QUEEN'S MATHEMATICAL COMMUNICATOR

SUMMER 1991

An aperiodical issued at Kingston, Ontario by the
Department of Mathematics and Statistics, Queen's University
Kingston, Ontario K7L 3N6

ATTRACTORS OF DYNAMICAL SYSTEMS - AND A FRACTAL ATTRACTOR

BY DAN NORMAN

Dan Norman has been a member of the department since 1965. In recent years, he has been interested in dynamical systems. From 1986 to 1989 he was Chair for Engineering Mathematics. Since 1986, he has chaired the Queen's University Pension Board. This article was originally presented as a Coleman-Ellis Undergraduate Colloquium in February 1991.

The attractors of a dynamical system provide important information about the behaviour of the dynamical system. Indeed, in some cases, the system may be nearly completely analysed in terms of its attractors. However, examples show that some appararently simple systems have very complex attractors. These ideas will be outlined, and an example due to Kaplan, Mallet-Paret and Yorke will be described. The example gives a family of systems such that some members of the family have a smooth torus as the attractor, while others have an attractor which is nowhere differentiable. In fact, it can be seen that the attractor does not have integer dimension - it is a fractal set.

A falling body, a mass bouncing at the end of a spring, and the solar system are familiar examples of dynamical systems. Roughly speaking, a dynamical system is something which can be described as having a "state" at time t and a rule under which the state changes or evolves with time. It will be useful to look at some examples.

<u>The undamped spring.</u>

Consider a mass M attached to a spring with spring constant k; the mass is free to slide horizontally in the direction of the the spring axis on a smooth (frictionless) surface. The other end of the spring is attached to a wall. Displace the mass from its rest position and release it. The state of the system at time t is described by the displacement $x(t)$ of the mass from its rest position. By Newton's law and Hooke's law for springs, the state x satisfies the differential equation (the dynamic for the system):

$$\mathrm{M}\ d^2x/dt \ = -k\ x$$

The general solution to this equation may be written:

$x = \mathrm{A}\ cos\ (\omega t + \phi)$, where $\omega^2 = k/\mathrm{M}$, and the amplitude A and the phase ϕ depend on the initial conditions $x(0)$ and $x'(0)$.

<u>The undamped spring as a first-order system in the plane.</u>

For some purposes it is more convenient to rewrite the spring equation as a pair of first order differential equations. Let $y(t) = dx/dt$; then $dy/dt = d^2x/dt^2 = -(k/\mathrm{M})x$, so that we may write the dynamical rule as

$$\frac{d}{dt}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -(k/\mathrm{M})x \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k/\mathrm{M} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Then the general solution can be written in the form:

$$(x(t), y(t)) = (\mathrm{A}\ cos\ (\omega t+\phi),\ -\omega \mathrm{A}\ sin\ (\omega t+\phi)).$$

We sometimes speak of these solution curves as "orbits" of the dynamical system; for this system the orbits are ellipses in the xy-plane:

$$x^2 + (y^2/\omega^2) = \mathrm{A}^2.$$

This is sometimes called the "phase-plane" picture of the system.

The damped spring.

Suppose that there is a damping force, perhaps due to friction between the mass and the surface on which it rests. Assume that the damping is proportional to the velocity. Then the differential equation is

$$M\, d^2x/dt^2 + \alpha\, dx/dt + k\, x = 0,$$

or, as a first order system, $\frac{d}{dt}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k/M & -\alpha/M \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$

The orbits can be written $(x,y) = Ae^{-pt}\left(cos(qt+\phi),\ -q\ sin\ (qt+\phi^*)\right)$, where p and q are constants depending on M,α, and k, where A and ϕ depend on initial conditions, and ϕ^* depends on ϕ and p and q. Details may be found in many textbooks. Two numerically computed orbits for this system are shown in Figure 1, for the case $k/M=9$, $\alpha/M=0.6$. Notice that the orbits all spiral asymptotically toward the point (0,0). This point is the attractor for this dynamical system. The next example has a more interesting attracting set.

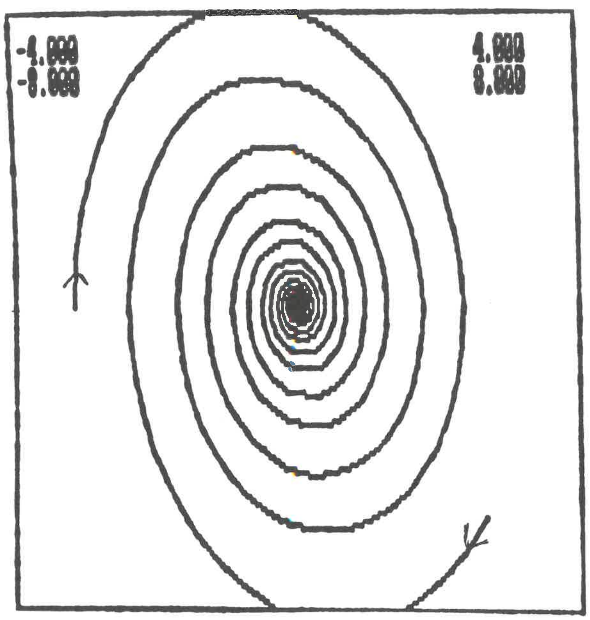


Figure 1. Phase-plane orbits of the damped spring

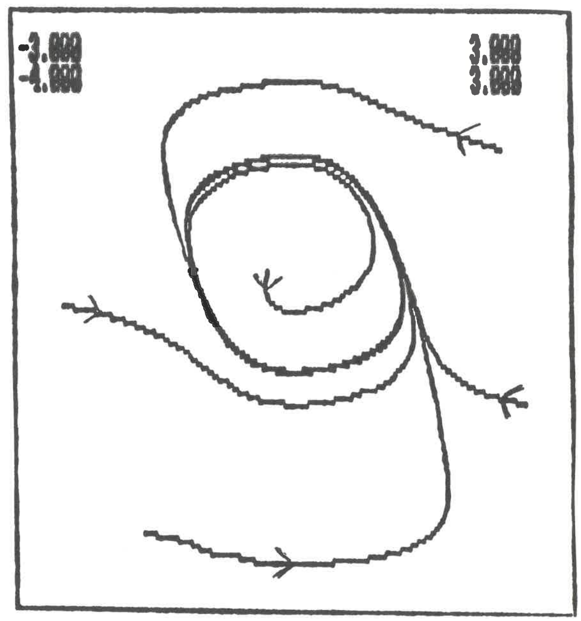


Figure 2. Orbits of the van der Pol equation

The van der Pol oscillator.

