THE DYNAMICAL SYSTEMS APPROACH
TO DIFFERENTIAL EQUATIONS

BY MORRIS W. HIRSCH[1]

> This harmony that human intelligence believes it discovers in nature —does it exist apart from that intelligence? No, without doubt, a reality completely independent of the spirit which conceives it, sees it or feels it, is an impossibility. A world so exterior as that, even if it existed, would be forever inaccessible to us. But what we call objective reality is, in the last analysis, that which is common to several thinking beings, and could be common to all; this common part, we will see, can be nothing but the harmony expressed by mathematical laws.
>
> H. Poincaré, *La valeur de la science*, p. 9

> ...ignorance of the roots of the subject has its price—no one denies that modern formulations are clear, elegant and precise; it's just that it's impossible to comprehend how any one ever thought of them.
>
> M. Spivak, *A comprehensive introduction to differential geometry*

INTRODUCTION

The mathematical subject we call dynamical systems was fathered by Poincaré, developed sturdily under Birkhoff, and has enjoyed a vigorous new growth for the last twenty years. As I try to show in Chapter I, it is interesting to look at this mathematical development as the natural outcome of a much broader theme which is as old as science itself: the universe is a sytem which changes in time.

Every mathematician has a particular way of thinking about mathematics, but rarely makes it explicit. Yet such perspectives, in any particular field, can be of great value to nonexperts who must apply the results of the field, or who want to learn them. In some fields the points of view of its practitioners cluster around several comparatively clearly defined concepts and insights. Dynamical systems theory is one of them; the basic concepts go back to Poincaré. The main thrust of Chapters I and II is to explain, rather discursively, some of

---

these basic themes in dynamical systems theory (as they appear to me), together with their backgrounds and some exemplifying theorems. I try to be comprehensible but not comprehensive; there are many topics which are not discussed, and points of view that are not mentioned. But what is covered is at least representative of much current work in the field.

The last chapter presents some recent results which illustrate the themes developed in the first two chapters and which can be better understood in the light of those themes. They concern some classes of systems in which most trajectories can be expected to approach equilibrium—unlike many currently fashionable systems.

I have obtained valuable insight and knowledge from many people. Particularly, helpful have been Giles Auchmuty, John Franks, David Fried, Jack Hale, Xavier Mora, Jenny Harrison, Nancy Kopell, Charles Pugh and especially Stephen Smale. I thank Felix Browder for encouraging me to write this article.

CHAPTER I: HISTORICAL BACKGROUND

1. Systems, dynamics, and the origins of science.

> Although it may be fashionable to acknowledge that everything is connected to everything else in principle, some things are more tightly connected to each other than to all the rest. Such a little knot of causal interactions goes by the name of a system.
>
> A. Winfree, *The geometry of biological time*
>
> All models divide naturally...into two a priori distinct parts: one kinematics, whose aim is to parameterize the forms of the states of the process under consideration, and the other is *dynamic*, describing the evolution in time of these forms.
>
> R. Thom, *Structural stability and morphogenesis* (1975), p. 4

It is the nature of minds to organize experience. Human minds can recognize individual objects and perceive relations between them, which in turn become objects; lump similar objects together into more abstract concepts; and analyze complex objects in terms of simpler ones. (Human minds may do other things; other minds may do these things.)

One result of this activity is that we perceive, construct, and analyze systems: A *system* is something having parts which is perceived as a single entity. Not everything is a system: Euclid defined a point as that which has no parts; in most theologies God has no parts; the empty set has no parts. But most things can usefully be seen as systems of some kind. Some well-known systems are: the solar system, the capitalist system, the decimal system, the federal reserve system, and the cardiovascular system.

Wittgenstein said of mathematical systems:

> A system is, so to speak, a world.

> In mathematics we cannot talk of systems in general, but only *within* systems. They are just what we cannot talk about.

Gödel's theorems can be interpreted as a precise formulation of Wittgenstein's meaning.

The parts making up a system may be clearly or vaguely defined. The interesting thing about a system is the way the parts are related to each other. For the systems studied in mathematics, the parts and their relations must be so clearly defined that we can single out a particular set of these relations as completely characterizing the *state* of the system. The mathematician then identifies the system with the collection of all its conceivable states.

It seems to be necessary, at least in the current state of mathematical art, that the state space be clearly and unambiguously defined. Unfortunately this usually means that the mathematical system is drastically oversimplified in comparison with the natural system being modelled. In an ecological system, for example, the state of the system is often taken to be the list of the populations of a fixed set of species. But in reality species may die out and new ones may appear through migration, mutation, or evolution. Or we may be unclear as to how many distinct species are present. A related problem comes up in economic models, in which firms enter and leave the market. In physics similar problems arise in studying phase transitions. To take another biological example, there are grave difficulties in modelling a developing embryo: the number of cells change, new tissues and organs appear whose configurations suddenly become an important part of the state. When such problems cannot be ignored, we usually have no recourse except simply to switch attention to a different system—a gas instead of a liquid, an adult organism instead of an embryo. (There is an interesting analogy to the mathematical operation of changing coordinates—except that here we do not know how to describe the coordinate changes!)

A *dynamical* system is one which changes in time; what changes is the state of the system. The capitalist system is dynamical (according to Marx), while the decimal system is (we hope) not dynamical. A mathematical dynamical system consists of the space of states of the system together with a rule called the *dynamic* for determining the state which corresponds at a given future time to a given present state. Determining such rules for various natural systems is a central problem of science. Once the dynamic is given, it is the task of mathematical dynamical systems theory to investigate the patterns of how states change in the long run.

From the earliest myths to general relativity, the most important dynamical system has been the natural system of the cosmos. The heavenly bodies are the parts, the states are their possible configurations; the central problem is to find the dynamic. While some aspects of the dynamic are easily grasped, others, such as the movements of the planets and the timing of the solstices, are far from obvious. It has been suggested that the purpose of the large stone monuments, such as Stonehenge, was to accurately determine the solstices; if so they must embody untold eons of observation of the dynamic of the cosmos.

Effective as Stonehenge may be as a calculating device, like any such device it affords its users no conceptual grasp of the dynamic it is calculating. Far better were the geometric pictures of the cosmos, such as the rolling spheres of Eudoxus and the epicycles of Ptolemy. The mind can seize as a whole the

picture of a family of spheres rolling on each other, one for each planet and one for the stars, and play with it. The picture is adaptable; extra eccentric spheres can be inserted to account for occasional retrograde planetary motion until a reasonable fit with observation is achieved. Most important was the capability of making theoretical calculations and predictions. By contemplating these rolling spheres the ancients took a momentous step: they studied a natural system by means of a *mathematical model*.

The various geometrical models of Ptolemy, Eudoxus, Copernicus, and Brahe had more or less success at calculating the dynamic of the solar system, but each suffered from the defect of being *too* adaptable. One could always insert an extra epicycle or two to fix up each newly found discrepancy from observation, but there were no guiding principles which might point the way to the systematic development of better models. While Copernicus' sun-centered model might carry tremendous philosophical implications, it did not lead to significantly more accurate predictions than the earth-centered ones (to which it is mathematically equivalent).

Kepler's dynamical method was fundamentally different. Instead of looking only for a mathematical system that would closely fit observations, he also sought inner laws governing the observations, and systems that would exhibit these laws.

> Kepler's principal goal was to explain the relationship between the existence of five planets (and their motions) and the five regular solids. It is customary to sneer at Kepler for this.... It is instructive to compare this with the current attempts to "explain" the zoology of elementary particles in terms of irreducible representations of Lie groups.
>
> S. Sternberg (1969), p. 95

The existence of five planets and five regular solids is a mathematical regularity of the universe; Kepler tried to explain it. Believing that there must be regularities inherent in the distances of planets from the sun and their periods and velocities of their revolution about the sun, he searched for them in the mass of Brahe's observations. Instead of building his system on the preconceived regularity of the circle, he took the modern approach of looking for regularities in the data—and found elliptical orbits, the proportionality of planetary year to the two-thirds power of distance from the sun, and the first conservation law: that the radius vector from sun to planet sweeps out equal areas in equal times.

Galileo observed the heavens through a telescope, but he also spent thirty years studying the humble terrestrial system of a ball rolling down an incline. Dwarfed by the grandeur of the cosmos, this simple system had a crucial scientific advantage: Galileo could perform experiments. (They may have been only thought experiments—but at least he could think about them.) These could not have been easy as there was no accurate way to measure time. Through his discovery of the laws relating to time, distance, and velocity—a more subtle problem than planetary forecasting—Galileo founded the first field of modern science, dynamical systems as a branch of mathematical physics.

We drill into every calculus student that in the study of motion the independent variable is time; but this was not so in Galileo's day.

> It was more natural at first to think of speeds and times as contraries and to relate each separately to distances which he could measure.
>
> S. Drake (1978)

Another advantage calculus students have over Galileo is their familiarity with the concept of continuously changing instantaneous speed. Drake says:

> The concept of "speed at a point" long appeared to Galileo as a mere fiction, if not a contradiction in terms. Speed required motion, which could not take place at a point.... Lines and distances were universally granted to be continuous magnitudes. Speed was not and Galileo was slow to follow in this regard the direction in which his mathematics pointed. It is a mistake to assume from the outset that mathematics governed nature and physics must conform to it; rather, mathematics gradually forced his hand in this thorny question of literally continuous change.
>
> S. Drake (1978), p. 116

The greatest of Galileo's achievements—without which not only would the mathematical study of dynamical systems be impossible, but all of modern science—was to teach us to consider time, motion, and velocity not as mysterious qualities or essences, but rather as mere variables to be externally measured and mathematically computed. In *The edge of objectivity* (1960), p. 42, C. C. Gillispie writes:

> What was original, therefore was Galileo's ultimate conception of motion, not his criticism of Aristotle. Indeed, so original was it that it may be taken as one of those exceedingly rare events, a true mutation in ideas, a break with the past. It altered man's consciousness of a real world outside himself in nature. This new world is to be grasped rather by measurement than by sympathy. In it an Archimedean science is possible, not just of statics contemplating things at rest, but of dynamics studying to know things in motion. One may say with confidence what was revolutionary in Galileo's law of falling bodies: It was that he treated time as an abstract parameter of a purely physical event. This enabled him to do what no Greek had done, to quantify motion. Galileo spent twenty years wrestling with the problem before he got free of man's natural biological instinct for time as that in which he lives and grows old. Time eluded science until Galileo.

In his desire to unify the cosmos, Galileo unfortunately did not correctly join his terrestrial dynamics to celestial physics. He paid no attention to Kepler's ellipses, and even though he knew that projectiles follow parabolic trajectories, he thought that bodies falling from rest on a rotating earth follow circular paths. And although Galileo thought of motion as persistent, he never formulated the principle of inertia. According to Gillispie (p. 51):

Motion is again the problem, the silver cord of all who could not make the break with Greece...when forced to choose between cosmic order and the absolute mathematization of nature, he too chose order. For the function of science in Greece was to explain the universe in a single rationale, and not simply to generalize some limited set of phenomena. A universe which can be explained, a cosmos which we can fit, must be finite, and Galileo never quite confronted the prospect of infinity. Consequently natural motion for Galileo, inertial motion, is that motion which neither rises nor falls, which is equidistant from the center of the earth. It is, therefore, circular motion. No longer the center of the cosmos, the earth remained the center of motion.

It was Descartes who stated the law of inertia: a body tends to stay in the same state, changing its state only on impact; and unconstrained motion follows a straight line. He also bequeathed to us the infinitude of space, the methods of analytic geometry, and the conception of the world, including nature, as a machine. These, and not his physics of whirling vortices of ethereal matter, were Descartes' real contributions to dynamics.

Newton and Leibniz each invented the basic dynamical tool, calculus. But Newton did much more: he set out a *physical* hypothesis—the law of universal gravitation—in sufficiently precise mathematical form that from it he could *derive* the cosmic dynamic. Kepler's and Galileo's laws became rigorous mathematical consequences.

It is curious that although Newton must have discovered his theorems in dynamics through the use of calculus, his exposition of them in *Principia* uses classical (and tedious) geometrical methods. "Newtonian" mechanics was developed later by Euler, Laplace, Lagrange, and others.

Newton contributed greatly to the philosophical program, begun by Galileo, of ridding science of metaphysical arguments. Modern mathematicians, with their distaste for any kind of philosophical thinking in connection with mathematics, probably are unaware of the great debt we owe to Galileo and Newton. Thanks to their struggles we are free to get on with the mathematics and let others worry about the metaphysics—and also the physics. For they worked hard to establish two principles of scientific methodology which we take for granted: first, that mathematical deductions from a mathematically formulated scientific law have the same scientific validity as the law; and second, that this is the best way to do theoretical science.

The profession of framing mathematical hypotheses and drawing mathematical conclusions is not always a comfortable one. The weapons used by Galileo's opponents were not purely intellectual ones: for publishing his *Dialogue concerning the two chief world systems* he was threatened with torture and sentenced to life imprisonment. A century earlier Bruno was burned for his heretical speculations on infinite space. Today, fortunately, the courts impose milder punishments for the dissemination of banned mathematics; yet it is sobering to think that recent advances in secret codes may render it a crime to publish the prime factorization of certain numbers.

2. Of time and mathematicians.

> I do not define time, space, and motion, as being well known to all.... Absolute, true and mathematical time, of itself and from its own nature, flows equally without relation to anything external, and by another name is called duration.
>
> Newton (1687)
>
> It may be remarked that the notion of time series has been the basis for an almost infinite amount of philosophic speculation.
>
> Birkhoff (1943)

   In an abstract sense, time need not, perhaps, be considered in mathematics. But if (as we should) we take mathematics to be what mathematicians actually do rather than what they say they do, then time becomes the most important variable in mathematics. Many eminent mathematicians and philosophers have perceived various intimate connections between time and mathematics. It is amusing and perhaps instructive to consider some of their ideas.
   No one can explain what Newton's famous dictum means, but having uttered it he was free to get on with the mathematics. Most modern mathematicians would never dream of putting anything about time in their scholarly writings, nor would they have anything of interest to say. Nevertheless the subject keeps popping up in their less formal pronouncements, often in curious ways.
   Two centuries ago when mathematicians and philosophers paid each other more attention, Kant found that

> Arithmetic achieves its concept of number by the successive addition of units in time.

This was echoed a century later by Schopenhauer:

> ...arithmetic rests on pure intuition of time...

> Time makes arithmetic possible. Space makes geometry possible.

Many later philosophers held similar views.
   The mathematician Hamilton saw not arithmetic but algebra (which hardly existed in Kant's day) as the science of time, and wrote at some length about it:

> ...the subject matter of algebraic science is the abstract notion of time...the thought of *possible* succession, or of pure ideal progression.

His biographer Hankins (1980) suggests that because he saw the reality of mental acts behind the symbols of algebra, as opposed to viewing the symbols as merely standing for unspecified numbers, Hamilton was able to free algebra from the necessity of obeying the law of ordinary arithmetic. Metaphysical speculation may play a more important role than we suspect in the development of mathematics.

Hamilton saw his quaternions as being "time plus space". This early conception of a four-dimensional space-time continuum had been anticipated by d'Alembert, as was pointed out to me by A. Borel. In the *Encyclopédie* article on dimension, d'Alembert (1754) wrote:

> J'ai dit plus haut qu'il n'étoit pas possible de concevoir plus de trois dimensions. Un homme d'esprit de ma connoissance croit qu'on pourroit cependant regarder la durée comme une quatrième dimension, & que le produit du temps par la solidité seroit en quelque maniere un produit de quatre dimensions; cette idée peut être contestée, mais elle a, ce me semble, quelque merite, quand ce ne seroit que celui de la nouveauté.

Was the "homme d'esprit" d'Alembert himself?

Hamilton characterized mathematics as the science of time and space. This view was supported by Augustus De Morgan, who wrote for a broad audience on many mathematical subjects; but it was ridiculed by C. S. Peirce (the son of the Harvard mathematician and the founder of pragmatism):

> I must say that it is rare to meet with a careful definition of a science so extremely objectionable as this.... Hamilton's intention probably was, by means of this definition, to throw a slur upon the introduction of imaginaries into geometry, as a false science; but what De Morgan...could have had in view, it is hard to comprehend, unless he wished to oppose Boole's theory of logic. Not only do mathematicians study hypotheses which...no otherwise relate to time and space than do all hypotheses whatsoever, but we now see clearly that, since the non-Euclidean geometry has become familiar to us, that there *is* a real science of space and a real science of time, and that these sciences are...branches of physics...

In an early work Peirce saw time and attention as related through the property of continuity:

> attention is a matter of continuous quantity; for continuous quantity, so far as we know it, reduces itself in the last analysis to time.
> (1868)

Later he suggested that the continuity of time might be an illusion:

> Why may there not be a succession of stationary states, say a milliasse or so of them or perhaps an infinite multitude per second,...

and in a remarkable anticipation of quantum theory he goes on to ask:

> ...and why may states of things not break down abruptly from one to the next?
> (1902)

Why indeed?—except that time would not then be a suitable variable for calculus.

L. E. J. Brouwer's profound topological theorems have been enormously useful in dynamics. A common way to produce periodic solutions to differential equations is to use Brouwer's results to find a fixed point of a related transformation. Brouwer was also an iconoclastic philosopher of science and mathematics; his intuitionism paradoxically caused him to reject his main mathematical works as unproved. For him mathematics was constructed in the human mind by an intrinsic natural faculty growing directly from our intuition of the passage of time:

> This neo-intuitionism considers the falling apart of moments of life into qualitatively different parts, to be reunited only while remaining separated by time, as the fundamental phenomenon of mathematical thinking, the intuition of the bare two-oneness.

<div align="center">*   *   *</div>

> In this way the apriority of time does not only qualify the properties of arithmetic as synthetic a priori judgments, but it does the same for those of geometry....
>
> <div align="right">(1913)</div>

For Brouwer all mathematical relations, in practical applications, become relations in time:

> Thus for instance Euclidean geometry when applied to reality gives the causal relations between the results of different measurements executed with the aid of the group of the rigid bodies.

<div align="center">*   *   *</div>

> *The only a priori element in science is time.*
>
> <div align="right">(1907)</div>

On the other hand, experience, as contrasted with intuition, is independent of both mathematics and time:

> Not only does...mathematics exist independently of all experience, but all experience is also independent of mathematics. Human experience is not passively subjected to any single mathematical system; not even to the coordinate of time, not even to the time continuum devoid of measure.
>
> <div align="right">(1907)</div>

and also:

> <div align="center">*Mathematics is independent of logic.*</div>
>
> <div align="right">(1907)</div>

Brouwer had little faith in "scientific truths":

> ...scientific thinking is nothing but a fixation of the will within the confines of the human head, a scientific truth no more than an infatuation of desire restricted to the human mind.
>
> <div align="right">(1907)</div>

or even in mathematical reasoning as a source of truth about nature:

> It is true that from certain relations among mathematical entities,
> which we assume as axioms, we deduce other relations according to
> fixed laws, in the conviction that in this way we derive truths from
> truths by logical reasoning, but this non-mathematical conviction of
> truth or legitimacy has no exactness whatever and is nothing but a
> vague sensation of delight arising from the knowledge of the
> efficacy of the projection into nature of these relations and laws of
> reasoning.
>
> (1913)

Bertrand Russell was concerned with the role of time in physics:

> Sir Isaac Newton's "absolute" time, although it remained embedded in
> the technique of classical physics, was not generally accepted....
> A frequent ground of objection to Newton's "absolute" time has been
> that it could not be observed. This objection, on the face of it, comes oddly
> from men who ask us to believe in electrons and protons and neutrons,
> quantum transitions in atoms, and what not, none of which can be
> observed. I do not think physics can dispense with inferences that go
> beyond observation. The fact that absolute time cannot be observed is not,
> by itself, fatal to the view that it should be accepted; what is fatal is the
> fact that physics can be interpreted without assuming it.
>
> (1948), p. 286

Russell goes on to point out that physics needs some interpretation of the
variable t, and of "instant". He then sketches a way of defining "instant" in
terms of overlapping events, similar to the construction of real numbers from
nested intervals (and probably suggested by the latter).

Quantum theory and relativity have between them completely overturned
our old concepts of time and space, but rather than go into their profound
influence on dynamics, let me quote from Eddington:

> There is a quantity, unrecognized in pre-relativity physics, which more
> directly represents the time known to consciousness. This is called
> proper-time or *interval*. It is definitely separated from and unlike
> proper-space. Your protest in the name of common sense against a mixing
> of time and space is a feeling which I desire to encourage. Time and space
> ought to be separated. The current representation of the enduring world as
> a three-dimensional space leaping from instant to instant through time is
> an *unsuccessful* attempt to separate them. Come back with me into the
> virginal four-dimensional world and we will carve it anew on a plan which
> keeps them entirely distinct. We can then resurrect the almost forgotten
> time of consciousness and find that it has a gratifying importance in the
> absolute scheme of nature.
>
> (1927), p. 37

While recently the mathematical community has evinced a renewed interest
in philosophical issues, most mathematicians still prefer to ignore them and
limit their public writings to technical exposition. Yet metaphysical intuitions
break into print in the form of jargon terms. Currently popular expressions like

"chaotic dynamics" and "strange attractor" reveal something about the world outlook of those who use them, as do "nonconstructive proof", "numerical simulation" and "evolution equation". Today's perfectly precise and technical words "irrational", "negative", and "imaginary" reflect yesterday's philosophical commitments.

Even titles of books can carry philosophical messages. I leave it as an exercise for the reader to elucidate the philosophical stances implicit in these recent titles: *The geometry of biological time*, by A. Winfree; *Dynamics—The geometry of behavior*, by R. Abraham and C. Shaw; *The mathematics of time: Essays on dynamical systems, economic processes, and related topics*, by S. Smale.

