \sin \theta \end{pmatrix},$$

and Eq.(24) we get the correspondence,

$$\begin{aligned}
\bar{e}_1' &= \bar{e}_1 \\
\bar{e}_2' &= \cos \theta \bar{e}_1 + \sin \theta \bar{e}_2
\end{aligned}$$

Example: Spherical coordinates. A more familiar transformation arises in the change from Cartesian to spherical coordinates on any chart that does not include the origin. Reverting to the most common notation, we have

$$\begin{aligned}
x &= r \sin \theta \cos \varphi \\
y &= r \sin \theta \sin \varphi \\
z &= r \cos \theta
\end{aligned}$$

³This statement is reminiscent of the "Erlangen programme" announced by Felix Klein in 1872. According to Klein, what distinguished one kind of geometry from another was the group of transformations under which its propositions remained valid. This idea was used to classify geometrical theorems until the advent of Riemannian geometry, which evidently did not fit into the scheme.

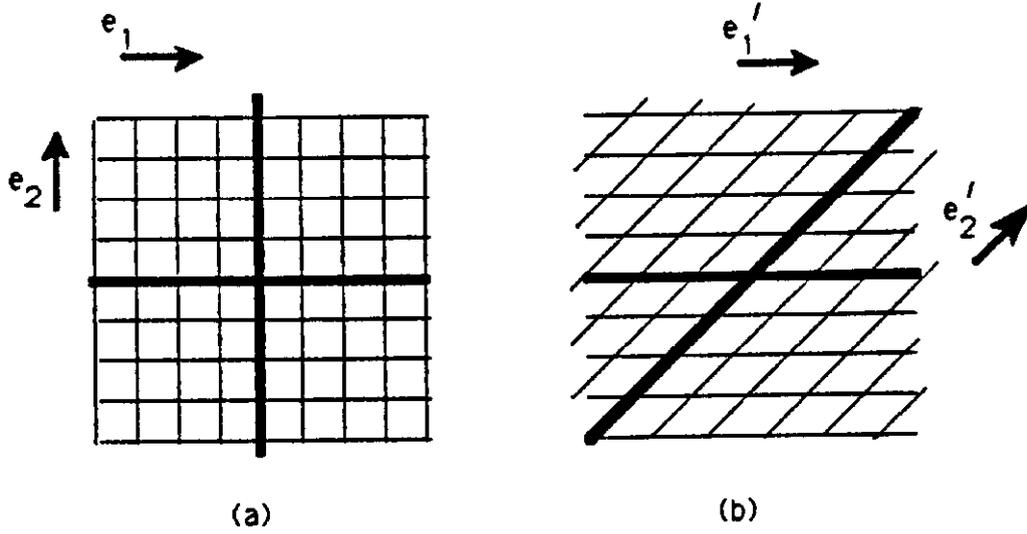


Figure 4: Natural basis vectors transform like $\partial/\partial z$.

from which we get

$$\frac{\partial(x, y, z)}{\partial(r, \theta, \varphi)} = \begin{pmatrix} x/r & r \cos \theta \cos \varphi & -r \sin \theta \sin \varphi \\ y/r & r \cos \theta \sin \varphi & r \sin \theta \cos \varphi \\ z/r & -r \sin \theta & 0 \end{pmatrix} .$$

Combined with Eq.(24) , this yields the usual result.

$$\begin{aligned} \vec{e}_r &= \frac{1}{r}(x\hat{i} + y\hat{j} + z\hat{k}) = \hat{r} \\ \vec{e}_\theta &= r(\cos \theta \cos \varphi \hat{i} + \cos \theta \sin \varphi \hat{j} - \sin \theta \hat{k}) = r\hat{\theta} \\ \vec{e}_\varphi &= r(-\sin \theta \sin \varphi \hat{i} + \sin \theta \cos \varphi \hat{j}) = r\hat{\varphi} \end{aligned}$$

Had we used $(r, r\theta, r\varphi)$ as the new coordinates, $\vec{e}_{r\theta}$ and $\vec{e}_{r\varphi}$ would have been dimensionless as well.

Components of a vector transform contravariantly to the basis.

$$\begin{aligned} \vec{v} &= v^j \vec{e}_j = v'^k \vec{e}'_k = v'^k (\partial z^j / \partial z'^k) \vec{e}_j \\ v^j &= (\partial z^j / \partial z'^k) v'^k \quad , \text{ or in matrix form,} \\ \underline{v} &= (\partial \underline{z} / \partial \underline{z}') \cdot \underline{v}' \end{aligned}$$

Comparing this to Eq.(10) we see that Hamilton's equations do indeed transform as the components of a vector provided that \mathbf{J} is identified as a tensor rather than a constant (scalar) matrix. Similarly, the Poisson bracket, as defined in Eq.(11) , is a true geometrical scalar. We emphasize again: seen in this way, an ordinary differential equation is the component representation of a vector field, a geometric object, associated with a particular

chart. All the ways, and more, in which we wrote differential equations for the harmonic oscillator come about from describing one and the same vector field from different points of view.

Having recognized that the objects which appear in Hamiltonian dynamics are actually geometrical constructs, the programme for fully geometrizing the theory proceeds by simultaneously (a) turning this development on its head and (b) making the notation and language as obscure as possible. Rather than starting from Hamilton's equations of motion, we abstractly postulate the existence of (1) an even-dimensional manifold, with its associated tangent and cotangent bundles, (2) a "symplectic structure," which is a closed, anti-symmetric, non-degenerate two-form (quadratic form) acting on tangent vectors, and (3) a flow⁴ which is a symmetry of the symplectic structure. The symplectic structure eventually becomes identified with Poisson brackets, and the flow is, by definition, locally Hamiltonian. A deep theorem, due to Darboux, then brings us full circle by assuring us that with these geometrical structures in place, there must exist some chart in which the components of the flow's vector field can be obtained from a single function, the Hamiltonian, H , according to the usual prescription; that is, there is a set of coordinates for which Hamilton's equations take their usual form. This "representation theorem" is similar, both in statement and proof, to the one we formulated earlier on the inner product. Readers who wish to pursue these ideas in detail may enjoy reading the third chapter of Abraham and Marsden.[1]

Comment 12: There are trivial examples of symplectic structure. Consider the complex plane, $C \simeq R^2$, as a vector space over the reals. On it we define the two-form $\gamma(a, b) \equiv \text{Im}[ab^*]$. Because γ is bilinear, nondegenerate, and anti-symmetric (verify!) it qualifies as the two-form of a symplectic manifold, (C, γ) , with symmetry group $O(2) \simeq SU(1)$, $T_\theta : z \mapsto e^{i\theta} z$.

Comment 13: Amazingly, it is possible for a dynamical system to be locally Hamiltonian everywhere but not globally Hamiltonian. We cannot describe here what this means, much less why it is so; once again I refer you to Abraham and Marsden[1] for details.

1.4 The geometry of resonance.

The nature of the orbits on an invariant torus specified by Eq's.(18) is critically important and depends on how close the $\underline{\omega}$ of the torus is to a resonance, the *analytical* definition of which is given by the equation

$$\underline{m} \cdot \underline{\omega} + n = 0 \tag{25}$$

where n is some integer and $\underline{m} = (m_1, m_2, \dots, m_N)$ is a multiplet of integers. In most of the textbooks, this is written with n set to zero, $\underline{m} \cdot \underline{\omega} = 0$. This latter form is relevant to either autonomous or, equivalently, "averaged" Hamiltonians; the one that we adopt, Eq.(25), appears in the analysis of Hamiltonians which are periodic in "time." It appears in accelerator theory while studying Hamiltonians which describe transverse motion of a particle in a periodic structure, such as a synchrotron: after the usual Floquet transformation, the linearized dynamics becomes equivalent to a harmonic oscillator, but the periodicity of the environment survives in the transformed nonlinear terms. Here the independent variable is not "time," t , but an azimuthal angle, θ , whose natural period is 2π . All the discussions in this paper are carried out with this application lurking in the background. We therefore implicitly assume 2π periodicity. The components of $\underline{\omega}$ are the winding numbers of orbits on the torus; put another way, they are the number of oscillations (i.e., the tune) undergone by the corresponding angle coordinate while increasing t (or θ) by 2π .

⁴The one-parameter family of mappings obeying the Chapman-Kolmogorov equation.

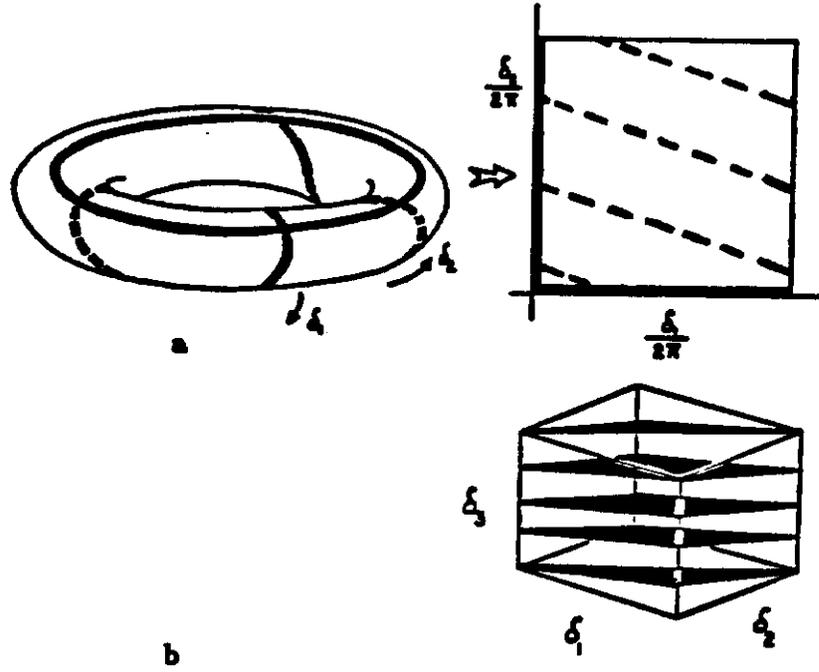


Figure 5: Regularly sampled points on a resonant orbit lie on an $N - 1$ -dimensional sub-manifold of the N -dimensional torus: (a) an $\omega_1 + 3\omega_2$ resonance. (b) $-\omega_1 + \omega_2 + 4\omega_3$, a six-dimensional phase space resonance.

In order to visualize the implications of this condition in phase space it is better to sample the orbit at the regular intervals $t_k = 2\pi k$ than to follow it continuously as it winds its way around the torus.

$$\begin{aligned} \underline{\delta} |_{t=2\pi k} &= \underline{\delta} |_{t=0} + 2\pi k \underline{\omega} \\ \frac{1}{2\pi} (\underline{\delta} |_{t=2\pi k} - \underline{\delta} |_{t=0}) &= k \underline{\omega} \simeq [k \underline{\omega}] \end{aligned} \quad (26)$$

Here, $[z]$ represents the fractional part of z , and for any N -tuple, $\underline{\alpha}$,

$$\begin{aligned} [\underline{\alpha}] &\equiv ([\alpha_1], [\alpha_2], \dots, [\alpha_N]) \\ &= \underline{\alpha} \bmod U^N, \end{aligned}$$

where $U^N \equiv \otimes^N(0, 1)$ is the N -dimensional unit cube.

If the $\underline{\omega}$ of an invariant torus obeys a resonance condition, as in Eq.(25), then from Eq.(26) we get,

$$\frac{1}{2\pi} (\underline{m} \cdot \underline{\delta} |_{t=2\pi k} - \underline{m} \cdot \underline{\delta} |_{t=0}) \simeq [k \underline{m} \cdot \underline{\omega}] = [-k n] = 0.$$

This means that the sampled points of the orbit all lie on an $N - 1$ -dimensional sub-torus parametrized by an equation of the form, $\underline{m} \cdot \underline{\delta} = \text{constant}$, as illustrated in Figure 5. If $\underline{\omega}$ satisfies two inequivalent resonance conditions, then they are contained within the intersection of two $N - 1$ -dimensional sub-tori, the connected parts of which are $N - 2$ -dimensional

sub-tori. In general, if k independent resonance conditions are satisfied, orbit samples lie on a family of $N - k$ -dimensional tori; if $k = N$, the orbit is periodic.

Comment 14: There is no substitute for sitting down with a pencil and paper and sketching a handful of these resonant orbits in $(\delta_1/2\pi, \delta_2/2\pi)$ space. What happens, for example, when m_1, m_2 , and n are not co-prime?

On the other hand, if ω obeys no resonance condition, then the sampled points will fill the N -dimensional invariant torus: the closure of this set is, in fact, the torus. If you watched this set develop for an ω that was "close to" but not exactly "on" resonance, then you would see a speckled band form in the vicinity of the resonance line and slowly expand until it filled the torus. For a small number of samples, the influence of the nearby resonance would be evident, but it would eventually wash out as the number increased without bound. That is a consequence of the following fundamental theorem, whose proof originates with Weyl[26]:

Zeroth ergodic theorem. If ω is non-resonant, then the set of points

$$S \equiv \{ [k\omega] \mid k \in \mathbb{Z} \}$$

is "uniformly distributed" over U^N .

Comment 15: Interestingly, if no resonance condition is satisfied, then the orbits are said to be "quasi-periodic," a definition which seems completely backwards: assuredly, the resonant orbits are the ones which should be given this appellation, but history and mathematicians have deemed it otherwise.

To understand and interpret this theorem, we must first describe a little more carefully what is meant by a "uniform distribution" of points in phase space. This requires that we first specify a measure over phase space, relative to which densities can be gauged. Fortunately, Hamiltonian systems come equipped with a natural, dynamically invariant measure, the Liouville measure of Eq.(20). Since the transformation to the action-angle coordinates of Eq.(15) is canonical, and since canonical transformations preserve the Liouville measure⁵, on a torus we must have

$$d\mu|_{\text{torus}} = d\delta_1 \wedge d\delta_2 \wedge \dots \wedge d\delta_N$$

which, apart from $(2\pi)^N$ normalization, is the usual Lebesgue measure on U^N . This is fortunate: it means that a set of points on an invariant torus in phase space will be uniformly distributed relative to the Liouville measure if and only if the corresponding points on the angle chart, U^N , are uniformly distributed relative to ordinary Lebesgue measure. To be definite, we shall assume a normalization: $\mu(U^N) = 1$, which means that we associate a point in the unit cube, U^N , with the coordinates $\delta/2\pi$.

Consider finite sections of the set S ,

$$S_l \equiv \{ [k\omega] \mid n = 1 \dots l \} \subset S,$$

and let $B \subset U^N$ be any μ -measurable subset of U^N . We denote the number of elements in a set W as $\text{card}(W)$. Then the statement "S is uniformly distributed" means that

$$\text{for all } B: \lim_{l \rightarrow \infty} l^{-1} \text{card}(S_l \cap B) = \mu(B)$$

⁵I have not proved or even demonstrated this, but you know it already.

Counting the number of elements in a set, W , can be turned into an analytic problem by introducing the characteristic functions of W .

$$K_W(\underline{x}) \equiv \begin{cases} 0, & \underline{x} \notin W \\ 1, & \underline{x} \in W \end{cases}$$

The exact condition defining uniform distribution is then written:

$$\text{for all } B : \lim_{l \rightarrow \infty} l^{-1} \sum_{k=1}^l K_B(\lfloor k\underline{\omega} \rfloor) = \mu(B)$$

In words, the fraction of points falling within a subset is equal to its size, or at least proportional to it for non-compact phase spaces.