Van der Pol introduced this system in 1927 in studying an electrical circuit with a triode valve. We consider the equations in a transformed form, with special choices of the parameters, but other forms, with other choices of constants, would give qualitatively the same picture.

$$dx/dt = -y + x(1-x^2); \qquad dy/dt = x.$$

The orbits cannot be given in closed form, but orbits for various choices of intial condition can be calculated numerically and plotted, as shown in Figure 2. Notice that all orbits "home in" on an "attracting limit cycle", which is a simple smooth closed curve, looking like a distorted ellipse. Orbits starting outside spiral in to this closed orbit, while points inside spiral out. Notice also that in the picture all orbits very quickly become indistinguishable from the attractor. Theory tells us that the solution curves must remain forever distinct, but they get so close to the cycle that they cannot be distinguished numerically. For practical purposes, the dynamics of this system can be thought of as being only the dynamics on the limit cycle. This is the prototypical "non-linear oscillator", and models like this are thought to explain how biological systems can have stable "clocks" regulating their behaviour: even if perturbed, the system will quickly revert to its stable limit cycle, with a well-defined period.

Notice that in this example the state space is two-dimensional, but the attracting set is a smooth one-dimensional curve. In the next example, the attractor is again a point attractor, but the state space is infinite-dimensional.

The heat equation.

Suppose a rod is situated between $x = 0$ and $x = L$. Let $u(x,t)$ be the temperature at x at time t; it will sometimes be convenient to write $u_t(x)$ to denote $u(x,t)$. The function u satisfies the heat equation:

$$\partial u/\partial t = K\ \partial^2 u/\partial x^2$$

with boundary conditions $u_t(0) = u_t(L) = 0$. To determine an orbit of this system, we must specify an initial state $u_0(x) = f(x)$. Notice that the "state space" of this system is the space of continuous functions on [0,L] which vanish at x=0 and x=L: the state space is infinite-dimensional. This problem may be solved by the method of Fourier series to determine the evolution of the state in this infinite-dimensional space of functions:

$$u_{t_0}(x) \longrightarrow u_{t_1}(x).$$

It is a standard result for this system that as $t \longrightarrow \infty$, whatever the initial state, the solution always approaches the "steady state" solution

$$u_{ss}(x) = 0 \text{ for all } x \text{ in } [0,L].$$

This single "point" u_{ss} in the infinite dimensional state space is the attractor for all orbits of this system. The evolution for one particular initial state is shown in Figure 3. (The rate at which $u_t \longrightarrow 0$ depends on the constant K.)

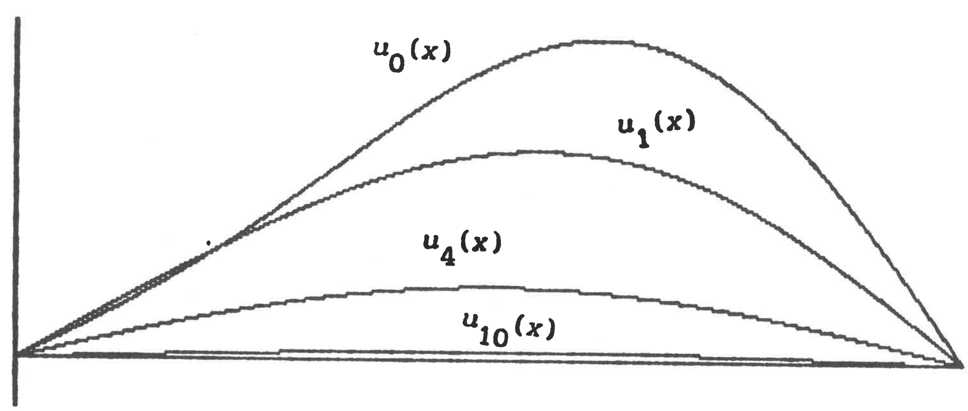


Figure 3. "Points" on an orbit of the heat equation

These examples illustrate that a dynamical system has the property that given the state of the system "now": $(x(0),y(0))$, we can predict the state at a later time T: $(x(T),y(T))$. For the examples so far, it is possible to make general statements about what happens to all orbits as $T \longrightarrow \infty$; in some cases these general statements are possible because we have a general solution, but for cases such as the van der

Pol oscillator the numerical calculations must be supplemented by some general theory. There are more complicated dynamical systems for which scientists believe they fully understand the dynamical laws, but where detailed prediction must be done numerically. In some problems, such as the problem of planetary orbits in the solar system, such numerical prediction may be fantastically accurate for thousands of years - but it does not allow us to answer questions about the long-term behaviour of the system with absolute confidence. The earth's weather is another dynamical system: meteorologists believe that they understand the dynamical laws, but inevitable limitations in obtaining initial data (essentially, the weather conditions at every point in the atmosphere), and the fact that solutions "depend sensitively on initial data" seems to make prediction for more than a few days impossible.

One of the objectives of dynamical systems theory is to try to make general statements about all orbits of each dynamical system. This somewhat grandiose goal has to be tempered: one is often content to be able to say interesting things about "most" or even "many" orbits of some dynamical systems having particular properties.

There are some systems which are governed by partial differential equations which seem to have something like stable limit cycles (some chemical reactions, for example, exhibit periodic behaviour). Encouraged by examples such as the van der Pol oscillator and the heat equation, some researchers speculated, "Maybe for dynamical systems with the "right" properties, even of high (infinite?) dimension, it will turn out that all the important dynamics show up on a smooth low-dimensional attractor." (The "right" properties certainly require that the system be "dissipative", which means that a typical region in state space shrinks in volume as the system evolves in time.) If true, this speculation would be particularly valuable for dealing with infinite- dimensional problems like the heat equation, and more complicated non-linear diffusion problems. The final example of this article shows that unfortunately, attractors can fail to be smooth even for apparently simple systems, so that the speculation cannot capture all the cases we might hope to include.

Discrete dynamical systems.

In the examples we have considered so far time has been a continuous variable. Some problems can be described by assigning a state space U (which could be the real numbers $\mathbb{R}$, or $\mathbb{R}^2$, or some space of functions, for example) and a "rule" (function) f:U$\to$U. The orbit of a point x_0 in U under this "dynamic" is defined to be $\{x_0$, f(x_0), f(f(x_0)), f(f(f(x_0))),...$\}$. We think of this as representing the state at time 0, then after time interval 1, then after time interval 2, and so on.