3. The uses of dynamics: prediction and insight.

> But, on reflection, very few phenomena depend on mathematically simply expressed laws.... Furthermore, even when system is controlled by explicit laws of evolution, it often happens that its qualitative behavior is still not computable and predictable....
>
> R. Thom (1975), p. 322.

The most spectacular use of mathematics, and especially dynamics, has been to predict accurately and successfully observations of planets and discoveries of new ones, atomic explosions, landings on the moon, and other events. But much of science concerns not prediction so much as understanding—how do galaxies form, how did species arise, how do economies develop? Dynamical systems theory plays an important part in trying to answer such questions, but its role is rather different from its more familiar one of a predictive tool. Especially as we move away from physics toward the biological and social sciences, we find that dynamics is used more as a source of qualitative insight than for making quantitative predictions. Its great value is its adaptability for constructing models of natural systems, which models can then be varied and analyzed comparatively easily.

Models in biological and social sciences are rarely "blueprints for constructing the world; they are more often similes for understanding it" (N. Georgescu-Roegen (1966), p. 116). Now a conceptual model of this kind must be much simpler than the complex reality it is modelling, or else there would be no point to the simile. So how can it be useful? The answer is that the model may shine a clear light on certain tendencies or relationships—not necessarily "laws"—that would otherwise be obscure. The value of such conclusions is often negative: if we think that the conclusions drawn from the model does not in fact hold for reality we are then led to look further for the reason. On the other hand we may be convinced by the plausibility or generality of the model that we have indeed learned something new about reality.

For such conclusions to be convincing the mathematical deductions from the model must be *robust*—insensitive to mathematical detail. This vague but

important concept refers to the persistence of some specified feature of the mathematical system when under small perturbations of the system. (The interpretation of "small perturbation" is often obscure, subtle, and controversial.) For example consider systems composed of a circle and a line in the plane. The property of there being two point of intersection is a robust one, while that of there being exactly one point of intersection is not robust.

A very strong form of robustness is *structural stability* of a system of differential equations. Roughly speaking this means that the entire topological and dynamical structure of the space of solutions (phase portrait) is unchanged by sufficiently small perturbations. This concept, emphasized by R. Thom in his *tour de force* of dynamical model-building, *Structural stability and morphogenesis* (1972, 1975), seems to occur too rarely to be useful as a general tool. Nevertheless the underlying idea, which is that *mathematical models should be robust in their important features*, is widely accepted. This "stability dogma" is discussed in the first edition of Abraham's and Marsden's *Foundations of mechanics* (1967) and in the recent book by Guckenheimer and Holmes (1983).

The end result of a successful mathematical model may be an accurate method of prediction. Or it may be something quite different but not necessarily less valuable: *a new insight*..

4. An early example of dynamical reasoning in economics: Malthus' insight.

> In one staggering intellectual blow Malthus undid all the roseate hopes of an age oriented toward self-satisfaction and a comfortable vista of progress.
>
>   Robert Heilbroner (1953)

> I found Malthus very sexy.
>
>   Paul Samuelson (quoted by L. Silk (1983))

In 1798 Malthus published *An essay on the principle of population as it affects the future improvement of society*. His gloomy message was that "the principle of population" is "conclusive against the perfectability of the mass of mankind" because

> population when unchecked, increases in a geometric ratio. Subsistence increases only in an arithmetical ratio. A slight acquaintance with numbers will show the immensity of the first power in comparison with the second.

The tendency toward geometric (exponential, in current jargon) growth of population is familiar to every calculus student, but the arithmetic (or linear) growth of subsistence may be less palusible today. We think of increasing agricultural output through the use of chemical fertilizers, heavy machinery, and genetic engineering. But in Malthus' day these methods did not exist; the only way to increase harvest significantly was to bring more land under cultivation—the limitations of which are obvious. Malthus argued that only

war, famine, and pestilence (and as Darwin later emphasized, infanticide) prevented the human population from outrunning its means of subsistence. He concluded that any effort to relieve the misery of the poor (who included practically everyone) was bound to fail, because any temporary improvement would only lead to an increase in population, which in turn would mean less food and worsened living conditions in general.

Malthus' thesis was enormously influential from the start, and has remained so up to the present. R. Heilbroner says:

> No wonder that after he read Malthus, Carlyle called economics "the dismal science", and that poor Godwin complained that Malthus had converted friends of progress into reactionaries by the hundreds. In one staggering intellectual blow Malthus undid all the roseate hopes of an age oriented toward self-satisfaction and a comfortable vista of progress.

From Malthus' day to now, economists have taken Malthus' argument very seriously. His friend Ricardo, the classical economist, debated him endlessly. A biographer wrote

> He was the best abused man of his age. . . . From the first, Malthus was not ignored. For thirty years it rained refutations.

Marx despised his reasoning while admitting that it had some validity under capitalism. Keynes justified the immense accumulation of capital by the wealthy classes as an antidote to Malthusian immersation:

> If only the cake were not cut but was allowed to grow in the geometrical proportion predicted by Malthus of population, but not less true of compound interest, perhaps a day might come when there would at least be enough to go around. . . . One geometrical ratio might cancel another, and the nineteenth centry was able to forget the fertility of species in a contemplation of the dizzy virtues of compound interest.
>
> *The economic consequences of the peace*
> (quoted by J. Strachey, *Contemporary capitalism*, p. 83)

If we judge the success of a theory by the amount of discussion it provokes, then Malthus' success was huge, and on its own merit his argument has an undeniable impact. Yet this is not due to any predictive power, nor to any kind of verification—in fact there is no conceivable way of testing it by experiment of observation. (Malthus' argument is in the subjunctive mood, his conclusion in the imperative; such statements are neither verifiable nor falsifiable.) Whence, then, his success?

I think the answer lies in the mathematical form of his reasoning, and in the robustness of the conclusion. It is not nearly so striking merely to argue, "population tends to increase faster than subsistence"; the punch comes from the contrast between exponential and linear growth. And we accept this

contrast as relevant to the real world, despite its absurd simplicity, precisely because it is so robust.

If Malthus were writing a century later, he might have used continuous instead of discrete dynamics. That is, he might have cast his arguments in terms of differential equations: "Because of previously explained sociological factors (S) that in the absence of war, etc., we are led to postulate the population $X(t)$ at time t which satisfies a differential equation of the type $dx/dt = cx$, with c a positive constant; and that the food supply $y(t)$ satisfies $dy/dt = k$, another positive constant. It follows that the *per capita* food supply $y/x = (y_0 + kt)/x_0 e^{ct}$ must go to zero as time proceeds. Therefore, misery is inevitable."

To the criticism that these differential equations are absurdly simplistic and unrealistic, Malthus could reply "They are really meant only as a simple illustration. Actually (S) implies the following inequalities: $dx/dt > cx$ and $dy/dt < k$, for some unknown positive constants c and k. From this it still follows that y/x goes to zero."

This argument is not easy to attack. The only reasonable strategy is to analyze the economic and sociological arguments used to derive the inequalities governing the growth rates. (Such analyses have been made; see e.g. J. Robinson (1956).)

The role of mathematics in Malthus' theory is very different from its role in mathematical physics, say Newton's work. In the latter the mathematical outcome *is* the physical conclusion, for example that planetary orbits are ellipses. Mathematical details are crucial to the success of the argument, which is a statement of how things actually are. The theory makes accurate predictions. With Malthus the mathematics only exhibits a tendency; and the details are not important. The theory does not purport to describe how things are, but how they would be under certain conditions which seem impossible to realize (no war, famine or pestilence). The purpose of Malthus' theory is not prediction but insight.

5. Dynamical insight in nonquantitative science: Malthus and evolution.

> See the ever memorable "Essay on the Principle of Population" by the Rev. T. Malthus...
>
> C. Darwin, *The descent of man*, p. 428n
>
> Evolution is the history of a system undergoing irreversible change.
>
> A. Lotka (1956), p. 24

Strong as was his influence on economics, Malthus' effect on biology has been more far reaching: his *Essay* was credited by both Darwin and Wallace as the immediate inspiration for their independent discoveries of natural selection as the driving force being evolution. In his *Autobiography* (1892) Darwin tells us:

> In October, 1838, that is, fifteen months after I had begun my systematic enquiry, I happened to read for amusement Malthus on *Population*, and being well prepared to appreciate the struggle for existence which everywhere goes on from long-continued observation of the habits of animals and plants, it at once struck me that under these circumstances favourable variations would tend to be preserved and unfavourable ones destroyed. The result of this would be the formation of new species. Here, then, I had at last got a theory by which to work....
>
> p. 40

Wallace wrote in his famous letter to Darwin that, remembering Malthus' book,

> it suddenly flashed on me...in every generation the inferior would inevitably be killed off and the superior would remain—that is, *the fittest would survive.*

What the evolutionists got from Malthus was an insight—and the use they made of it was very different from Malthus':

> though he accepted Malthus' premises, Darwin wanted to stand Malthus' conclusion on its head: Whereas Darwin saw struggle as leading to change, Malthus essentially saw struggle as ruling out change!
>
> M. Ruse, *The Darwinian revolution,* p. 175

Ruse goes on to say:

> Darwin found Malthus' *Essay* new and striking not because of the truth of his premises—nearly everyone in the 1830s knew these and accepted them as indubitable—but because Malthus presented his ideas in a lawlike, quantitative form, with a deductive approach. This was just what Darwin, soaked in the contemporary [Newtonian] philosophy of science, was looking for.... As soon as Darwin had read Malthus, he started to think in terms of forces and pressures that were pushing organisms into the available and not-so-available gaps in the economy of nature...insofar as Malthus was able to show him something akin to a force at work on organisms, Darwin was becoming aware of something that would be a prime candidate for a scientific evolutionary mechanism. Because of the philosophical lens through which he viewed matters, he was highly receptive to the way Malthus presented the struggle for existence.

Newton's dynamical theory was so successful that by Darwin's day it had become the standard model for all scientific theories: it wasn't enough for something (e.g. evolution) to happen, there had to be a force-like dynamic driving the action. Darwin and Russell found their dynamic in Malthus.

Just as "Newtonian" mechanics was the explication of the dynamic of the cosmos, so the theory of natural selection is the unfolding of the basic dynamic of the living world. But while evolution is clearly a dynamical process, it is not easy to say what the *system* is. Are the parts of the system the individual

organisms, or species, or genes, or chromosomes? Such questions merely reflect the chasm between the simplicities of mathematics and the inherent complexities of biology. (Perhaps they merely reflect the inadequacy of my definition of dynamical system.) But there seems to be something about evolution that is stubbornly nonquantifiable—or should we say nonmathematizable?

It is of course true that various subjects related to evolution can be profitably investigated mathematically—genetics for example (to which dynamical systems has contributed). It is interesting that Darwin expressed regret at not having learned more mathematics (1892), p. 18.

The resistance of evolution to mathematical treatment may be traced to three basic aspects of evolution. The first pertains to all the biological and social sciences: the active and essential role played by the observer. It is, after all, the human observer who intuitively and prescientifically classifies things into living and nonliving, and who recognizes—but can scarcely define—instinct, learning, intelligence, mobility, adaptation, social groupings, etc. Only in the deepest and most recent parts of physical science is it necessary to bring in the observer in an essential theoretical way; but in the biological and social disciplines the observer's part is crucial from the beginning—so basic that we may overlook it. And it is hard to mathematically model a human being.

The second aspect of evolution that tends to make it nonmathematical is the *uniqueness* of evolution on earth—unrepeatable and only very partially observable,

And finally, evolution is about many small but significant concrete differences in the details of living things—and only the human observer can assess their significance. Whereas physics has progressed by abstracting basic properties of matter from irrelevant differences between actual objects, evolutionary theory proceeds just the opposite way—by paying the closest attention to these differences. It is not possible to say in advance of observation just what the differences are, or which will prove important; thus it is hard to treat them abstractly.

The biologist J. Z. Young makes a similar point in psychological terms:

> The study of variation presents certain special difficulties because it involves the opposite of our basic tendency to compare things, to find likenesses, and to give names. The point of examining all the variety of individuals in a population, say of men, is to show the respects in which they are *not* alike. This is an activity that is difficult and distasteful for the kind of brain that looks for the easy and obvious solutions achieved by classification. For many people the only satisfactory way of speaking about the world is in terms of a series of sharply defined categories, the properties of which are exactly known.
>
> *Doubt and certainty in science* (1951), p. 147

His last sentence may well refer to mathematicians.

It is ironic that Malthus' crude, elementary, but robust mathematical argument inspired the crucial insight into the least quantitative of the great scientific theories, evolution.

6. Ordinary differential equations and dynamical systems.

> Among all mathematical disciplines the theory of differential
> equations is the most important.... It furnishes the explana-
> tion of all those elementary manifestations of nature which
> involve time.
>
> Sophus Lie (1895)

> At a time when no physical theory can properly be termed
> fundamental...it may be asserted with confidence that ordinary
> differential equations in the real domain, and particularly
> equations of dynamical origin, will continue to hold a position
> of the highest importance.
>
> G. D. Birkhoff (1927)

Felix Klein, never one to shrink from magisterial pronouncements, held even more sweeping views on the significance of differential equations:

> It is well known that the central problem of the whole of modern
> mathematics is the study of transcendental functions defined by
> differential equations.
>
> F. Klein (1911)

One need not accept Klein's definition of the central problem—or even believe that there is such a problem—in order to recognize the importance of differential equations in mathematics. Besides being the most important tool for applications, the study of differential equations has spawned whole fields of abstract mathematics: Poincaré, who contributed so much to differential equations, invented algebraic topology in order to solve problems arising in differential equations; Lie invented the groups named for him for the same purpose; Fourier invented his series to study the heat equation, and then Cantor was led to topology and set theory by convergence problems in Fourier series; much of functional analysis grew out of partial differential equations. And of course mathematical dynamical systems theory was originally a branch of ordinary differential equations.

It is interesting that each of these different descendants of differential equations is not only an important field of study by itself, it is also a main tool for studying differential equations.

In his illuminating reviews of Poincaré's work, Hadamard (1912, 1912a) speaks of the earlier "golden age" of differential equations; the period during which the equations of interest could be solved by finding "integrals"—func-tions which are constant on solution curves. Since every solution curve lies in a level surface of such a function, the existence of an integral means that the dimension of the state space can be reduced by one: we simply replace the original state space by the level surface containing the initial value of the solution we are interested in computing. Given enough independent integrals, in this way we can reduce the dimension of the state space to one. Now we have (in new coordinates) one differential equation in one variable, for which

the solution can be computed (approximately) by standard methods. But "the list of these simple cases was quickly exhausted. In general the number of known integrals is insufficient". In particular, says Hadamard, "the n-body problem offers all the difficulties of the general problem of differential equations. These difficulties reside in the heart of things...the very conclusions obtained by Poincaré...explain to us why these general problems require methods not only distinct, but profoundly different from those which had at first sufficed". (1912, my translation)

These new methods are those that now make up the subject of dynamical systems theory. First of all, since Poincaré was interested only in real solutions, he had to abandon the use of complex variables which had proved so successful in analysis. And as mentioned above, he could not rely on finding integrals; on the contrary, he found strong evidence that in the three-body problem there simply aren't enough.

Instead Poincaré introduced what he called the *qualitative* study of solution curves, as opposed to the more common *quantitative* study. He draws an analogy with the study of an algebraic equation: *First* one discovers the number of real roots, using Sturm's theorem; that is the qualitative part. *Then* one can begin to calculate them numerically—the quantitative part. He draws a similar analogy to the study of an algebraic curve: first one "constructs" the curve, that is, one discovers which branches are closed, which are infinite, etc.; *after* this qualitative step one determines exactly a certain number of points on the curve. It is the same with differential equations, says Poincaré: the first step is: "construire les courbes définies par des équations différentielles" (1881).

Hadamard (1912) says that with a few isolated exceptions this point of view was almost totally absent among Poincaré's predecessors. The reason, he suggests, is that the great successes of complex function theory had completely turned analysts away from the real domain, while at the same time "Science found itself completely disarmed in the face of the great difficulty of the questions thus posed, the first for which this theory of analytic functions brought not a single solution".

What methods did Poincaré use? Hadamard suggests an illuminating historical parallel: In the older study of algebraic equations, attention was focussed on the discovery of a single root of the given equation. But the theory was able to pass from this somewhat empirical state to "logical perfection" when Galois and others considered all the roots simultaneously. By examining the relations which exist between the roots, "all became clear". (The analogy goes much deeper: just as Galois theory investigates permutations of the roots of an algebraic equation, Lie theory looks at symmetries of solutions to a differential equation.)

Similarly with differential equations: earlier analysts, with few exceptions, had studied individual solutions in isolation. It was Poincaré who systematically studied the mutual relations between all the solutions. In particular, Hadamard says, he made the crucial innovation of considering the value of unknown not as a function of the independent variable (time, in dynamics),

but *as a function of the initial conditions*. The modern mathematical notion of dynamical system is an abstract formalization of this point of view.

A provisional definition of *dynamical system* (which from now on we take in the mathematical sense) is a pair (X, ϕ) comprising a topological space X called the *state space* and a collection $\phi = \{\phi_t\}$ of maps, called the dynamic or *flow*. Here the index t runs over the set $\mathbf{R}^+$ of positive real numbers, or sometimes the set $\mathbf{R}$ of all reals; and each ϕ_t maps an open subset of X into X. These data are subjected to a few axioms to be given later. The motivating question is: what happens to $\phi_t(x)$ as $t \to +\infty$?

From a vector field F on Euclidean n-space $\mathbf{R}^n$ we obtain a dynamical system with state space $X = \mathbf{R}^n$ as follows. Assume F is C^1 (continuously differentiable) so that every initial value problem

$$dx/dt = F(x), \qquad x(0) = v,$$

has a unique solution $x(t, v)$ for t in some maximal interval I_v of the form (a, b) with $a < 0 < b$. The map ϕ_t has for its domain the (perhaps empty) set of v for which $t \in I_v$, and is defined by

$$\phi_t(v) = x(t, v).$$

More generally F could be a vector field on a smooth manifold M (without boundary). By expressing F in a local coordinates and solving the resulting differential equation in $\mathbf{R}^n$, we obtain a dynamical system whose state space is M.

A good part of classical mechanics consists in deriving, from Newton's laws and later developments such as conservation of energy, differential equations expressing the "laws of motion" of various physical systems having a finite number of degrees of freedom. (Infinitely many degrees of freedom lead to partial differential equations.) These equations usually have extra structure (symmetries, Hamiltonian form, conservtion laws) which aid in studying their solutions, but which also make the study of their perturbations especially delicate.

Faced with a differential equation we can take several points of view. Up to the time of Poincaré what was usually sought was a formula or series expansion, convergent if possible, for a particular solution. In his address to the International Congress of 1908 Poincaré said:

> In the past an equation was only considered to be solved when one had expressed the solution with the aid of a finite number of known functions; but this is hardly possible one time in a hundred. What we can always do, or rather what we should always try to do, is to solve the *qualitative* problem so to speak, that is to try to find the general form of the curve representing the unknown function.
>
> It then remains to find the *quantitative* solution to the problem; but if the unknown cannot be determined by a finite calculation, one can always represent it by a convergent infinite series which permits us to calculate it. Can this be regarded as a true solution? It is said that Newton communicated to Leibniz an anagram somewhat like this: aaaaabbbeeeeii, etc.

Leibniz naturally understood nothing at all of this; but we who have the key know that this anagram meant, translated into modern language: I know how to integrate all differential equations; and this suggests to us that either Newton was very lucky or else he held peculiar illusions. He meant quite simply that he could form a power series (by the method of undetermined coefficients) which formally satisfies the proposed equation.

Such a solution today would not satisfy us, for two reasons: because the convergence is too slow, and because the terms succeed each other without obeying any law at all....

But then there are no longer some problems which are solved and others which are not; there are only problems which are *more or less* solved, according as the series converges more or less rapidly, or has terms governed by a law which is more or less harmonious.... Sometimes the series converges so slowly that the calculation is impractical and we have only succeeded in demonstrating the possibility of the problem.

And then the engineer finds that ludicrous, and for good reason.... He is not very interested to know that it will be useful to the engineers of the 22nd century; we, on the other hand, think otherwise and we are sometimes happier to have saved on day of work for our small children than one hour for our contemporaries.

Poincaré (1908), my translation

In his classic memoirs (1881, 1885) and his treatise *Les méthodes nouvelles de la mécanique céleste*, (1899) Poincaré initiated the qualitative study of differential equations. Besides finding many basic results and methods, he established what are still the main themes of dynamical systems theory: stability, periodicity, and recurrence. The motivating question for *Les méthodes nouvelles*, as for all celestial mechanics, is the stability of the solar system—a qualitative question *par excellence*. If there is only one planet and the sun, then Newton's equations of motion are easy to solve explicitly; the orbits are the Keplerian ellipses, and all questions are easily answered. When there are two or more planets the situation is quite obscure. Very little is known about such questions as these: will two of the planets ever collide? Do the mutual distances stay bounded away from zero and infinity? Is it probable that a planet can escape to infinity, or be captured from infinity? If we slightly alter the mass of a planet, how does that change the trajectories?

Such questions have simple and precise interpretations mathematically as questions about the differential equations. In trying to answer them, Poincaré frequently found himself in a situation where a certain result could be proved, except for certain kinds a trajectories which, he felt, represented totally improbable configurations, analogous to a randomly chosen real number being an integer.