Because of the importance of the theorem, we shall give the

Skeleton of a proof: The characteristic function of any measurable subset of U^N can be expanded in Fourier series.

$$K_B(\underline{x}) = \sum_{\underline{m}} c_{\underline{m}} e^{2\pi i \underline{m} \cdot \underline{x}} \text{ almost everywhere over } U^N$$

We then evaluate

$$\begin{aligned} \lim_{l \rightarrow \infty} \frac{1}{l} \sum_{k=1}^l K_B(\lfloor k\underline{\omega} \rfloor) &\stackrel{\text{a.e.}}{=} \lim_{l \rightarrow \infty} \frac{1}{l} \sum_{k=1}^l \sum_{\underline{m}} c_{\underline{m}} e^{2\pi i \underline{m} \cdot \lfloor k\underline{\omega} \rfloor} \\ &= \sum_{\underline{m}} c_{\underline{m}} \left[\lim_{l \rightarrow \infty} \frac{1}{l} \sum_{k=1}^l e^{2\pi i k \underline{m} \cdot \underline{\omega}} \right] \end{aligned}$$

Since $\underline{\omega}$ is non-resonant, by hypothesis, for all $\underline{m} \neq 0$ we have,

$$\begin{aligned} \left| \frac{1}{l} \sum_{k=1}^l (e^{2\pi i \underline{m} \cdot \underline{\omega}})^k \right| &= \frac{1}{l} \left| \frac{e^{2\pi i(l+1)\underline{m} \cdot \underline{\omega}} - e^{2\pi i \underline{m} \cdot \underline{\omega}}}{e^{2\pi i \underline{m} \cdot \underline{\omega}} - 1} \right| \\ &\leq \frac{1}{l} \frac{2}{|e^{2\pi i \underline{m} \cdot \underline{\omega}} - 1|} \\ &= \frac{1}{l} \frac{1}{|\sin \pi \underline{m} \cdot \underline{\omega}|} \\ &\rightarrow 0, \text{ as } l \rightarrow \infty \end{aligned}$$

Therefore, the only term which survives is c_0 . But

$$c_0 = \int_{U^N} d^N \underline{x} K_B(\underline{x}) = \mu(B)$$

So we get the desired result:

$$\lim_{l \rightarrow \infty} \frac{1}{l} \sum_{k=1}^l K_B(\lfloor k\underline{\omega} \rfloor) = \mu(B) \quad - \text{ QED } -$$

We have here a purely geometric description of resonance based on the dimensionality of the closure of a set of orbit samples. Resonance corresponds to dimensional collapse, if you will: resonant orbits lie on lower dimensional tori, non-resonant orbits fill invariant N -tori.

Comment 16: The reader who wishes to verify the details of this skeleton and flesh it out into a *legitimate* proof will find that more than enough theorems exist to help him do so. However, I suspect that a few of these require a strong application of the dreaded Axiom of Choice. It is an interesting point: Is this ergodic theorem still valid if we accept only the weakest form of the Axiom of Choice? Weyl and Poincaré themselves had little patience with people who posed such questions.

Comment 17: There is a delightful application of this theorem to the “first digits problem” of number theory. Let M and b be fixed integers with M not a rational power of b and $b \geq 3$. If the first digit of M^k in base b is d , then

$$\exists p \geq 0: d \times b^p \leq M^k < (d+1) \times b^p$$

Taking the logarithm yields the inequalities:

$$p \leq k \cdot \log_b M - \log_b d < p + \log_b(1 + 1/d)$$

which in turn means that

$$\{ k \cdot \log_b M - \log_b d \} < \log_b(1 + 1/d)$$

Provided M is not a rational power of b , the fraction of numbers M^k , $k = 1, 2, \dots$ whose base b expansion begins with the digit d is therefore $\log_b(1 + 1/d)$. Notice that these numbers sum to 1, as they should.

$$\begin{aligned} \sum_{d=1}^{b-1} \log_b(1 + 1/d) &= \log_b \left(\prod_{d=1}^{b-1} \frac{d+1}{d} \right) \\ &= \log_b b \\ &= 1 \end{aligned}$$

If anyone thinks that he knows something, he has not yet known as he ought to know.

— St. Paul
1 Corinthians 8, 2

The thing I am going to try to explain . . . may be ahead of me. I may be thinking I have got there when I have not. I can only ask [experts] to watch very carefully, and tell me when I go wrong; and others to take what I say with a grain of salt — as something offered, because it may be a help, not because I am certain that I am right.

— C. S. Lewis
Mere Christianity

2 SEPARATRICES.

In the previous lecture we tried to motivate the geometric approach to the theory of dynamical systems, in which ordinary differential equations describing a system are interpreted as components of a vector field. Qualitative analysis of systems proceeds by identifying invariant submanifolds of different dimensions. A good deal of attention was paid to the harmonic oscillator, with special emphasis on the idea that there exist dynamically invariant N -dimensional tori, T^N , imbedded in the $2N$ -dimensional phase space. These tori foliate the phase space, by which it is meant that (a) they are disjoint from one another, and (b) their union is the entire phase space. As such, they can act as the level sets of N coordinates, (action) which thereby serve to label an individual torus; N more coordinates (angle) are then needed to locate position on a torus. Finally, we identified resonance conditions for the winding numbers, $\underline{\omega}$, of orbits. By sampling an orbit with the period advance map, we can recognize *geometrically* whether these *analytic* conditions are satisfied: If $\underline{\omega}$ is off-resonance, then the samples fill a uniformly dense subset of T^N ; their closure is T^N itself. If k resonance conditions are satisfied, then their closure is a sub-torus, $T^{N-k} \subset T^N$. In the limiting case, where N resonance conditions are satisfied, the orbit is periodic.

The natural question to ask at this point is, What other Hamiltonian systems admit such useful foliations?

2.1 Liouville-Arnold theorem.

A simple example of one is a straightforward generalization of Eq's.(18) ; it also describes the world's most benign nonlinear system, the shearing Hamiltonian. Assume that H is a function of \underline{I} only, and is independent of the $\underline{\delta}$ coordinates.

$$\begin{aligned} H &= H(\underline{I}) \equiv \underline{\nu} \cdot \underline{I} + H_s(\underline{I}) & (27) \\ \dot{\underline{I}} &= -\partial H / \partial \underline{\delta} = 0 \\ \dot{\underline{\delta}} &= \partial H / \partial \underline{I} \equiv \underline{\omega}(\underline{I}) = \underline{\nu} + \partial H_s / \partial \underline{I} \end{aligned}$$

Mathematicians call this system a "twist map"; accelerator physicists refer to H_s as the "detuning" or "shear" terms; such a Hamiltonian is sometimes also said to be "in normal form." Their only important effect is to produce amplitude dependent tunes: unlike the harmonic oscillator, in which all tori possessed the same $\underline{\omega}$, now $\underline{\omega}$ depends explicitly on \underline{I} .

Comment 18: The simplest physical example of a shearing Hamiltonian is provided by the pendulum. Its period of oscillation is amplitude dependent, despite Galileo and freshman physics. (Following through on the consequences of that — and the consequences of the consequences, and so forth — can lead to a pleasant life's work in the theory of elliptic functions, Riemann surfaces, and algebraic topology.) If the apocryphal story about Galileo's observations on a swinging lamp is true, I suspect that the reason he believed its period was unvarying was that his pulse rate increased as he became more excited by the discovery.

The N coordinates, I_1, I_2, \dots, I_N , viewed as observables, have three important properties:

(a) They are dynamical invariants; $dI_k/dt = 0$, for all k . Since time does not appear explicitly in the definition of I_k , this is equivalent to saying that $[I_k, H] = 0$, so that each I_k "commutes" with H .

(b) They are in involution, which means that they all "commute": $\forall i, j: [I_i, I_j] = 0$.

(c) They are independent. Formally, the differential forms $\{dI_k, k = 1 \dots N\}$ are linearly independent everywhere. More intuitively, observables are independent when their level surfaces are nowhere tangent: they intersect transversely.

It turns out that these three conditions are all that is required to assure the existence of a local twist map. This is incorporated into the

Liouville-Arnold theorem: Let G_1, G_2, \dots, G_N be N independent, invariant observables in involution defined over a $2N$ dimensional phase space, \mathcal{M} . From dynamical invariance, orbits must lie on level sets,

$$M_{\underline{g}} \equiv \{p \in \mathcal{M} \mid \underline{g} = \underline{G}(p)\}.$$

Each $M_{\underline{g}}$ is a smooth, invariant manifold. If $M_{\underline{g}}$ is compact and connected, then it is diffeomorphic to T^N , an N -torus. Orbits on $M_{\underline{g}}$ are "conditionally periodic." That is, it is possible to introduce a chart of angle coordinates, $\underline{\xi}$, on the torus so that, for some N -tuple of real numbers, $\underline{\omega}_{\underline{g}}$, $\dot{\underline{\xi}} = \underline{\omega}_{\underline{g}}$.

Comment 19: A Hamiltonian over a $2N$ dimensional phase space possessing N invariant observables in involution is called "integrable." This is the classical analog of the quantum mechanical concept of a "complete" set of commuting observables.

Two closed curves on the torus are homotopic if there is a way of continuously deforming one into the other without leaving the torus. Let $\{\gamma_1, \gamma_2, \dots, \gamma_N\}$ be a set of N mutually non-homotopic closed curves in T^N , and define the line integrals,

$$I_k \equiv \frac{1}{2\pi} \oint_{\gamma_k} \underline{p} \cdot d\underline{x}, \quad k = 1, 2, \dots, N.$$

Then it is difficult to show⁶ that: (a) if $\underline{z} = (\underline{x}, \underline{p}) \mapsto \underline{z}' = (\underline{x}', \underline{p}')$ is a canonical transformation, the line integrals evaluate to the same numbers using either set of coordinates, (b) if we change the curve γ_k to another which is homotopic to it, then the value of I_k remains

⁶Here is where the full power of the topological methods and theorems alluded to in the first section is very helpful.

the same,⁷ and (c) the transformation from $(\underline{x}, \underline{p})$ to $(\underline{\delta}, \underline{I})$ is itself canonical. These are, of course, the celebrated action coordinates; the corresponding angle coordinates are then fixed by the bracket conditions, $[\underline{\delta}, \underline{I}] = 1$. Action coordinates possess the additional property of being *adiabatic invariants*. If we adiabatically change the Hamiltonian, say $H_1 \rightarrow H_2$, then orbits on a torus of H_1 will slowly deform into orbits on a torus of H_2 . Which one? The one with the same values of the action coordinates.

This construction does not determine *unique* action-angle coordinates: there is more than one way to slice a torus so that it can be laid out flat onto the unit cube, U^N , or to select a family of homotopically inequivalent curves. We shall see an application of this when we “straighten out” a resonance in a later subsection.

2.2 Theorem of Kolmogorov, Arnold, and Moser.

One of the most fundamental problems of classical mechanics is finding a coordinate chart in which a given Hamiltonian is at least locally representable by a twist map, as in Eq.(27). The Liouville-Arnold theorem tells us that such charts exist if there is a complete, commuting set of observables — i.e., N invariant observables in involution. But this is essentially never the case: almost all dynamical systems are not integrable, a fact not usually given the attention it deserves in academic curricula. Nonetheless, if the system of interest is “near-integrable,” most orbits will still exist on invariant tori and generally act as though they were governed by an integrable vector field. The proof of this remarkable statement was first accomplished by Kolmogorov, refined and publicized by Arnold, and further developed by Moser. It is thus known as the KAM theorem.

Suppose that the (nonintegrable) Hamiltonian H is obtained from a shearing Hamiltonian H_0 via the addition of a “small” perturbation.

$$H(\underline{\delta}^*, \underline{I}^*) = H_0(\underline{I}^*) + \epsilon H_1(\underline{\delta}^*, \underline{I}^*)$$

Here, ϵ is an order parameter which gives the relative scale between the two terms, and $\underline{\delta}^*$ and \underline{I}^* are action-angle coordinates on a chart appropriate for the tori of H_0 . Under what conditions will H possess invariant N -dimensional tori of its own in the vicinity of $\epsilon \approx 0$? The KAM theorem, which partially answers this question, is not easy to state, but the gist of it is this: when H_0 is itself a nonlinear dynamical system (of a certain kind), most invariant tori of H_0 whose winding numbers are sufficiently off resonance will survive small perturbations, they will be deformed into invariant tori of H .

Let us try to be more precise. For this section only I am going to set $n = 0$ in Eq.(25). This is tantamount to saying that we are dealing with autonomous Hamiltonians; a generalization of the KAM theorem exists for the non-autonomous case, but we shall not look into it here.⁸ Let $\underline{\omega}(\underline{I}^*) = \partial H_0(\underline{I}^*) / \partial \underline{I}^*$ be the N -tuple of winding numbers associated with the invariant torus of H_0 labelled by \underline{I}^* , and suppose that H_0 is “sufficiently” nonlinear so that

$$\text{either } |\partial \underline{\omega} / \partial \underline{I}^*| \neq 0 \quad \text{or} \quad \begin{vmatrix} \partial \underline{\omega} / \partial \underline{I}^* & \underline{\omega} \\ \underline{\omega}^T & 0 \end{vmatrix} \neq 0 . \quad (28)$$

The torus is said to be “off-resonance” if there exist real positive γ and η such that for all N -tuples of integers, \underline{m} ,

$$|\underline{m} \cdot \underline{\omega}(\underline{I}^*)| \geq \gamma \|\underline{m}\|^{-\eta} , \quad (29)$$

$$\|\underline{m}\| \equiv \sum_{k=1}^N |m_k| .$$

⁷ I_k is thus identified with a homotopic equivalence class of curves, i.e., an element of the homology group of the torus.

⁸For the purely personal reason that I am not yet comfortable with it. The notation and statement of the theorem which I use here have been combined from Arnold[3] and Thirring[24].

The numbers γ and η are not completely arbitrary: among other things, η is constrained by the number of degrees of freedom and γ depends on ϵ . There are numerous such details which we shall not go into here. The key point is the meaning of "off-resonance" conveyed by the inequality. The first assertion of the theorem is that as $\epsilon \rightarrow 0$ the phase space volume occupied by H_0 tori whose $\underline{\omega}(\underline{I}^*)$ do not obey such a condition becomes arbitrarily small; essentially all H_0 tori will be off-resonance for small ϵ . The second is that those tori for which Eq.(29) is satisfied will be preserved, in the sense that

(a) there exists an invariant torus of H and a local $(\underline{\delta}, \underline{I})$ chart such that

$$\dot{\underline{\delta}} = \underline{\omega}(\underline{I}^*) \equiv \underline{\omega}^0 \quad \text{and} \quad \dot{\underline{I}} = 0,$$

(b) the connection between the two charts is a near-identity transformation:

$$\begin{aligned} \underline{I}^* &= \underline{I} + \underline{v}(\underline{\delta}, \underline{I}, \epsilon, \underline{\omega}^0) \\ \underline{\delta}^* &= \underline{\delta} + \underline{u}(\underline{\delta}, \underline{I}, \epsilon, \underline{\omega}^0) , \end{aligned}$$

(c) \underline{u} and \underline{v} are smooth functions of their variables, possibly excluding $\underline{\omega}^0$, and are zero-average, periodic functions of $\underline{\delta}$, and

(d) \underline{u} and \underline{v} vanish as ϵ approaches zero: $\lim_{\epsilon \rightarrow 0} \underline{u}, \underline{v} = 0$

This collection of assertions comprise a statement of the theorem; the invariant tori whose existence is thus assured are called the KAM tori of H . The proof is partially constructive and includes a procedure for estimating the volume of phase space containing KAM tori.[3]

Comment 20: From a heuristic standpoint, the theorem assures us that most invariant tori of a twist map will not be destroyed by a perturbation *provided that* there exist sufficiently large nonlinearities to detune resonances; that is the reason for the condition in Eq.(28). But suppose that this condition is not satisfied, as is the case, for example, with the harmonic oscillator. Is it possible for a perturbation to be so destabilizing that it could tear apart all the tori even for arbitrarily small ϵ ? The answer is yes: one example, originally studied by Contopoulos and others, can be found in Lichtenberg and Lieberman.[14, p.164]

Comment 21: For autonomous systems up to two degrees of freedom, the existence of KAM tori is enough to insure stability of orbits bounded by them and level surfaces of the Hamiltonian; for non-autonomous systems with two degrees of freedom (the so-called $2\frac{1}{2}$ degree of freedom problems) or for phase spaces of larger dimensions, this is no longer the case. This is, of course, the essence of "Arnold diffusion."