Such discrete time systems are helpful in understanding continuous time models. Suppose $y(t)$ is the orbit of a continuous time problem such as the damped spring. Consider the points on that orbit only for integer values of t: $\{y(0), y(1), y(2),...\}$. The continuous time dynamic gives rise to a discrete rule $y(n)\to y(n+1)$, which we may call f, so that $f(y(n)) = y(n+1)$. This is called the "time 1 map" corresponding to the continuous system, and it often reveals much of the interesting information about the general system. The remaining examples will all be discrete dynamical systems.

Linear mappings of the real line $\mathbb{R}$.

The only mappings of this kind are of the form $f(x) = kx$, where $x \in \mathbb{R}$, and k is an arbitrary real multiplier. Then a typical orbit is $\{x,kx,k^2x,...\}$, and it is easy to see that the orbit approaches 0 as $n\to\infty$ if and only if $0 < |k| < 1$. In this case we say that the origin is an attractor for the system. If $|k| > 1$, the origin is a repellor for the system. In this and subsequent cases, the possibility that k is negative means that the mapping involves a reflection; it does not affect whether a set is attracting or repelling, so from now on, we take $k > 0$.

Linear mappings of the plane $\mathbb{R}^2$ with distinct real eigenvalues.

By elementary linear algebra, such a mapping $f:\mathbb{R}^2 \to \mathbb{R}^2$ can be represented by a diagonalized matrix mapping:

$$\begin{bmatrix} x \\ y \end{bmatrix} \to \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad \text{so that } f^n(x,y) = (a^n x, b^n y).$$

Case 1. Suppose that $0<a<b<1$. Then it is easy to see that the orbit of any point (x,y) tends to $(0,0)$ as $n \to \infty$, so the origin is the attractor for this dynamical system. Since $a<b$, any orbit will approach the y-axis (where $x=0$) more rapidly than it approaches the x-axis. Draw a few orbits for the case $a=0.1$, $b=0.9$.

Case 2. Suppose that $1<a<b$. Then every orbit is repelled from the origin, with y-coordinates growing faster than x-coordinates as we follow a single orbit. In fact, this is just the inverse of Case 1.

Case 3. Suppose that $a<1<b$. Then as $n \to \infty$, the x-coordinate of $f^n(x,y) = (a^n x, b^n y)$ tends to 0, while the y-coordinate tends to $\pm\infty$. The points of a single orbit will all fall on one branch of a curve resembling a hyperbola, with the coordinate axes as its asymptotes. In this case, the origin is a saddle point for the dynamical system.

An important special case.

Consider the mapping $F:\mathbb{R}^2 \to \mathbb{R}^2$ defined by $\begin{bmatrix} x \\ y \end{bmatrix} \to \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$. From linear algebra, we know that we can understand the mapping in terms of eigenvalues and eigenvectors. For this mapping, the eigenvalues are

$$\lambda_1 = (3-\sqrt{5})/2 \cong 0.382, \text{ and } \lambda_2 = (3+\sqrt{5})/2 \cong 2.618$$

with corresponding eigenvectors

$$\underline{v}_1 = (2,-(1+\sqrt{5})) \cong (2,-3.24), \text{ and } \underline{v}_2 = (2,-1+\sqrt{5}) \cong (2,1.24).$$

For later reference, note that $\underline{v}_1$ determines a line in $\mathbb{R}^2$ with irrational slope $-(1+\sqrt{5})/2$.

If we introduce a new coordinate system, with the axes determined by the two eigenvectors $\underline{v}_1$ and $\underline{v}_2$, we see that this linear mapping F is a special case of Case 3 above, because $0<\lambda_1<1<\lambda_2$. Thus (0,0) is a saddle point for the mapping F.

The mapping F has two important special properties. First, because the entries in the matrix are all integers, it follows that if the coordinates of (x,y) are integers, then the coordinates of $F(x,y)$ are also integers. We say that "F maps the integer lattice to the integer lattice". Second, the determinant of the matrix of F is easily calculated to be 1. It follows by standard linear algebra that F is an "area-preserving" map: it maps a square of area 1 to a parallelogram of area 1. Also, $\lambda_1\lambda_2 = 1$.

The torus $\mathbb{T}^2$ obtained by identification from the plane $\mathbb{R}^2$.

Recall that x_1 is congruent to x_2, modulo 1 if the difference $x_1 - x_2$ is an integer multiple of 1; we write $x_1 \equiv x_2 \pmod 1$. For example, $4.35 \equiv 0.35 \pmod 1$. It is easy to see that this is an equivalence relation on the real numbers ($x_1 \equiv x_1$; if $x_1 \equiv x_2$ then $x_2 \equiv x_1$; if $x_1 \equiv x_2$ and $x_2 \equiv x_3$, then $x_1 \equiv x_3 \pmod 1$). We can put the set of real numbers into congruence classes, modulo 1: for example, the congruence class of 0.35, denoted [0.35] consists of {...,-1.65,-0.65,0.35,1.35,...}. We speak of any one of these numbers as a "representative" of the class [0.35] We often think of "identifying" all the points in any one equivalence class, and then the real line is "rolled up" into a circle, with 0 coinciding with 1, with 2, and so on. It is important to note that we can do the usual arithmetic with congruence classes: for example, $[x]+[y]\equiv[x+y]$.

In terms of this congruence, we define an equivalence relation on the points of the plane $\mathbb{R}^2$. Let $(x,y)\sim(z,w)$ if $x \equiv z \pmod 1$ and $y \equiv w \pmod 1$. Then identify points which are equivalent under this relation, so, for example, identify (0.35,0.67) with (3.35,5.67). This "rolls up" the plane in the x-direction, and the y-direction, and the result is called the (2-)torus $\mathbb{T}^2$. To visualize it, first roll up the plane in the x-direction: this gives rise to a cylinder, with an equivalence class of vertical lines in the plane corresponding to one vertical line on the cylinder. Now make imaginary cuts at $y=0$ and $y=1$, giving circular cross-sections. These two circles are to be identified, giving rise to the torus, which looks like the surface of an old-fashioned donut. There are of course many points in $\mathbb{R}^2$ corresponding to each point in $\mathbb{T}^2$. It is convenient to think of the points in the half-closed, half-open square $0\leq x<1$, $0\leq y<1$ in the plane as being the "fundamental representatives" of the points on the torus. With this understanding, we may treat the coordinates (0.3,0.56), for example, as the coordinates of a point on the torus.

Anosov diffeomorphisms (hyperbolic toral automorphisms).