Such questions can be interpreted as qualitative questions about the differential equations used to model the solar system. Of course the answers, provided one can find them, will depend on the details of the model; Poincaré was careful to point out that the equations of celestial mechanics neglect phenomena such as tidal friction, whose cumulative effect must, in the long

run, predominate (Poincaré (1896)). Regardless of the eventual physical significance of the results, however, the purely mathematical questions are of great interest in themselves.

Poincaré concentrated his efforts on finding periodic trajectories in the restricted planar three-body problem and determining their stability—do all orbits near a given periodic one stay near for all future time?

In developing stunning new methods for attacking this problem, Poincaré laid the foundations of what has become dynamical systems theory. This theory comprises not only methods and theorems, but also characteristic ways of looking at things. One of the dominant viewpoints today is that what is important is *generic* behaviour. This idea, which comes directly from Poincaré's work, has two aspects: on the one hand, it is not *sufficient* (i.e., not sufficiently interesting) to exhibit phenomena, e.g. a periodic solution, which is of such a special nature that it seems totally unlikely to be observed; and on the other hand, it is not *necessary* to prove things about all conceivable objects under consideration (e.g. trajectories of a given system, or systems in a given class); one can exclude a set of objects which appears to be extremely small in some sense (e.g. of measure zero).

Hadamard, referring to a certain stability theorem of Poincaré's about "Poisson stability" of trajectories, wrote:

> ...his result has a significance totally different from those which had previously been obtained. It governs not all trajectories without exception, but only *excepting certain exceptional trajectories*.
>
> The words "exceptional trajectories" are here to be interpreted in terms of the calculus of probabilities: they mean that when a trajectory is chosen at random, the probability that it is one of the counterexamples to the theorem is *infinitely small* (and not merely very small).
>
> In other words, it is not absolutely certain that an arbitrary trajectory possesses Poisson stability, but the odds are infinitely small that it doesn't.
>
> Hadamard (1912), p. 270, my translation

The question of what sets of mathematically interesting objects can be considered "exceptional" in this sense is an interesting one. Poincaré's criterion of measure zero (in modern terms) is an obvious choice, but there are other natural possibilities. If the objects lie in some topological space it may be reasonable to consider a nowhere dense subset as negligible; thus the integers would be a negligible subset of the real numbers. But so would a Cantor set—and a Cantor set can have positive measure. Thus various definitions of "exceptional" are in conflict. A simple dynamical setting where the conflict appears is discussed in Chapter II, §3.

Poincaré and later Birkhoff ran into situations where their intuition convinced them that the counterexamples to what they were trying to prove were indeed exceptional, but they were unable to formulate this precisely and had to be content with merely stating their opinion. Sometimes later work vindicated their guesses—as when Birkhoff felt it was unlikely that the stable and

unstable manifolds of a periodic orbit could coincide (Birkhoff (1932), p. 379; see Chapter II, §4 below). It is interesting that Poincaré simply ignored this point—he just assumed they did not coincide and made further deductions, which he heavily relied on.

On the other hand, intuition could occasionally deceive even a Birkhoff. He conjectured that the ergodic hypothesis held for "most" Hamiltonian systems with two degrees of freedom (which lead to flows on three-dimensional energy surfaces); that is, except for a set of measure zero, all points have dense orbits. "Conjectured" is too weak a word—he was certain enough to base an entire section of his important book *Dynamical systems* (1927) on this hypothesis. But the opposite was shown to be the case by the famous Kolmogorov-Arnold-Moser "twist theorem" (see e.g. Moser (1973)): generically (in a certain topological sense) the state space is separated by infinitely many invariant closed hypersurfaces which trajectories cannot cross. (See the notes by J. Moser in the 1966 reprinting of Birkhoff's book.)

In recent years mathematical dynamical systems have become increasingly common as models in biology, chemistry, economics, and other fields. These systems are often quite different from the more familiar ones of physics; they seldom have Hamiltonian form, for example, but they have interesting features of their own and frequently lead to very challenging problems. They rarely can be used to make predictions as accurate as those in celestial mechanics, but are often used in quite a different way. Rather than being an expression of the theory as they usually are in physics, in other sciences differential equations are more often derived in an ad hoc way and used to illustrate or test a theory by deriving consequences from it. Many times the main conclusion is qualitative rather than quantitative, but none the less valuable for that.

So successful have been differential equations for calculating trajectories and making long-term predictions in natural systems that other methods are now used only rarely, and the mathematical study of dynamical systems is virtually a subfield (or at least an outgrowth) of differential equations theory.

It is a curious fact that while the subject of differential equations is often taught in the driest "cookbook" way and loathed by students as the dullest drudgery, at the same time those who have penetrated to its inner core are frequently moved to rhapsodies of praise. Referring to Helmholtz, C. C. Gillispie (1960) wrote:

> ...he couched his discussion in the most sophisticated language known to physics, not all weighted down by lumps of data like Joule's heavy-handed laboratory reports, nor confined to the primitive numerical equivalents of Mayer, but in the graceful, taut, and lissome differential equations of classical dynamics,

Einstein (1930) pointed out that Kepler's laws

> are concerned with the movement as a whole, and not with the question of *how the state of motion of a system gives rise to that which immediately follows it in time*. They are, as we should say now, integral and not differential laws.

The differential law is the only form which completely satisfies the modern physicist's demand for causality. The clear conception of the differential law is one of Newton's greatest achievements.

Rene Thom (1975), p. 4, went so far as to say:

The possibility of using the differential model is, to my mind, the final justification for the use of quantitative models in the sciences.

So vast is the subject, however, and so formidable are its problems that a pessimistic tone sometimes appears in the writings of its most experienced and knowledgeable practitioners. In his address to the International Congress in 1958, G. Temple said:

The group of problems which I propose to describe belong to that Cinderella of pure mathematics—the study of differential equations. The closely guarded secret of this subject is that it has not yet attained the status and dignity of a science, but still enjoys the freedom and freshness of such a pre-scientific study as natural history compared with botany. The student of differential equations —significantly he has no name or title to rank with the geometer or analyst—is still living at the stage where his main tasks are to collect specimens, to describe them with loving care, and to cultivate them for study under laboratory conditions. The work of classification and systematization has hardly begun.

Now only *nonlinear* differential equations have interesting dynamics. Concerning them, Professor Temple went on to say:

An inviting flora of rare equations and exotic problems lies before a botanical excursion into the nonlinear field.

It is true that we have made considerable progress in the past three decades, but while the subject may now have more "status and dignity", we are farther away than ever from a unified approach. (I see this as a reason for enthusiasm, however, rather than pessimism!) In a recent survey of nonlinear control systems, J. Casti wrote:

All current indications point toward the conclusion that seeking a completely general theory of nonlinear systems is somewhat akin to the search for the Holy Grail: a relatively harmless activity full of many pleasant surprises and mild disappointments, but ultimately unrewarding. A far more profitable path to follow is to concentrate upon special classes of nonlinear problems, usually motivated by applications, and to use the structure inherent in these classes as a guide to useful (i.e., applicable) results.

And indeed all the recent advances in dynamical systems theory have been made by exploiting the structure inherent in special classes of systems. In Chapter III we shall look at certain systems with a high degree of structure, which can be usefully exploited.

7. Partial differential equations and dynamical systems.

> Before Maxwell people conceived of physical reality—insofar
> as it is supposed to represent events in nature—as material
> points, whose changes consist exclusively of motions, which are
> subject to total differential equations. After Maxwell they
> conceived of physical reality as represented by continuous
> fields, not mechanically explicable, which are subject to partial
> differential equations. This change in the conception of reality
> is the most profound and fruitful one that has come to physics
> since Newton; but it has at the same time to be admitted that
> the program has by no means been completely carried out yet.
>
> Albert Einstein (1931)
>
> ...that subtlest of all instruments for putting Nature and
> Reason to the question—a partial differential equation.
>
> J. Sylvester (1877)

Even before Maxwell—at least as far back as Euler—partial differential equations were used to describe the movements of continuous media like fluids or vibrations strings. Such motions constitute the changing state of physical systems which differ in one crucial way from systems of particles: to describe the state we need not merely a finite set of variables, but a function. The state of a bar of iron as regards temperature is described by the function which to each point of the bar assigns its temperature on some convenient scale. As time passes this function changes: physical theory leads to a mathematical law governing this change. With suitable boundary conditions, and in an appropriate function space, one obtains a dynamical system whose state space is the function space. The dynamic is expressed by a differential equation of the form $\partial u/\partial t = F[u]$, where F is some nonlinear operator on the function space. In this equation, $u = u(x, t)$ denotes the temperature at point x at time t, while $F[u]$ means that at time t the operator F is applied to the function $u(\cdot, t)$, t being held fixed. Often $\partial u/\partial t$ depends only on the value of u and its x-derivatives, so that F is a differential operator.

As a typical example consider a biological model of a single numerous species (e.g. bacteria) occupying a long narrow tube. We model the tube by a closed interval $[a, b]$. At time t the population density at point x of the interval is $u(x, t)$. Suppose at first that the bacteria neither die nor procreate, but simply wander at random over the interval without crossing the endpoints. The population at time t in an interior subinterval $[x_0, x_0 + h]$ is $\int_{x_0}^{x_0+h} u(y, t)\, dy$. The only way this quantity can change is through migration across endpoints of the subinterval. In a small time interval only individuals very near an endpoint can cross it. If in the vicinity of x_0 there are more individuals to the right of x_0 than to the left, then there will be a net migration across x_0 out of the subinterval, and it is plausible to assume that the rate of migration out of the subinterval across x is proportional to the x-gradient of u, that is, to $-u_x(x_0, t)$. Similarly the rate of migration *into* the subinterval across the other endpoint $x + h$ is assumed proportional to $u_x(x_0 + h, t)$. Thus we find that the time-rate of change of the total population in $[x_0, x_0 + h]$ is

$$\frac{d}{dt}\int_{x_0}^{x_0+h} u(y,t)\,dy = c\big(u_x(x_0+h,t) - u_x(x_0,t)\big),$$

where the proportionality factor c may depend on x. Dividing by h and taking limits as h goes to 0, we find

$$u_t = c(x)u_{xx}$$

with Neumann (or "no-flux") boundary conditions

$$u_x(a) = u_x(b) = 0.$$

Now suppose that at each point x there is an intrinsic growth rate g depending only on x and the current population density at x. Thus to the population changes caused by migration we must add the change owed to the growth. We obtain the equation

$$u_t = cu_{xx} + g(x,u)$$

with the same no-flux boundary conditions.

If the bacteria live not in a tube but rather in a region Ω of the space, similar reasoning leads to the following "reaction-diffusion" equations:

(1) $$u_t = c(x)\Delta u + g(x,u), \qquad \partial u/\partial n = 0.$$

Here Δ is the Laplace operator $\sum_{i=1}^{3}(\partial/\partial x_i)^2$, while $\partial/\partial n$ denotes the directional derivative at boundary points of Ω in the outward normal direction, and $x = (x_1, x_2, x_3)$ is a vector.

There is an extensive literature on equations such as (1). There is no unique "natural" function space to use, but there are several different ones, convenient for various purposes, in which it can be proved that (1) has solutions satisfying sufficiently strong theorems of existence, uniqueness, and continuity to yield a dynamical system. If X is such a function space, then the dynamic $\{\phi_t\}$ is obtained as follows. Let $u_0 \in X$ be given. Let $u(x,t)$ be the solution to equations (1) such that $u(x,0) = u_0(x)$. Then $\phi_t(u_0)$ is the function which sends x to $u(x,t)$:

$$\phi_t(u_0)(x) = u(x,t).$$

In other words, for each $t > 0$, ϕ_t sends the function $u(\cdot,0)$ to $u(\cdot,t)$.

One could model several interacting species by a system of reaction-diffision equations. One could assume that the growth rate g depends also on the spatial gradient of u. Further, interesting complications will occur to the reader.

In Chapter III some dynamical theorems applying to equation (1) will be presented, but in general very little is known about the dynamics of partial differential equations. There is one method, however, which in many applications yields useful information. It frequently happens that a dynamical system in a function space X contains a smooth *finite-dimensional* subsystem, that is, a finite-dimensional manifold $M \subset X$ which is invariant under the flow. Moreover, in many cases M is an attractor (see §II.1) and is known to contain

periodic orbits or other dynamical features of interest. In such a case the methods and results from the much more accessible dynamics of ordinary differential equations can be applied to the subsystem. For applications of this important technique see, for example, D. Ruelle and F. Takens (1971), J. Marsden and M. McCracken (1976), and J. Guckenheimer and P. H. Holmes (1983).

CHAPTER II: CONVERGENCE, CHAOS, AND STABILITY

> The recent development of the qualitative theory of differential equations, originated by Poincaré, led to the realization that similar to the fact that the explicit integration of differential equations is generally impossible, the qualitative study of general differential equations with a multidimensional phase space turns out to be impossible.
>
> V. Arnold (1983)

> More than 10 years ago I posed the problem of finding a dense open set U (or at least a Baire set), of Dyn(M) such that the elements of U could somehow be described qualitatively by discrete numerical and algebraic invariants.... The problem posed in this way is too simple, too rough, and too centralized. I believe now that the main problems of dynamical systems can't be unified so elegantly.
>
> S. Smale (1969)

The development of mathematical dynamical systems theory can be viewed as the simultaneous pursuit of two lines of research: on the one hand, the quest for *simplicity*, comprehensibility, stability; on the other hand, the discovery of *complexity*, instability, chaos. When new complexities are discovered we try to tame them through analysis and classification. This enterprise has had some notable, even spectacular successes; but as other sciences advance they make increasing use of dynamical systems, and we are currently flooded with interesting systems from many fields, about which little has been proved. R. Abraham's statement of what he calls the yin-yang problem is still valid:

> Some large (yin) sets of differential equations with generic proper-
> ties are known, some small (yang) sets which can be classified are
> known, but in general the two domains have not yet met.
>
> (1969)

The general goal of dynamical systems theory is to find useful ways of answering the question, what happens in the long run? In this chapter we first set out the precise terminology for discussing this question—the language of topological dynamics (flows, equilibria, periodic orbits, limit sets). Next a look at some highlights of the very successful theory of flows on surfaces (the one place where yin = yang). Finally, a brief indication of a major source of chaotic behavior —homoclinic points. (Only the first section will be needed for Chapter III.)

1. Flows.

In the classical sense, a dynamical system is a system of
ordinary differential equations with at least sufficient condi-
tions imposed to insure continuity and uniqueness of the
solutions. As such, a dynamical system defines a (one-parame-
ter or continuous) flow in a space. A large body of results for
flows which are of interest for classical dynamics has been
developed, since the time of Poincaré, without reference to the
fact that flows arise from differential equations.

W. Gottschalk and G. Hedlund (1955)

A *dynamical system* consists of a pair (X, ϕ) where X is a topological space,
and $\phi = \{\phi_t\}_{t \in \mathbf{R}_+}$ is a collection of maps indexed by the set $\mathbf{R}_+$ of nonnegative
reals. The *state space* X is assumed to be Hausdorff; usually X is a manifold or
an open subset of some topological vector space. The *dynamic* or *flow* ϕ is
made up of continuous maps $\phi_t \colon X_t \to X$ where each domain X_t is a (possibly
empty) open subset of X. These maps satisfy the following two conditions:

(1) *Continuity.* The set

$$X^{\sharp} = \{(t, x) \in \mathbf{R}_+ \times X : x \in X_t\}$$

is open in $\mathbf{R}_+ \times X$, and the map

$$\phi \colon X^{\sharp} \to X, \qquad (t, x) \to \phi_t x,$$

is continuous.

(2) *Determinism.* For each $x \in X$ the set

$$I(x) = \{t \in \mathbf{R}_+ : x \in X_t\}$$

is a half-open interval of the form $I(x) = [\,0, \tau_x)\,, 0 < \tau_x \leqslant \infty$. If $t \in I(x)$ and
$s \in I(\phi_t x)$, then $s + t \in I(x)$ and

$$\phi_s \phi_t x = \phi_{s+t} x,$$

while

$$\phi_0 x = x \quad \text{for all } x \in X.$$

It often happens that the *escape-time* τ_x is ∞ for all x, although to prove this
can be a major challenge. Such a system is called *complete*. The dynamic ϕ (or
the map $\phi_{\sharp}$) is called a *flow* (more precisely, a local semiflow).

Where ϕ is understood we may write $\phi_t x = x(t)$. The map $I(x) \to X$,
$t \to x(t)$, is the *trajectory* of $x \in X$. Its image is the *orbit* $O(x)$, and the closure
of $O(x)$ is the *orbit closure* $\overline{O}(x)$.

It is easy to prove that if $\overline{O}(x)$ is compact, then the trajectory of x is
complete, i.e. $\tau_x = \infty$.

The long-term behavior of the trajectory of x is conveniently studied in
terms of the *limit set* (or *omega limit* set)

$$\omega(x) = \bigcap_{t \in I(x)} \overline{O}(x(t)).$$

Thus

$$y \in \omega(x) \Leftrightarrow y = \lim_{k \to \infty} x(t_k)$$

for some sequence $\{t_k\}_{k \in \mathbf{Z}_+}$ converging to τ_x in $I(x)$.

Also of interest, although primarily for technical reasons, is the *alpha limit set* $\alpha(x)$ comprising those points y which are limits of a sequence $\{y_k\}$ in X such that there exists a sequence $t_k \to \infty$ in $\mathbf{R}_+$ with $t_k \in I(y_k)$ and $y_k(t_k) = x$.

Certain flows (usually coming from ordinary differential equations) are *reversible*: each ϕ_t maps X_t homomorphically onto an open set. In this case we set $X_{-t} = \phi_t(X_t)$ for $t > 0$ and define

$$\phi_{-t} = (\phi_t)^{-1} : X_{-t} \to X.$$

In this way we extend the dynamic to a set of maps $\{\phi_t\}$ indexed by $t \in \mathbf{R}$ satisfying axioms analogous to Continuity and Determinism. In particular, for each x the set of $t \in \mathbf{R}$ such that $x \in X_t$, is an open interval about 0.

If ϕ is reversible, a new flow ψ is obtained by *time-reversal*: we define $\psi_t = \phi_{-t}$. These flows have the same equilibria and periodic orbits, but omega limits for one are alpha limits for the other.

Let us say that a point x (or its trajectory) is *attracted* to a set K, or *approaches* K, if $\overline{O}(x)$ is compact and $\omega(x) \subset K$. This implies that the trajectory of x eventually remains in any neighborhood of K. We also say that a set S is attracted to K if every point in s is attracted to K.

We can now begin to discuss long-term behavior of trajectories, that is, limit sets and the way trajectories approach them.

The simplest case is a *stationary point* or *equilibrium*: this means $x(t) = x$ for all $t \in I_x$. It follows (from the axioms of Continuity and Determinism) that $I_x = [0, \infty)$. Other terms for equilibrium include "rest point" and, especially for flows derived from vector fields, "singular point". (The term "critical point" is also unfortunately used; it should be restricted to its standard meaning of a point where the differential of a function vanishes.)

The set of equilibria is always denoted by E; it is a closed subset of X.

The simplest limiting behavior for a nonstationary point x is to be attracted to an equilibrium p; such a point x is called *convergent*.

If x is attracted to the set E of equilibria, I call x *quasiconvergent*. In practice (e.g., in numerical experiments) is it difficult to distinguish such trajectories from convergent ones. Moreover, if E is totally disconnected (as when X is metrizable and E is countable), then quasiconvergence implies convergence, because $\omega(x)$ is always connected when $\overline{O}(x)$ is compact.

We call x *periodic* if there exists $T > 0$ such that $x(T) = x$. Then $x(t + T) = x(t)$ for all $\tau \geqslant 0$. In this case $O(x)$ is called a *closed* or *periodic orbit*. A periodic orbit which is not a stationary point is called a *cycle*.

Dynamicists have always been fascinated (not to say obsessed) by periodicity. Poincaré devoted much of his seminal treatise on celestial mechanics to proving the existence of infinitely many periodic solutions to the three-body problem. He saw these solutions as the only breech in this difficult problem:

ce qui nous rend ces solutions périodiques si précieuses, c'est qu'elles sont, pour ainsi dire, la seule brèche par où nous puissions essayer de pénétrer dans une place jusqu'ici réputée inabordable.

Les méthodes nouvelles de la mécanique celeste, Vol. I, p. 82

Birkhoff (1920) suggested a psychological necessity for periodicity:

> From early times the mind of man has persistently endeavored to characterize the properties of the motions of the stars by means of periodicities. It seems doubtful whether any other mode of satisfactory description is possible. The intuitive basis for this is easily stated: any motion of a dynamical system must tend with lapse of time towards a characteristic cyclic mode of behavior.

The last statement, while true for many systems, is false if taken literally (even for the Hamiltonian systems which Birkhoff had in mind). Nevertheless, it expresses one of our basic intuitions about long term behavior of systems. One can view the entire development of dynamics as the attempt to find an accurate restatement of it.

Besides asking "what set attracts a given trajectory" we can also ask, "which trajectories are attracted to a given set?" An interesting case is that of an *attractor*: a compact nonempty set K which attracts some neighborhood N of itself. It is also assumed that K is *invariant*, that is, it contains the orbits of all its points. The neighborhood N can always be chosen to be invariant also—simply replace it with the union of the orbits of all its points. The largest such N—the set of all points attracted to K—is called the *basin* of K.

The notion of attractor first appeared, as far as I know, in the work of Thom and Smale in the 1960s. Thom attributes to attractors a basic significance as models in science:

> Every object, or physical form, can be represented as an *attractor C*
> of a dynamical system on a space M of internal variables.
> *Structural stability and morphogenesis* (1975), p. 320.