Comment 22: In many ways the KAM theorem possesses sociological similarities to Gödel's famous theorem in logic: (a) Both are widely known and talked about, yet many people are rather vague on what the theorems actually state, and very few have actually read the proofs, much less validated them.⁹ (b) Each has been called, by different mathematicians, the most important theorem of the twentieth century. (c) Neither is useful for practical calculations: almost by definition, Gödel's theorem provides no hint on how to recognize undecidable propositions, and the stable phase space estimated by the KAM theorem is typically too conservative to be of value.

⁹I apologize that the present discussion will not improve that situation and may even exacerbate it.

Comment 23: The KAM theorem is tied up with the question of ergodicity in classical statistical mechanics. Recall that the ergodic hypothesis was required in order to equate time averages with ensemble averages: it was necessary that a system sample all the phase space available to it within the constraint of energy conservation. If it possessed dynamical invariants other than energy, however, this could not be the case. The question then is, Are non-integrable systems ergodic? The KAM theorem says no, not generally: even in the absence of a complete set of invariant observables, measurable regions of phase space will contain orbits confined to N -dimensional tori, just as though the invariants were still there.

Comment 24: I am not certain of this, but let me state it anyway: In generalizing the KAM theorem to non-autonomous systems, the condition Eq.(29) should be replaced by,

$$|(\underline{m}, n) \cdot (\underline{\omega}^0, 1)| \geq \gamma \|(\underline{m}, n)\|^{-\eta} .$$

2.3 Resonance topology: higher dimensions.

The KAM theorem guarantees that off-resonant tori survive; this does not by itself mean that on-resonant ones do not. Nonetheless, that is generally the case: typically, on-resonant tori are replaced by separatrices, and we shall devote the rest of this lecture to a discussion of these objects.

It is regrettable that the concept "resonance" is generally introduced as a disease of perturbation theory. The usual scenario is this: An attempt is made to construct a transformation (canonical or not, it does not matter) which will put a Hamiltonian into normal form, as in Eq.(27). This attempt eventually founders: some terms in the perturbation series may become arbitrarily large because of division by "small denominators," of the form $\sin[\pi(\underline{m} \cdot \underline{\omega} + n)]$. This problem is associated with the existence of nearby resonances, and the connection is completed. The overall effect is to suggest that a resonance has more to do with the way things are calculated than with real, physical phenomena—the sort of (equally false?) feeling one sometimes gets about renormalization in quantum field theory. This characterization ignores what should be the central geometric features of a resonance: the dimensional collapse of the torus, which we have already seen in the first lecture, and the existence of a *separatrix*, the utility of which is its ability to organize phase space via partitioning, enabling the simultaneous classification of all orbits and their relationships. In keeping with our geometric point of view, let me emphasize that a separatrix is a topological object: no continuous transformation, whether constructed perturbatively or inspired by God, can deform phase space so as to make it disappear. Small denominators are not the real stumbling block but only *its manifestation* within the context of perturbation theory. The *real* problem is that we are attempting something fundamentally impossible.

Comment 25: Consider the following thought experiment, which forces one to think in a coordinate-free way. Suppose that you are given a one-to-one symplectic mapping, F , defined over some four-dimensional phase space and realized in an *unspecified system of coordinates*. (Think of F , for example, as a tracking program that returns 4-tuples of real numbers and models the period map of a $2\frac{1}{2}$ degree of freedom Hamiltonian system.) You are given the ability to calculate forward or backward iterates of F infinitely quickly, so you can generate as many as you want without any problem; further, you have the capability for visualizing these samples in four dimensions. Given even these extraordinary tools, how would you test the hypothesis, "This system exhibits a first order $\omega_1 + 2\omega_2$ sextupole resonance"? What *topological* features must one search for in the "data" in order to

confirm or deny such a statement?

Fortunately, there do exist integrable dynamical systems possessing separatrices: the "single resonance" models. These allow us to study "regular" (as opposed to "chaotic") separatrices analytically and thereby to discover their main features. In the next few sections we shall define a methodology for doing this and illustrate it with a concrete example. The analytical model to be employed is the Hamiltonian,

$$H(\underline{\delta}, \underline{I}) = \underline{\nu} \cdot \underline{I} + H_s(\underline{I}) + H_r(\underline{I}) \cos[\underline{m} \cdot \underline{\delta} + n\theta + \phi(\underline{I})] , \quad (30)$$

which describes a single, isolated resonance, the sort of model that might be filtered out of a low order perturbative expansion. The phase $\phi(\underline{I})$ appears in lieu of including both sine and cosine terms in the Hamiltonian.¹⁰ I have changed the symbol for the "independent" variable from t (time) to θ (angle) to conform more exactly with the way this Hamiltonian appears in applications to periodic accelerators.

Global analysis of instances of this model proceeds along the following lines.

2.3.1 Transformations.

The first step is to take advantage of the fact that angle coordinates appear only in the linear combination $\underline{m} \cdot \underline{\delta}$. This suggests employing a canonical transformation of the form Eq.(13) or Eq.(21) to a new set of coordinates, $(\underline{\delta}, \underline{I}) \rightarrow (\underline{\eta}, \underline{J})$, which we shall call the "resonance projected" coordinates. To simplify the presentation, we shall assume a four-dimensional phase space, but everything that we do here can be generalized easily to larger dimensional problems. The equations of transformation are written as follows:

$$\begin{aligned} \underline{I} &= \underline{M} \underline{J} & \underline{\eta} &= \underline{M} \underline{\delta} \\ \underline{J} &= \frac{1}{m_1^2 + m_2^2} \underline{M} \underline{I} & \underline{\delta} &= \frac{1}{m_1^2 + m_2^2} \underline{M} \underline{\eta} \end{aligned} \quad (31)$$

where the matrix \underline{M} is,

$$\underline{M} \equiv \begin{pmatrix} m_1 & m_2 \\ m_2 & -m_1 \end{pmatrix} .$$

Redefining angle-action coordinates in this way amounts to choosing a new set of homotopically inequivalent closed curves, $\{\gamma_k\}$, one which conforms more closely to the way resonant orbits actually wind around their N -tori. Resonant orbits are thus straightened out on this chart.

The representation of the Hamiltonian on the new chart is a function $Q(\underline{\eta}, \underline{J})$,

$$\begin{aligned} Q(\underline{\eta}, \underline{J}) &= \underline{m} \cdot \underline{\nu} J_1 + \underline{m} \times \underline{\nu} J_2 \\ &+ Q_s(\underline{J}) + Q_r(\underline{J}) \cos[\eta_1 + n\theta + \tilde{\phi}(\underline{J})] \end{aligned}$$

where $\tilde{\phi}(\underline{J}) \equiv \phi(\underline{I})$, $Q_{s,r}(\underline{J}) \equiv H_{s,r}(\underline{I})$, and $\underline{m} \times \underline{\nu} \equiv m_2 \nu_1 - m_1 \nu_2$. There are two important points to be noted here:

(1) η_2 is an ignorable coordinate, so J_2 is a constant of motion. Our method of visualizing the flow will be to slice phase space along the three-dimensional surfaces $J_2 = \text{constant}$. We then can consider J_2 either as a dynamical variable of the original Hamiltonian or as a

¹⁰In retrospect, this may not have been the best way of going about this, but I must draw the line on rewriting somewhere.

control parameter of the "projected Hamiltonian."

(2) η_1 and η_2 are coordinates whose modular range is $(|m_1| + |m_2|)2\pi$. Since Q is of period 2π in η_1 the flow consists of $|m_1| + |m_2|$ identical copies of a fundamental domain, $\eta_1 \in [0, 2\pi)$.

A second transformation $(\underline{\eta}, \underline{J}) \rightarrow (\underline{\xi}, \underline{J}')$ gets rid of the explicit dependence of the Hamiltonian on θ . One generating function which accomplishes this is:

$$F(\underline{\eta}, \underline{J}') = (\eta_1 + n\theta)J'_1 + \eta_2 J'_2 + \int^{J'_1} dJ''_1 \tilde{\phi}(J''_1, J'_2) . \quad (32)$$

This produces the following equations of transformation.

$$\begin{aligned} \underline{J} &= \underline{J}' \\ \xi_1 &= \eta_1 + n\theta + \tilde{\phi}(\underline{J}) \\ \xi_2 &= \eta_2 + \int^{J'_1} dJ''_1 \partial_2 \tilde{\phi}(J''_1, J_2) \end{aligned} \quad (33)$$

The final form of the Hamiltonian is given by

$$K = J_1 \Delta + \underline{m} \times \underline{\nu} J_2 + K_s(\underline{J}) + K_r(\underline{J}) \cos \xi_1 \quad (34)$$

where $\Delta \equiv m_1 \nu_1 + m_2 \nu_2 + n$. It is expected that Δ is a small quantity. Indeed, for this Hamiltonian to be at all interesting Δ must be small enough so that $J_1 \Delta$ is comparable in magnitude to $K_s(\underline{J})$ and $K_r(\underline{J})$.

The independent variable θ no longer appears explicitly in Eq.(34), so K is a constant of motion. That K and J_2 form a pair of dynamical invariants means, by the Liouville-Arnold theorem, that this four-dimensional (now autonomous) system is integrable.

Comment 26: These developments can be generalized trivially to more degrees of freedom. A single resonance Hamiltonian, of the form in Eq.(30), depends on a single linear combination of phases. There are therefore $N - 1$ linearly independent combinations which are ignorable and whose conjugate variables are therefore constants of the motion. After the second transformation, which makes the Hamiltonian autonomous, the Hamiltonian itself becomes the N th dynamical invariant. The system is thus integrable, all compact orbits lie on invariant tori, and, perhaps most importantly, there are *only two coordinates* which are not invariant: J_1 and ξ_1 . Therefore, regardless of the number of dimensions, the single resonance problem collapses down to a two-dimensional (or one degree of freedom) autonomous system.

2.3.2 Behavior at infinity.

The easiest thing to examine is the behavior of the flow at infinity. Dividing through by K_r gives us an equation for $\cos \xi_1$.

$$\frac{K_s(\underline{J})}{K_r(\underline{J})} + \cos \xi_1 = \frac{K - J_1 \Delta - \underline{m} \times \underline{\nu} J_2}{K_r(\underline{J})} \quad (35)$$

Under normal circumstances $K_r \rightarrow \infty$ as $J_1 \rightarrow \infty$. Thus, unbounded motion is only possible for finite, invariant K only along the asymptotic phases ξ_1^∞ which satisfy

$$\cos \xi_1^\infty = - \lim_{J_1 \rightarrow \infty} [K_s(\underline{J})/K_r(\underline{J})] \quad (36)$$

Whether the right hand side (rhs) of this expression lies within $[-1, 1]$ is determined easily by examining the highest powers of $\sqrt{J_1}$ which appear in the polynomials defining K_s and K_r . Let these be respectively n_s and n_r . If $n_s > n_r$, then rhs unbounded, which means that the flow must be confined; no orbit can go to infinity. If $n_r > n_s$, then $\text{rhs} \rightarrow 0$, and unbounded flows are possible at phases $\xi_1^\infty \simeq \pm\pi/2$. If $n_r = n_s$, then the situation is far more interesting: values of the phase asymptotes will depend on the parameters of the problem, including the value of J_2 (and the other invariant J 's, in higher dimensions).

2.3.3 Resonant orbits, regular and irregular.

Resonant orbits of the original Hamiltonian, Eq.(30), correspond to fixed points of the projected Hamiltonian, Eq.(34). They are obtained by setting \dot{J}_1 and $\dot{\xi}_1$ to zero.

$$\begin{aligned} \dot{J}_1 &= -\partial K / \partial \xi_1 \\ &= K_r \sin \xi_1 \end{aligned} \quad (37)$$

$$\begin{aligned} \dot{\xi}_1 &= \partial K / \partial J_1 \\ &= \Delta + \partial_1 K_s + \partial_1 K_r \cos \xi_1 \end{aligned} \quad (38)$$

Setting $\dot{J}_1 \doteq 0$ gives us the possibility of two types of fixed points: (1) regular fixed points (*reg fp*) are those for which $\sin \xi_1 = 0$, and (2) irregular fixed points (*irreg fp*) are those for which $K_r = 0$. Let the coordinates of a fixed point be symbolized as $(\xi_1^{(0)}, J_1^{(0)})$. In the case of a regular fixed point, $\xi_1^{(0)} \simeq 0$ or π so that $\cos \xi_1^{(0)} = \pm 1$ and the condition $\dot{\xi}_1 = 0$ simplifies to

$$\Delta + \partial_1 K_s \pm \partial_1 K_r = 0 \quad \begin{cases} + \xi_1^{(0)} \simeq 0 \\ - \xi_1^{(0)} \simeq \pi \end{cases} \quad (39)$$

This equation, which is generally a polynomial in $\sqrt{J_1}$, must then be solved for the values of $J_1^{(0)}$.

In the case of irregular resonant orbits, we must first solve the equation $K_r = 0$ for allowed values of $J_1^{(0)}$ and then use Eq. (38) to get the corresponding values of $\xi_1^{(0)}$. Unlike the regular resonant orbits, the phase of an irregular resonant orbit is not necessarily pinned to a particular value; this must be treated on a case-by-case basis.

Which resonant orbits are stable and which are unstable? The linearized equations of motion in the tangent space, $T\mathcal{M}_p$, near a regular fixed point are written by expansion to first order in $d\xi_1$ and dJ_1 .

$$\dot{J}_1 = K_r(J_1^{(0)}) \cos \xi_1^{(0)} d\xi_1 \quad (40)$$

$$\dot{\xi}_1 = [\partial_1^2 K_s(J_1^{(0)}) + \partial_1^2 K_r(J_1^{(0)}) \cos \xi_1^{(0)}] dJ_1 \quad (41)$$

Therefore, if $K_r(J_1^{(0)}) \cos \xi_1^{(0)}$ and $\partial_1^2 K_s(J_1^{(0)}) + \partial_1^2 K_r(J_1^{(0)}) \cos \xi_1^{(0)}$ have the same sign then the fixed point is unstable; if they have opposite signs, the fixed point is stable. (See Fig. (6).)

The set of all orbits which approach an unstable resonant orbit as $\theta \rightarrow +\infty$ is called its "stable" manifold; those which approach it as $\theta \rightarrow -\infty$ is its "unstable" manifold. The union of the unstable resonant orbits along with their stable and unstable manifolds is the *separatrix* of the system.

2.4 A model: the $\nu_1 + 2\nu_2$ sextupole resonance.

To illustrate all of this, we shall find the separatrix for the first order (1,2) sextupole

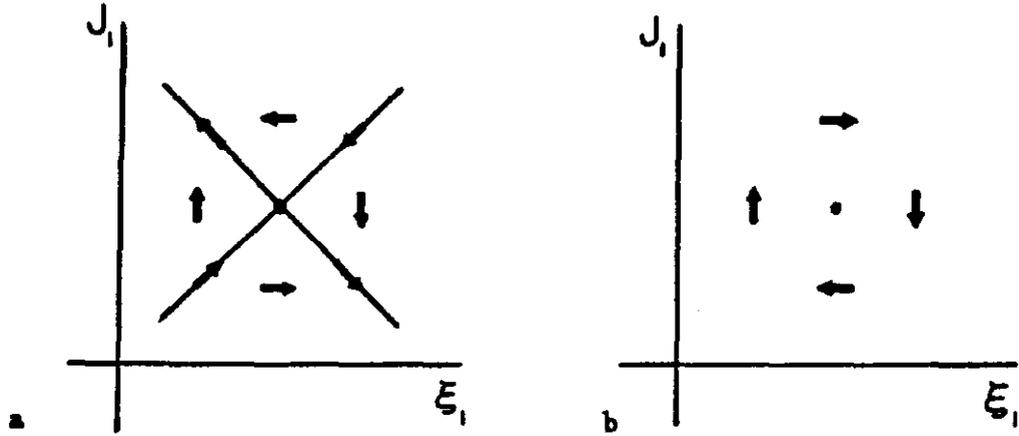


Figure 6: The nature of a regular fixed point depends on whether the coefficients of $d\xi_1$ and dJ_1 have (a) the same sign or (b) opposite signs.

resonance, that is, the $\nu_1 + 2\nu_2$ resonance excited by sextupoles to first order in the sextupole strength. Visualizing a four-dimensional object like this is a little involved, but not impossible. One method is to take a sequence of three-dimensional slices, much as one might present a cube to a two-dimensional creature by slicing it from bottom to top. Of course, we must take some care in arranging the slices; our two-dimensional friend would form a distorted concept of a cube were it presented sliced along a diagonal. We shall obtain a good representation of the four-dimensional dynamics by drawing the separatrix within three-dimensional surfaces specified by the condition $J_2 = \text{constant}$.