Now recall the linear map F from $\mathbb{R}^2$ to $\mathbb{R}^2$ with matrix $\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}$. Since the matrix has integer entries, we can define a mapping of $f:\mathbb{T}^2\to\mathbb{T}^2$ by letting $f([x],[y]) = \begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}$ (coordinates modulo 1). For example,

$$f([2/3],[2/3]) = \begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}2/3\\2/3\end{bmatrix} = \begin{bmatrix}6/3\\4/3\end{bmatrix} \sim \begin{bmatrix}0\\1/3\end{bmatrix} \text{ (modulo 1)}$$

so f maps the point (2/3,2/3) on the torus to the point (0,1/3) on the torus. We should check that the mapping is properly defined on the torus: the image of a point on the torus should not depend on the representative chosen to represent the point on the torus. In our example, it is easy to see that if (2/3,2/3) is represented by (5/3,8/3), for example, it would be mapped to $(18/3,13/3)\sim(0,1/3)$, so that the image point is the same. In fact, the image point depends only on the point on the torus, and not on which representative we use, because the matrix has integer entries.

Because the matrix $\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}$ has determinant 1, the mapping $F:\mathbb{R}^2\to\mathbb{R}^2$ is an area preserving linear map. It follows that the mapping f induced on the torus is an invertible mapping: f^{-1} must exist: it can be obtained from the inverse matrix for F, but we do not need an explicit description.

We thus have an invertible mapping f defined on the torus $\mathbb{T}^2$, which is called a hyperbolic toral automorphism (or Anosov diffeomorphism). It is easy to see that any 2 by 2 matrix with integer entries, determinant 1, and distinct real eigenvalues would give a mapping with the same properties. We now consider f as a discrete dynamical system with $\mathbb{T}^2$ as state space. We denote a point on the torus by the coordinates of its fundamental representative. Some properties follow easily.

(1) (0,0) is a fixed point of f (obvious), in fact it is the only fixed point of f. It is a saddle point for the system.

(2) Any point with rational coordinates is a periodic point of f. For example, the orbit of (2/3,2/3) is

$$(2/3,2/3)\to f(2/3,2/3)=(0,1/3)\to f(0,1/3)=(1/3,1/3)\to f(1/3,1/3)=(0,2/3)$$
$$\to f(0,2/3)=(2/3,2/3).$$

Since this required four iterations of f, (2/3,2/3) is a point of period 4.

Generally, if a point on the torus has rational coordinates, by bringing the coordinates to a common denominator and recalling that we may always represent a point by its fundamental representative, we may assume that it is of the form $(p/q, r/q)$ where $0 \leq p < q$ and $0 \leq r < q$. If we calculate the orbit of this point, images of every order will all have denominator q and numerators between 0 and q. Since there are only finitely many such possible image points, the orbit must eventually start repeating, and therefore every rational point belongs to a periodic orbit. Since there are infinitely many possible denominators for rational points, there are infinitely many periodic orbits.

(3) Every point on the torus is the limit of a sequence of points with rational coordinates, so the periodic points of f are dense in the torus.

(4) Consider the eigenvector $\underline{v}_2$ of the mapping F of the plane. $F(\underline{v}_2)=\lambda_2\underline{v}_2$, $F(F(\underline{v}_2))=\lambda_2^2\underline{v}_2$, and so on. These images of $\underline{v}_2$ under iterations of F all lie on a line in the plane through (0,0) which has irrational slope (because of the components of $\underline{v}_2$). The set on the torus corresponding to this line in the plane is a "line" on the torus; in the fundamental region of the plane, this "line" is represented by an infinite number of line segments all with the same slope. The "line" never closes in $\mathbb{T}^2$. (Contrast a line such as $2y=3x$ in the plane which gives a curve on the torus which does "close" on itself, since (2,3) is on the line, and $(2,3)\sim(0,0)$.) One can show that the sequence of points $\{f^n(\underline{v}_2)\}$ gets arbitrarily close to any point in the torus infinitely often; this is a dense orbit.

(5) It follows from above that the smallest set in $\mathbb{T}^2$ which can be thought of as an attractor for the dynamical system f is the entire torus. This is already a very complex dynamical system.

The example of Kaplan, Mallet-Paret and Yorke
- a nowhere differentiable torus as an attractor.

Let $\mathbb{M}$ denote the set $\{(x,y,z)\colon (x,y)\in\mathbb{T}^2,\ z\in\mathbb{R}\} = \mathbb{T}^2\times\mathbb{R}$. You may find it helpful to think of this as the part of $\mathbb{R}^3$ sitting above and below the fundamental region in the xy-plane corresponding to the torus. Introduce a dynamical system g on this state space: let $g(x,y,z) = (f(x,y), \mu z + p(x,y))$, where f is the hyperbolic toral automorphism above, where $0<\mu<1$, and where $p(x,y)$ is differentiable and periodic in x and in y with period 1, so that p is properly defined as a function on $\mathbb{T}^2$. (For example, you might take $p(x,y) = \sin 4\pi x \cos 6\pi y$.)

We find the attracting set for this dynamical system on $\mathbb{M}$ by the method of the graph transform. Consider the graph of any smooth function over $\mathbb{T}^2$, say $\phi(x,y)$. Let $\mathbb{N}_0 = \{(x,y,z)\colon z = \phi(x,y)\}$, so that $\mathbb{N}_0$ is the set of points making up the graph of $\phi\colon\mathbb{T}^2\to\mathbb{R}$. We can visualize this as a graph in 3-space sitting over the fundamental region in the xy-plane. We now consider the image of this set of points under the dynamic g; let $\mathbb{N}_1$ be the image of $\mathbb{N}_0$:

$$\mathbb{N}_1 = g(\mathbb{N}_0) = \{g(x,y,z)\colon (x,y,z)\in\mathbb{N}_0\} = \{(f(x,y),\ \mu\phi(x,y) + p(x,y))\}.$$

Since f is an invertible (one-to-one) mapping of the torus, we see that $\mathbb{N}_1$ is again a graph over the torus (visualized again as the fundamental region).