(Some writers, including Thom, require more internal structure of an attractor, for example that it should be the closure of some orbit; but for this paper the broader definition is sufficient.)

An attractor enjoys a kind of stability: any trajectory starting near it may wander away, but eventually returns to approach it asymptotically. A different kind of stable set K is one which is *orbitally stable*: this means that K is a compact nonempty invariant set having arbitrarily small invariant neighborhoods. For example, in simple harmonic motion (considered as a flow in the plane), every orbit is orbitally stable, but none is an attractor.

A set which is orbitally stable and is also an attractor is called *asymptotically stable*.

Of special interest are the asymptotically stable equilibria, called *sinks*, and the asymptotically stable cycles, sometimes called *limit cycles*. According to the stability dogma these are the only periodic orbits that are physically meaningful or observable.

At the opposite extreme from a sink is the kind of equilibrium called a *source*. This is generally defined only for reversible flows: p is a source if it is a sink for the time-reversed flow. Thus p has a neighborhood N such that every trajectory starting in $N \setminus p$ eventually leaves N, never to return.

"Strange attractor" is a vague, but currently fashionable, term used to indicate an attractor which is neither a sink nor a limit cycle; it is usually assumed to have some internal structure (e.g. cycles are dense) and to be robust. Even since the beginning of the modern era in dynamical systems (roughly twenty years ago) such attractors have been well known to experts, and not considered "strange" any more than imaginary numbers are "imaginary". Some of the most interesting were constructed by Smale (1967, 1980).

Scientists in other fields, however, tend to consider strange attractors as quite exotic, perhaps because they contradict the natural expectation that any system tends towards "a characteristic cyclic mode of behavior". It is in fact exceedingly difficult to decide, either theoretically or practically, whether a particular system has a strange attractor. Systems possessing strange attractors are often described as "chaotic", and it is true that individual trajectories usually appear to be randomly distributed. Recently, J. Guckenheimer (1976, 1982, 1983) has studied the interesting problem of trying to distinguish between deterministic and random behavior on the basis of observed time series.

A seminal and controversial paper by D. Ruelle and F. Takens (1971) suggests that the presence of strange attractors characterizes turbulent hydro-dynamics.

2. Vector fields and diffeomorphisms.

Newton's fundamental discovery, the one which he considered necessary to keep secret and published only in the form of an anagram, consists of the following: *Data aequatione quotcunque fluentes quantitae involvente fluxiones invenire et vice versa.* In contemporary mathematical language this means: "It is useful to solve differential equations".

V. Arnold (1983), p. iii

An *equivalence* (or topological equivalence) between two differential equations is a homeomorphism preserving several trajectories (or orbits). The qualitative problem of differential equations is to obtain information on equivalence classes of differential equations on a given manifold.

S. Smale (1962)

Let $X \subset \mathbf{R}^n$ be a nonempty open set and let $F = (F_1, \ldots, F_n) : X \to \mathbf{R}^n$ be a C^1 (continuously differentiable) vector field. For each $x \in X$ let $I(x) = [0, \tau_x)$ be the maximal interval of this form on which there is a solution $u : I(x) \to X$ to the *initial value problem*

(1a) $$du/dt = F(u),$$

(1b) $$u(0) = x.$$

Setting $\phi_t(x) = u(t)$, we obtain a reversible flow in X; the axioms of Continuity and Determinism are standard theorems in ordinary differential equations.

Uniqueness of solutions implies that x is stationary precisely when $F(x) = 0$. If $\overline{O}(x)$ is compact (in X) then x is quasiconvergent precisely when $F(x(t)) \to 0$ as $t \to \infty$.

Now let F be a C^1 vector field on a (finite-dimensional) manifold M and assume $F(x)$ is tangent to the boundary ∂M whenever $x \in \partial M$. Then one obtains a reversible flow in M by solving in local coordinates the differential equation determined by F. If M is compact then the flow is complete. Flows of this kind I call *smooth*. (If instead one assumes that $F(x)$ is transverse to ∂M and points into M for $x \in \partial M$, then F determines a flow in M which is *not* reversible if $\partial M \neq \varnothing$.)

It is not easy to determine dynamic behavior directly from the vector field. One of the few things that *can* be so determined, in most cases, is the nature of the equilibria. The following standard theorem is basic:

THEOREM. *Let $p \in M$ be an equilibrium of a C^1 vector field F on M. Let $\Sigma \subset \mathbf{C}$ be the spectrum of the linear operator $DF(p)$ on the tangent space M_p. Then*

(a) *If p is orbitally stable or an attractor, then* $\operatorname{Re} \lambda \leqslant 0$ *for* $\lambda \in \Sigma$;
(b) *If* $\operatorname{Re} \lambda < 0$ *for all* $\lambda \in \Sigma$ *then p is a sink.*
(c) *If* $\operatorname{Re} \lambda > 0$ *for all* $\lambda \in \Sigma$ *then p is a source.*

An important kind of equilibrium p is one for which $\operatorname{Re} \lambda \neq 0$ for all $\lambda \in \Sigma$; in this case p is called *hyperbolic*. Their importance stems from the fact that the *local* dynamics near p is very simple, topologically equivalent to the Cartesian product of two linear flows satisfying (b) or (c) at the origin.

If $0 \notin \Sigma$ then p is called *simple*. Simple equilibria are isolated.

A *nonautonomous* differential equation

$$(2) \qquad\qquad du/dt = G(t, U),$$

where $G\colon \mathbf{R} \times X \to \mathbf{R}^n$ is C^1, also leads to a flow; the state space is not X, however, but $\mathbf{R} \times X$. This flow is constructed by considering the autonomous system

$$(3a) \qquad\qquad ds/dt = 1,$$

$$(3b) \qquad\qquad du/dt = G(s, u).$$

The flow in $\mathbf{R} \times X$ defined by (3a), (3b) evidently has the form

$$(4) \qquad\qquad \phi_t(s, x) = (t + s, \psi(t, s, x)).$$

Then for any $(t_0, x_0) \in \mathbf{R} \times X$, the solution to (2) such that $u(t_0) = x_0$, is

$$t \mapsto \psi(t - t_0, t_0, x_0),$$

defined for all t in some open interval about t_0, as seen by differentiating. Here again X could be any smooth manifold M; we then take $G\colon \mathbf{R} \times M \to TM$ to

be a one-parameter family of vector fields on M (a "time-dependent vector field").

An important case is that called *forced oscillations*, when $G(t, x)$ is periodic in t, say of period $T > 0$:

$$G(t + T, x) \equiv G(t, x).$$

In this case we can interpret (3a), (3b) as a vector field on $S^1 \times M$, where S^1 is the circle $\mathbf{R}/T\mathbf{Z}$, and (4) defines a flow on $S^1 \times M$.

Another important way of obtaining smooth flows is by *suspension*. Starting from a diffeomorphism $h: Y \to Y$ of a compact manifold Y, let Y_h be the quotient space of $Y \times \mathbf{R}$ under the equivalence relation $(y, s) \sim (f^n(y), x - n)$. In other words Y_h is the space of orbits of the action of $\mathbf{Z}$ on $Y \times \mathbf{R}$ generated by $(y, s) \mapsto (f(y), s - 1)$. (Topologically X is a bundle over S^1 with fiber Y.) We obtain a smooth flow $\{\phi_t\}_{t \in \mathbf{R}}$ on Y_h by setting $\phi_t[y, s] = [y, s + t]$, where $[y, s] \in Y_h$ denotes the equivalence class of $(y, s) \in Y \times \mathbf{R}$. The generating vector field on $Y \times h$ is that lifting to $0 \times \partial/\partial t$ on $Y \times \mathbf{R}$.

Properties of the flow correspond naturally to those of the diffeomorphism. For example $[y, s]$ belongs to a closed orbit of period $n \in \mathbf{Z}_+$ if and only if $h^n(y) = y$.

This suspension construction is important in two ways. First, many smooth flows have a *global section* Y, that is, a compact hypersurface Y meeting every orbit infinitely often, and transverse to the defining vector field. The "first-return map" $h: Y \to Y$ obtained by following trajectories, has for its suspension a flow having the same orbit structure (ignoring parameterization) as the original flow. Secondly, many interesting examples of flows can be constructed as suspensions.

An example of a suspension is the flow (4) on $S^1 \times M$ obtained from a time-periodic vector field $G(t, x)$ on a compact manifold M. As above, we assume $G(t + T, x) \equiv G(t, x)$ and $S^1 = \mathbf{R}/T\mathbf{Z}$. This flow has a global section $\{p\} \times M$ for any $p \in S^1$. Taking p to be the coset of 0 in S^1, we find the corresponding homeomorphism h of M to be $h(x) = \psi(T, 0, x)$. This is just the time-T map of the flow (4) restricted to $\{p\} \times M$. The suspension of h is exactly the same as (4) if $T = 1$.

The close connections between time-periodic vector fields, global sections, and diffeomorphisms have stimulated a huge body of literature on so-called discrete dynamical systems, i.e. diffeomorphisms. The whole subject was started by Poincaré in his studies of celestial mechanics. Besides its implications for differential equations,

> there is a second and more important reason for studying the diffeomorphism [conjugacy] problem (besides its great natural beauty). That is, the same phenomena and problems of the qualitative theory of ordinary differential equations are present in their simplest form in the diffeomorphism problem. Having first found theorems in the diffeomorphism case, it is usually a secondary task to translate the results back into the differential equations framework.
>
> S. Smale, (1967), p. 747

3. Flows on surfaces: stability versus pathology.

> A system which completely lacks stability would be a poor
> model for reality, as reality is *always* a perturbation of what we
> think it is. Thus some kind of stability is crucial.
>
> R. Williams (1983)
>
> ...the crude speculation that all dynamical systems are
> periodic or nearly so presents itself irresistibly to the human
> mind.
>
> G. Birkhoff (1941)

Questions of long-term behavior arise in various degrees of generality. Depending on circumstances and tradition, more or less rigor and detail may be required. While individual trajectories can usually be approximately computed to high accuracy, this procedure is not easily applied to such questions of the global structure of a system as: Do most trajectories converge? Are cycles dense? Do nearby systems have similar behavior?

Two kinds of *stability* are of great importance in applications of differential equations: stability respecting perturbation of initial values for a fixed equation, and stability respecting perturbations of the equation itself. The terms "persistent" or "robust" denote this second type of stability.

Suppose, for example, that we are interested in finding equilibria of a system (X, ϕ). Having located an equilibrium p, the first stability question is whether p is orbitally (or better, asymptotically) stable. The second question is whether p is persistent, that is, whether every system with state space X and flow sufficiently "close" to ϕ has an equilibrium in any given neighborhood of p. Of course, the meaning of "close"—the appropriate topology on the set of flows under consideration—must be made precise.

For flows determined by a C^1 vector field F on a manifold M, a natural topology is that induced by the "weak C^1 topology" on the set $V(M)$ of C^1 vector fields on M; in this topology a sequence of vector fields converges precisely when, in each coordinate system, the sequence of vector fields and the sequence of their derivatives converge uniformly on compact sets. (But there are many interesting problems where more differentiability is needed and the appropriate topology must take higher derivatives into account.)

The question of the persistence of a singular point under C^1 small perturbations of a vector field translate, via local coordinates, to the persistence of a root $p \in \mathbf{R}^n$ of the equation $F(x) = 0$, where $F: W \to \mathbf{R}^n$ is a C^1 map defining the vector field in an open set $W \subset \mathbf{R}^n$. A well-known sufficient (but not necessary) condition for this is that the linear operator $DF(p)$ be invertible. Thus the same useful condition that guarantees asymptotic stability of p—that the spectrum of $DF(p)$ lie in the open left half-plane—also guarantees persistence.

After contemplating the hard fact that, in practice, we can never know the dynamic or the initial data without error, we are led to seek results that are *generic* in the sense that they hold for most systems, and for most initial values of a given system. To this end Andronov and Pontryagin (1937) introduced the

notion of structural stability. A C^1 vector field F on a compact manifold M (or the flow it determines) is called *structurally stable* provided F has a neighborhood $\mathscr{V}$ in $V(M)$ such that if $G \in \mathscr{V}$ then some homeomorphism h_G of M carries each orbit of G onto an orbit of F, and as G approaches F in $V(M)$ then h_G converges to the identity map of M, in the compact-open ($= C^0$) topology.

It is clear that such an F has exactly the same dynamical properties as any G in the neighborhood v. For any compact surface M it turns out that not only is the set of structurally stable fields dense and open in $V(M)$, but also that such fields are characterized by very simple dynamics.

In fact the following three conditions together are equivalent to structural stability of a C^1 vector field F on a compact surface M:

(a) If p is an equilibrium then $DF(p)$ has no eigenvalue on the imaginary axis;

(b) If p belongs to a cycle of period $T > 0$, then the spectrum of $D\phi_T(p)$ meets the unit circle in 1 as a simple eigenvalue.

(c) No trajectory joins equilibria of saddle type. That is, if $\alpha(x) = p$ and $\omega(x) = q$ then either p is a source or q is sink.

This "General Density Theorem" is due to M. Peixoto (1962). Moreover, it turns out that there are only finitely many periodic orbits in any such system.

There are, of course, surface vector fields that are not structurally stable, and which exhibit exotic limiting behavior. For example, consider the torus T^2 as the coset space $\mathbf{R}^2/\mathbf{Z}^2$. For any $\alpha \in \mathbf{R}$ let F_α be the vector field on T^2 covered by the constant vector field $(1, \alpha)$ on $\mathbf{R}^2$. If α is rational, every trajectory of F_α has period equal to the denominator of α. If α is irrational then every orbit is dense in T. These fields F_α are thus quite exceptional among all the fields in $V(T)$—they do not share the behavior which typifies the dense (and open) set of structurally stable fields.

Even worse (i.e. more interesting) behavior can occur for vector fields which are C^1 but not C^2. A. Denjoy (1932) constructed a field of this type (obtained by suspension of a C^1 diffeomorphism of the circle) without periodic orbits, *all* of whose trajectories are attracted by a certain weird, nowhere dense continuum. While this cannot happen for C^2 fields, as was shown by Poincaré, T. M. Cherry (1938) showed there exist real analytic (even algebraic) fields on the torus for which some trajectories have exotic limit sets containing isolated equilibria.

All such behavior in surface flows can be considered pathological in the light of the General Density Theorem—it disappears under arbitrarily C^1 small perturbations of the vector field, while a field which lacks this behavior has a whole neighborhood which lacks it. Thus from the generic point of view we need only consider structurally stable fields, and these are characterized by simple dynamics. The enumeration of the dynamically different types of structurally stable fields on a compact surface appears to be a manageable topological-combinatorial task—although I don't know if it has been carried out.

However, the generic point of view is not the only one! In practice we ordinarily have to deal with a *restricted class* of fields—in which case we will

not accept arbitrary perturbations, but only those within the class. This requirement, however, prevents us from applying the General Density Theorem; indeed, the whole class may consist of systems which, from the generic point of view, are pathological.

For example, suppose we will consider only flows on the torus which preserve area and have no equilibria. Such a flow cannot possibly be structurally stable: it is known that either every orbit is dense or every orbit is periodic. The linear vector fields F_α discussed above show that no reasonable kind of dynamics stability can hold in this class of area-preserving flows. Instead there are more delicate analytic questions, for example, the existence of a diffeomorphism taking trajectories onto those of some F_α.

Another reservation about the significance of the density of structurally stable systems has been raised by V. Arnold (1983), p. 108. It concerns the meaning of "generic". One plausible definition of a generic property (say of C^1 vector fields on some manifold) is one which is true of all members of some dense and open set of fields. But often we are dealing with *finite-dimensional families* of fields, parametrized for example by a unit k-dimensional cube. It then is plausible to interpret "generic" to mean "true with probability one" in the natural measure on the k-cube. As Arnold points out, *these two definitions conflict for very simple and natural examples.*

Arnold considers flows on the torus obtained by suspending homeomorphisms of the circle which are close to rotations. Let y be the angular coordinate on the circle $\mathbf{R}/2\pi\mathbf{Z}$, and let $a(y)$ be any 2π-periodic analytic function. For each pair of real numbers (r, s) Arnold defines a diffeomorphism $h_{r,s}$ of the circle by the formula

$$h_{r,s}(y) = y + r + sa(y).$$

For sufficiently small fixed $s > 0$, Arnold says that the set

$$D_s = \left\{ r \in [0, 2\pi] : h_{r,s} \text{ has dense orbits} \right\}$$

has measure close to 2π in $[0, 2\pi]$. Now a diffeomorphism has dense orbits if and only if its suspension is a flow with dense orbits, and such flows correspond to a nowhere dense set of structurally unstable vector fields on the torus. Thus we see a conflict between two notions of "generic".

Putting it another way, we can say that the map sending (r, s) to the corresponding vector field takes a set of positive measure into a nowhere dense set. There is nothing contradictory about this in the mathematical sense, but the point is that the family $\{h_{r,s}\}$ is a rather *natural* one. Arnold writes:

> ...numerical experiments usually lead to (at least apparently) everywhere dense orbits...the idea of structural stability is not the only approach to the notion of a generic system. The metric approach indicated above is more appropriate for the description of the actually observable behavior of the system in some cases.

There are limits to the accuracy of any numerical experiment; note Arnold's parenthetical escape clause! One might also ask: Is Arnold's 2-parameter family of diffeomorphisms of the circle a very special family, or is it typical of

such families? In other words, is it a *generic* family! While this line of inquiry may seem to be somewhat philosophically circular, it leads to mathematics of great richness, the theory of *bifurcations*—a subject of major importance for applications, to which Arnold has made important contributions.

4. Flows in higher dimensions: homoclinic tangles and stable chaos.

... "simple differential equations of dimension three or greater can possess solutions of stunning complexity. Moreover, since such systems, in the form of forced oscillators or autonomous evolution equations, play an important role in the modelling of nonlinear processes, an understanding of typical structures of their solutions is essential. ... such an understanding is best achieved from a geometric or topological view point. Reams of computer simulations, without some form of explanation and analysis, are not very helpful.

J. Guckenheimer and P. Holmes (1983), p. 116

Il paraît donc que tout système dynamique non intégrable qui admet une seule solution homocline de cette espèce, doit admettre une hiérarchie presque inconcevable de solutions dans le voisinage étendu correspondant.

G. Birkhoff (1935)

Poincaré, among his many fundamental discoveries in dynamics, found one of the primal sources of chaotic dynamics: the homoclinic orbit. He was working in the "restricted planar three-body problem", a class of what we today call Hamiltonian flows in four-dimensional state spaces. These flows have an invariant function representing energy; fixing the energy gives a three-dimensional state space. Poincaré dealt with a special case having a global section; therefore the orbit structure of the flow is the same as that of the suspension of a diffeomorphism of a two-dimensional manifold—in Poincaré's case, a half-plane. Moreover, this diffeomorphism preserves area. This situation has stimulated much work on surface transformations, continuing unabated to the present.

Clearly, Poincaré worked in a very special system—yet his basic insight, it is now known, is universally applicable.

Poincaré considered trajectories asymptotic, in positive or negative time, to a fixed cycle γ. Using modern terminology (which he would not have liked), we call the set of points attracted by γ the *stable manifold* $S(\gamma)$; and the stable manifold of γ for the time-reversed flow is the *unstable manifold* $U(\gamma)$. These sets are, under fairly generic conditions, immersed submanifolds in the state manifold M; their respective dimensions s and u satisfy $s + u = \dim M + 1$. In the systems Poincaré studied, M is an open set in $\mathbf{R}^3$ and $s = u = 2$, and $S(\gamma)$ is generically an immersed cylinder or Möbius band, and likewise for $U(\gamma)$.

Now $S(\gamma)$ and $U(\gamma)$ intersect at all points of γ, but they may also intersect at other points. A point $y \in S(\gamma) \cap U(\gamma) \backslash \gamma$ is called a *homoclinic point*

belonging to γ; its orbit (or trajectory) is also called homoclinic. Clearly y is *doubly asymptotic* to γ, that is, $\omega(y) = \alpha(y) = \gamma$. The whole orbit of y (in positive and negative time) consists of homoclinic points. We call a homoclinic point y *transverse* if the stable and unstable manifolds intersect transversely at y.

Poincaré's fundamental discovery was that *if y is a transverse homoclinic point then there are infinitely many different homoclinic trajectories in $S(\gamma) \cap U(\gamma)$ and they approach γ in very different ways.* For example, there is a neighborhood N of γ with the following property. Let U be any very small neighborhood of γ and let $t_0 > 0$ be such that $y(t) \in N$ for all $t \geqslant t_0$. Let $t_1 > t_0$; then every neighborhood of y contains another homoclinic point z such that $z(t) \in N$ for $t_0 \leqslant t \leqslant t_1$, but $z(t_2) \notin N$ for some $t_2 > t_1$.

The significance of this is that in every neighborhood of the homoclinic point, y contains other homoclinic points exhibiting infinitely many different kinds of limiting behavior, and the same applies to each of those other homoclinic points, etc. The system thus exhibits *extreme instability with respect to initial values.*

Poincaré ((1899), pp. 389, 391) was struck by the complexity of such dynamics, as well as he might be: the possibility of classifying or even describing the dynamics seems remote. (Especially for Poincaré, who strongly insisted that all mathematical objects be definable in a finite number of terms referring only to previously defined objects. He would probably have rejected even our definition of $\omega(x)$, as it involves infinitely many not-yet-defined sequences. I believe his distaste for set theory may have hampered him.)

The phenomenon of transverse homoclinic points was studied intensely by Birkhoff, who proved that for three-dimensional systems *every neighborhood of a transverse homoclinic point meets infinitely many cycles*, thus demonstrating a new kind of complexity.