Our single resonance model Hamiltonian is,

$$H = \nu_1 I_1 + \nu_2 I_2 + g I_1^{1/2} I_2 \cos(\delta_1 + 2\delta_2 + n\theta + \phi) .$$

where the numbers g and ϕ are functionals of the sextupole distribution. The transformation to the resonance projected coordinates yield

$$\begin{aligned} J_1 &= (I_1 + 2I_2)/5 \\ J_2 &= (2I_1 - I_2)/5 \\ \xi_1 &= \delta_1 + 2\delta_2 + n\theta + \phi \\ \xi_2 &= 2\delta_1 - \delta_2 \end{aligned} \quad (42)$$

On this chart the projected Hamiltonian is represented,

$$K = J_1 \Delta + J_2 \Gamma + g I_1^{1/2} I_2 \cos \xi_1 , \quad (43)$$

where $\Delta \equiv \nu_1 + 2\nu_2 + n$ and $\Gamma \equiv 2\nu_1 - \nu_2$. It is expected that Δ is a small quantity. Indeed, for this Hamiltonian to be at all interesting Δ must be small enough so that $J_1 \Delta$

is comparable in magnitude to the resonant term. Invariant manifolds must run parallel to ξ_2 , since ξ_2 does not appear in K . The Hamiltonian flow, projected along ξ_2 , is given by the vector field

$$\begin{aligned} \dot{J}_1 &= gI_1^{1/2}I_2 \sin \xi_1 \\ \dot{\xi}_1 &= \Delta + gI_1^{-1/2}\left(\frac{1}{2}I_2 + 2I_1\right) \cos \xi_1 . \end{aligned} \quad (44)$$

Resonant orbits of K are projected into fixed points of Eq.s(44). The regular ones are those for which $\sin \xi_1 = 0$; the irregular ones are those for which either $I_1 = 0$ or $I_2 = 0$.

Symmetries of the projected flow will allow us to confine our attention to the parameter quadrant: $\Delta > 0$, $g > 0$. Clearly, if we simultaneously change the sign of both these quantities, the flow simply changes direction. Changing the sign of g alone can be compensated for by the transformation $\xi_1 \rightarrow \xi_1 + \pi$. Finally, changing the sign of Δ alone amounts to performing both previous transformations in succession. In fact there are no essential parameters in this problem: both Δ and g can be made to vanish by a simple scaling transformation. Let us define $\kappa \equiv \Delta/g$, and scale the amplitude variables by κ^2 .

$$j_{1,2} \equiv J_{1,2}/\kappa^2 \quad i_{1,2} \equiv I_{1,2}/\kappa^2 \quad (45)$$

Then the level sets—which determine the topology of the flow—of the function

$$\mathcal{K} \equiv g^2(K - J_2\Gamma)/\Delta^3 = j_1 + i_1^{1/2}i_2 \cos \xi_1$$

are identical to those of K . Further, \mathcal{K} can act as a true Hamiltonian for the scaled variables provided we simultaneously rescale $\theta \rightarrow \theta\Delta^3/g^2$.

For reference purposes, we shall present the answer first and then go through its development. The separatrix is sketched in Figure 7. Each frame shows its intersection with a single three-dimensional J_2 leaf projected along the ξ_2 direction onto the (ξ_1, J_1) plane. A few points should be kept in mind while scanning these pictures. First, the ξ_1 axis corresponds not to $J_1 = 0$ but to $J_1 = -2J_2$ ($I_1 = 0$), when $J_2 < 0$, and to $J_1 = \frac{1}{2}J_2$ ($I_2 = 0$), when $J_2 > 0$. Second, the dynamical range of ξ_1 is 6π : we are viewing only one-third of the full projection; each picture is repeated twice. Third, remember that a "fixed point" in the diagram is the projection of a resonant orbit, which is a 1-torus, a closed curve corresponds to a 2-torus, and an open (unbounded) curve corresponds to a two-dimensional surface.

The fixed point equation, Eq. (39), which we use to find the resonant orbits is written:

$$\Delta/g \pm I_1^{-1/2}\left(\frac{1}{2}I_2 + 2I_1\right) = 0 \quad (46)$$

Clearly if $\Delta/g > 0$ ($\Delta/g < 0$) then the $-$ ($+$) sign is indicated, which means that $\xi_1^{(0)} \simeq \pi$ ($\xi_1^{(0)} \simeq 0$). To be definite, let us assume in everything that follows that $\Delta > 0$ and $g > 0$, and rely on the previously mentioned symmetries to extrapolate results to other regions. In any case we shall set $\kappa \equiv |\Delta/g|$. The fixed point equation (46) can be written

$$\frac{1}{2}I_2 + 2\left(I_1^{1/2} - \frac{\kappa}{4}\right)^2 = \frac{\kappa^2}{8}$$

which describes an ellipse in $(\sqrt{I_1}, \sqrt{I_2})$ space. We want to express this in terms of the constant of motion J_2 . Substituting from Eq's.(42) we get the following.

$$3I_1 - \kappa I_1^{1/2} - \frac{5}{2}J_2 = 0 \quad (47)$$

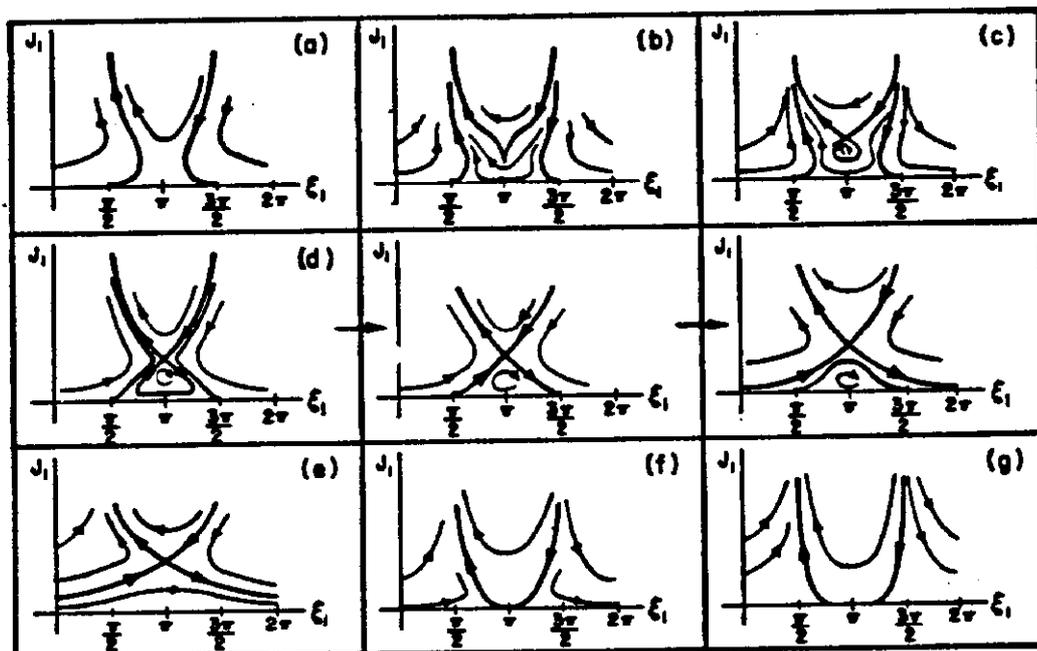


Figure 7: Flow diagrams for the projected Hamiltonian of the (1,2) resonance.

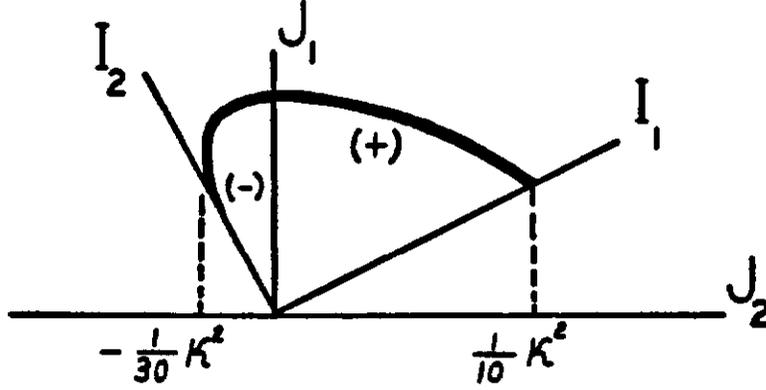


Figure 8: Track of the projected resonant orbits (fixed points).

with solution

$$\sqrt{I_1^{(0)}} = \frac{1}{6}(\kappa \pm \sqrt{\kappa^2 + 30J_2}) \quad (48)$$

This is sketched in Fig. (8).

Applying the tangent space Eq. (41) to Eq. (44) provides us with the localized flow.

$$\dot{J}_1 = gI_1^{(0)1/2}I_2^{(0)} \cos \xi_1^{(0)} d\xi_1 \quad (49)$$

$$\dot{\xi}_1 = g/4 I_1^{(0)-3/2} (6J_1^{(0)} + 17J_2) \cos \xi_1^{(0)} dJ_1 \quad (50)$$

Therefore, unstable resonant orbits are those for which $6J_1^{(0)} + 17J_2 > 0$. This inequality must be satisfied within the region $J_2 > 0$, and since the + branch is the only one to exist in this region, this is enough to establish that the + branch is the unstable one and the - branch the stable one. As a double-check, we examine the region $J_2 < 0$. First note that

$$6J_1^{(0)} + 17J_2 = \frac{1}{3} [\kappa^2 + 30J_2 \pm \kappa\sqrt{\kappa^2 + 30J_2}]$$

so, as expected, since $\kappa^2 + 30J_2 > 0$ it automatically follows that the + branch is unstable. Further, since $\sqrt{\kappa^2 + 30J_2} < \kappa$ we have immediately that

$$\kappa^2 + 30J_2 < \kappa\sqrt{\kappa^2 + 30J_2}$$

which means that the - branch traces the path of a stable fixed point.

Recall that the irregular resonant orbits are found by setting K_r to zero in Eq.(37). This provides the value of the amplitude variable at the resonant orbit, after which the phase value is obtained from Eq.(38) upon setting ξ_1 to zero. In our model, $K_r = 0$ means that either I_1 or I_2 must vanish. Consider first approaching the surface $I_1 = 0$. Eq.(44) is then

dominated by the $I_1^{-1/2}$ term:

$$I_1^{1/2} \dot{\xi}_1 \approx \frac{1}{2} g I_2 \cos \xi_1$$

$\dot{\xi}_1$ is indefinite at $\xi_1 \approx \pm\pi/2$. This is the location of the irregular resonant orbit, whose phase, in this case, happens to be pinned. (See Figures (7a-d).) At any other value of ξ_1 ,

$$\dot{\xi}_1 = \text{sgn}[g \cos \xi_1] \cdot \infty .$$

This type behavior will obviously occur whenever K_r contains a term whose the highest power of I_1 is $\frac{1}{2}$, and similarly for J_2 .

Now consider the surface $I_2 = 0$. In this case Eq.(44) simplifies to the following.

$$\dot{\xi}_1 = \Delta + 2gI_1^{1/2} \cos \xi_1$$

The resonant orbit thus sits at

$$\cos \xi_1^* = -\frac{1}{2} \kappa I_1^{1/2}$$

and is not pinned, but varies with I_1 . Real solutions are possible only for $I_1 > \kappa^2/4$. Notice also that $\xi_1^* \rightarrow \pm\pi/2$ as $I_1 \rightarrow \infty$. In Figure 8, the point ($I_1 = \kappa^2/4, I_2 = 0$) corresponds to the intersection of the *reg f p* track with the I_2 axis. What happens is that the unstable regular fixed point is pushed down into the surface $I_2 = 0$ where it splits into two irregular fixed points. (See Figures 7).

Note that ξ_1 remains finite for $I_2 = 0$, whereas $\dot{\xi}_1$ was infinite on $I_1 = 0$. Nearby orbits approach the instability in more leisurely fashion. This is intuitively appealing: one would expect a normal sextupole resonance to destabilize an orbit with zero vertical emittance more slowly than one of having zero horizontal emittance.

The J_1 amplitudes at which the regular resonant orbits occur on a particular J_2 surface are found by intersecting the fixed point track of Figure(8) with the corresponding line, $J_2 = \text{constant}$. The merger that occurs at $J_2 = -\frac{1}{30}\kappa^2$, $J_1 = \frac{17}{180}\kappa^2$ represents a local bifurcation¹¹ which, following Thom[25], is called a "catastrophe." As exhibited in Figures 7a-c, the topological character of the flow changes when κ^2 crosses this threshold: below it, all orbits diverge; above it there is a "pocket" of stable orbits. Precisely at catastrophe, stable and unstable resonant orbits merge to form a cusp, which annihilates both of them. If one thinks of J_2 as a control variable for the projected Hamiltonian—so that the control space is the set of doublets (J_2, κ) —then the subset $\{(J_2, \kappa) | J_2 = -\kappa^2/30\}$, which marks the control points at which bifurcation occurs, is called the "catastrophe surface."

Catastrophes are *local* bifurcations: it is possible to observe the transition by viewing the flow locally, in the vicinity of the cusp. In contrast, *global* bifurcations cannot be observed locally: the separatrix as a whole undergoes a transition; locally, nothing interesting happens. One type of global bifurcation, a *saddle switch*, can occur in our example when the Hamiltonian takes on the same value at the unstable regular and irregular resonant orbits, so that a branch of the separatrix can connect them. Assigning $\xi_1 = \pi/2$ and $I_1 = 0$ in Eq.(43) gives us the value of the reduced Hamiltonian at the irregular fixed point.

$$K_{irr, f p} = -2\Delta J_2 + \underline{m} \times \underline{\nu} J_2$$

Its value at a regular fixed point is

$$K_{reg, f p} = \Delta J_1^{(0)} + \underline{m} \times \underline{\nu} J_2 - g \sqrt{I_1^{(0)} I_2^{(0)}}$$

¹¹It is a bifurcation of the projected Hamiltonian; it is a fold in the separatrix of the original, four-dimensional system.