Repeat:

$$\mathbb{N}_2 = g(\mathbb{N}_1) = \{g\ (f(x,y),\ \mu\phi(x,y) + p(x,y)\ )\}$$

$= \{ (f(f(x,y)), \mu(\mu\phi(x,y) + p(x,y)) + p(f(x,y))\}$

$= \{ (f^2(x,y), \mu^2\phi(x,y) + \mu p(x,y) + p(f(x,y))\}.$

By repeating k times, we get

$$\mathbb{N}_k = g(\mathbb{N}_{k-1}) = \{ (f^k(x,y), \mu^k\phi(x,y) + \mu^{k-1}p(x,y) + \mu^{k-2}p(f(x,y)) + \ldots$$
$$\ldots + \mu p(f^{k-2}(x,y)) + p(f^{k-1}(x,y))) \}$$
$$= \{ (f^k(x,y), \mu^k\phi(x,y) + \sum_{j=0}^{k-1} \mu^j p(f^{k-1-j}(x,y))) \}.$$

Now we want to know what happens as $k \to \infty$.

(1) Since $0<\mu<1$, $\mu^k \to 0$, so that $\mathbb{N}_\infty = \lim_{k\to\infty} \mathbb{N}_k$ is independent of the function ϕ: every function ϕ leads to the same limiting set. It follows that $\mathbb{N}_\infty$ is the <u>attractor</u> for our dynamical system.

(2) Let $(u,v) = f^k(x,y)$; note that $f^{k-1-j}(x,y) = f^{-1-j}(u,v)$. Then

$$\mathbb{N}_k = \{ (u,v, (*) + \sum_{j=0}^{k-1} \mu^j p(f^{-1-j}(u,v))) \}, \text{ where } (*) \to 0 \text{ as } k\to\infty.$$

Let $\psi(u,v) = \lim_{k\to\infty} \sum_{j=0}^{k-1} \mu^j p(f^{-1-j}(u,v)) = \sum_{j=0}^{\infty} \mu^j p(f^{-1-j}(u,v))$. Since p is differentiable on $\mathbb{T}^2$, it is certainly continuous and bounded on the closed unit square $0\le x,y \le 1$. Since $0<\mu<1$, it follows by a standard theorem (the "Weierstrass M-test") that the series for $\psi(u,v)$ converges uniformly and the sum $\psi(u,v)$ is continuous on $\mathbb{T}^2$. Thus the attractor $\mathbb{N}_\infty$ is the graph of the continuous function $\psi(u,v)$ defined on $\mathbb{T}^2$. This means that the attractor is itself at least approximately like the torus - it is homeomorphic to the torus.

(3) Is the attractor a smooth (differentiable) set, like our usual version of $\mathbb{T}^2$? To answer this we need to be able to decide whether $\psi(u,v)$ is a differentiable function. Note that the terms in the sum for ψ involve the function f^{-1-j}. Since the exponents here are negative, we will have most difficulty trying to differentiate in the direction of $\underline{v}_1$, the eigenvector corresponding to $\lambda_1<1$. (To see why this should be so, note that *sin kt* is "bumpier" than *sin* (t/k) for $k>1$, and recall that $F^{-1-j}(\underline{v}_1) = (\lambda_1)^{-1-j}\underline{v}_1$ by standard linear algebra.)

Let $(u,v) = t\underline{v}_1$, and consider

$$\psi(t\underline{v}_1) = \sum_{j=0}^{\infty} \mu^j p\left(\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}^{-1-j} t\underline{v}_1\right) = \sum_{j=0}^{\infty} \mu^j p(\lambda_1^{-1-j} t\underline{v}_1).$$

Since $\underline{v}_1$ is fixed, let $q(s)=p(s\underline{v}_1)$, and then $\psi(t\underline{v}_1) = \sum_{j=0}^{\infty} \mu^j q(\lambda_1^{-1-j}t)$. Now try to differentiate term-by-term, noting the chain rule:

$$\frac{d}{dt}\psi(t\underline{v}_1) = \sum_{j=0}^{\infty} \mu^j \lambda_1^{-1-j} q'(\lambda_1^{-1-j}t) = \sum_{j=0}^{\infty} \left(\frac{\mu}{\lambda_1}\right)^k \lambda_1^{-1} q'(\lambda_1^{-1-j}t).$$

Such term-by-term differentiation is justified provided the series <u>after</u>

differentiation is uniformly convergent - and the resulting series in this case is uniformly convergent if $\mu < \lambda_1$. Thus ψ can be shown to be differentiable if the factor μ which represents contraction onto the attracting torus is smaller (more strongly contracting) than the eigenvalue λ_1 which describes contraction within the torus.

On the other hand, if $\mu > \lambda_1$, the series does not converge, and in general the torus is nowhere differentiable. However, the theorem we have just used is a "one-way" theorem: if the series does not converge, the theorem says nothing about differentiability. To be sure, one would have to examine the function q, which is "almost periodic". A considerable part of the paper of Kaplan, Mallet-Paret and Yorke is concerned with the extraordinarily restrictive conditions on $p(x,y)$ under which the torus is differentiable even when $\mu > \lambda_1$.

If ψ is differentiable, then $\mathbb{N}_\infty$ is a smooth surface, and it must have dimension 2. We shall see that the dimension is greater than 2 in the nowhere differentiable case where $\mu > \lambda_1$.

Fractal Dimension.

Our usual notions about dimension are based on experience with simple familiar sets in $\mathbb{R}^3$. For example, a smooth curve is one-dimensional, a smooth surface is two-dimensional while a solid cube is 3-dimensional. One way to try to capture these ideas is to ask: how many open balls (solid spheres) of radius ε are required to cover the set? Call the minimum number of balls $N(\varepsilon)$. Simple pictures suggest that for a smooth curve, $N(\varepsilon)$ is roughly proportional to $(1/\varepsilon)$: write $N(\varepsilon) \tilde{} (1/\varepsilon)^1$. For a smooth surface, $N(\varepsilon) \tilde{} (1/\varepsilon)^2$, while for a simple cube, $N(\varepsilon) \tilde{} (1/\varepsilon)^3$. It appears that the exponent in these expressions gives the dimension, so for any set S in $\mathbb{R}^3$, let $N(\varepsilon)$ be the minimum number of open balls of radius ε required to cover S, and define

$$d(S) = \lim_{\varepsilon \to 0} \frac{\log (N(\varepsilon))}{\log (1/\varepsilon)}.$$

This is the "Kolmogorov capacity" of S, sometimes called the "fractal dimension" of S. For curves, surfaces, and solids, it does give the expected dimensions, but there are sets which have non-integer dimension, and they are sometimes called "fractal sets". I digress to describe the simplest and most famous such set.

The Cantor middle third set.