Poincaré's study of transverse homoclinic orbits is all the more remarkable because he did not have any examples, or even a proof that one existed! In fact, I do not know whether anyone has ever constructed the kind of example he discussed: A one-parameter family $(S_\mu)_{\mu \geqslant 0}$ of restricted planar three-body problems, such that for some fixed energy, S_0 is an integrable system (all trajectories periodic) and for some $\mu_0 > 0$, every S_μ with $0 < \mu < \mu_0$ has a transverse homoclinic point; and, in addition, the associated vector fields are real analytic in (μ, x) where x is the state variable. Birkhoff (1932), p. 385, admits he cannot construct such an example. This was pointed out by T. M. Cherry (1968), who constructs a single such system but not a one-parameter family. (Poincaré, but not Birkhoff, relied heavily on analyticity and the area preserving property. Both proved their results only for three-dimensional systems.)

Neither Poincaré nor Birkhoff discussed the robustness of their results, although they frequently discounted "infinitely improbable" situations. Nevertheless, it is intuitively plausible that a transverse homoclinic point is a robust phenomenon; and indeed this has been rigorously proved by Smale (1961, 1963) and independently by I. Kupka (1963). In Smale (1961), Birkhoff's theorem on the existence of infinitely many cycles near a transverse homoclinic

point is neatly extended to all dimensions. The proof used symbolic dynamics in a new and extremely fertile way. *Symbolic dynamics* refers to the "shift" homeomorphism $T: C \to C$ on the Cantor set C of all bi-infinite sequences in some finite alphabet; T sends the sequence $c = \{c_j\}_{j \in \mathbf{Z}}$ to the sequence $Tc = \{d_j\}$ where $d_j = c_{j+1}$. Smale produces an embedding f of C into any neighborhood of a homoclinic orbit in such a way that if $f(c) = x$ then $f(Tc) = x(t)$ for some $t > 0$. It follows that if c is a periodic sequence then $f(c)$ is a periodic point. Since periodic sequences are dense in C, and by construction $f(c)$ meets the homoclinic orbit, it follows that every homoclinic point is a limit of periodic points.

J. Moser (1973), p. 100, says that "these ideas, with less detail and rigor, were already described by Birkhoff [in (1935)], where also the concept of sequences of symbols can be found". For that matter, the concept (in this context) goes back to Poincaré (1899), p. 389.

Poincaré's intuition that transverse homoclinic points exist very commonly has proved to be fully justified. In recent years they have been shown to exist in many "natural" dynamical systems.

A very simply described and historically important system is the *suspension of the hyperbolic toral automorphism*. An *automorphism* means a diffeomorphisms f_A of the n-torus $T^n = \mathbf{R}^n/\mathbf{Z}^n$ covered by a linear automorphism of $\mathbf{R}^n$ defined by an n by n integer matrix A of determinant ± 1. We say f_A is *hyperbolic* if A has no eigenvalue on the unit circle. It is easy to see that the periodic points of f_A are exactly the points of T^n corresponding to points of $\mathbf{R}^n$ having rational coordinates; such points are dense in T^n.

The suspension of f_A is a real analytic flow ϕ_A on a certain compact 3-manifold M_A in which *cycles are dense*.

In 1960 Thom asked whether f_A (and thus ϕ_A) is structurally stable for the special case $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. (A diffeomorphism $f: M \to M$ is called *structurally stable* if it has a neighborhood v in the space of C^1 diffeomorphisms of M of which every element is conjugate to f in the group of homeomorphisms of M.) S. Smale (unpublished) and D. Anosov (1962) proved that f_A is structurally stable for any hyperbolic automorphism of T^2; Anosov generalized this to T^n, and to the much wider class of systems now called after him. An important kind of Anosov flow is the geodesic flow on the unit tangent bundle of a compact Riemannian manifold of negative sectional curvature (Anosov (1967)).

Since f_A is structurally stable, *sufficiently nearby diffeomorphisms g of the torus have exactly the same dynamic behavior as f_A*—a highly nontrivial fact. Moreover, the proof of structural stability yields explicit estimates as to how close g must be. Thus we obtain important information on the dynamics of a whole open set of diffeomorphisms. (The same applies to any Anosov system.) For example, it is comparatively easy to prove that f_A has a dense orbit, the union of its periodic orbits is dense, the stable and unstable manifolds of each periodic orbit are dense, and the set of homoclinic points is dense. It follows that the perturbations g all have these same properties.

We can now construct a lovely "strange attractor". Fix a hyperbolic automorphism A of the n torus and let F_A be the vector field on M_A defining the flow ϕ_A obtained by suspending A. Set $V = M_A \times S^1$, where S^1 is the circle

R/Z. On S^1 let G be any vector field with a sink $p \in S^1$. The vector field $F_A \times G$ on V generates a flow ψ having $M_A \times p$ as an attractor. The restriction of ψ to $M_A \times p$ is the same as ϕ_A, whence we are justified in calling $M_A \times p$ a strange attractor.

It can be shown that the flow ψ is structurally stable. Its dynamics are exotic: $\omega(x) = M_A \times p$ for almost every $x \in V$. Most trajectories wind around $M_A \times p$ in a seemingly random way.

These and other examples revealed a wealth of robust dynamic behavior in higher dimensions that does not exist in surface flows. The hope that structurally stable vector fields might be dense in $V(M)$ was dashed with Smale's construction (1965) of a diffeomorphism of the three-torus which cannot be C^1 approximated by structurally stable ones. The suspension of this diffeomorphism is a vector field on a four-dimensional manifold which cannot be C^1 approximated by structurally stable fields. In 1971 S. Newhouse found similar examples in dimension three.

It is interesting that in all such examples, as well as in the proofs of the structural stability theorems referred to above, the key points are intersection properties of stable and unstable manifolds. Thus we see that Poincaré's original insight into their significance continues to be fruitful.

In a recent book J.Guckenheimer and P. Holmes (1983) outline Newhouse's construction and give an interesting discussion of its implications for bifurcation theory and also for numerical experiments in specific systems such as Duffing's equation and the Hénon attractor.

The crucial idea in Smale's and Newhouse's examples is to go beyond Poincaré by considering homoclinic points at which the stable and unstable manifolds are *not* transverse. A system with such a *tangential homoclinic point* cannot be structurally stable, nor could a limit of such systems be structurally stable. Smale and Newhouse showed such systems can be dense in a neighborhood in $V(M)$.

Newhouse (1979, 1980) also showed that if p is a fixed point of a surface diffeomorphism f for which the stable and unstable manifolds are tangent, and $Df(p)$ has negative determinant, then arbitrarily C^r near f there is a diffeomorphism having infinitely many *asymptotically stable* periodic orbits for any $r > 0$. By suspension one gets the analogous result for flows. Poincaré would have been horrified.

While such examples destroyed hope of proving an analog of Peixoto's General Density Theorem for higher dimensions, it appeared plausible that structurally stable systems, while not dense, could at least be characterized by a convenient set of properties. Smale put forward such a list and revised it as counterexamples were found. Currently the gap between known necessary conditions and known sufficient ones is small but still unbridged.

One such condition, known to be sufficient in conjunction with other properties, and conjectured to be necessary, has been the basis for all further work on structural stability. It is Smale's

AXIOM A. *The nonwandering set has a hyperbolic structure.*

For a flow $\{\phi_t\}$ on a compact manifold M this means the following. A point x of M is *wandering* if it has a neighborhood U such that $(\phi_t U) \cap U$ is empty for all sufficiently large $t > 0$; otherwise x is *nonwandering*. The set of nonwandering points has a *hyperbolic structure* in case for every nonwandering point the following holds: The tangent space to M at x is the direct sum of three linear subspaces. One is spanned by the vector field. The elements of the other two are, respectively, exponentially expanded and exponentially contracted, under the induced action of the flow on the tangent bundle of M, as $t \to +\infty$.

For a diffeomorphism the definition is analogous, involving only two linear subspaces.

It is usually exceedingly difficult to find the nonwandering set of a given system, and equally difficult to determine whether it has a hyperbolic structure. For a vector field given by explicit formulas, such as the Lorenz system (see below), these tasks appear to be hopeless. But in many cases the system in question is not given in that way; rather it is described in more geometric or algebraic terms. I have in mind examples such as the geodesic flow on the unit tangent bundle of a compact Riemannian manifold: Anosov (1962) proved that if all sectional curvatures are negative then not only the nonwandering set but the whole state space has a hyperbolic structure.

There are two extreme cases of smooth flows on compact manifolds having easily determined nonwandering sets. If the flow preserves a smooth measure then *every* point is nonwandering, as follows trivially from the definition of wandering. And if every omega limit set is a hyperbolic periodic orbit then the union of these orbits is the nonwandering set, and it has a hyperbolic structure.

In this latter case J. Palis and S. Smale have proved (1970) that a necessary and sufficient condition for structural stability is that all intersections between stable and unstable manifolds be transverse.

It is fitting to close this chapter with an explicit *system of differential equations*, *the celebrated Lorenz system*:

$$dx/dt = -10x + 10y,$$
$$dy/dt = 28x - y - xz,$$
$$dz/dt = -8z/3 + xy.$$

This is an extreme simplification of a system arising in hydrodynamics. By computer simulation E. Lorenz (1963) found that most trajectories seem to wander back and forth between two particular stationary states in an unpredictable way. Trajectories which start out very close together eventually diverge, with no relationship between their long-run behaviors.

But this chaotic behavior has not been *proved*; and in fact, almost nothing has been proved about this particular system, although it has become very well known to dynamicists. J. Guckenheimer (1976) and R. Williams (1979) proved that there do indeed exist many systems which exhibit, in a rigorous sense, the observed behavior of Lorenz's system; but no one knows if Lorenz's system is one of them. It is of no particular importance to answer this question; but our ignorance is a sharp challenge to dynamicists. Considering all the attention paid to Lorenz's system, this situation is something of a scandal.

   In this chapter the emphasis has been on complexity and stability, neglecting many other topics of great importance. These include the highly developed theory of bifurcations, or finite-dimensional families of sytems, and the related theory of normal forms; Hamiltonian systems, important for mechanics and geometry; ergodic theory, closely allied to probability theory and statistical mechanics; and the subtle regularity problems and deep topological questions, basic to any structural understanding of the space of systems on a manifold. Very interesting work has been done on specific classes of systems arising from various applications; these often point the way to general results. We have not touched on the vast field of infinite-dimensional systems, a subject of great interest for applications, but almost completely undeveloped from the dynamical perspective; some new results for parabolic partial differential equations are presented in the next chapter. And except for mentioninig the Hénon attractor, no attention has been paid to the rapidly expanding field of the dynamics of noninvertible maps (or "difference equations").

   For reviews and expositions of some of the recent work in these subjects the reader may consult the books of J. Moser (1973), V. Arnold (1983), S. Smale (1980), J. Palis and W. Demelo (1982), R. Abraham and J. Marsden (1978), J. Hale (1963), J. Marsden and M. McCracken (1976), J. Franks (1982), J. Walker (1980), B. Hassard, N. Kazarinoff and Y.-H. Wan (1980), and D. Henry (1981). The 1983 book by J. Guckenheimer and P. Holmes contains, in addition to useful discussions on many topics, an unusually complete bibliography. See also G. Iooss et al (1983), G. Barenblatt et al (1983).

CHAPTER III: CONVERGENCE AND STABILITY
IN MONOTONE FLOWS

> Peut-être le but final de la dynamique théorique est de faire "l'intégretion logique". A vrai dire, le développment caractéristique d'une théorie mathématique semble être achevé quand nous pouvons passer librement de la forme purement quantitative à la forme purement qualitative et inversement. En faisant ainsi, nous en obtenons toute la structure logique. Dans le cas des systèmes dynamiques, nous partons de la forme quantitative (le système donné des équations différentielles), et nous cherchons à déterminer les propriétés qualitatives des mouvements et leur relations...

> G. D. Birkhoff (1935), p. 268

> Most differential equations admit neither an exact analytic solution nor a complete qualitative description.

> V. Arnold (1983), p. 142

   Introduction. It is currently fashionable to study dynamical systems on manifolds, and indeed these are the proper setting for many questions. In practice, however, there are usually certain natural variables which are related by a differential equation; and it is natural, at least as a first step, to choose

these variables as coordinates, thus obtaining a differential equation in some Cartesian space $\mathbf{R}^n$. Invariant functions, or constraints, may subsequently suggest restricting attention to a submanifold, and symmetries or periodicities may also lead to manifolds through identifications.

The class of systems to be described in this chapter are those whose flow preserves a (partial) ordering on the state space. In the examples arising in practice the state space is either $\mathbf{R}^n$ with its vector ordering (given by inequalities between corresponding components), or else it is a space of real-valued functions with the natural ordering. It is technically convenient to consider also open subsets of such spaces.

For the results below to be valid the ordering must have a few special properties which hold in $\mathbf{R}^n$ and in many, but by no means all, of the popular function spaces. The linear structure plays no other role; various frequently used properties of function spaces, such as completeness of a norm, reflexiveness, local convexity, etc., are not needed for the proofs of the main results, although such properties are often present in the settings of the main examples.

For these reasons it seems natural to proceed at a fairly high level of generality, and in the last section this will be technically useful. The state spaces will thus be ordered spaces having a few special properties. The reader who prefers a more concrete approach can assume the state space to be either $\mathbf{R}^n$ or the C^1 functions on a compact Riemannian manifold which vanish on the boundary, or altenatively, whose normal derivatives vanish at the boundary. At a higher level of abstraction the state space can be taken to be an open set in any ordered Banach space whose positive cone has nonempty interior.

The results of this section mean that for the class of flows considered here one can reasonably expect most trajectories to converge to stable stationary states. For these flows there are no strange attractors, no chaotic dynamics. These "strongly monotone" flows form a small but not unnatural class having the further advantage that it is usually easy to determine whether a given vector field has a flow in the class. It also enjoys a healthy robustness, although I discuss this here only for flows in $\mathbf{R}^n$.

§1 contains basic definitions and examples. A basic convergence criterion, Theorem 2.3, is proved in §2; this is applied in §§3 and 4 to prove existence of equilibria in attractors. §5 contains results indicating that for strongly monotone flows convergence of trajectories to equilibria is to be expected for most initial states. §6 reconsiders the epidemiological model given in §1. The final section explains in what sense these results remain valid under perturbations of the flow.

1. Monotone flows in strongly ordered spaces.

> We may formulate the aim of dynamics as follows: to characterize completely the totality of motions of dynamical systems by their qualitative properties.
>
> G. D. Birkhoff (1920)

An *ordered space* consists of a topological space X together with a (partial) order relation $R \subset X \times X$ which is a closed subspace. We write:

$$x \leqslant y \quad \text{if } (x, y) \in R,$$
$$x < y \quad \text{if } x \leqslant y \text{ and } x \neq y,$$
$$x \ll y \quad \text{if } (x, y) \in \text{Int } R$$

where Int indicates the interior of a set. Notations such as $x > y$ have the obvious meanings.

The ordered space X is said to be *strongly ordered* if it has the following two properties:

(SO 1) If $U \subset X$ is open and $x \in U$ then there exist a, b in U with $a \ll x \ll b$.

(SO 2) If $U \subset X$ is open and a, b are in U with $a \ll b$ then there exists $x \in U$ with $a \ll x \ll b$.

Notice that (SO 1) implies that Int R is dense in R.

EXAMPLE 1.1. Let V be a topological vector space and $V_+ \subset V$ a closed convex cone (with vertex at the origin) such that if both x and $-x$ are in V_+ then $x = 0$. We make V into an ordered space by defining $x \leqslant y$ to mean $y - x \in V_+$. This kind of ordered space is called an *ordered topological vector space*. Every open subset is an ordered space in the induced ordering. It is easy to see that such spaces are strongly ordered if and only if Int V_+ is nonempty. In this case $y \gg x$ means $y - x \in \text{Int } V_+$.

The property of being strongly ordered is rather delicate. Consider for example the following spaces, where M is a compact smooth manifold:

$$C^r(M) = \{ C^r \text{ maps } M \to \mathbf{R} \},$$

$$C_0^r(M) = \{ u \in C^r(M) : u \mid \partial M = 0 \}.$$

Here r is any nonnegative integer. These are topological vector spaces when given the weak C^r topology; they have Banach norms. The cone of nonnegative functions makes them into ordered topological vector spaces. It is clear that $C^r(M)$ is strongly ordered, with $u \gg 0$ if and only if $u(x) > 0$ for all x. But if ∂M is nonempty then $C_0^0(M)$ is *not* strongly ordered. On the other hand, for $r > 0$, $C_0^r(M)$ *is* strongly ordered, with $u \gg 0$ if and only if u is strictly positive on $M \setminus \partial M$ and the gradient of u (in any Riemannian metric) is transverse to ∂M at every boundary point.

EXAMPLE 1.2. The *vector ordering* on $\mathbf{R}^n$ is defined by the following cone, called the *closed positive orthant*,

$$\mathbf{R}_+^n = \{ x : x_j \geqslant 0 \text{ for } j = 1, \ldots, n \}.$$

This is a strong ordering, with $x \gg 0$ if and only if all $x_j > 0$. Unless otherwise indicated $\mathbf{R}^n$ is always given this ordering.

EXAMPLE 1.3. Let X be an open set in a strongly ordered topological vector space V. For x, y in X define $x \leqslant_p y$ in case there is a *monotone path* h: $[0, 1] \to X$ from x to y, that is, h is continuous, $h(s) > h(t)$ if $s > t$, and $h(0) = x$, $h(1) = y$. It is easy to see that the relation $\leqslant_p$ is a strong ordering of X which agrees with the ordering of V in convex subsets of X. In fact, it agrees in any subset U of X, which is what I call *p-convex*: U contains the segment

joining any two of its points related by $<$. This ordering $\leqslant_p$ is the *monopathic* ordering of X.

A map f between ordered spaces is called *monotone* if $x \leqslant y$ implies $f(x) \leqslant f(y)$. It is *strongly monotone* if $x < y$ implies $f(x) \ll f(y)$.

A real matrix $A \geqslant 0$ (all entries $\geqslant 0$) defines a monotone linear map and a strongly monotone linear map if $A \gg 0$ (all entries positive). The following example is a nonlinear generalization:

EXAMPLE 1.4. Let $X \subset V$ be as in Example 1.3 and let $f \colon X \to X$ be a C^1 map (Gateaux differentiability suffices). Suppose that each linear map $Df(x)$: $V \to V$ is monotone. Then f is monotone for the monopathic ordering in X. To see this suppose x and y are joined by a monotone path in X. Replace this path by a piecewise linear monotone path. By induction on the number of vertices it suffices to consider the case where x and y are endpoints of a segment in X, and $x < y$. Computing $f(y) - f(x)$ by Taylor's formula shows that $f(x) < f(y)$; since this applies to any two points of the segment it follows that the image of the segment is a monotone path in X joining $f(x)$ to $f(y)$.

A similar argument proves that f is strongly monotone for the monopathic ordering provided each $Df(x)$ is strongly montone. When X is p-convex these results hold for the ordering inherited from V.

Now let $\phi = \{\phi_t\}_{t \geqslant 0}$ be a flow (in the sense of Chapter II, §1) in an ordered space X. I call ϕ a *monotone flow* when each map ϕ_t is monotone, and a *strongly monotone flow* if ϕ_t is strongly monotone for all $t > 0$.

EXAMPLE 1.5. Let A be a real $n \times n$ matrix such that $A + cI \geqslant 0$ for some real number c. The linear flow $\{e^{tA}\}_{t \geqslant 0}$ is monotone, as is seen by writing $A = -cI + (A + cI)$. Thus $e^{tA} = e^{-ct}e^{t(A+cI)}$, and the exponential of the matrix $t(A + cI)$ is $\geqslant 0$ because the coefficients of the exponential series are positive.

Recall that the matrix A is *irreducible* if it maps into itself no proper nonzero subspace defined by equations $x_{i_1} = \cdots = x_{i_k} = 0$. It can be shown that if A in Example 1.5 is irreducible then $e^{tA} \gg 0$ for $t > 0$, and thus the linear flow is strongly monotone.

The following is a nonlinear generalization of the preceding example:

EXAMPLE 1.6. Let $F \colon X \to \mathbf{R}^n$ be a C^1 vector field in an open set $X \subset \mathbf{R}^n$. When the off-diagonal terms of the Jacobian matrices $DF = [\partial F_i / \partial F_j]$ are all $\geqslant 0$, I call F (and its flow) *cooperative*. If each Jacobian matrix is irreducible then F is called *irreducible*. Using a theorem of Kamke on differential inequalities it can be shown (Hirsch (1983a)) that if F is cooperative then its flow ϕ has the property that each matrix $D\phi_t(x)$ is $\geqslant 0$, and if also F is irreducible then $D\phi_t(x) \gg 0$ for $t > 0$. Using the result in Example 1.4 we obtain

THEOREM 1.7. *Let ϕ be the flow generated by a cooperative vector field F in an open set $X \subset \mathbf{R}^n$. Then ϕ is monotone for the monopathic ordering in X, and strongly monotone when F is irreducible.*

The negative of a cooperative vector field is called a *competitive* field G; it is characterized by $\partial G_i / \partial x_j \leqslant 0$ for $i \neq j$. The flows of cooperative and competitive fields correspond under time-reversal—replacing the independent variable

t by $-t$. For some purposes the direction of time doesn't matter and one can pass freely from competitive to cooperative fields as convenient. The limit sets of such fields have special topological and dynamical properties (Hirsch (1982)), some of which will be discussed in subsequent sections.