Now, the algebra is made simpler by expressing everything in terms of I_1 and J_2 .

$$\begin{aligned} I_2^{(0)} &= 2I_1^{(0)} - 5J_2 \\ J_1^{(0)} &= I_1^{(0)} - 2J_2 \end{aligned} \quad (51)$$

Making these substitutions gives us the following.

$$K_{resfp} = \Delta(I_1^{(0)} - 2J_2) + \underline{m} \times \underline{\nu} J_2 - 2gI_1^{(0)3/2} + 5gI_1^{(0)1/2} J_2$$

Now use the fixed point equation, Eq.(47), to write

$$5gI_1^{(0)1/2} J_2 = 6gI_1^{(0)3/2} - 2\Delta I_1^{(0)}$$

Making this substitution and simplifying a little gives us the result:

$$K_{resfp} = K_{irrfp} + 4gI_1^{(0)3/2} - \Delta I_1^{(0)}$$

The necessary condition for a global bifurcation, $K_{resfp} = K_{irrfp}$, can thus be written as follows.

$$\sqrt{I_1^{(0)}} = \frac{1}{4}\kappa$$

Substituting for $I_1^{(0)}$ from Eq.(48) gives us the value of J_2 where the bifurcation occurs.

$$J_2 = -\frac{1}{40}\kappa^2$$

The "control space" surface $\{(J_2, \kappa) | J_2 + \frac{1}{40}\kappa^2 = 0\}$ is called the "Maxwell surface" by Gilmore.[8]

We have now finished our global analysis of the (1, 2) resonance model and are ready to recapitulate the complete description of its flow (Refer to Figure 7): (a) For J_2 large and negative all orbits are unbounded except the irregular resonant orbits, which are pinned to the surface $I_1 = 0$ at phases $\xi_1 \simeq \pm\pi/2$. (b) As J_2 increases, a local bifurcation, or catastrophe, occurs on the leaf $J_2 = -\frac{1}{30}\kappa^2$. It is heralded by the appearance of a new branch of the separatrix connected non-transversally (forming a cusp) to a new resonant orbit, a 1-torus. (c) That orbit splits, and for $-\frac{1}{30}\kappa^2 < J_2 < -\frac{1}{40}\kappa^2$ there is a single class of bounded orbits. (d) A global bifurcation, a saddle-switch, occurs on the surface $J_2 = -\frac{1}{40}\kappa^2$. At this precise value, the surface $I_1 = 0$ is stable for phases that are 2π -equivalent to the range $\pi/2 < \xi_1 < 3\pi/2$. On the surfaces $-\frac{1}{40}\kappa^2 < J_2 < 0$ there are two classes of bounded orbits. The first, say Class A, is as before and is characterized by a bounded phase, $\pi/2 < \xi_1 < 3\pi/2$. The second, Class B, has an unboundedly increasing phase ξ_1 . (Another way of saying this: Class A orbits exist in "islands.") The entire surface $I_1 = 0$ is now locally stable. (e) For $0 < J_2 < \frac{1}{10}\kappa^2$ the Class A orbits have disappeared; Class B orbits are still bounded. (f) When $\frac{1}{10}\kappa^2 < J_2$ Class B has disappeared as well. All orbits are once more unbounded, except the two unpinned irregular resonant orbits in the plane $J_2 = 0$ which begin at $\xi_1 \simeq \pi$ at $J_2 = \frac{1}{10}\kappa^2$ and (g) wander to $\xi_1 \simeq \pm\pi/2$ as $J_2 \rightarrow \infty$.

2.4.1 Adiabatic resonance width.

Except for the irregular resonant orbits pinned on $J_1 = 0$ and $J_2 = 0$, the (1,2) resonance possesses no bounded orbits on the leaves for which $J_2 < -\frac{1}{30}\kappa^2$ or $\frac{1}{10}\kappa^2 < J_2$, whereas between these leaves bounded orbits fill some volume of phase space. This is the general behavior of all resonances, except the quadrupole resonances for which all orbits are either bounded or unbounded: the region of bounded orbits slowly shrinks as the resonance is approached. One quantitative measure of this approach to global instability is the "resonance width." We take this term to mean the size of the smallest strip in tune space which is centered on the resonance line, Eq.(25), and outside of which a beam is stable. This definition remains ambiguous, because it depends on the size and shape of the beam as well as on the experimental setup—e.g., on whether the resonance is approached adiabatically or the beam is suddenly injected into the resonant situation. In order to avoid beam parameters entirely, we shall associate an "adiabatic resonance width" with each individual orbit. That is, we imagine initializing an orbit in phase space with control parameters set far from resonance, then approaching the resonance very slowly, and finally noting when the orbit becomes unbounded.

For the (1,2) resonance of our example this means beginning with $\kappa \approx \infty$ and letting $\kappa \rightarrow 0$ on a time scale much greater than $\max(1/\nu_1, 1/\nu_2)$. At $\kappa = \infty$ all orbits are harmonic oscillator orbits, the variables I_1 , I_2 , J_1 and J_2 are conserved separately, and we can label an orbit with any two of the four initial values, J_1^{in} , J_2^{in} , J_1^{in} and J_2^{in} .¹² According to the usual adiabatic theorems the variation of an orbit as κ approaches zero will be regulated by the adiabatic invariance of the action integrals. Because J_2 is a constant of motion for fixed κ , we can take $J_2 = \frac{1}{2\pi} \oint J_2 d\xi_2$ itself as the first adiabatic invariant. To the second we attach the symbol $\mathcal{A} \equiv \oint J_1 d\xi_1$, whose value is $\mathcal{A}^{in} = 6\pi J_1^{in}$.

What happens to an orbit as κ slowly decreases depends critically on the sign of J_2^{in} . For $J_2^{in} > 0$ the diagrams of Figure 7e-g are the relevant ones, and we now must think of them as flow diagrams for the projected Hamiltonian (see Eq.(43)) rather than mapping diagrams of the function F . As κ decreases the separatrix pushes downward. Each orbit remains on its leaf, $J_2 = J_2^{in}$, it maintains its value of \mathcal{A} , and it crosses the separatrix, thus becoming unbounded, when the area under the separatrix has decreased to \mathcal{A}^{in} .

For $J_2^{in} < 0$ the situation is much more interesting, as the separatrix contains two branches. Figures (7a-e) are now the relevant ones, but they must be traversed in reverse order. As κ decreases from ∞ the upper branch pushes downward, as before, but simultaneously a bubble, representing the lower branch of the separatrix, forms and begins to grow. As these two branches grow closer, approaching their merger at the saddle-switch ($\kappa^2 = -40J_2^{in}$), orbits either are captured by the island or pass through the upper branch, depending on their values for \mathcal{A}^{in} . The total area under the saddle-switch is $\mathcal{A}_s = -(15 + 33\pi/4)J_2^{in}$. If $\mathcal{A}^{in} > \mathcal{A}_s$, the orbit passes through the upper branch of the separatrix; if $\mathcal{A}^{in} < \mathcal{A}_s$, then it is captured by and subsequently leaks through the lower branch. If the latter happens, \mathcal{A} undergoes a discontinuous change upon passage through the separatrix, since only one of the three islands can capture the orbit. (Remember, the period 3 property refers to the phase space mapping, not the transformed flow.) As κ continues to decrease, the orbit will retain its new value for \mathcal{A} as the island lifts and shrinks. Eventually—at some point before $\kappa^2 = -30J_2^{in}$ —the island becomes too small to contain the orbit.

Figure 9 contains a "master curve," drawn in the normalized (j_1^{in}, j_2^{in}) coordinates of Eq.(45), which uses this scenario to assign resonance widths to individual orbits. The curve was computed by numerically integrating the area under the upper branch of the separatrix when $-1/40 < j_2 < 1/10$ and within the island when $-1/30 < j_2 < -1/40$. It is used in the following way. Suppose one starts an orbit at $\kappa \approx \infty$ with initial amplitude variables J_1^{in}

¹²Because the system is linear for $\kappa = \infty$ we can legitimately associate J_1^{in} and J_2^{in} with the initial horizontal and vertical emittances divided by 2π . [19]

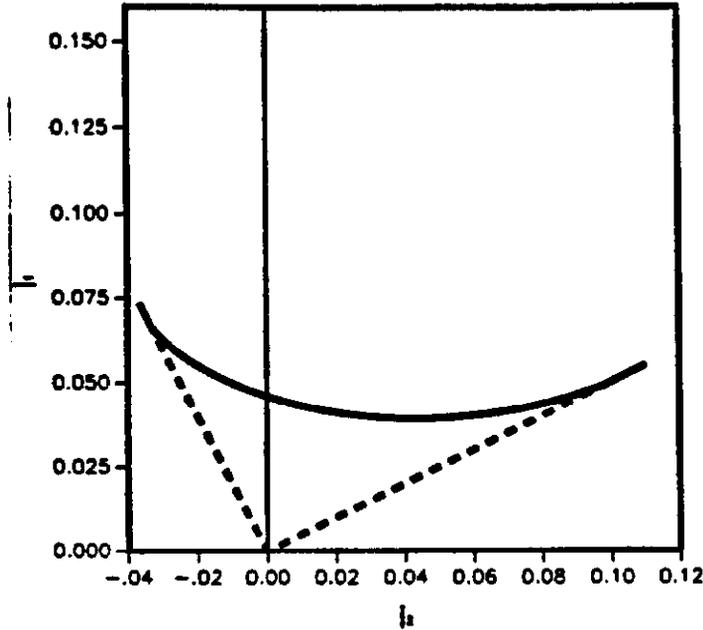


Figure 9: Resonance width master curve.

and J_2^{in} . To find the value of κ at which the orbit becomes unbounded, first calculate J_1^{in} and J_2^{in} , using Eq.s(42), and take their ratio. The intersection of the ray $j_1^{in}/j_2^{in} = J_1^{in}/J_2^{in}$ with the “master curve” is now read off; call that point (j_1^{int}, j_2^{int}) . The value of κ at which the orbit becomes unbounded is

$$\kappa \equiv \sqrt{J_1^{in}/J_1^{int}} .$$

For a given resonant coupling, the adiabatic resonance width of the orbit is then determined according to $2\Delta = 2g\kappa$.

A more dynamic picture is obtained by removing the $1/\kappa^2$ normalization: the curve of Figure 9 would be no longer static but sweep through the (J_1^{in}, J_2^{in}) space, converging on the origin as κ approaches zero and making orbits unbounded as it passes their initial conditions.

2.5 Resonance topology, revisited.

We can generalize our observations on the isolated resonance of a four-dimensional, integrable, periodic system to larger dimensional phase spaces. The wonderful thing is that, regardless of the number of dimensions, the problem always collapses to an autonomous Hamiltonian acting on a “projected” two-dimensional phase space. What we are trying for here is a generic, geometric description of separatrices of integrable Hamiltonian systems. The usefulness of such a structure is that it partitions phase space into disconnected regions, of which it is the boundary, thereby organizing the flow. It is built up in pieces from invariant submanifolds of various dimensions. Here is what we should expect (the term “orbit” should be interpreted as the set of discrete samples obtained by applying the period advance map to a periodic system):

1. At the highest level of structure, there is a way of slicing $2N$ -dimensional phase space along disjoint $(N+1)$ -dimensional adiabatically invariant sub-manifolds. Analytically,

these would be labelled by the action variables, J_2 through J_N . These slices are the “leaves” of a foliation. The invariance property means that every orbit is confined to a single leaf: J_2 through J_N are constants of the motion.

2. At the next level of structure, bounded orbits lie on invariant N -tori (N -dimensional tori), T^N . (Liouville-Arnold theorem) Almost all of these orbits are non-resonant and will uniformly and densely fill their tori under repeated application of the period advance map.
3. Orbits whose winding numbers obey k resonance conditions will be confined to $N - k$ dimensional sub-tori, T^{N-k} . We are especially interested in the case $k = 1$.
4. Within each leaf, each resonant T^{N-1} that is unstable — or more exactly, whose orbits are unstable — forms a cluster set for orbits lying on zero-measure, N -dimensional manifolds. They are the “alpha and omega limit sets” of these orbits, which generalizes the concept of “stable” and “unstable” manifolds attached to fixed points. We shall risk abusing the terminology and call them by the same name.
5. The “separatrix” is the union of all the stable and unstable manifolds along with the unstable $(N - 1)$ -tori (i.e., tori made up of unstable resonant orbits) to which they are attached. Since the section of a separatrix within each leaf is an N dimensional surface, and since the leaves themselves have codimension $N - 1$, the full separatrix is a $(2N - 1)$ -dimensional surface, which are enough dimensions to enable it to partition the $2N$ -dimensional phase space.

One really needs to let these images simmer for awhile before they fall into place. The topological description of any particular resonance consists of listing the resonant tori, the T^{N-1} , and describing how the branches of the separatrix connect them together, much as we have done with the $\nu_1 + 2\nu_2$ sextupole resonance.¹³ Separable resonances — those whose separatrices remain “far” from each other — in more complicated, non-integrable dynamical systems are then associated with the existence of similar structures, at least on the macroscopic scale.

2.6 Resonance seeding.

We are led to a conceptual picture of near-integrable Hamiltonian systems much like the one in Arnold’s famous sketch, shown in Figure 10. It refers either to the flow of an autonomous system or to the period advance mapping of a periodic system (or, perhaps, to the Poincaré map of an arbitrary system). As we move out from the “origin,” which is actually a fixed point of the map (or flow) we pass through a series of layers of invariant tori shearing past each other. Even this is an over-simplification, however, because the tori corresponding to resonant tunes (winding numbers) will, upon more detailed inspection, be seen to be not tori at all but very thin separatrices sheltering sub-harmonic tori within their islands. These can be complicated objects, but the complexity does not end there. The separatrix which we sketched in the last section was of an integrable resonance, one consequence of which was that the unstable manifold of one resonant orbit joined tangentially with the stable manifold of another, thereby assuring that the separatrix was itself a smooth surface. However, the generic Hamiltonian is not integrable, and there is no reason to expect this phenomenon to occur: the generic behavior is that the unstable manifold of one resonant orbit intersects the stable manifold of its partner transversally, not tangentially. Now, all points on the intersection belong both to a stable manifold and to an unstable one — that is, they get mapped back into these manifolds under the period advance mapping. It then

¹³For another example, see reference [17].

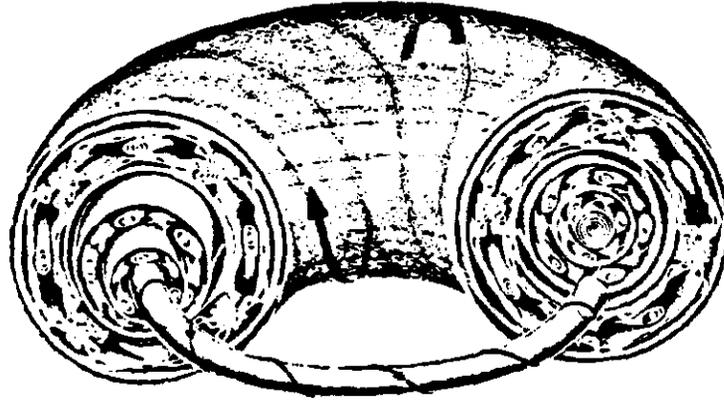


Figure 10: V. I. Arnold's conceptual drawing of a Hamiltonian system.

follows that these surfaces must intersect not only once but an infinite number of times. Further, the phase space volume bounded by the regions between successive intersections must be preserved. The result is that the separatrix, rather than being the bare, smooth surface depicted in the last section, is clothed with a complicated layer of "chaotic" orbits, the exact meaning of which will be the subject of the next lecture. The situation is sketched in Figure 11 for a two-dimensional phase space; please keep in mind that four-dimensions can be vastly more complex, and we live in a six-dimensional world.

Close to the origin, however, all this complexity exists on very tiny scales; on a macroscopic scale, all we would see are tori shearing past each other, much like an integrable system. As we move away from the origin, resonances may become broader and their chaotic separatrices thicker. Eventually, the chaotic layers from different resonances may begin to overlap each other, resulting in macroscopic chaos. For any system with $2\frac{1}{2}$ degrees of freedom or more — say a non-autonomous Hamiltonian on a four-dimensional phase space, or an autonomous one in six dimensions — this can result in "Arnold diffusion" around KAM tori. In lower dimensional systems there typically will be an outer stability boundary, a "shoreline" of chaotic orbits which marks the edge of the connected region of phase space occupied by bounded orbits. Even beyond this, however, there may be isolated pockets, or "islands," of stability. The full description is seldom simple.