Let C_1 be the set obtained by deleting from the interval [0,1] its open middle third (1/3,2/3). Let C_2 be the set obtained from C_1 by deleting the middle third of each of the pieces of C_1; that is, delete (1/9,2/9) and (7/9,8/9). Continue in this fashion, obtaining $C_2, C_3, \ldots, C_n \ldots$. The Cantor set C is the limit of this sequence of sets. The Cantor set is the standard example of a "totally disconnected perfect" set; here "perfect" means that the set has no isolated points. Note also following properties.

(1) C has "measure zero": we started with an interval of length 1 and subtracted an infinite set of intervals of total length 1, since

$1/3 + 2/9 + 4/27 + \ldots + 2^{k-1}/3^k + \ldots = 1$. Nevertheless, by a fairly simple argument it can be shown that the points of C are in one-to-one correspondence with the points of the interval [0,1].

(2) To determine the fractal dimension of C, use balls of radius $\varepsilon = (1/3)^k$; we need

2^{k+1} balls to cover C (since we must cover all end-points of subintervals of C_k). It follows that $d(C) = \lim_{k\to\infty} \frac{\log 2^{k+1}}{\log 3^k} = \frac{\log 2}{\log 3} \cong 0.63$. so the Cantor set has non-integral dimension.

<u>The Lyapunov dimension of the attracting torus.</u>

It is not always easy to construct the coverings required to compute the Kolmogorov capacity, so other definitions of dimension have been introduced. These give the same dimension in many reasonable(?) cases, but give differing results for some examples. The <u>Lyapunov dimension</u> can be calculated for the attractor of some dynamical systems and Kaplan, Mallet- Paret and Yorke show that it gives the same result as the Kolmogorov capacity for the nowhere differentiable torus.

Suppose that g is a differentiable mapping from a 3-dimensional state space to itself. Let us suppose (to over-simplify) that its linear approximation, the differential dg, is a diagonalized matrix of the form

$\begin{bmatrix} L_1 & 0 & 0 \\ 0 & L_2 & 0 \\ 0 & 0 & L_3 \end{bmatrix}$, where $L_1 \geq L_2 \geq L_3 > 0$, and $L_1L_2L_3 < 1$. The fact that the product is less than 1 means that the mapping shrinks volumes, so that it is reasonable to expect the system to have an attracting set. The L_i are the "Lyapunov numbers" of the system; in general, these have to be defined rather more elaborately. In our simplified case the matrix is diagonal, so g stretches lengths in the first coordinate direction by factor L_1, in the second coordinate direction by L_2, and in the third coordinate direction by L_3. Thus the greatest rate of growth of lengths under the mapping g is L_1. By considering two-dimensional subspaces, we see that the greatest rate of growth of area is L_1L_2, and of volume, $L_1L_2L_3$.

Now consider a candidate for the attracting set of g. It must neither grow nor shrink overall under the action of g: if it grows, it is too small to be the attractor, if it shrinks it is too big. This is a heuristic introduction to the (partial) definition of Lyapunov dimension:

if $L_1=1$, $L_1L_2<1$, $\dim_L$(attractor of g) = 1;

if $L_1L_2=1$, $L_1L_2L_3<1$, $\dim_L$(attractor of g) = 2;

if $L_1L_2>1$, $L_1L_2L_3<1$, $\dim_L$(attractor of g) = $2 + \frac{\log (L_1L_2)}{(-\log L_3)}$.

(This last line is not exactly obvious, but because of the inequalities, it is easy to see that in this case $2 < \dim_L < 3$.)

To apply this to the example of the attracting torus, we use the fact that the Lyapunov numbers are the eigenvalues λ_1, λ_2 and the factor μ which causes the contraction onto the attractor. Note that all are positive, and that since $\lambda_1\lambda_2=1$ and $\mu<1$, they have product less than 1 as required.

<u>Case 1</u>. $\mu<\lambda_1<1<\lambda_2$. Then $L_1=\lambda_2$, $L_2=\lambda_1$, $L_3=\mu$, so $L_1L_2=1$, $L_1L_2L_3<1$, and the dimension of the attractor is 2. This is the case where the attracting torus is differentiable.

<u>Case 2</u>. $\lambda_1<\mu<1<\lambda_2$. Then $L_1=\lambda_2$, $L_2=\mu$, $L_3=\lambda_1$, so $L_1L_2>1$, $L_1L_2L_3<1$, and the dimension of the attractor is between 2 and 3. The attractor cannot be a

differentiable surface in this case. (Note that this calculation of Lyapunov dimension does not give the correct dimension of the attracting torus in the truly exceptional cases where $\lambda_1<\mu$, yet the torus is nevertheless differentiable.)

It is quite striking that this change size of μ relative to the eigenvalue λ_1 can change the attractor from "nice" to "nasty". John Mallet-Paret likened it to ironing a starched shirt: the starch causes some stickiness which causes the shirt to "contract" (that is, wrinkle) along the surface unless the force flattening the shirt is stronger (that is, unless μ is small enough so that the contraction onto the surface is faster than the contraction along the surface).

Bibliography

An excellent readable popular book about dynamical systems is *Chaos* by James Gleick. A book written for undergraduate mathematicians which describes one-dimensional dynamical systems quite thoroughly and also introduces the byperbolic toral automorphisms and other interesting dynamical systems is*An Introduction to Chaotic Dynamical Systems* by Robert L. Devaney, published by Benjamin/Cummings. The paper by J.L. Kaplan, J. Mallet-Paret, and J.A.Yorke is entitled *The Lyapunov dimension of a nowhere differentiable attracting torus;* it appeared in **Ergodic Theory and Dynamical Systems, volume 4 (1984), pp.261-281.**

The figures were produced using a PC and a dot matrix printer. Figures 1 and 2 were generated with *PhasePlane* by B. Ermentrout, published by Brooks/Cole. Figure 3 was produced with *Calculus Pad* by Ian Bell,Jon Davis, and Steve Rice (at Queen's), also distributed by Brooks/Cole.

NEWS

Prof. A. V. Geramita was recently elected Vice-President of the Canadian Mathematical Society.

Prof. A. M. Herzberg becomes President of the Statistical Society of Canada July 1st.

Prof. J. A. Whitley recently won the Frank Knox Award for Excellence in Teaching.

Profs. R. Giles, K. Oberai, N. Pullman and **W. Woodside** have completed 25 years of service to the Department.

Profs. L. Broekhoven, H. A. Still, R. C. Willmott and **E. J. Woods** have retired.

Dr. R. M. Erdahl has been promoted to full professor.

Drs. D. de Caen, E. Kani and **M. A. Maes** have been promoted to associate professor.

NEWS OF MTHE GRADUATES

Mike Lazier ('76) is a Mathematics and Physics teacher at the new Holy Cross School in Kingston Township.