EXAMPLE 1.8. A. Lajmanovich and J. Yorke (1976) used a class of irreducible cooperative flows to model the spread of gonorrhea. Their equations can be generalized as follows. Consider a disease which, like gonorrhea, confers no immunity: anyone who doesn't have it is susceptible. Assume the population divided into n disjoint classes. Let P_i be the cardinality of class i, assumed constant, and x_i the number of infecteds in class i. Then $P_i - x_i$ is the number of susceptibles in class i. Both x_i and $P_i - x_i$ are nonnegative integers, but by the usual handwaving we treat x_i as a continuous variable and assume it depends differentiably on time. Let R_i be the rate of infection of class i and C_i the cure rate, and assume (unrealistically) that R_i is a function only of the vector $x = (x_1, \ldots, x_n)$, and that C_i depends only on x_i. The key assumption, and a reasonable one, is $\partial R_i / \partial x_j \geq 0$. Finally, we assume that each group can directly or indirectly infect every group.

These assumptions are plausibly modeled by the system of equations

$$dx_i / dt = R_i(x) + C_i(x_i) \equiv F_i(x) \qquad (i = 1, \ldots, n).$$

Evidently, $\partial F_i / \partial x_j \geq 0$ for $i \neq j$, so the system is cooperative, and the assumption of mutual indirect infectibility means that we can assume the system is also irreducible. Thus it leads to a strongly monotone flow. (The equations of Lajmanovich and Yorke have the special form

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} A_{ij} x_j (P_i - x_i) K_i x_i \qquad (i = 1, \ldots, n)$$

with real constants $A_{ij} \geq 0$ *and* $K_i > 0$, *and* $[A_{ij}]$ *irreducible*.) The general theory of monotone flows is thus applicable to these equations, which are discussed in more detail in §6 below.

For further discussion of the use of cooperative or competitive systems as biological models the reader may consult, among many other works, A. Rescigno and I. Richardson (1967), A. N. Kolmogorov (1936), S. Grossberg (1978), S. Smale (1976), W. Leonard and R. May (1975), P. Waltman (1983), H. Freedman (1980). S. Hsu et al (1978), H. Freedman and P. Waltman (1984), J. Coste et al (1979).

EXAMPLE 1.9. Let $M \subset \mathbf{R}^n$ be a compact n-dimensional submanifold with smooth boundary ∂M. (In this example "smooth" means "of class C^4".) Consider a semilinear parabolic equation on M:

(1) $\partial u / \partial t = Au + f(x, u, \nabla u) \qquad (t > 0, x \in M)$

with boundary conditions either of the form

(2) $u \mid \partial M = 0 \quad$ (Dirichlet)

or

(3) $\nabla u \cdot v = 0 \quad$ (Neumann).

Here A is a uniformly elliptic second order differential operator of the form

$$A = \sum_{i,j=1}^{n} a_{ij} D_i D_j,$$

with $[a_{ij}]$ a matrix of smooth real-valued functions on M, positive definite at each point, and $D_i = \partial/\partial x_i$; the vector function ∇u denotes the gradient $(D_1 u, \ldots, D_n u)$ of $u(x, t)$ in x, and $v: \partial M \to \mathbf{R}^n$ is a smooth vector field transverse to ∂M. The map $f: M \times \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}$ is smooth and satisfies certain compatibility conditions in the case of Dirichlet boundary conditions. The unknown function $u: M \times [0, \tau) \to \mathbf{R}$ is to be continuous for some $\tau > 0$, and is to satisfy (1), and (2) or (3).

It has long been known that under appropriate technical assumptions, for any reasonable initial function $w: M \to \mathbf{R}$ there is a unique solution to (1) with either boundary condition such that $u(x, 0) = w(x)$. Putting $\phi_t(w) = u(\ , t)$ we seem to obtain a flow, but we face the problem of specifying the function space in which it acts. It is currently fashionable to look for some sort of generalized solutions in L^p spaces. The resulting flows are monotone, but they cannot be strongly monotone since such spaces are not strongly ordered. If we impose too *much* smoothness we run the risk that the "flow" is not continuous.

Fortunately, recent work by X. Mora (1983) neatly finesses these difficulties and produces flows in spaces of C^0 or C^1 functions satisfying the boundary conditions. Standard maximum principles imply that these flows are monotone or strongly monotone (e.g. M. Protter and H. Weinberger (1967), Theorem 12 on p. 187 and Theorem 14 on p. 190). From Mora's work (which treats more general kinds of parabolic equations) we obtain the following result:

THEOREM 1.10. (a) *Solutions of the Neumann problem* (1), (3) *determine a strongly monotone flow* ϕ *in* $C^0(M)$ *and also in the space* $C_v^1(M)$ *of all* C^1 *functions satisfying* (3). *The latter flow is* C^1, *that is, each map* ϕ_t *has a continuous Fréchet derivative.*

(b) *Solutions of the Dirichlet problem* (1), (2) *determine a monotone flow in* $C_0^0(M)$ *provided* $f(x, 0, y) = 0$ *for all* $x \in \partial M, y \in \mathbf{R}^n$.

(c) *Solutions of the Dirichlet problem* (1), (2) *determine a strongly monotone flow in* $C_0^1(M)$ *provided* $f(x, 0, y) = 0$ *for all* $x \in \partial M$, $y \in \mathbf{R}^n$. *If also* $Df(x, 0, y)(0, 0, z) = 0$ *for all* $(x, y, z) \in \partial M \times \mathbf{R}^n \times \mathbf{R}^n$, *then this flow is* C^1.

Other examples of monotone flows may be found in Hirsch (1983).

2. Convergence criteria for monotone flows.

> May the God who watches over the right use of mathematical symbols in manuscipt, print, and on the blackboard, forgive me this and my many other sins.
>
> H. Weyl, *The classical groups*, p. 289n.

Throughout this section (and the rest of the paper) X denotes a strongly ordered space. The following notation will be used. If $A, B \subset X$ are subsets then $A < B$ means $a < b$ for all $a \in A$, $b \in B$; and similarly for $A \ll B$, $A \leqslant B$, etc. If $x \in X$, notations such as $x < B$ have the obvious meaning.

The *open order interval* $[[x, y]]$ is

$$[[x, y]] = \{z \in X : x \ll z \ll y\}.$$

The *closed order interval* $[x, y]$ is

$$[x, y] = \{z \in X : x \leqslant z \leqslant y\}.$$

(If $x \ll y$ then $[x, y]$ is the closure of $[[x, y]]$.) Also define

$$[[A, B]] = \{z \in X : A \ll z \ll B\}.$$

Let $S \subset X$ be any set. A *maximal element* of S is a point $s \in S$ such that if $x \in S$ and $x \geqslant s$ then $x = s$. *Minimal elements* are similarly defined. It follows easily from Zorn's lemma that *every nonempty compact subset of X contains a maximal and a minimal element*.

X is a *lattice* if every pair of elements has a least upper bound and a greatest lower bound. It is then easily proved that every compact subset has a least upper bound and a greatest lower bound.

To the strongly ordered space X we associate another topological space $\hat{X}$. The underlying set of $\hat{X}$ is X, while the topology of $\hat{X}$ is generated by all open order intervals $[[a, b]]$. The open sets of this *order topology* are also open in the original topology. The original ordering makes $\hat{X}$ into a strongly ordered space. Neighborhoods in $\hat{X}$ are called *order neighborhoods*. The identity map $X \to \hat{X}$ is continuous and monotone. Notice that $x \ll y$ means the same in X and $\hat{X}$.

(If X is a strongly ordered topological vector space then the order-topology can be defined by a norm as follows: Fix $e \gg 0$ in X and define

$$\|x\|_e = \inf\{t > 0 : t^{-1}x \in [[-e, e]]\}.$$

It is left as an exercise to verify that this is a well-defined norm which induces the order topology. If M is a compact manifold and $X = C^1(M)$ then the order-topology is the same as the C^0 topology on $C^1(M)$. But if $\partial M \neq \varnothing$ and $X = C_0^1(M)$, then the order topology is finer than the C^0 topology. Unfortunately, we do not make use of these interesting subtleties in the present paper.)

Since the identity map from X to X is continuous, it follows that *for any compact $K \subset X$ the order topology on K is the same as the original topology*. This has the following useful corollary:

LEMMA 2.1. *Let $\{x_i\}$ be a sequence in a compact subset K of a strongly ordered space X. Suppose $p \in X$ has the property that whenever $a \ll p \ll b$ there exists i_0 such that $a \ll x_i \ll b$ for all $i > i_0$. Then $\lim_{i \to \infty} x_i = p$.*

PROOF. Since $K \cup \{p\}$ is compact it suffices to establish convergence in the order topology; and that is the hypothesis. Q.E.D.

The following result says that a monotone flow ϕ in X is continuous in $\hat{X}$. (The same maps ϕ_t would define a flow in $\hat{X}$ except for the possibility that its domain $X_{\sharp}$ might not be an open set in $[0, \infty) \times \hat{X}$.)

LEMMA 2.2. *Let $X^{\#} \subset [0, \infty) \times X$ be the domain of a monotone flow ϕ in X. Let $\hat{X}^{\#}$ denote the subspace of $[0, \infty) \times \hat{X}$ occupied by the set $X^{\#}$. Then ϕ is continuous as a map $\hat{X}^{\#} \to \hat{X}$.*

PROOF. Suppose $\phi(t_0, x_0) \in [[a, b]]$. Since ϕ is continuous there is a neighborhood $I \times W \subset [0, \infty) \times X$ of (t_0, x_0) such that $\phi(I \times W) \subset [[a, b]]$. Fix u, v in W such that $u \ll x_0 \ll v_0$ (using property (SO 1) of strongly ordered spaces). Since each ϕ_t preserves order we have, for $(t, x) \in I \times [[u, v]]$,

$$a \ll \phi(t, u) \ll \phi(t, x) \ll \phi(t, v) \ll b.$$

Thus

$$\phi(I \times [[u, v]]) \subset [[a, b]]. \quad \text{Q.E.D.}$$

Throughout the remainder of this section ϕ denotes a monotone flow in X.

The following convergence criterion is useful as an existence theorem for equilibria:

THEOREM 2.3. *Let $x \in X$ have compact orbit closure, and assume there is a real $T > 0$ with $x(T) \gg x$ or $x(T) \ll x$. Then $x(t)$ converges (necessarily to an equilibrium) as $t \to \infty$.*

Before proving 2.3 we draw some consequences.

COROLLARY 2.4. *A monotone flow does not have an attracting cycle.*

PROOF. Let $C \subset X$ be a periodic orbit which attracts a neighborhood N of C. Pick any $p \in C$ and any $x \in N$ such that $x \gg p$. Since x is attracted to C it follows that $\omega(x) = C$; in particular, $p \in \omega(x)$. Therefore there exists $T > 0$ such that $x(T)$ belongs to the following neighborhood W of p: $W = \{z \in X : z \ll x\}$. Since $x(T) \ll x$, Theorem 2.1 implies that $\omega(x)$ is a singleton; thus C reduces to $\{p\}$ and cannot be a cycle. Q.E.D.

COROLLARY 2.5. *A monotone flow in an open set $X \subset \mathbf{R}^3$ cannot have a knotted cycle or two linked cycles.*

PROOF. Let $C \subset \mathbf{R}^3$ be a cycle for ϕ. Let $E^2 \subset \mathbf{R}^3$ be the plane perpendicular to the vector $(1, 1, 1)$ and let $\pi : \mathbf{R}^3 \to E^2$ be the orthogonal projection. Now $\pi | C$ must be injective. For if $\pi(x) = \pi(y)$ with x, y distinct points of C, then $y = x(T)$, some $T > 0$, and since $y - x$ is a scalar times $(1, 1, 1)$, either $x(T) \gg x$ or $x(T) \ll x$. But then $x(t)$ would converge so C could not be a cycle. And injectivity of π means C is unknotted. The proof of unlinking is left as an exercise. Q.E.D.

These results are generalized in Hirsch (1982).

The proof of Theorem 2.3 is based on a useful test for stationarity.

LEMMA 2.6. *Suppose $x \ll y$ and there exists a sequence $t_i \to \infty$ such that $x(t_i)$ and $y(t_i)$ converge to p in the order topology. Assume there exists $z \in [[x, y]]$ such that $z(t)$ is defined for all $t \geqslant 0$. Then p is stationary.*

PROOF. Fix $\delta > 0$ so small that $z(t) \in [[x, y]]$ for all $t \in [0, \delta]$. Then $x(t_i) \leqslant z(t_i + t) \leqslant y(t_i)$ for all $t \in [0, \delta]$ and all i. It follows that as $i \to \infty$ the

sequence $z(t_i + t)$ converges to p in the order topology for all $t \in [0, \infty]$. In particular, $z(t_i) \rightarrow p$ in X. Since each ϕ is continuous in X we also have $z(t_i + t) \rightarrow p(t)$ in X for all $t \in [0, \delta]$. Thus $p(t) = p$ for all $t \in [0, \delta]$, whence p is stationary.

PROOF OF THEOREM 2.3. We assume $x \ll x(T)$, the other case being similar. Let p be a limit point of $\{x(mT): M \in \mathbf{Z}_+\}$. Then the sequence $\{x(mT)\}$ evidently converges to p in the order topology as $m \rightarrow \infty$, by monotonicity of the map ϕ. Applying Lemma 2.6 to x and $y = x(T)$, with $t_i = iT$ for $i \in \mathbf{Z}_+$, we find that p is stationary. To prove that $x(t) \rightarrow p$ in X it suffices to prove convergence in $\hat{X}$ since $\overline{O}(x)$ is compact. For this it is enough to prove that for every unbounded sequence $\{s_j\}$ in $\mathbf{R}_+$ the sequence $\{x(s_j)\}$ has a subsequence (relabeled as $\{s_j\}$) so that $s_j = n_j T + r_j$, with $n_j \in \mathbf{Z}_+$ and $r_j \in [0, T)$, such that there exists $\lim_{j \rightarrow \infty} r_j = r \in [0, T)$. Set $y_j = x(n_j T)$ and observe that continuity of ϕ in the order topology implies $y_j(r_j) \rightarrow p(r)$ in $\hat{X}$. Since $y_j(r_j) = x(s_j)$ and $p(r) = p$, the proof is complete.   Q.E.D.

The equilibrium p exhibited in the foregoing proof has a special property, namely that $y(t) \rightarrow p$ if $y \in [[x, x(T)]]$ and y has compact orbit closure. Thus when all orbit closures are compact—a not uncommon situation—p attracts all elements in some nonempty open set. I call such a p a *trap*. If X is a Banach space and each map ϕ_t is differentiable this implies that the spectrum of the linear operator $D\phi_t(p)$ on X is contained in the closed unit disk. If ϕ is a smooth flow in $\mathbf{R}^n$ generated by a smooth vector field F then $\operatorname{div} F(p) = 0$. In this case, under generic assumption on F, p is a sink.

3. Equilibria in attractors of monotone flows.

> An idea which can be used once is a trick. If it can be used
> more than once it becomes a method.
>
> G. Polya and S. Szegö (1971)

In this section ϕ is a monotone flow in the strongly ordered space X and $K \subset X$ is an attractor for ϕ (see Chapter II, §1).

THEOREM 3.1. *K contains an equilibrium.*

The proof uses Birkhoff's concept of a *nonwandering point* z—a point such that there exist sequences $z_i \rightarrow z$ in X and $t_i \rightarrow \infty$ in $\mathbf{R}_+$ such that $z_i(t_i) \rightarrow z$. These include all omega limit points. Every compact invariant set contains at least one nonwandering point. The set of nonwandering points is denoted by Ω.

The set $\Omega \cap K$ is compact, invariant and nonempty. For any $z \in \Omega \cap K$ there are maximal and minimal elements p, q, of $\Omega \cap K$ such that $p \leqslant z \leqslant q$.

Theorem 3.1 follows from the following more precise information.

THEOREM 3.2. *Let $p \in \Omega \cap K$ be any minimal element. Then p is an equilibrium and there exists $y \ll p$ such that $y(t) \rightarrow p$. A similar result holds for maximal elements.*

PROOF. Choose sequences $z_i \to p$ in X and $t_i \to \infty$ in $\mathbf{R}$ such that $z_i(t_i) \to p$. Let y in the basin of K be such that $y \ll p$ (using property (SO 1) of §1). Then $y \ll z_i$ for large i. Passing to a subsequence if necessary we assume $y(t_i)$ converges to some $x \in K$. Now $y(t_i) \leqslant z_i(t_i)$ by monotonicity, so by continuity of the flow, $x \leqslant p$; and $x \in \Omega \cap K$, so $x = p$ by minimality of p.   Q.E.D.

The proof shows that maximal and minimal equilibria in K are traps (see end of §2). If all equilibria in K are known to be simple then these extremal ones are sinks; this implies that the dynamic in K cannot be very chaotic.

Another application of 3.2 is

THEOREM 3.3. *If K contains only one equilibrium p then every trajectory attracted to K converges to p.*

PROOF. Let x be in the basin of K. Let $z \in \omega(x)$ be arbitrary. By 3.2 there are equilibria p_1, p_2 in K with $p_1 \leqslant z \leqslant p_2$. But $p_1 = p_2 = p$, so $\omega(x) = \{ p \}$. Q.E.D.

4. Stable equilibria in strongly monotone flows.

> ...even the structurally stable attractors in rather simply given examples contain a vast mixture of periodic, almost periodic, homoclinic, and other kinds of phenomena.
>
> S. Smale (1971)

The goal of this section is to prove that every attractor of a strongly monotone flow contains an equilibrium enjoying a certain kind of stability (Theorem 4.1).[2]

Throughout this section ϕ is a *strongly* monotone flow in the strongly ordered space X. It is further assumed that all orbits have compact closure (and therefore each ϕ_t is globally defined). Such compactness is often satisfied in the examples considered in §1. When $X = \mathbf{R}^n$ it means that every orbit is bounded (in forward time). In the case of the parabolic semilinear equations of Example 1.9 and Theorem 1.10, well-known smoothing properties of the solution flow in the appropriate function space X imply that any orbit which is bounded in the $L^2(M)$ norm has compact closure in X.

Stability—the over-arching theme of dynamical systems theory—is a protean concept which inevitably gives rise to many different definitions in technical exposition. Before fixing precise terminology it may be helpful to describe roughly two important general notions of stability of an equilibrium p. If points near p have trajectories which do not stray very far from p then p is called simply "stable". If, in addition, these trajectories all have p as a limit then p is called "asymptotically stable". The exact meanings of these terms depend on the topology of the state space X. In the present context we may want to employ the order topology, in which case we use the modifier "order". And it will be technically advantageous to consider weaker types of stability in which only trajectories $\geqslant p$ or $\leqslant p$ are required to stay near p; for this we use the modifiers "upper" and "lower".

---

[2] A similar result has been independently proved by H. Matano, who kindly sent me a preprint.

An efficient master definition of stability is the following. Let $\mathscr{S}$ be some family of subsets of X. We call ϕ $\mathscr{S}$-*stable* if for any $Q_0 \in \mathscr{S}$ there exists $Q_1 \in \mathscr{S}$ such that $\phi_t(Q_1) \subset Q_0$ for all $t \geqslant 0$.

By specifying $\mathscr{S}$ we obtain various kinds of stability, including the following. An equilibrium p is called

> *stable* if $\mathscr{S} = \{\text{neighborhoods of } p\}$,
>
> *order-stable* if $\mathscr{S} = \{\text{order neighborhoods of } p\}$,
>
> *upper order-stable* if $\mathscr{S} = \{[p, v] : v \gg p\}$,
>
> *lower order-stable* if $\mathscr{S} = \{[u, p] : u \ll p\}$.

Given $\mathscr{S}$, we call p *asymptotically $\mathscr{S}$-stable* whenever p is $\mathscr{S}$-stable and there exists $Q_* \in \mathscr{S}$ with the following property: for any $Q \in \mathscr{S}$, $\phi_t(Q_*) \subset Q$ for all sufficiently large $t > 0$. Thus each of the four types of stability listed above has its stronger asymptotic form.

For flows in $\mathbf{R}^n$ the concepts of stability and order-stability are equivalent, and so are their asymptotic versions. This is not always true in more general strongly ordered spaces, but it holds under the additional hypothesis that ϕ is *order-compact*. This means that $\phi_t[[a, b]]$ has compact closure for all $t > 0$ and every open order interval $[[a, b]]$.

The proof that order-compactness of ϕ makes order-stability of p equivalent to stability, is based on the following idea. Order compactness implies that $\phi_r(X)$, for fixed $r > 0$, inherits the same topology from X and $\hat{X}$ (see §2). Therefore the two stability notions coincide for the restriction of ϕ to $\phi_r(X)$, and this can be used to finish the proof.

The parabolic equation $\partial u / \partial t = Au + f(x, u, \nabla u)$, with Neumann or Dirichlet boundary conditions, generates strongly monotone flow in function spaces $C_v^1(M)$ and $C_0^1(M)$, respectively, when the hypotheses of Example 1.9 and Theorem 1.10(a) or (c) are satisfied. Owing to the smoothing properties of these flows, they can be shown to be order-compact if $f = f(x, u)$. This is discussed further in Hirsch (1983).

THEOREM 4.1. *Every attractor K for the strongly monotone flow ϕ contains an order-stable equilibrium. When ϕ is order-compact K contains a stable equilibrium.*

The proof is based on the observation that the *equilibria constructed by the convergence criterion Theorem 2.3 are asymptotically upper or lower order-stable*, as follows easily from the definitions. In particular, any minimal equilibrium in K, promised by 3.2, is asymptotically lower order-stable, and analogously for maximal ones.