We have mentioned that resonances appear in perturbation theory as "small denominators" which arise while trying to convert the Hamiltonian into normal form perturbatively. Now, there are two things which are constructed in perturbation theory: (a) a transformation which changes the phase space chart in which one represents the dynamics, and (b) the representation of the Hamiltonian on the new chart. Small denominators appear while building the transformation. They can be sidestepped by relaxing the constraints on the new Hamiltonian representation. Rather than demanding that it be a shearing Hamiltonian, we can filter the most offending resonances out of the transformation and absorb them into the new Hamiltonian.[19,18,20] If there is only one of these, and typically the lowest order

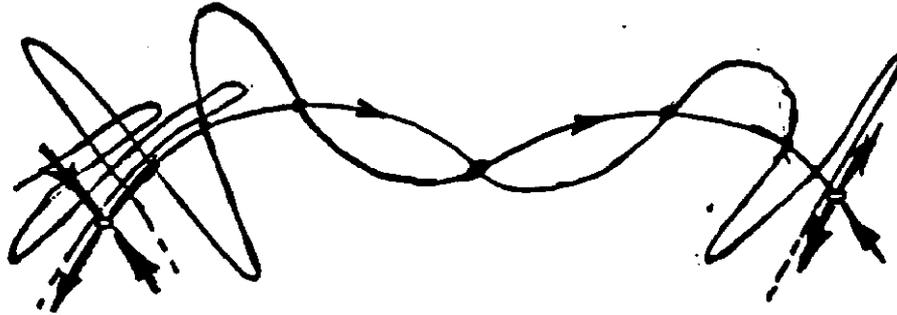


Figure 11: Transverse intersection of invariant manifolds as a mechanism for chaos.

one will be the most important, then the new Hamiltonian is still integrable; if there is more than one, it is not.

In light of all the complexity inherent in an arbitrary system, what information can we really hope to glean from a perturbative expansion in which one keeps only a handful of terms? A surprising answer emerges from numerical experiments on two-dimensional maps.[21,11] It turns out that the outer "shoreline" of stability generally appears in the vicinity of the separatrix associated with the lowest order resonances filtered out of a perturbative expansion. This is, to my simple and rather naive mind, astonishing. There is no reason to expect the two to be connected, and yet the stability boundary appears to grow on this separatrix, much like a crystal of salt will grow on a string immersed in a saturated solution: it is, if you will, "seeded" by the underlying low order resonance.

This is illustrated in Figure 12 for the simple case of a nonlinear kick arising from a single, thin sextupole.¹⁴ We confine our attention to horizontal motion only. The two-dimensional phase space mapping is expressed as follows.

$$\begin{pmatrix} z \\ p \end{pmatrix} \leftarrow \begin{pmatrix} \cos 2\pi\nu & \sin 2\pi\nu \\ -\sin 2\pi\nu & \cos 2\pi\nu \end{pmatrix} \begin{pmatrix} z \\ p - \lambda z^2 \end{pmatrix}$$

Here, λ measures the integrated strength of the sextupole. This mapping is called the Hénon map, named after the man generally credited with first studying its properties.¹⁵ We can set $\lambda \equiv 1$ without loss of generality by rescaling, $z \rightarrow z/\lambda$ and $p \rightarrow p/\lambda$. This is in keeping with Hénon's observation that any area preserving quadratic map can be put into a one-

¹⁴I apologize for the difference in size between the two halves of this figure. These black against white images were made from color slides, and the studio personnel misunderstood their instructions. It is too late to attempt a fix.

¹⁵If this were a just world, which it is not, this mapping would be named after one of the accelerator physicists who already had been working with it in connection with sextupoles: perhaps the "Laslett map."

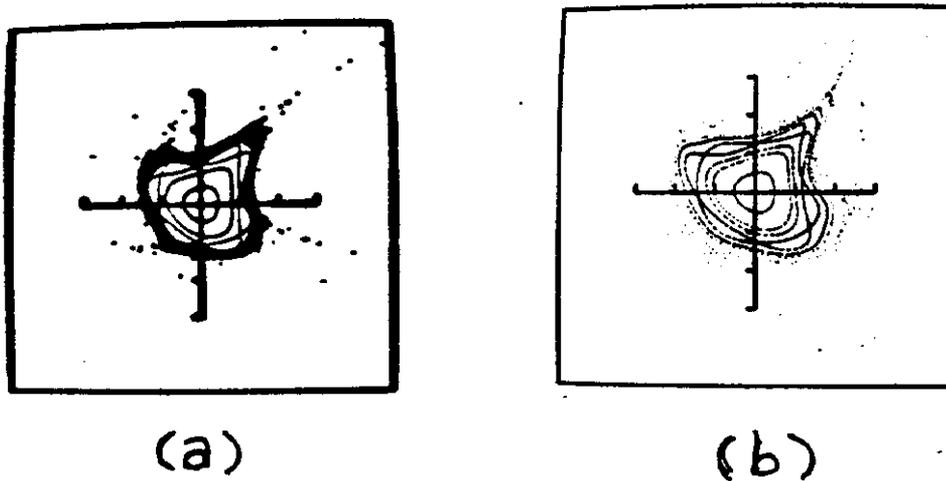


Figure 12: Perturbation theory suggests that the outer stability limits of a Hamiltonian system are "seeded" by low order resonances. (The tic marks on the axes are separated by 0.5.)

parameter form.[12] Part (a) of Figure 12 shows an orbit of the exact map for a tune value $\nu = 0.29$. Its most dramatic feature is the very large $2/7$ resonance which produces a system of seven islands. Seventh "order" resonances (i.e., resonances with winding number seven) should not appear until fifth order in the perturbation expansion, while the island chain is certainly more than a fifth order effect. In fact it is due to an *interference* between the $1/3$ resonance, which appears at first order in the perturbation expansion, and the $1/4$ resonance, which appears at second order. This is confirmed in part (b) which shows the perturbation theoretic prediction when those two resonances are explicitly taken into account. The $2/7$ resonance, which was *not* explicitly put into the Hamiltonian, nonetheless surfaces. Even more important, the stability shoreline appears in approximately the correct location and with approximately the correct shape. Figure 13 shows a similar correspondence at a tune value of $\nu = 0.32$. Here the dominant resonance is the third integer, and it once again seeds the shoreline. The rather chaotic collection of points comes from the exact mapping; the rather strange looking curve from second order perturbation theory with the (first order) third integer resonance filtered out of the transformation. The agreement is terrible for large amplitudes — more terms are required — but the shoreline surrounding the central stable region is again approximated extremely well, both in size and shape. Similar tests at other values of the tune and with octupoles (cubic kicks) indicate that this "resonance seeding" hypothesis can generally predict the shoreline of stability to within 5-15%. Of course, these examples are just for two-dimensional mappings; resonance seeding should be tested on four-dimensional maps as well.

We should not be overly confident in our conceptual model, as in Figure 10, of the behavior of a near-integrable system. It is, after all, limited by our own visualization abilities, and it may break down when the system is far from integrable. Figure 14, for example,

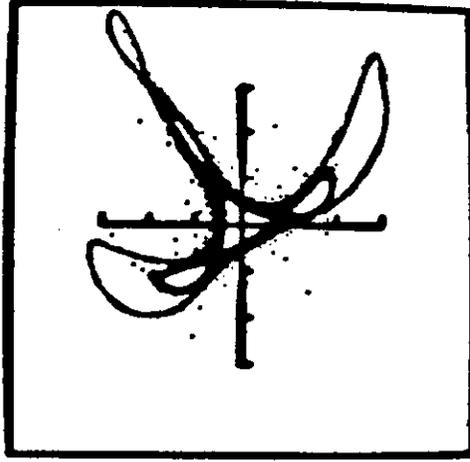


Figure 13: Resonance seeding in the Hénon map at $\nu = 0.32$

displays a three-dimensional projection of a “tangled” orbit arising in a model of the beam-beam interaction. Shown are orbit samples taken using the period advance map. They certainly do not lie on a torus, but neither do they have the randomly scattered appearance we have come to expect from chaotic orbits. My belief, at the moment, is that it is indeed a chaotic orbit but one with a very low entropy — which brings us to the next topic.

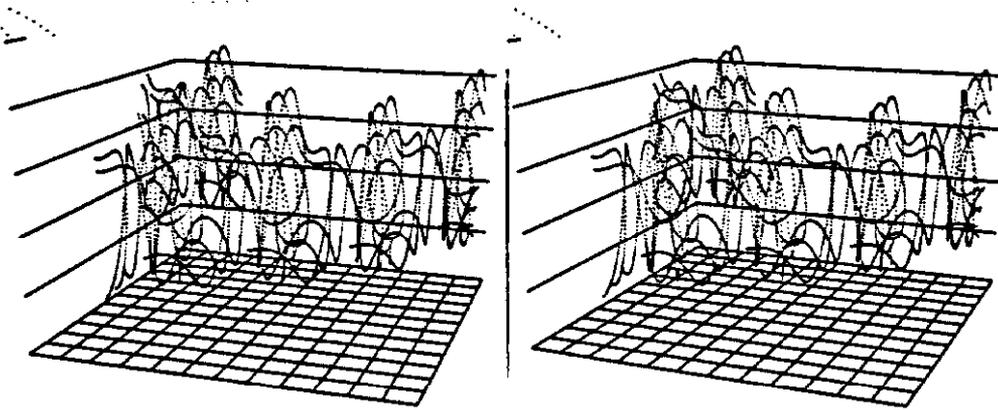


Figure 14: Stereo views showing a three-dimensional projection of a tangled orbit belonging to a beam-beam interaction model; the coordinates are δ_1 , δ_2 , and I_1 .

The essential character of Enlightenment thinking was to allow the clear light of reason to play upon an objective and determinate world. Scarcely a feature of that description now survives intact. ... The world ... is a good deal curiouser and more shadowy than the eighteenth and nineteenth centuries could have conceived. That in itself is no great cause for ... rejoicing. The ancient Hebrews knew well the dangers of the waters of chaos.

— John Polkinghorne
One World

3 CHAOS.

Integrable systems are exceptional: in the metric space of vector fields they occupy a set of measure zero. Nonintegrable systems are everywhere; they cannot be ignored — despite the fact that they *have* been ignored by Physics departments in the United States until very recently. In particular, chaotic orbits can be created in the vicinity of an integrable separatrix by arbitrarily small perturbations. The subject of chaos has achieved almost cult popularity, much like catastrophe theory in the 1970's.¹⁶ We shall not attempt to overview the subject in this section. Our goal is modest: we shall address the question, "What is chaos?" Exactly what objective, quantifiable property of dynamical systems is this word supposed to describe?

¹⁶There is some danger in this, and I am tempted to deliver an amateur essay on the sociology of science but shall resist.

Answering this is not a trivial matter. The fundamental property of a classical dynamical system, after all, is its predictability. The past determines the future. The quintessential statement of this was expressed by Laplace who announced that given the initial positions and velocities of all the particles in the universe, he could calculate the future, thus giving rise to the phrase “Laplacian determinism.” The advent of quantum mechanics pulled us out of this trap: Nature was suddenly seen to be nondeterministic, although this was more a matter of interpretation than of mathematical formalism. The theory predicted deterministic evolution of wave functions, but the interpretation of those wave functions led to a probabilistic model of events in space-time. Laplacian determinism had died, at least among mainstream physicists.

However, it was not necessary to supercede classical mechanics in order to kill determinism. That our understanding has changed dramatically since the time of Laplace is exemplified by the following statement, in which Chirikov speculates on the deeper consequences of chaotic systems.

“It is worth noting that such a motion has been searching for since long ago with the purpose of foundation of the statistical mechanics. . . . Could it be that in Nature there are no such ‘genuine’ random processes as we fancy them? In my opinion . . . it is not excluded that no ‘more random’ processes than . . . the motion of a K-system do really exist.”[6]

Despite the labored English, his message is clear: classical mechanics has changed so drastically that Chirikov suggests that classical chaos can account even for quantum phenomena. Whether one believes him or not — and I do not — it is important to look at classical dynamics in this new light.

The discussion in this section follows along the lines of several authors, but especially Billingsley[5], Sinai[23], and Arnold and Avez[4].

3.1 Phase space partitions.

Period advance maps and Poincaré maps are instances of discrete dynamical systems (DDS), which formally consist of four pieces,

$$DDS \equiv (\mathcal{M}, T, \gamma, \mu) .$$

Here, \mathcal{M} is the phase space, $T: \mathcal{M} \rightarrow \mathcal{M}$ is a mapping, γ is a σ -algebra of measurable subsets of \mathcal{M} , and μ is a measure defined on γ . If \mathcal{M} has finite measure, then we normalize so that $\mu(\mathcal{M}) = 1$. For Hamiltonian systems, it is natural to assume that μ is the invariant Liouville measure, as in Eq.(20) .

Ergodic theory deals with the properties of discrete dynamical systems when T preserves the measure, as is the case with Hamiltonian systems.

$$\forall a \in \gamma: T^{-1}a \in \gamma \quad \& \quad \mu(a) = \mu(T^{-1}a)$$

By writing the condition in this way, we do not require T to be invertible: T^{-1} is defined on the power set of \mathcal{M} , and the symbol $T^{-1}a$ represents the subset of points which map into points in a under T .

We now come to the heart of the matter: chaos and unpredictability can exist in deterministic systems simply because it is impossible for us to measure quantities with infinite accuracy, and any observations which we make must be finitely expressible. Thus, the model appropriate for our observations of a classical system is not \mathcal{M} but a *partition* of \mathcal{M} .

Def: Partition. “ α is a measurable partition of \mathcal{M} ” means: (1) α is a set of measurable subsets of \mathcal{M} , $\alpha \subseteq \gamma$, (2) these subsets are disjoint: $\forall a, a' \in \alpha: \text{either } a = a' \text{ or } a \cap a' = \emptyset$,

and (3) the union of all the sets in the partition is phase space itself: $\mathcal{M} = \bigcup_{a \in \alpha} a$.

Constructing a partition amounts to breaking phase space up into "cells," each of which is a measurable, recordable "macrostate" of the system.

Partitions naturally form a partially ordered set by reason of inclusion.

$$\alpha \geq \beta \text{ means } \forall a \in \alpha \exists b \in \beta : a \subseteq b$$

This is read, "the partition α is a refinement of the partition β ." Thus, $\alpha \geq \beta$ if α is obtained by chopping the cells of β into smaller pieces. The most refined and the least refined (or most coarse) partitions are:

$$\begin{aligned} \text{God's partition : } \varepsilon &\equiv \{\{z\} \mid z \in \mathcal{M}\} \\ \text{the universe : } U &\equiv \{\mathcal{M}\} \end{aligned}$$

In a sense, it is unfair to include ε as a possible partition, since it is nondenumerable, but it serves as an upper bound for all partitions. Clearly, any other partition α of \mathcal{M} must satisfy $\varepsilon \geq \alpha \geq U$. Therefore, any two partitions possess both an upper bound and a lower bound, and it is not surprising that they will also have a least upper bound (lub) and a greatest lower bound (glb). The binary operations corresponding to finding the lub and glb impose an algebraic structure on the set of partitions, transforming it into what algebraists call a "lattice,"¹⁷ whose operations are called "meet" and "join."

$$\begin{aligned} \text{join : } \alpha \vee \beta &\equiv \text{lub}(\alpha, \beta) \\ &= \{a \cap b \mid a \in \alpha \ \& \ b \in \beta\} \\ \text{meet : } \alpha \wedge \beta &\equiv \text{glb}(\alpha, \beta) \end{aligned}$$

These operations are illustrated in Figure 15 .

Comment 27: Unlike the join, there seems to be no binary operator, O , acting on sets, such that

$$\alpha \wedge \beta = \{a \ O \ b \mid a \in \alpha \ \& \ b \in \beta\} .$$

Is this truly the case?