Stewart Crozier ('80) has completed his Ph.D. in Electrical Engineering at Carleton University and is working at the Communications Research Centre in Ottawa.

Jamie McLellan ('81) is an Assistant Professor in the Dept. of Chemical Engineering at Queen's.

Marilyn Lightstone ('85) and **John Mackinnon** ('84) were married at Queen's last August. Both are nearing completion of their Ph.D.'s in Fluid Dynamics at the University of Waterloo. Among the guests at their wedding were former classmates Daood Al-Aidroos, Charles Arnoldi, Ian Bell, Peter Giardetti, Bob Pronk and Karen Rudie.

Mark Green ('87) recently completed his Ph.D. in Structural Engineering at Cambridge University and is now a Post-doctoral Fellow in the Dept. of Civil Engineering at Queen's.

MATHEMATICS AND ENGINEERING SEMINAR

For the past two years, fourth year Math & Engineering students have been required to attend a weekly seminar in the Fall Term. The speakers are mostly from outside the university - from business, industry or government laboratories, and they speak about various aspects of engineering careers. Topics have included: the professional responsibilities of an engineer; how to succeed as an entrepreneur; an application of statistics in process control; control of rolling mills; design of electrical generators; a consulting problem in civil engineering. Students are required to write a brief report on each talk, so that the course serves to reinforce writing skills as well as providing insight into a variety of engineering careers. We thought you might enjoy reading one of this year's reports. It is based on a seminar presented by Gillian Woodruff, who graduated from Mathematics and Engineering at Queen's in 1980.

An Invitation to Mathematics and Engineering Graduates

Students naturally find talks from Mathematics and Engineering graduates particularly interesting. So far, in addition to Gillian, Bob Lyons (Sc.'67), John Dorland (Sc.'68), Vijay Bhargava (Sc.'70), John Redding (Sc.'73), Mike Dick (Sc.'77), Don McLaren (Sc.'80), Ross Ethier (Sc.'80), James McLellan (Sc.'81), and Amanda Hubbard (Sc.'83) have come to present seminars. We invite all Mathematics and Engineering graduates to consider contributing in this way to the education of your successors. Write to Dr. Dan Norman, Department of Mathematics and Statistics.

Performance Modelling for the Telecommunications Industry

by Brad McFarlane

A prediction for the next few decades is that everyone will be connected to a vast network, linking personal computers, phones, cable TV, and just about any other data service. Usually omitted in the prediction is how the network is going to work.

On Friday, November 16, 1990, Ms. Gillian Woodruff, a manager at Bell-Northern Research currently on leave to do a Ph.D. in Electrical Engineering, spoke to a class of Queen's Mathematics & Engineering students about the way the phone system currently works, and how it will be made better.

When a normal call is made, a complex series of messages is sent back and forth along branch lines, through switches and along trunk (main) lines, both to request channels and to indicate the phone has been picked up, or that the phone should ring. This message exchange (which happens before you can talk to the other end) causes two important delays: dial-tone delay, which is the time between picking up the phone and when the dial tone sounds (which is usually negligible; however, on some older systems there is a noticeable delay); and post-dialing delay, which is the time between when the last digit is entered and when the phone on the other end rings (most noticeable on long-distance calls).

In order to minimize the delays (while at the same time not overspending on equipment), the phone company models the system. The simulation involves injecting messages into the model, and queueing them up if the server which is to handle them is busy. The time a message spends in the queue and the time it takes to be processed are measured and analyzed. The difficult part of the model is the assumptions made about incoming messages: whether arrivals are independent (which they may not be, as some messages come in bunches), whether messages are processed first-come first-serve or whether there are priorities, and the statistical model used for arrival times. Things are further complicated when a network is modelled, as the output from one server becomes the input to another. Models are also used to investigate call blocking, which can happen when a line is needed but none is available.

In order to gauge traffic, Bell used to use an old rule of thumb, which involved measuring traffic during a predetermined "busy month", averaging over the season and time of day, and applying some fudging. The new estimation method ("extreme value engineering") takes the average daily peak traffic and its standard deviation, giving much more reliable and standard data.

Ms. Woodruff is using these models to figure out things like how big message buffers must be and how many trunk lines, switching modules, etc. have to be put in place now to meet the traffic needs in the future.

Future needs are definitely going to increase. Currently, there are dedicated separate lines and switching systems for voice and digital (computer) data. The Integrated Services Digital Network (ISDN), projected to be the next revolution in communications, combines these, allowing you to plug a computer into a standard home jack to transfer data over normal phone lines without using a modem. Further in the future, Broadband ISDN will allow sending things such as full motion video over phone lines. These new developments are more flexible, faster, and more integrated than the present service. However, they mean more different kinds of data are going over it, making for a more complex and vulnerable system (more connections carrying more data means more chances for broken connections, and more complex means of figuring out how to survive a break).

That's the challenge for the future: defining new rules for how robust, complex, flexible, and survivable the system will be, while providing the service the customers want.

OLD PROBLEMS

SOLUTIONS TO MARTIN KREUZER'S PROBLEMS - BY THE PROPOSER

PROBLEM 1: If every point in the plane has one of n colours, show that there is a rectangle with vertices all of the same colour.

SOLUTION: Draw a line ℓ and consider $n^2 + 1$ of its points. At least $n + 1$ of them have the same colour c. Starting from those $n + 1$ points draw perpendicular lines $\ell_1,\ldots,\ell_{n+1}$. Also draw parallels $p_1,p_2,\ldots,$ of ℓ. If two or more intersection points of some p_i with $\ell_1,\ldots,\ell_{n+1}$ have the colour c, then they form together with the corresponding points of ℓ a rectangle with c-coloured vertices. Otherwise, if on each parallel p_i at most one of the intersection points with $\ell_1,\ldots,\ell_{n+1}$ has colour c, a different colour has to occur at least twice.

Those two equicoloured intersection points can be any of $\binom{n+1}{2} = \frac{1}{2}n(n+1)$ pairs and any of $n - 1$ colours. Hence, if we draw $N = \frac{1}{2}(n-1)n(n+1) + 1$ parallels $p_1,\ldots,p_N$, one pair of non-c-coloured intersection points with $\ell_1,\ldots,\ell_{n+1}$ has to be repeated, and those two pairs form the set of vertices of an equicoloured rectangle.

QED

PROBLEM 2: If every point of the plane is coloured either red or blue, show that there is a square with equicoloured vertices.