For every $p \in K \cap E$ define

$$L(p) = \operatorname{clos} \bigcup \{\omega(x) : X \gg p\}.$$

It is easy to see that $L(p)$ is fully invariant, i.e. $\phi_t L(p) = L(p)$ for all $t \geqslant 0$.

LEMMA 4.2. (a) *If $p \in L(p)$ then p is upper order-stable.*

(b) *If $p \notin L(p)$ there exists an equilibrium $q \gg p$ in K which is asymptotically lower order-stable.*

PROOF. (a) Suppose $p \in L(p)$. If $p \in \omega(x)$ for some $x > p$, then 2.3 implies p is upper order-stable. Otherwise for every $v \gg p$ there exists $x > p$ and $y \in \omega(x) \cap [[p, v]]$ such that $p \notin \omega(x)$. Then $p < \omega(x)$, and by full invariance and strong monotonicity $p \ll \omega(x)$. Now fix $u \in [[p, \omega(x)]]$. By strong monotonicity we have, for all $t \geqslant 0$,

$$\phi_t[p, u] \subset [p, \omega(x)]$$

and thus

$$\phi_t[p, u] \subset [p, v].$$

This proves that p is upper order-stable.

(b) Suppose $p \notin L(p)$. Then $p < L(p)$ and, by full invariance, $p \ll L(p)$. Pick any $z \in [[p, L(p)]]$. Then $\omega(z) \subset L(p)$ whence $z \ll \omega(z)$. By 2.3 $z(t)$ converges to an equilibrium $q \gg z \gg p$, and such a q must be asymptotically lower order-stable.   Q.E.D.

PROOF OF THEOREM 4.1. By 3.2 K contains a minimal equilibrium which is asymptotically lower order-stable. Let p be a maximal element of the compact set (nonempty because it contains p_0):

$$H = \operatorname{clos}\{ y \in K \cap E : y \geqslant p_0 \text{ and } y \text{ is lower order-stable}\}.$$

Then $p \in K \cap E$ and one proves readily that p is lower order-stable. Maximality of p precludes alternative (b) of Lemma 4.2, so 4.2(a) holds. Thus p is both upper and lower order-stable.   Q.E.D.

As with Theorem 3.1 the proof yields more information. For example, if K contains only one order-stable equilibrium then every trajectory in the basin of K must converge. This is a consequence of the following result:

THEOREM 4.3. *Let K be an attractor for the strongly monotone flow ϕ. Suppose z is attracted to K but is not quasiconvergent. Then K contains two order-stable equilibria p, q such that $p \ll \omega(z) \ll q$.*

PROOF. Set $B = \{x \in \text{basin of } K : x \ll \omega(z)\}$. Then B is nonempty. For by hypothesis there exists $y \in \omega(z) \setminus E$. Since y is nonwandering K contains an equilibrium $b < y$, and hence $\ll \omega(z)$, by Theorem 2.3. Clearly, $b \in B$.

Set $K_0 = \operatorname{clos} \bigcup \{\omega(x) : x \in B\}$. Then K_0 is an attractor whose basin is B, and $K_0 \ll \omega(z)$ by full invariance. Similarly there exists an attractor $K_1 \gg \omega(z)$. Now apply Theorem 4.1 to K_0 and K_1.   Q.E.D.

The results of this section are further evidence that monotone flows do not have chaotic dynamics: an attractor K containing a stable equilibrium cannot contain a dense orbit (unless it is a single equilibrium), nor can cycles in K be dense.

5. Convergence almost everywhere in strongly monotone flows.

In his discussion of recurrent motion, H. Poincaré introduces the fundamental notion of a property which, without being true for all possible motions, has a probability of one of being realized.

G. D. Birkhoff and B. O. Koopman (1932)

In this section ϕ is a strongly monotone flow in the strongly ordered space X; the set of equilibria is as usual denoted by E. In the interest of brevity many proofs are omitted.

The main working tool is the following *limit set dichotomy*:

THEOREM 5.1. *Suppose $x < y$. Let x and y have compact orbit closures. Then exactly one of the following holds*:
  (a) $\omega(x) \ll \omega(y)$;
  (b) $\omega(x) = \omega(y) \subset E$.
In particular, (a) *must hold unless both x and y are quasiconvergent.*

A proof for the case where X is an open set in $\mathbf{R}^n$ is given in Hirsch (1983a).

Theorem 5.1 means intuitively that limit sets containing nonstationary points take up a lot of space. This idea can be exploited under various additional hypotheses to show that "most" trajectories are quasiconvergent.

For the following theorems let $S \subset X$ be a nonempty simply ordered subset, all of whose elements have compact orbit closure. Let $L \subset X$ be a set containing $\bigcup_{x \in S} \omega(x)$. (For example, X might be an ordered Banach space with L an attractor and S a line segment parallel to a positive vector.) Let $S_0 \subset S$ be the subset of nonquasiconvergent points.

The hypothesis of the following result, the main one of this section, is satisfied when X is a separable Banach space, e.g. a subspace of $C^r(M)$.

THEOREM 5.2. *If the topology of L has a countable basis then S_0 is countable.*

PROOF. Suppose S_0 is uncountable. Let $z(x) \in \omega(x)$ for each $x \in S_0$ (using the axiom of choice). By the limit set dichotomy (5.1), the set $X = \{ z(x) : x \in S_0 \}$ is simply ordered by $\ll$. Be second countability B contains one of its accumulation points, say $z_0 = z(x_0)$ for some $x_0 \in S_0$. We assume z_0 is a limit point of $\{ z(x) : x \in S, x \gg x_0 \}$, the other case being similar. It follows that $z_0 \geqslant \omega(x_0)$, since $\omega(x) \gg \omega(x_0)$ for all $x \in S_0, x \gg x_0$. But since $z_0 \in \omega(x_0)$ it follows that x_0 is convergent, a contradiction. Q.E.D.

The intuitive meaning of Theorem 5.2 is that almost every point with compact orbit closure has a quasiconvergent trajectory. When X is an open set in $\mathbf{R}^n$ this statement is true in the technical sense:

THEOREM 5.3. *Let ϕ be a strongly monotone flow in an open set $X \subset \mathbf{R}^n$ in which all orbits have compact closures. Then the set Y of nonquasiconvergent points with compact orbit closures has measure zero.*

PROOF. Let $E^{n-1} \subset \mathbf{R}^n$ be the hyperplane perpendicular to some vector $v \gg 0$. Theorem 5.2 implies that every line perpendicular to E^{n-1} intersects Y in a set which is countable and therefore of measure zero. Since Y is easily proved to be a Borel set it follows from Fubini's theorem that Y has measure zero. Q.E.D.

With slightly stronger (but reasonable) assumptions, S_0 in Theorem 5.2 can be proven to be even sparser:

THEOREM 5.4. *If L is compact, or is a lattice in the induced ordering, then S_0 is discrete in its relative topology and every point of $S \setminus S_0$ is convergent.*

THEOREM 5.5. *If E has finite cardinality $|E|$ then so has S_0, and $|S_0| \leqslant |E| + 1$.*

For attractors containing only finitely many equilibria we can improve on Theorem 4.1. Using the limit set dichotomy and reasoning similar to that in the proof of Theorem 4.1, one can prove that any isolated order-stable equilibrium in an attractor must be asymptotically order stable. In this way one proves

THEOREM 5.6. *Let $K \subset X$ be an attractor such that $K \cap E$ is finite. Then K contains an asymptotically order-stable equilibrium.*

6. Applications to the gonorrhea model and to a parabolic equation.

> Successful science is abstractive: of all that is of interest, only a fragment is susceptible to explanation, and only a fragment of that fragment to serious mathematical analysis.
>
> Power and scope... are often *inversely* related.
>
> D. Berlinski (1976), p. 85

As an illustration of some of the preceding results we return to the strongly monotone system of Example 1.8,

$$(1) \qquad dx_i/dt_i = R_i(x) - C_i(x_i) \equiv F_i(x) \qquad (i = 1, \dots, n),$$

generalizing the special case studied by Lajmanovich and Yorke:

$$(2) \qquad \frac{dx_i}{dt} = \sum_{j=1}^{n} A_{ij}(P_i - x_i)x_j - K_i x_i \equiv H_i(x) \qquad (i = 1, \dots, n).$$

Equations (2) have the properties that $dx_i/dt \leqslant 0$ if $x_i = 0$ and $dx_i/dt < 0$ if $x_i = P_i$. These conditions must be satisfied if the system is to make biological sense, since the number x_i of infecteds cannot go below 0 or rise above P_i. It is therefore plausible to assume that equations (1) have the same properties. This makes the compact set $W = \{x \in \mathbf{R}^n : 0 \leqslant x_i \leqslant P_i, i = 1, \dots, n\}$ invariant under the flow; thus all orbit closures in W are compact. Trajectories outside of W have no biological meaning.

With no further assumptions we see that the flows of (1) and (2) in W have a strong tendency towards convergence: from Theorem 5.3 we see that starting from almost any initial level $x \in W$ of infections the disease will progress toward the set E of stationary states—practically speaking, infection levels will stabilize (if this model is valid).

Let us consider the case where the infection rates R_i and cure rates C_i are such that very low infection levels increase, that is $F(x) \gg 0$ if $x > 0$ and $|x|$ (the Euclidean norm) is sufficiently small. It follows easily that Int W must contain an attractor, and thus by Theorem 4.1 there is in this case a stable equilibrium at which all infection levels are positive.

Equations (2), however, have a much more interesting property. Lajmanovich and Yorke proved that either *all* trajectories in $\mathbf{R}^n_+$ tend to 0, or else there is a unique equilibrium $p \gg 0$ and all trajectories in $\mathbf{R}^n_+ \backslash \{0\}$ tend to p. In particular, infection levels do not oscillate indefinitely, no matter what the initial levels are. (This is true for *almost* every initial level in the more general model (1).).

Actual data in fact show a yearly oscillation. In a more sophisticated model, Aronsson and Mellander (1980) allow the contact rates A_{ij} to be periodic functions of time; he shows that either all trajectories die out or they approach a unique cycle in $\mathbf{R}^n_+$. Thus it appears that oscillations merely reflect periodic changes in the environment rather than subtle dynamic behavior.

The biological interpretation of the Lajmanovich-Yorke result is clear: in the absence of periodic environmental fluctuations either the disease dies out, or it stabilizes at a positive infection level in each of the n groups. Moreover, the vector of stabilized levels is unique (because there is no other equilibrium in $\mathbf{R}^n_+ \backslash \{0\}$).

This nontrivial result is not at all obvious, either from equations (1) or the biological assumptions which motivated them. Its chief interest is not as a detailed description of reality (which it is not) but as an insight into the behavior of the disease, and as a link between biological hypotheses and conclusions.

One can use the mathematical model to learn what effect various changes in the data might have on infection levels. In particular, different strategies for lowering these levels have been investigated by Hethcote et al (1982).

One of the surprising features of the model is the uniqueness of the equilibrium $p \gg 0$. It is not biologically evident why there could not be several such equilibria, so that, for example, low initial levels of infection would eventually stabilize at an equilibrium $p_1 \gg 0$, but sufficiently high initial levels would stabilize at $p_2 \gg p_1$.

As far as I know the uniqueness of the stable infection levels has not been investigated in the literature; it is merely an unexplained mathematical consequence of equations (2).

It is possible to "explain" this uniqueness in the following sense: it is a consequence of a particular mathematical property of equations (1), and this property has a simple biological interpretation. To be precise, the vector field $F = (F_1, \ldots, F_n)$ in $\mathbf{R}^n$ has the property that

$$(3) \qquad \frac{\partial^2 F_i}{\partial x_k \, \partial x_j} \leqslant 0 \quad \text{for all } (i, j, k).$$

We can interpret $\partial F_i / \partial x_j$ as the *marginal* rate at which group j infects group i. Thus (3) means: *marginal infection rates decrease as infection levels rise.* Whether this is true, or plausible, is an interesting biological question; but it is the mathematical reason for uniqueness of stable infection levels. This is a consequence of the following result, whose hypotheses are easily verified for (1):

THEOREM 6.1. *Let F be a C^1 vector field in $\mathbf{R}^n$, whose flow ϕ preserves $\mathbf{R}^n_+$ for $t \geqslant 0$ and is strongly monotone in $\mathbf{R}^n_+$. Assume that the origin is an equilibrium and that all trajectories in $\mathbf{R}^n_+$ are bounded. Suppose the matrix-valued map DF: $\mathbf{R}^n \to \mathbf{R}^{n \times n}$ is strictly antimonotone, in the sense that*

$$(4) \qquad \qquad \textit{if } x > y \quad \text{then} \quad DF(x) < DF(y).$$

Then either all trajectories in $\mathbf{R}^n_+$ tend to the origin, or else there is a unique equilibrium $p \in \text{Int } \mathbf{R}^n_+$ and all trajectories in $\mathbf{R}^n_+ \backslash \{0\}$ tend to p.

PROOF. Consider the case that the trajectory of some $x \in \mathbf{R}^n_+$ does not tend to 0. Then $\omega(x)$ has a least upper bound $y \in \mathbf{R}^n_+ \backslash \{0\}$. It is easy to see that $y(t)$ is also an upper bound for $\omega(x)$ for all $t \geqslant 0$, whence $y(t) \geqslant y$. By strong monotonicity and Theorem 2.3, $y(t)$ converges to an equilibrium $q \geqslant y$. Thus, $q > 0$, so by strong monotonicity $q \gg 0$.

Now fix any equilibrium $p \gg 0$. We claim that:

(5) *If $0 < x < p$ then $x(t) \to p$.*

This is based on the following:

(6) *Let z be in the open line segment having endpoints 0 and p. Then $F(z) > 0$.*

To prove (6) consider, for each $i = 1, \ldots, n$, the map $g_i: [0,1] \to \mathbf{R}$, $g_i(s) = F_i(sp)$. Then $g_i(0) = g_i(1) = 0$; and the assumption on DF implies that $g_i(s)$ is nonincreasing and not identically zero for some i. It follows from the mean value theorem that $g_i(s) \geqslant 0$ for all i and all $s \in [0,1]$, while for some i we have $g_i(s) > 0$ for $0 < s < 1$. This proves (6).

To prove (5) consider the closed order intervals (n-cubes) $B_s = [sp, p]$, $0 < s < 1$. It follows from strong monotonicity and the facts that $F(sp) > 0$, $F(p) = 0$, that if $x \in B_s \backslash \{p\}$ then $F(x) \neq 0$ and $F(x)$ points *into* B_s.

This implies that $\phi_t(B_s \backslash \{p\}) \subset \mathrm{Int}\, B_s$ for all $t > 0$, $0 < s < 1$. In particular, $sp \leqslant \phi_t(sp) < p$. Therefore the trajectory of sp converges to an equilibrium in B_s, which can only be p.

Now fix any $y \in [0, p] \backslash \{0\}$. There exists s, $0 < s < 1$, such that $sp \leqslant y \leqslant p$. Since $\phi_t(sp) \to p$, $\phi_t(p) = p$, and the flow is monotone, it follows that $\phi_t(y) \to p$. This proves (5).

It follows from (5) that there is a *unique* equilibrium $p \gg 0$. Fix $y \in \mathbf{R}^n_+ \backslash \{0\}$. *It is impossible that* $y(t) \to 0$, for this would entail $y(t_0) \in [0, p] \backslash \{0\}$ for some $t_0 > 0$. But then the trajectory of $y(t_0)$ would approach p, contradicting $y(t) \to 0$. The argument at the beginning of the proof shows that some equilibrium—which must be p—is $\geqslant \omega(y)$. If $p \in \omega(y)$ then using 2.3 one proves that $y(t) \to p$. If $p > \omega(y)$ then strong monotonicity implies $p \gg \omega(y)$. But then $p \gg y(t_0)$ for some $t_0 > 0$, and again $y(t) \to p$. This completes the proof of 6.1.   Q.E.D.

Theorem 6.1 has a generalization to certain parabolic equations (Example 1.9 and Theorem 1.10). Consider an equation of the form

(7a) $$\partial u / \partial t = Au + f(u),$$

(7b) $$u \,|\, \partial M = 0,$$

a special case of equations (1), (3) of Example 1.9. Here we assume $f: \mathbf{R} \to \mathbf{R}$ is C^4 and $f(0) = 0$. The resulting strongly monotone flow ϕ in $C^1_0(M)$ is C^1 by Theorem 1.10(c).

THEOREM 6.2. *Suppose $w \in C^1_0(M)$ is a stationary solution to (7a), (7b), that is, a solution of the elliptic equation $\partial u / \partial t = Au + f(u)$, $u|\partial M = 0$, and assume $w \gg 0$. Let $\alpha > 0$ be the maximum value of w. Assume either (a) $f''(y) > 0$ for $0 < y < \alpha$ or (b) $f''(y) < 0$ for $0 < y < \alpha$. Let $v \in [0, w]$ be any initial value different from 0 and w. Then $\phi_t(v) \to 0$ in case (a), and $\phi_t(v) \to \alpha$ in case (b).*

PROOF. The proof is similar in outline to that of Theorem 6.1, the main point being that for each $x \in M$ the map

$$g_x \colon [0,1] \to \mathbf{R}, \qquad s \mapsto A(sw)(x) + f(sw(x))$$

is increasing (in case (a)) or decreasing (in case (b)). This follows because $g_x''(s) = f''(sw(x))$. This implies in case (b) that each order interval $[sw, w]$ is invariant; in case (a) the order intervals $[0, sw]$ are invariant. The rest of the proof is formally the same.   Q.E.D.

A different approach to Theorem 6.2 and other applications to parabolic equations may be found in Hirsch (1983).

7. Robustness.

In every mathematical investigation the question will arise whether we can apply our results to the real world.... Consequently, the question arises of choosing those properties which are not very sensitive to small changes in the model and thus may be viewed as properties of the real process.

V. I. Arnold (1983), p. 87

In this section X denotes an open set in $\mathbf{R}^n$ and ϕ is a smooth flow on X generated by a C^1 vector field F in X.

The most general hypothesis so far exhibited, which guarantees strong monotonicity of ϕ, is that F be cooperative and irreducible (§1). Now the property that a real $n \times n$ matrix be irreducible persists under small perturbations of the matrix. Therefore, any vector field sufficiently C^1 near F will have irreducible matrices $DF(x)$ for x in any given compact set. But the cooperativity property—$\partial F_i / \partial x_j \geqslant 0$ for $i \neq j$—does not enjoy similar robustness. While it is true that the stronger hypothesis that DF have strictly positive off-diagonal terms is robust, such a stringent condition excludes many interesting systems that have strongly monotone flows.

To escape this dilemma we need to look closely at the reason why an irreducible cooperative field generates a strongly monotone flow, viz., because each matrix $D\phi_t(x)$ is $\gg 0$ (that is, $D\phi_t(x)$ has strictly positive entries) for all $t > 0$.

Now let G denote a vector field which generates a flow ψ. Then for any fixed $t_0 > 0$ and any compact $C \subset \mathbf{R}^n$ we can make $\psi_t(x)$ close to $\phi_t(x)$ and $D\psi_t(x)$ close to $D\phi_t(x)$ for all $t \in [0, t_0]$ and $x \in K$, by taking G sufficiently C^1 close to F in the compact set $\bigcup \{\phi_t(K) \colon 0 \leqslant t \leqslant t_0\}$. When C is *invariant* under ϕ_t for all $t \geqslant 0$ it turns out that F has a neighborhood $\mathcal{N} \subset V(M)$ such that there exists $t_0 > 0$ with the following property: If $x \in C$ and $t_2 \geqslant t_0$ are such that $\psi_t(x) \in K$ for all $t \in [0, t_2]$, then $D\psi_t(x) \gg 0$ for all $t \in [t_0, t_2]$. Moreover, t_0 can be taken in any neighborhood of 0 provided $\mathcal{N}$ is sufficiently small.

It is not unreasonable to assume further that C is also invariant under ψ_t for $t \geqslant 0$; in this case it follows that

$$(1) \qquad D\psi_t(x) \gg 0 \quad \text{for all } t \geqslant t_0, \; x \in C.$$

We shall define a new ordering in the open set $X_0 = \text{Int } C$. To this end, for each $x \in X_0$ define

$$\Gamma(x) = \bigcap \left\{ D\psi_t(x)^{-1} \mathbf{R}_+^n : t \geq t_0 \right\} = \bigcap \left\{ D\psi_t(x)^{-1} \mathbf{R}_+^n : 2t_0 \geq t \geq t_0 \right\}.$$

One readily verifies that $\Gamma(x)$ is a closed convex cone containing $\mathbf{R}_+^n$, and $\Gamma(x) \cap (-\Gamma(x)) = \{0\}$. Moreover, (1) implies

$$(2) \qquad\qquad D\psi_t(x)\Gamma(x) \subset \text{Int } \Gamma(\psi_t x) \quad \text{for all } t \geq 0.$$

For $x, y \in X_0$ define $x \prec y$ to mean that there is a piecewise smooth path $h: [0,1] \to X_0$ such that $h(0) = x$, $h(1) = y$, and for all $s \in [0,1]$, $h'(s) \in \Gamma(h(s))$ if $h'(s)$ exists. This defines a strong ordering in X_0. (This ordering can also be described as the monopathic ordering (Example 1.3) associated to the ordering R defined in X_0 by

$$x \, R \, y \Leftrightarrow \psi_t(x) \leq \psi_t(y) \quad \text{for all } t \geq t_0.)$$

It follows from (2) that *the flow* $\psi | X_0$ *is strongly monotone for the ordering* $\prec$.