Tracking the evolution of a DDS means observing the state of the system after each "time-step," after each iteration of the mapping. Let z represent the initial state of the system. After k sampling intervals the state will be $T^k z$. Postulate an apparatus for observing the system whose mathematical model is a partition, α . After each iteration we make an observation and record which cell in α the system occupies. The data from such a sequence of measurements can be encoded into a string of symbols, $a_0 a_1 a_2 a_3 \dots a_n$, which is interpreted:

$$\begin{aligned} z \in a_0 \in \alpha \text{ and } Tz \in a_1 \in \alpha \text{ and } T^2 z \in a_2 \in \alpha \\ \dots \text{ and } T^n z \in a_n \in \alpha \end{aligned}$$

where each a_k is in the partition α . This is equivalent to

$$\begin{aligned} z \in a_0 \in \alpha \text{ and } z \in T^{-1} a_1 \in T^{-1} \alpha \text{ and } z \in T^{-2} a_2 \in T^{-2} \alpha \\ \dots \text{ and } z \in T^{-n} a_n \in T^{-n} \alpha \end{aligned} \tag{52}$$

¹⁷ Which is very different from what physicists call a lattice.

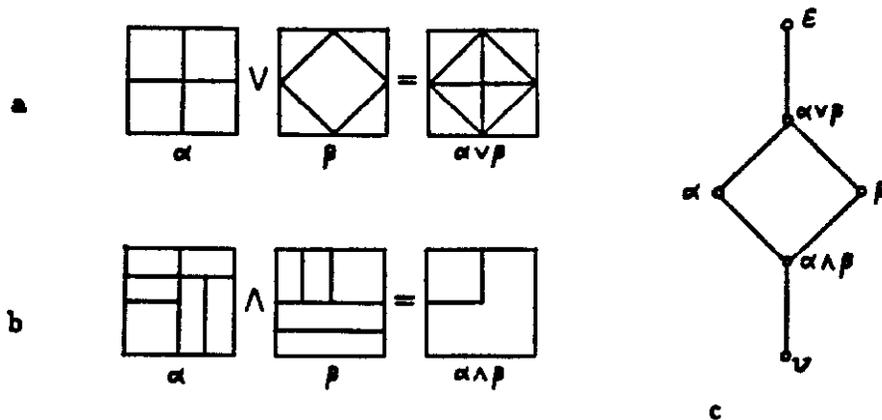


Figure 15: The algebra of partitions.

which, in its turn, is equivalent to

$$x \in \bigcap_{k=0}^n T^{-k} a_k \in \bigvee_{k=0}^n T^{-k} \alpha$$

Our sequence of observations is therefore equivalent to making a *single observation of the initial state* of the system using a partition that is more highly refined than the one provided by our apparatus, α . (Of course, we are assuming classical mechanics throughout: our observations do not disturb the system in any way.) Provided that we design α intelligently, the larger the number of iterations the more refined is the super-partition, $\bigvee_{k=0}^n T^{-k} \alpha$, and the greater the precision with which we know x .

3.2 Information and entropy.

The critical question is this: What is the expected (or average) *information* contained in such a sequence of observations?

Before attacking this, we shall review quickly what is meant by the term "information." The concept was introduced in 1948 by Shannon[22], who used the idea that the information in a message should depend not on the message alone but also on the state of the receiver. Specifically, the information in a message, m , should be a function of the a priori probability, p_m , of receiving that particular message, an assertion which we shall write symbolically as,

$$\text{infor}(m) \equiv f(p_m) .$$

A severe constraint is placed on the function f by requiring that the information content of independent messages be additive. Let m_1 and m_2 be two independent messages with a priori probabilities p_{m_1} and p_{m_2} . Then the joint probability is $p_{m_1} p_{m_2}$, and we have the following line of reasoning.

$$\text{infor}(m_1 \text{ and } m_2) = \text{infor}(m_1) + \text{infor}(m_2)$$

$$\begin{aligned}\Rightarrow f(p_{m_1}, p_{m_2}) &= f(p_{m_1}) + f(p_{m_2}) \\ \Rightarrow f(p) &= -\log p\end{aligned}$$

The negative logarithm is used so that information is a positive quantity; if the logarithm is expressed in base 2, then the unit of information is called a "bit." This definition is intuitively reasonable: less probable messages hold more information than more probable messages, and in the limit $p = 1$ the information content goes to zero.

The average, or expected, information in a set of possible messages, $\{m_1, m_2, m_3, \dots\}$, is called the "entropy" of the set.

$$H(\{m\}) \equiv - \sum_m p_m \log p_m$$

It is intuitively obvious, and easy to verify, that for a finite number of messages, $N = \text{card}[\{m\}]$, $H(\{m\})$ is maximized by making each $p_m = 1/N$.

$$H_{\max} = \log N = \log \text{card}[\{m\}] \quad (53)$$

Comment 28: It is natural to ask if there is any connection between this information theoretic notion of entropy and the one which appears in thermodynamics. In fact, Shannon leaned heavily on statistical mechanics in developing his ideas, and the two concepts are virtually identical. Recall from statistical mechanics that the change in thermodynamic entropy, dS , of a system undergoing an infinitesimal isothermal expansion satisfies

$$\begin{aligned}TdS &= dQ = -\langle dE \rangle + d\langle E \rangle \\ &= -\frac{1}{Z} \sum_m dE_m e^{-\beta E_m} + d\langle E \rangle \\ &= \frac{1}{\beta Z} d \left(\sum_m e^{-\beta E_m} \right) + d\langle E \rangle \\ &= d[kT \ln Z + \langle E \rangle]\end{aligned}$$

where T now represents temperature, not a mapping, and $\beta = 1/kT$. Integrate this equation with the boundary condition that $S = 0$ at $T = 0$.

$$\begin{aligned}S &= k[\ln Z + \beta \langle E \rangle] \\ &= k \left[\ln Z + \sum_m p_m \beta E_m \right] \\ &= k \left[\ln Z - \sum_m p_m \ln(Z p_m) \right] \\ &= -k \sum_m p_m \ln p_m\end{aligned}$$

Apart from Boltzmann's constant, which only normalizes the expression and can be absorbed into the base of the logarithm, this is the average information in an observation of the system's state.

3.3 KS entropy.

A partition of phase space defines a set of messages:

$$m = a \Leftrightarrow z \in a \in \alpha$$

This message states that one observation was made on the system and it was observed to be in the cell labelled "a." To calculate its information, we must assign it an a priori probability. The "natural" one — and one that enjoys the property of being dynamically invariant — uses the phase space measure.

$$p_m \equiv \mu(a)$$

If \mathcal{M} has finite measure, then upon renormalization to $\mu(\mathcal{M}) = 1$ p_m becomes a legitimate probability; if \mathcal{M} has infinite measure, then we must be content with p_m as a relative probability. From this, the entropy (the information expected from a single observation) associated with the partition (or measuring device) α is

$$H(\alpha) = - \sum_{a \in \alpha} \mu(a) \log \mu(a)$$

What, then, is the expected information in the message stream: $a_0 a_1 a_2 \dots a_n$? (see Eq.(52)) This composite message states that the state of the system, z , is in a particular cell of the partition $\bigvee_{k=0}^n T^{-k}\alpha$. The expected information in the message is therefore the entropy of this partition.

$$\begin{aligned} a_0 a_1 a_2 \dots a_n &\rightarrow z \in \bigcap_{k=0}^n T^{-k} a_k \in \bigvee_{k=0}^n T^{-k} \alpha \\ (\text{infor}(a_0 a_1 \dots a_n)) &= H\left(\bigvee_{k=0}^n T^{-k} \alpha\right) \end{aligned}$$

Increasing the number of observations gives us more information about the initial state — the partitions becomes more refined, and the initial state is determined to greater precision. The limiting rate at which information about the system increases with the number of observations is the crucial quantity.

$$h(\alpha, T) \equiv \lim_{n \rightarrow \infty} \frac{1}{n} H\left(\bigvee_{k=0}^n T^{-k} \alpha\right)$$

This number depends on the initial partition, α , or equivalently, on the experimental apparatus. Nothing prevents us from choosing poorly. For example, choosing $\alpha = U$, results in zero information per observation, unless the system suddenly disappears. A better choice of apparatus will optimize $h(\alpha, T)$, thereby providing the maximum rate of increase of information about the system, a quantity we shall symbolize as $h(T)$.

$$h(T) \equiv \sup_{\alpha} h(\alpha, T)$$

The supremum is taken over all possible measurable partitions. This limiting value is called the Kolmogorov-Sinai (KS) entropy of the system.¹⁸

We are finally ready to give an exact meaning to "chaos":

$$\begin{aligned} \text{regular motion} &\text{ means } h(T) = 0 \\ \text{chaotic motion} &\text{ means } h(T) > 0 \end{aligned}$$

¹⁸ More exactly, of the mapping T .

If the system evolves “regularly,” then the information gain per observation approaches zero. Its future is *predictable* in the sense that as time goes by we could continue to increase the interval between observations without loss of information. This is not the case when the motion is chaotic. There is a lower bound on the rate of gaining information. Conversely, there is an upperbound on the interval between observations that can be tolerated without losing information. This has nothing to do with quantum mechanics; it would be true even if we violated the quantum uncertainty principle. It depends only on the fact that no matter how good the measuring apparatus is, it is still necessarily finite.

3.4 Kolmogorov’s theorem.

The definition of KS entropy is not computationally useful. We cannot search through all possible partitions to find the supremum. Fortunately, most sensible partitions attain this limiting value and Kolmogorov’s theorem tells us how to recognize them. We first must define what is meant by a “generating partition.” Loosely speaking, a partition, α , is “generating for T ” if an infinite number of observations will determine \mathbf{x} completely. Heuristically, this means that the size of the cells in $\bigvee_{k=0}^n T^{-k}\alpha$ can be made arbitrarily small by taking n large enough. A more precise way of putting it is this: if γ is the σ -algebra of sets over which μ is defined, then $\bigvee_{k=0}^n T^{-k}\alpha \rightarrow \gamma$ as $n \rightarrow \infty$. A partition which violates this is particularly poor in that it will have a limiting coarseness to it. The fundamental theorem which we need is then stated as follows.

Kolmogorov’s theorem: If (a) α is generating for T and (b) $H(\alpha)$ is finite, then

$$h(T) = h(\alpha, T).$$

Comment 29: This result seems plausible. It is remarkable, nonetheless, that all generating partitions, some of which can be very crude indeed, attain the same limiting $h(\alpha, T)$. If α is generating, then the original dynamical system is equivalent to shifts on strings of message symbols. This is the approach of “symbolic dynamics.”

EXAMPLE: Harmonic oscillator. We observe a harmonic oscillator at equally spaced times, say $0, \tau, 2\tau, \dots$ and record whether its velocity is positive or negative. If time $t = 0$ is set to correspond to maximum negative displacement of the oscillator, then the n^{th} measurement records $\text{sgn} \sin(n\omega\tau)$, where $\omega = 2\pi f$ is the angular frequency of the oscillator. Let us take the state variable, \mathbf{x} , to be the phase of the oscillator divided by 2π ; its value, mod 1, lies in the interval $U^1 = [0, 1)$ and the mapping corresponding to one sampling interval is

$$T : \mathbf{x} \mapsto \mathbf{x} + f\tau \pmod{1} .$$

Finally, the partition describing our apparatus is simply

$$\begin{aligned} \alpha &= \{a_0, a_1\} \\ a_0 &= [0, 1/2) \\ a_1 &= [1/2, 1) \end{aligned} \tag{54}$$

If $\mathbf{x} \in a_0$ the oscillator has positive velocity, and if $\mathbf{x} \in a_1$ its velocity is negative. Designing the experiment intelligently means that we choose τ so that $f\tau$ is irrational. Then, by the “zeroth ergodic theorem” of the first lecture, α will be a generating partition, and its entropy should equal the KS entropy of the oscillator.

To evaluate this, note that each observation adds only two cells to $\bigvee_{k=0}^n T^{-k}\alpha$, from which it follows that

$$\text{card} \left[\bigvee_{k=0}^n T^{-k}\alpha \right] = 2(n+1) .$$

By Eq.(53), we then have the following inequalities.

$$\begin{aligned} H\left(\bigvee_{k=0}^n T^{-k}\alpha\right) &\leq \log[2(n+1)] \\ h(\alpha, T) &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log[2(n+1)] \\ &= 0 \end{aligned}$$

Since necessarily $h(\alpha, T) \geq 0$, it follows that $h(\alpha, T) = 0$. Finally, since α is generating for T , Kolmogorov's theorem tells us that $h(T) = 0$.

Comment 30: To nobody's surprise, the harmonic oscillator undergoes regular motion. The key reason was that the number of cells in the super-partition did not increase quickly enough. Indeed, we can make the general observation that since $H \leq \log N$, N must grow at least exponentially fast to have $h(T) \neq 0$.

EXAMPLE: Doubling. We consider an "abstract" dynamical system defined by assuming the same partition as before, Eq.(54), but now we take the map to be $T: z \mapsto 2z \pmod 1$. (This is an example of an irreversible, measure preserving mapping.) Consider the first step in the evaluation of $\bigvee_{k=0}^n T^{-k}\alpha$. The partition $T^{-1}\alpha$ consists two sets, the first of which maps into α_0 under T , and the second into α_1 .

$$T^{-1}\alpha = \left\{ \left[0, \frac{1}{4}\right) \cup \left[\frac{1}{2}, \frac{3}{4}\right), \left[\frac{1}{4}, \frac{1}{2}\right) \cup \left[\frac{3}{4}, 1\right) \right\}$$

We now "join" this with α to get

$$\alpha \vee T^{-1}\alpha = \left\{ \left[0, \frac{1}{4}\right), \left[\frac{1}{4}, \frac{1}{2}\right), \left[\frac{1}{2}, \frac{3}{4}\right), \left[\frac{3}{4}, 1\right) \right\}$$

Going on to subsequent steps, at the n^{th} stage the unit interval is divided into subintervals of size $1/2^{n+1}$, giving us the following result.

$$\bigvee_{k=0}^n T^{-k}\alpha = \left\{ \frac{1}{2^{n+1}}[m, m+1) \mid m = 0, 1, \dots, 2^{n+1} - 1 \right\}$$

This is obviously a generating partition for T , and we can proceed to evaluate the KS entropy. Since every cell in this super-partition is the same size,

$$\forall a \in \bigvee_{k=0}^n T^{-k}\alpha : \mu(a) = \frac{1}{2^{n+1}} ,$$

this becomes an easy calculation.

$$\begin{aligned} H\left(\bigvee_{k=0}^n T^{-k}\alpha\right) &= -\sum \mu(a) \log \mu(a) \\ &= \log[2^{n+1}] \end{aligned}$$

$$= (n + 1) \log 2$$

$$\begin{aligned} h(T) = h(\alpha, T) &= \lim_{n \rightarrow \infty} \frac{1}{n} (n + 1) \log 2 \\ &= \log 2 \quad (1 \text{ bit}) \end{aligned}$$

Comment 31: What an eminently reasonable answer! It has a natural interpretation if we think of \mathbf{x} as represented by its binary expansion. The mapping T then corresponds to shifting the symbols left one position, and the partition α means that one measurement corresponds to noting whether the first digit to the right of the decimal is a 1 or a 0. A sequence of measurements just reproduces the original binary expansion. The information in a measurement that is skipped cannot be recaptured by subsequent observations: it is gone forever.

EXAMPLE: Linear automorphisms of a torus. Suppose that \mathcal{M} is a two-dimensional torus, which we consider as topologically equivalent to U^2 with identification of opposite edges.

$$\mathcal{M} \simeq U^2 = [0, 1) \times [0, 1)$$

Consider the mapping,

$$T : \mathbf{x} \mapsto \mathbf{A}\mathbf{x} \text{ mod } \mathcal{M}$$

where \mathbf{A} is an integer, 2×2 , unimodular matrix. Because of these conditions on \mathbf{A} , the mapping represents a true one-to-one mapping of the torus onto itself. Calculating the entropy by means of a generating partition is more difficult than in the two previous examples; details can be found in Sinai.¹⁹ [23] The answer is

$$h(T) = \log \lambda_+$$

where λ_+ is the eigenvalue of \mathbf{A} satisfying $\lambda_+ > 1$.

3.5 Lyapunov exponents.

Although the KS entropy is a well defined mathematical concept, using generating partitions to calculate its value is intractable in all but the simplest models. In addition, it assumes that all orbits are of the same character, whereas in most real problems the dynamical system will possess both regular and chaotic orbits. A different approach is needed.

Fortunately, one exists. Chaos is characterized by a phenomenon of divergence: two chaotic orbits that start out infinitesimally close to each other diverge exponentially rapidly. Consider, for example, the evolution of two infinitesimally close orbits, say with initial conditions \mathbf{x} and $\mathbf{x} + \epsilon$ under the doubling map. After n iterates,

$$\begin{aligned} T^n(\mathbf{x} + \epsilon) &= 2^n(\mathbf{x} + \epsilon) \\ &= 2^n\mathbf{x} + 2^n\epsilon \\ &= T^n\mathbf{x} + 2^n\epsilon \end{aligned}$$

Thus, the distance between the two orbits grows exponentially.

$$|T^n(\mathbf{x} + \epsilon) - T^n(\mathbf{x})| = e^{\Gamma n} |\epsilon|, \quad \text{where } \Gamma = \log 2$$

¹⁹ Although I think his proof is flawed.

Consider also the behavior of two infinitesimally close orbits under the torus automorphism.

$$T(\underline{x} + \underline{\epsilon}) = \mathbf{A}\underline{x} + \mathbf{A}\underline{\epsilon} \pmod{\mathcal{M}}$$

$$\begin{aligned} |T^n(\underline{x} + \underline{\epsilon}) - T^n(\underline{x})| &= |\mathbf{A}^n \underline{\epsilon}| \\ &\approx \lambda_+^n |\underline{\epsilon}| \\ &= e^{\Gamma n} |\underline{\epsilon}|, \quad \text{where } \Gamma = \log \lambda_+ \end{aligned}$$

Of course, doing only two examples does not prove anything, but in each case infinitesimally close orbits diverge exponentially at a rate, Γ , which is numerically equal to $h(T)$, the KS entropy of the map. This provides, then, an alternate, more computationally tractable method for calculating $h(T)$.

Making a giant leap forward, we now consider the divergence of infinitesimally close orbits under iteration of a generic map, T .

$$\begin{aligned} T(\underline{x} + \underline{\epsilon}) &= T(\underline{x}) + \underline{\epsilon} \cdot \underline{DT}(\underline{x}) \\ T^2(\underline{x} + \underline{\epsilon}) &= T(T(\underline{x}) + \underline{\epsilon} \cdot \underline{DT}(\underline{x})) \\ &= T(T(\underline{x})) + \underline{\epsilon} \cdot \underline{DT}(\underline{x}) \cdot \underline{DT}(T(\underline{x})) \\ &\equiv T^2(\underline{x}) + \underline{\epsilon} \cdot \underline{DT}^2(\underline{x}) \\ &\vdots \\ T^n(\underline{x} + \underline{\epsilon}) &= T^n(\underline{x}) + \underline{\epsilon} \cdot \underline{DT}^n(\underline{x}), \end{aligned}$$

where the \underline{DT}^n are computed recursively.

$$\underline{DT}^n(\underline{x}) = \underline{DT}^{n-1}(\underline{x}) \cdot \underline{DT}(T^{n-1}(\underline{x}))$$

If the orbits are diverging exponentially, then we can connect the rate to the norm of \underline{DT} .

$$\begin{aligned} |T^n(\underline{x} + \underline{\epsilon}) - T^n(\underline{x})| &\stackrel{?}{\approx} |\underline{\epsilon}| e^{\Gamma n} \\ &\approx |\underline{\epsilon}| \|\underline{DT}^n(\underline{x})\| \end{aligned}$$

This suggests both a definition,

$$\Gamma(\underline{x}) \stackrel{?}{\equiv} \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|\underline{DT}^n(\underline{x})\|$$

and an association, $h(T) \stackrel{?}{\equiv} \Gamma$.

This cannot be correct, however. In the first place $\Gamma(\underline{x})$ depends on the initial conditions, \underline{x} , while $h(T)$ is a global number, and in the second, the norm of \underline{DT} does not carry enough information to characterize the motion.

The correct answer is only a little more complicated.[14] Let $\lambda_1 \geq \lambda_2 \geq \dots \lambda_N$ be the N eigenvalues of $\underline{DT}^n(\underline{x})$ arranged in decreasing order. The "Lyapunov exponents" of the map T evaluated for the orbit passing through \underline{x} are the numbers

$$\Gamma_k(\underline{x}) \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \ln \lambda_k(\underline{x}) .$$

The fundamental result which then connects these quantities to the KS entropy is as follows.

$$h(T) = \int_{\mathcal{M}} d\mu(\underline{x}) \sum_{\Gamma_k > 0} \Gamma_k(\underline{x})$$

This expresses the global quantity, KS entropy, as an average of a local function, the Lyapunov coefficients. It reflects the fact that regular and chaotic motion can exist simultaneously but in different regions of phase space.

3.6 Fractal dimensions.

Nonzero entropy is not the only signature of chaotic orbits; for example, they are associated with broad band Fourier spectra and fractal dimensions as well. The latter quantity in particular has captured people's imagination, and we shall consider it briefly before closing.

Consider the problem of operationally determining the dimension of a set of points, such as a surface of some kind, given a procedure which uniformly samples points in the set. One approach, first suggested by Hausdorff, is to cover the set — or the generated points — with spheres of a given radius. The minimum number of spheres (or cubes, or ellipsoids, or whatever) with radius $\leq \tau$ needed to cover a set will grow like an inverse power of τ , for small τ .

$$N(\tau) \sim \tau^{-d}$$

The exponent is the "Hausdorff dimension" of the set.

$$d = - \lim_{\tau \rightarrow 0} \frac{\log N(\tau)}{\log \tau} \quad (55)$$

The characterization of this number as a dimension is confirmed by the observation that when the set is (an open subset of) a manifold, its value is indeed the dimension of the manifold: the Hausdorff dimension of a curve is 1, of a surface 2, of a volume 3, and so forth. To see this most trivially, consider packing a measurable subset of R^3 with small spheres of radius τ . If the volume of the set is V ,

$$V \approx N(\tau) \times \frac{4}{3} \pi \tau^3 ,$$

from which,

$$N(\tau) \sim \tau^{-3} ,$$

so that the Hausdorff dimension is three. However, the procedure defining this number makes no assumption about the point set being sampled; it need not be a manifold, and its value need not be an integer. This can happen, for example, with samples of a chaotic orbit confined to a compact subset of phase space, for strange attractors in dissipative systems, or for basin boundaries of systems with more than one attractor. A set with non-integer Hausdorff dimension is a **fractal**, and its dimension is then called a **fractal dimension**.

Example: Cantor's set. The patriarch of all fractals was devised by Cantor to demonstrate the existence of a set which had zero measure and the cardinality of the continuum. It is defined as follows. Begin with the sets,

$$\begin{aligned} C_0 &\equiv U^1 = [0, 1] \subset R \\ S_0 &\equiv \{U^1\} . \end{aligned}$$

Notice that C_0 is an interval of the real axis, while S_0 is a set whose element is C_0 . We now define a sequence of sets, S_k , $k = 0, 1, 2, \dots$, recursively by specifying the members of S_{k+1} in terms of those of S_k . This is described most easily with a pseudo-program:

$$\begin{aligned} S_0 &\equiv \{U^1\}; \\ \text{for } (k = 1 \dots \infty) \{ \end{aligned}$$

$$S_k \equiv \emptyset;$$

$$\text{for every } (x, y) \in S_{k-1} : S_k = S_k \cup \{(x, \frac{2}{3}x + \frac{1}{3}y)\} \cup \{(\frac{1}{3}x + \frac{2}{3}y, y)\};$$

$$\}$$

In words: S_k is obtained from the intervals in S_{k-1} by a process of chopping out their middle thirds. Thus, in particular,

$$S_0 = \{[0, 1]\}$$

$$S_1 = \{[0, \frac{1}{3}], [\frac{2}{3}, 1]\}$$

$$S_2 = \{[0, \frac{1}{9}], [\frac{2}{9}, \frac{1}{3}], [\frac{2}{3}, \frac{7}{9}], [\frac{8}{9}, 1]\}$$

and so forth. The k^{th} approximant to Cantor's set is the union of all intervals in S_k .

$$C_k \equiv \bigcup_{s \in S_k} s$$

Finally, the Cantor set itself is the intersection of all its approximants.

$$C \equiv \bigcap_{k=0}^{\infty} C_k$$

(This is a set-theoretic way of saying $C = \lim_{k \rightarrow \infty} C_k$.) It is easy to see that the set has measure zero, $\mu(C) = 0$, since it is covered by each approximant, C_k , and $\mu(C_k) = (2/3)^k$. Further, it has the cardinality of the continuum, since C comprises all those numbers in U^1 whose expansion in base three contains only the digits 0 and 2. By changing each 2 to a 1, and reinterpreting the string as a *binary* expansion, we can map Cantor's set one-to-one onto U^1 . Finally, its Hausdorff dimension is not an integer. To see this, note that each S_k contains 2^k intervals of radius $(1/3)^k$ and covers C . A moment's reflection is sufficient to convince that it is the smallest such set. Therefore, the dimension of C can be evaluated, after taking the limit in Eq.(55),

$$d = - \lim_{k \rightarrow \infty} \frac{\log[2^k]}{\log[(1/3)^k]} = \log 2 / \log 3$$

Comment 32: Cantor's set shares one more property with other fractals: it is self-similar. Multiplication by three maps the lower third of Cantor's set onto the full set.

$$C' \equiv 3C \cap U^1 = C$$

$$x \in C' \text{ iff } x \in U^1 \text{ and } \frac{1}{3}x \in C$$

Comment 33: The term "fractal" was coined by Benoit Mandelbrot, the man largely responsible for their reintroduction into the modern stream of collective consciousness.[15,16] Following his lead, researchers have found fractals arising in a wide variety of applications. In retrospect, this is not surprising. We began these lectures by talking about manifolds as the stage on which dynamical systems perform. This *idée fixe* that the background environment is smooth pervades all of physics, from freshman mechanics to general relativity.

We inherited it from the Greeks, who were principally interested in studying such things as circles and triangles. The stage which they employed was the Euclidean plane, and we have basically projected their concept forward through thirty, or so, centuries. Lost somewhere in that process was a fundamental observation: circles and triangles do not exist in Nature; they were mental constructs very useful for solving a particular class of problems. Fractals, also mental constructs, are far more appropriate to describe large classes of natural and mathematical phenomena which cannot be handled by the Greek models. One need only stare at a tree to see how Nature makes use of fractal structures and to appreciate their beauty. It has been suggested that fractal figures drawn by computers are the modern equivalent of the Greeks' circles and triangles, a first step beyond those smooth models which they created and we inherited; it is conceivable that fractals may eventually lead us to new geometries; it is possible that among these we may find one that serves us better in describing the way our world works. For now, however, this is little more than a popular speculation and likely to remain so for quite some time.

Let me close these lectures by offering two exercises which lead to fractals.

(a) **Fractal basin boundary.** Using bitmap graphics, draw the set of starting points in the complex plane for which Newton's method fails to find n^{th} roots of unity. In particular, find all complex z for which the iterative scheme $z \leftarrow (2z^3 + 1)/3z^2$ fails to converge to one of the three cube roots of unity: 1 and $e^{\pm 2\pi i/3}$. The roots of unity in this example are attractors of Newton's map. The set of all points in the complex plane which converge to one of them is called its "basin of attraction." The points for which Newton's method fail are on the boundary between adjoining basins; hence the term "fractal basin boundary."

(b) **Strange attractor.** Let $f_1, f_2, f_3 : U^2 \rightarrow U^2$, be three contractive mappings of the unit square into itself, defined as follows.

$$\begin{aligned} f_1 : (x, y) &\mapsto (x/2, y/2) \\ f_2 : (x, y) &\mapsto ((x+1)/2, y/2) \\ f_3 : (x, y) &\mapsto (x/2, (y+1)/2) \end{aligned}$$

Construct a stochastic process on U^1 by randomly choosing at each step from f_1, f_2 , and f_3 . That is, z_0, z_1, z_2, \dots will be a random sequence of points in U^2 , and for all k , $\text{Prob}[z_{k+1} = f_n(z_k)] = 1/3$, $n = 1, 2, 3$. Beginning anywhere in U^2 , plot an orbit of this process. The strange attractor which results has an obvious relation to Pascal's triangle mod 2. What is its Hausdorff dimension?

References

- [1] Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics*. Benjamin/Cummings, Reading, Massachusetts, 1982.
- [2] V. I. Arnold. *Ordinary Differential Equations*. The MIT Press, Cambridge, Massachusetts, 1980.
- [3] V. I. Arnold. Proof of a. n. kolmogorov's theorem on the preservation of quasi-periodic motions under small perturbations of the hamiltonian. *Russian Mathematical Surveys*, 18:9-36, 1963.
- [4] V. I. Arnold and A. Avez. *Ergodic Problems of Classical Mechanics*. W. A. Benjamin, New York, 1968.
- [5] Patrick Billingsley. *Ergodic Theory and Information*. John Wiley & Sons, New York, 1965.
- [6] Boris V. Chirikov. A universal instability of many-dimensional oscillator systems. *Physics Reports*, 52(5):263-379, 1979.
- [7] P. A. M. Dirac. *The Principles of Quantum Mechanics*. Oxford at the Clarendon Press, London, fourth edition, 1966.
- [8] Robert Gilmore. *Catastrophe Theory for Scientists and Engineers*. John Wiley & Sons, New York, 1981.
- [9] Herbert Goldstein. *Classical Mechanics*. Addison-Wesley, Reading, Massachusetts, 1959.
- [10] Victor Guillemin and Shlomo Sternberg. *Symplectic Techniques in Physics*. Cambridge University Press, New York, 1984.
- [11] Enrique Henestroza. *Perturbative Analysis of Nonlinear Betatron Oscillations*. Master's thesis, Massachusetts Institute of Technology, 1988.
- [12] M. Hénon. *Quart. App. Math.*, 27:291, 1969.
- [13] Serge Lang. *Differential Manifolds*. Addison Wesley, Reading, Massachusetts, 1972.
- [14] A. J. Lichtenberg and M. A. Lieberman. *Regular and Stochastic Motion*. Springer-Verlag, New York, 1983.
- [15] Benoit B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Company, San Francisco, 1983.
- [16] Benoit B. Mandelbrot. *Fractals: Form, Chance, and Dimension*. W. H. Freeman and Company, San Francisco, 1977.
- [17] Leo Michelotti. *Catastrophe and Maxwell Surfaces of the Half Integer Resonance Excited by Quadrupoles and Octupoles*. Fermi Note 393, Fermilab, 1983.
- [18] Leo Michelotti. Deprit's algorithm, green's functions and multipole perturbation theory. *Particle Accelerators*, 19:205-210, 1986.
- [19] Leo Michelotti. Introduction to the nonlinear dynamics arising from magnetic multipoles. In Melvin Month and Margaret Dienes, editors, *Physics of Particle Accelerators*, American Institute of Physics, 1987. AIP Conference Proceedings No. 153.

- [20] Leo Michelotti. Moser-like transformations using the lie transform. *Particle Accelerators*, 16:233, 1985.
- [21] Leo Michelotti. Perturbation theory and the single sextupole. In Y. S. Kim and W. W. Zachary, editors, *The Physics of Phase Space*, Springer-Verlag, 1987.
- [22] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379-423,623-656, 1948.
- [23] Y. G. Sinai. *Introduction to Ergodic Theory*. Princeton University Press, Princeton, New Jersey, 1976.
- [24] Walter Thirring. *Classical Dynamical Systems*. Springer-Verlag, New York, 1978.
- [25] R. Thom. *Structural Stability and Morphogenesis*. W. A. Benjamin, Reading, Massachusetts, 1975.
- [26] H. Weyl, C. L. Siegel, and K. Mahler. *Geometry of Numbers*. Princeton University Press, Princeton, 1950.