These considerations imply that all the preceding theorems about strongly monotone flows apply to the flow $\psi | X_0$ when we give X_0 the ordering $\prec$. In this sense the preceding results are robust. This can be summarized as follows:

THEOREM 7.1. *Let F be an irreducible, cooperative vector field in the open set $X \subset \mathbf{R}^n$. For any open set X_0 having compact closure in X there exists a C^1 neighborhood $\mathcal{N} \subset V(X)$ with the following property. Let $G \in \mathcal{N}$ generate a flow ψ which for all $t > 0$ preserves some open set $U \subset X_0$. Then $\psi | U$ is strongly monotone for some strong ordering of U that locally is an enlargement of the vector ordering.*

As an application here is a robustification of Theorem 4.1.

COROLLARY 7.2. *Let $F, X, \mathcal{N}$ be as in Theorem 7.1. Then for any $G \in \mathcal{N}$ the flow ψ of G has the following property: Every attractor for ψ, whose basin lies in X_0, contains a stable equilibrium.*

PROOF. Take U in 7.1 to be the basin of an attractor for ψ and apply Theorem 4.1.   Q.E.D.

The following corollary of 7.1 has a more convenient hypothesis:

COROLLARY 7.3. *Let $M \subset \mathbf{R}^n$ be a compact n-dimensional submanifold with piecewise smooth boundary. Let F be a cooperative irreducible vector field defined in a neighborhood of M, which at every boundary point is transverse to ∂M and points into $\text{Int } M$. Then F has a neighborhood $\mathcal{N} \subset V(M)$ with the following property: Any $G \in \mathcal{N}$ has a flow ψ for which M is invariant in positive time; and $\psi | \text{Int } M$ is strongly monotone for a certain strong ordering.*

PROOF. Apply 7.1 to X = the neighborhood of M mentioned in 7.3, taking $X_0 = \text{Int } M$. By choosing $\mathcal{N}$ small enough we ensure that any G in $\mathcal{N}$ is inwardly transverse to ∂M and thus its flow preserves $\text{Int } M$ for $t > 0$. We can now take $U = \text{Int } M$ in 7.1.   Q.E.D.

Robustness allows the introduction of generic properties of vector fields. For example, the set of fields on a compact manifold M having only hyperbolic

equilibria, is dense and open in $V(M)$. For such a field the equilibrium set is finite and every trap (see §2) is a sink, and quasiconvergent trajectories are convergent. In applying results of §§4 and 5 one obtains

THEOREM 7.4. *Let M, F and $\mathcal{N}$ be as in Corollary 7.3. Then there is a dense open subset $\mathcal{N}_0$ of $\mathcal{N}$ such that the flow of any field in $\mathcal{N}_0$ leaves M invariant for $t \geqslant 0$, and has the following properties in M:*

  (a) *all equilibria are hyperbolic;*

  (b) *the trajectory in forward time of almost every point of M tends to a sink;*

  (c) *in any monotone arc $I \subset M$, the set of points which are not in the basin of a sink, is finite;*

  (d) *every attractor contains a sink;*

  (e) *if M contains only one sink p then all trajectories tend to p.*

BIBLIOGRAPHY

R. ABRAHAM
  1971   *Predictions for the future of differential equations*, Lecture Notes in Math., vol. 206 (D. Chillingworth, ed.), Springer-Verlag.
R. ABRAHAM AND J. MARSDEN
  1967   *Foundation of mechanics*, Benjamin, New York.
  1978   *Foundation of mechanics*, 2nd ed., Benjamin/Cummings, Reading, Mass.
R. ABRAHAM AND C. SHAW
  1982   *Dynamics—The geometry of behavior*, Aerial Press, Santa Cruz, Calif.
M. AMMAN
  1976   *Fixed point equations and nonlinear eigenvalue problems in ordered Banach spaces*, SIAM Rev. **18**, 620–709.
A. A. ANDRONOV AND L. PONTRYAGIN
  1937   *Systèmes grossiers*, Dokl. Akad. Nauk. SSSR **14**, 247–251.
D. V. ANOSOV
  1962   *Roughness of geodesic flows on compact Riemannian manifolds of negative curvature*, Soviet Math. Dokl. **3**, 1068–1070.
  1967   *Geodesic flows on closed Riemannian manifolds with negative curvature*, Trudy Mat. Inst. Steklov **90**, 3–209; English transl. in Proc. Steklov. Inst. Math. (I. G. Petrovskiĭ and S. M. Nikol'skiĭ, eds.), Amer. Math. Soc., Providence, R. I., 1969.
V. I. ARNOLD
  1983   *Geometrical methods in the theory of ordinary differential equations*, Grundlehren Math. Wiss., vol. 250, Springer-Verlag.
G. ARONSSON AND I. MELLANDER
  1980   *A deterministic model in biomathematics: asymptotic behavior and threshold conditions*, Math. Biosci. **49**, 207–222.
D. BERLINSKI
  1976   *On systems analysis*, MIT Press, Cambridge, Mass.
G. D. BIRKHOFF
  1920   *Recent advances in dynamics*, Science (N.S.) **51**, 51–55 and *Collected mathematical papers*, vol. 2, Amer. Math. Soc., Providence, R. I., pp. 106–110.
  1927   *Dynamical systems*, Amer. Math. Soc. Colloq. Publ., vol. 9, Amer. Math. Soc., Providence, R. I.; reprinted 1966.
  1927a   *On the periodic motion of dynamical systems*, Acta Math. **50**, 359–379 and *Collected mathematical papers*, vol. 2, op. cit., pp. 333–353.
  1932   *Sur l'existence de régions d'instabilité en dynamique*, Ann. Inst. H. Poincaré **2**, 369–386 and *Collected mathematical papers*, vol. 2, op. cit., pp. 444–461.
  1935   *Sur le problème restreint des trois corps (première mémoire)*, Ann. Scuola Norm. Sup. Pisa (2) **4**, 267–306 and *Collected mathematical papers*, vol. 2, op. cit., pp. 466–306.

1941   *Some unsolved problems of theoretical dynamics*, Science **94**, 598–600 and *Collected mathematical papers*, vol. 2, op. cit., pp. 710–712.

1935a   *Nouvelles recherches sur les systèmes dynamiques*, Mem. Pont. Acad. Sci. Nov. Lyn. **1**, 85–216 and *Collected mathematical papers*, vol. 2, op. cit., pp. 530–661.

1943   *The mathematical nature of physical theories*, Amer. Sci. **31**, 281–310 and *Collected mathematical papers*, vol. 2, op. cit., pp. 890–919.

1950   *Collected mathematical papers*, vols. 1, 2, 3, Amer. Math. Soc., Providence, R. I.

G. D. BIRKHOFF AND B. O. KOOPMAN

1932   *Recent contributions to ergodic theory*, Proc. Nat. Acad. Sci. U.S.A. **18**, 279–282 and *Collected mathematical papers*, op. cit., vol. 3, pp. 462–465.

L. E. J. BROUWER

1907   *On the foundation of mathematics*, Collected Mathematical Works, Vol. 1, American Elsevier, New York.

1913   *Intuitionism and formalism*, Bull. Amer. Math. Soc. **20**, 81–96; *Collected works*, Vol. 1, op. cit., pp. 123–138.

1975   *Collected works*, Vol. 1 (A. Heyting, ed.), American Elsevier, New York.

J. L. CASTI

1982   *Recent developments and future perspectives in nonlinear systems theory*, SIAM Rev. **24**, 301–331.

T. M. CHERRY

1938   *Analytic quasi-periodic curves of discontinuous type on a torus*, Proc. London Math. Soc. (2) **44**, 175–215.

1968   *Asymptotic solutions of analytic Hamiltonian systems*, J. Differential Equations **4**, 142–159.

W. A. COPPEL

1965   *Stability and asymptotic behavior of differential equations*, Heath, Boston.

J. COSTE, J. PEYRAUD AND P. COULLET

1979   *Asymptotic behaviors in the dynamics of competing species*, SIAM J. Appl. Math. **36** (1979), 516–543.

D'ALEMBERT AND DIDEROT (Editors)

1754   *Encyclopédie*, Vol. 4, Briasson, David, LeBreton, Durand, Paris.

C. DARWIN

1892   *Life of Charles Darwin* (F. Darwin, ed.), John Murray, London.

1936   *The origin of species* and *The descent of man*, Modern Library, New York.

A. DENJOY

1932   *Sur les courbes définies par les équations différentielles à la surface du tore*, J. Math. Pures Appl. (4) **17**, 333–375.

S. DRAKE

1978   *Galileo at work*, Univ. of Chicago Press, Chicago, Ill.

A. S. EDDINGTON

1927   *The nature of the physical world*, Cambridge Univ. Press, Cambridge, England.

A. EINSTEIN

1930   *On the occasion of the three hundredth anniversary of Kepler's death*, Frankfurter Zeitung, November 9, 1930 and *Ideas and opinions*, Crown, New York, 1954.

1931   *On the one hundreth anniversary of Maxwell's birth*, James Clerk Maxwell: A Commemorative Volume, Cambridge Univ. Press, Cambridge, England, and *Ideas and opinions*, op. cit.

J. M. FRANKS

1982   *Homology and dynamical systems*, CBMS Regional Conf. Ser. in Math., Vol. 49, Amer. Math. Soc., Providence, R. I.

H. I. FREEDMAN

1980   *Deterministic mathematical models in population ecology*, Dekker, New York.

H. I. FREEDMAN AND P. WALTMAN

1984   *Persistence in models of three interacting species*, Math. Biosci. (in press).

N. GEORGESCU - ROEGEN

1966   *Analytical economics: Issues and problems*, Harvard Univ. Press, Cambridge, Mass.

C. C. GILLISPIE
    1960   *The edge of objectivity*, Princeton Univ. Press, Princeton, N. J.
W. H. GOTTSCHALK AND G. A. HEDLUND
    1955   *Topological dynamics*, Amer. Math. Soc. Colloq. Publ., vol. 36, Amer. Math. Soc., Providence, R. I.
S. GROSSBERG
    1978   *Competition, decision and consensus*, J. Math. Anal. Appl. **66**, 470–493.
J. GUCKENHEIMER
    1976   *A strange strange attractor*, The Hopf Bifurcation and Its Applications (J. E. Marsden and M. McCracken, eds.), Springer-Verlag, pp. 368–381.
    1982   *Noise in chaotic systems*, Nature **298**, 358–361.
J. GUCKENHEIMER AND G. BUZYNA
    1983   *Dimension measurements for geostrophic turbulence*, Phys. Rev. Lett. **51**, 1438–1441.
J. GUCKENHEIMER AND P. HOLMES
    1983   *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Applied Math. Sci., Vol. 42, Springer-Verlag.
J. HADAMARD
    1912   *L'oeuvres mathématique de Henri Poincaré*, Acta Math. **38**, 203–287; and *Oeuvres de Jacques Hadamard* 4, CNRS, Paris, pp. 1921–2005.
    1912a   *Henri Poincaré et le problème de trois corps*, Rev. Mois **16**, 385–418 and *Oeuvres de Jacques Hadamard*, op. cit., pp. 2007–2041.
    1968   *Oeuvres de Jacques Hadamard* 4, CNRS, Paris.
J. K. HALE
    1963   *Oscillations in nonlinear systems*, McGraw-Hill, New York.
T. L. HANKINS
    1980   *Sir William Rowan Hamilton*, Johns Hopkins Univ. Press, Baltimore.
B. D. HASSARD, N. D. KAZARINOFF AND Y.-H. WAN
    1980   *Theory and application of the Hopf bifurcation*, Cambridge Univ. Press, Cambridge, England.
R. HEILBRONER
    1953   *The worldly philosphers*, Simon and Schuster, New York.
D. HENRY
    1981   *Geometric theory of semilinear parabolic equations*, Lecture Notes in Math., vol. 840, Springer-Verlag.
H. W. HETHCOTE
    1973   *Asymptotic behavior in a deterministic epidemic model*, J. Math. Biol. **35**, 607–614.
H. HETHCOTE, A. NOLD AND L. YORKE
    1982   *Gonorrhea modelling: a comparison of control methods*, Math. Biosci. **58**, 93–109.
M. W. HIRSCH
    1983   *Differential equations and convergence almost everywhere in strongly monotone flows*, Contemporary Math., vol. 17, Amer. Math. Soc., Providence, R. I., pp. 267–285.
    1983a   *Systems of differential equations which are competitive or cooperative.* II: *Convergence almost everywhere*, SIAM J. Math. Anal. (in press); preprint, Center for Pure and Appl. Math., Univ. of California, Berkeley.
    1982   *Systems of differential equations which are competitive or cooperative.* I: *Limit sets*, SIAM J. Math. Anal. **13**, 167–179.
S. B. HSU, S. P. HUBBELL AND P. WALTMAN
    1978   *Competing predators*, SIAM J. Appl. Math. **35**, 617–625.
E. KAMKE
    1932   *Zur Theorie der Systeme gewöhnlicher differential-gleichungen.* II, Acta Math. **58**, 57–85.
F. KLEIN
    1911   *Lectures on mathematics* (Proc. Colloq., Evanston, 1893), Amer. Math. Soc., Providence, R. I.
A. KOLMOGOROV
    1936   *Sulla teoria di Volterra della lotta per l'esistenzia*, Giorn. Ist. Ital. Attuari **7**, 47–80.

I. KUPKA
    1963   *Contribution à la théorie des champs génériques*, Contributions to Differential Equations
            2, 457–484.
A. LAJMANOVICH AND J. YORKE
    1976   *A deterministic model for gonorrhea in a nonhomogeneous population*, Math. Biosci. **28**,
            221–236.
W. LEONARD AND R. MAY
    1975   *Nonlinear aspects of competition between species*, SIAM J. Appl. Math. **29**, 243–275.
S. LIE
    1895   *Zur allgemeinen Theorie der partiellen Differentialgleichungen beliebiger Ordnung*, Leip-
            ziger Berichte **47**, 53–128 and *Sophus Lie Gesammelte Abhandlungen* (F. Engel, ed.), Bd.
            4, Teubner, Leipzig, 1929, pp. 320–386.
E. N. LORENZ
    1963   *Deterministic non-periodic flow*, J. Atmospheric Sci. **20**, 130–141.
A. LOTKA
    1956   *Principles of mathematical biology*, Dover, New York.
J. E. MARSDEN AND M. MCCRACKEN (Editors)
    1976   *The Hopf bifurcation and its applications*, Springer-Verlag.
X. MORA
    1983   *Semilinear problems define semiflows in C^k-spaces*, Trans. Amer. Math. Soc. **278**, 21–55.
R. E. MORITZ
    1914   *Memorabilia mathematica*, Macmillan, New York; reprinted as *On mathematics and
            mathematicians*, Dover, New York, 1958.
J. MOSER
    1973   *Stable and random motions in dynamical systems*, Princeton Univ. Press, Princeton, N. J.
S. NEWHOUSE
    1971   *Nondensity of axiom $A(a)$*, Global Analysis, Proc. Sympos. Pure Math, vol. 14 (S.-S.
            Chern and S. Smale, eds.), Amer. Math. Soc., Providence, R. I., pp. 191–203.
    1979   *The abundance of wild hyperbolic sets and non-smooth stable sets for diffeomorphisms*, Inst.
            Hautes Études Sci. Publ. Math. **50**, 101–151.
    1980   *Lectures on dynamical systems*, Dynamical Systems (CIME Summer School, Bressanone,
            1978), Birkhäuser, Boston, pp. 1–114.
H. G. OTHMER
    1976   *The qualitative dynamics of a class of biochemical control circuits*, J. Math. Biol. **3**, 53–78.
J. PALIS AND W. DEMELO
    1982   *Geometric theory of dynamical systems: An introduction*, Springer-Verlag.
J. PALIS AND S. SMALE
    1970   *Structural stability theorems*, Global Analysis, Proc. Sympos. Pure Math., vol. 14 (S.-S.
            Chern and S. Smale, eds.), Amer. Math. Soc., Providence, R. I.
C. S. PEIRCE
    1955   *The philosophical writings of C. S. Peirce*, Dover, New York.
M. PEIXOTO
    1962   *Structural stability on 2-dimensional manifolds*, Topology **1**, 101–120.
    1973   *On the classification of flows on two-manifolds*, Dynamical Systems (M. Peixoto, ed.),
            Academic Press, New York.
H. POINCARÉ
    1881   *Mémoire sur les courbes définies par une équation différentielle*, J. Math. Pures Appl. **7**,
            375–422 and *Oeuvres*, Vol. 1, Gauthier-Villars, Paris.
    1880–1890   *Sur les courbes définies par les équations differentielles*. I–VI, *Oeuvres*, Vol. 1, op. cit.
    1890a   *Sur les équations de la dynamique et le problème de trois corps*, Acta. Math. **13**, 1–270.
    1892, 1893, 1899   *Les méthodes nouvelles de la mécanique céleste*, Vols. I, II, III, Gauthier-Vil-
            lars, Paris.
    1908   Atti IV Congr. Internaz. Mat. (Roma, 1908), Vol. I, Acad. dei Lincei, Rome, 1909.
    1914   *La valeur de la science*, Flammarion, Paris.
    1916–1956   *Oeuvres de Henri Poincaré*, Vols. 1–11, Gauthier-Villars, Paris.

1921 *Analyse des travaux scientifiques de Henri Poincaré faites par lui-même*, Acta Math. **38**, 1–135 and *Oeuvres*, Vol. 1, op. cit.

G. PÓLYA AND S. SZEGÖ
1971 *Problems and theorems in analysis*, Springer-Verlag.

M. PROTTER AND H. WEINBERGER
1967 *Maximum principles in differential equations*, Prentice-Hall, Englewood Cliffs, N. J.

A. RESCIGNO AND I. RICHARDSON
1967 *The struggle for life*. I: *Two species*, Bull. Math. Biophys. **29**, 377–388.

J. ROBINSON
1956 *The accumulation of capital*, Macmillan, London.

D. RUELLE AND F. TAKENS
1971 *On the nature of turbulence*, Comm. Math. Phys. **20**, 167–192; ibid. **23**, 343–344.

M. RUSE
1979 *The Darwinian revolution*, Univ. of Chicago Press, Chicago, Ill.

B. RUSSELL
1948 *Human knowledge, its scope and limits*, Simon and Schuster, New York.
1955 *Basic writings, 1903–1959* (R. Egner and L. Denner, eds.), Simon and Schuster, New York.

J. F. SELGRADE
1979 *Mathematical analysis of a cellular control process with positive feedback*, SIAM J. Appl. Math. **36**, 219–229.
1980 *Asymptotic behavior of solutions to single loop positive feedback systems*, J. Differential Equations **38**, 80–103.

L. SILK
1976 *The economists*, Basic Books, New York.

S. SMALE
1961 *On gradient dynamical systems*, Ann. of Math. (2) **74**, 199–206.
1962 *Dynamical systems and the topological conjugacy problem for diffeomorphisms*, Proc. Internat. Congress Math. (Stockholm, 1962), Inst. Mittag-Leffler, Djursholm, 1963, pp. 440–496.
1963 *A structurally stable differentiable homeomorphism with an infinite number of periodic points*, Proc. Internat. Sympos. Nonlinear Vibrations, Izdat. Akad. Nauk. Ukrain. SSR, Kiev.
1965 *Diffeomorphisms with many periodic points*, Differential and Combinatorial Topology (S. Cairns, ed.), Princeton Univ. Press, Princeton, N. J., pp. 63–80.
1966 *Structurally stable systems are not dense*, Amer. J. Math. **88**, 491–495.
1967 *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73**, 747–817.
1967a *Dynamical systems on n-dimensional manifolds*, Sympos. Differential Equations and Dynamical Systems (Puerto Rico), Academic Press, New York.
1967b *Stability and genericity of dynamical systems*, Sem. Bourbaki, no. 374, 1969–1970 and *The mathematics of time*, Springer-Verlag, 1980.
1976 *On the differential equations of species in competition*, J. Math. Biol. **3**, 5–7.
1980 *The mathematics of time*, Springer-Verlag.

M. SPIVAK
1970 *A comprehensive introduction to differential geometry*, Vol. 1, Publish or Perish, Waltham, Mass.

S. STERNBERG
1969 *Celestial mechanics*, Part 1, Benjamin, Reading, Mass.

J. STRACHEY
1956 *Contemporary capitalism*, Random House, New York.

J. SYLVESTER
1877 Appendix to Address on Commemoration Day at Johns Hopkins University, in *Collected mathematical papers of James Joseph Sylvester*, Vol. 3, Cambridge Univ. Press, Cambridge, England, 1908, pp. 85–87.

G. TEMPLE
    1958   *Linearization and delinearization*, Proc. Internat. Congress Math. (Edinburgh, 1958), Cambridge Univ. Press, Cambridge, England, pp. 233–247.

R. THOM
    1975   *Structural stability and morphogenesis*, Benjamin, Reading, Mass.; original ed., Paris, 1972.

W. WALTER
    1970   *Differential and integral inequalities*, Springer-Verlag.

J. A. WALKER
    1980   *Dynamical systems and evolution equations*, Plenum Press, New York.

H. WEYL
    1939   *The classical groups*: *Their invariants and representations*, Princeton Univ. Press, Princeton, N. J.

R. F. WILLIAMS
    1979   *The structure of Lorenz attractors*, Inst. Hautes Études Sci. Publ. Math. **50**, 101–152.
    1983   *Review of "Dynamical Systems on Surfaces" by C. Godbillon*, Amer. Math. Monthly (in press).

A. WINFREE
    1979   *The geometric theory of biological time*, Springer-Verlag.

J. Z. YOUNG
    1951   *Doubt and certainty in science*, Oxford Univ. Press, New York.

SCHOOL OF MATHEMATICS, INSTITUTE FOR ADVANCED STUDY, PRINCETON, NEW JERSEY 08540