SOLUTION: We proceed in two steps. First we show that there is a line which contains three equidistant points of the same colour.

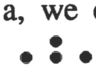
Draw any line and mark three points. At least two of them have the same colour, say red. Label them 0 and 1 and use this to identify the points of the line with the real numbers. Suppose the claim is false. Then the point 2 has to be blue, because otherwise {0,1,2} is an equidistant equicoloured set of points. In the same manner we conclude that -1 is blue, 5 is red, -4 is red, 3 is blue and 4 is red. But then the set {-4,0,4} contradicts our assumption. Therefore the claim is true.

In the second step we label those three equidistant, equicoloured points (0,1), (0,0) and (0,-1) and identify the plane with $\mathbb{R}^2$. W.ℓ.o.g. their colour is red. We suppose the claim of the problem is false, i.e. there is no square with equicoloured vertices. For the points (-1,0) and (1,0) there are two cases.

CASE a: Suppose that one of (-1,0) and (1,0) is blue and the other red. By symmetry, we can assume that (-1,0) is blue and (1,0) is red. We conclude that (1,1) and (1,-1) are both blue. For the pair {(2,1),(2,-1)} there are two subcases.

SUBCASE a1: Either (2,1) or (2,-1) is blue, and the other one is red. W.l.o.g. let (2,-1) be blue and (2,1) be red. Then also (0,-2) has to be red. Now, if (3,-1) is red, then (1,-2) must be blue, (2,-2) red, (2,-3) blue, (2,0) blue, (-1,-3) red, (1,-4) blue, (0,-3) red, (-1,-2) blue, (-1,-4) red, and (0,-4) gives us a contradiction. Similarly if (3,-1) is blue, then (3,1) has to be red, (3,-2) blue, (2,-2) red, (4,-1) blue, (4,-2) red, (1,-3) blue, (2,2) red, (3,2) blue, and (4,0) yields a contradiction.

SUBCASE a2: Suppose that (2,1) and (2,-1) are both blue. One of (-1,1) and (-1,-1) has to be red, so assume w.l.o.g. that (-1,1) is red. Then (0,2) has to be red, (0,-2) red, (-2,-1) blue, and (-1,2) blue. Now, if (-2,-2) is red, then (-2,0) has to be blue, (-3,0) blue, (-3,2) red, and (-3,-1) yields a contradiction. But if (-2,-2) is blue, then (0,-3) has to be red, (1,-2) red, (1,-3) blue, (-1,-2) blue, (-1,-1) red, (2,0) blue, (2,-2) blue, (3,-1) red, (2,-3) blue, (0,-4) red, and (-2,-3) gives us a contradiction.

CASE b: Suppose that both (1,0) and (-1,0) are blue. Since we are already done with case a, we can also assume that there is no constellation of equicoloured points of type  . Again we look at (0,2) and have two subcases.

SUBCASE b1: If (0,2) is red, then (1,1) and (-1,1) must both be blue, because otherwise we are in the situation of case a. Then one of (1,2) and (-1,2) is red and the other blue. Suppose that (1,2) is red and (-1,2) is blue. Then (-2,1) has to be red, (-2,-1) blue, (0,-2) red, (1,-1) blue, and (-1,-1) yields a contradiction.

SUBCASE b2: If (0,2) is blue, then at least one of (1,-1) and (-1,-1) is blue. Assume that (1,-1) is blue. Then (2,1) has to be red, (2,-1) blue, (2,0) red, (1,-2) red, (-1,-1) blue, and (-2,1) yields a contradiction.

In every case the assumption that there is no equicoloured square leads to a contradiction.

QED

NOTE: These two problems are closely related to the game of HIP described in Martin Gardner's "New Mathematical Diversions from Scientific American", Simon and Schuster, New York 1966. The game is played on a 6×6 checkerboard by two players. One player has 18 red counters; his (her) opponent has 18 blue counters. They take turns placing a counter on any vacant cell. Each tries to avoid placing his counters so that four of them mark the corners of a square. The square may be any size and tipped at any angle. A player wins when his opponent becomes a 'square' by forming one of the 105 possible squares.

NEW PROBLEMS

THE DARTBOARD PROBLEM I (proposed by M. A. Maes and W. Woodside)

Find the probability that n darts thrown at random at a circular dartboard leave at least half (any half) of the board dart-free. (Assume all n darts strike the board.)

DARTBOARD PROBLEM II

Find the probability that n darts thrown at random at the surface of a sphere leave at least half (any half) of the surface dart-free.

DR. L. B. JONKER - NEW HEAD OF MATHEMATICS AND STATISTICS

Dr. Leo Jonker has been appointed head of the department for a five-year term beginning July 1, 1990. He received a B.Sc. in mathematics in 1963 and a Ph.D. in 1967 from the University of Toronto. He was a post-doctoral fellow at the University of California before coming to Queen's in 1969 as an assistant professor. He was promoted to associate professor in 1977 and to full professor in 1985.

Dr. Jonker started his research career in differential geometry. In his Ph.D. thesis he attempted a classification of natural transformations betwen vector spaces of tensor fields over a differential manifold. Later work dealt with the global geometry of submanifolds of Euclidean space or Hilbert space implied by local assumptions on the curvature tensor.

After a sabbatical year in England at the University of Warwick he began work in dynamical systems. This started, in collaboration with Prof. David Rand of Warwick, on the dynamical structure of unimodal maps of the unit interval. It turns out that iterations of functions as simple as $f(x) = \lambda x(1-x)$ exhibit very interesting chaotic dynamics which vary enormously with the parameter λ. This family of functions is an excellent source of examples of chaotic behaviour of the kind studied and applied widely today.

His interest in one-dimensional dynamical systems continues, currently centering on the dynamics of circle maps and relating such things as the rotation number, the degree of degeneracy of critical points, and the Hausdorf dimension of the resulting fractal sets.

'My work as head poses a new and formidable challenge, calling on talents I did not know I had as well as some I know I do not have. My teaching has had to be reduced to a minimum, and so far the administrative duties have left little or no time for research'.

While his other hobbies of reading and windsurfing have had to be curtailed, he maintains his daily routine swim 'provided it is not preempted by a business luncheon'.

STRUCTURAL ASPECTS OF CHESS PROBLEMS
BY ANDREW KALOTAY

Andrew Kalotay (Arts'64, M.Sc.'66, Ph.D. Toronto) represented Canada in the 1966 Chess Olympiad in Cuba. He has worked at Bell Laboratories in New Jersey, and has been a director of research at Salomon Brothers Inc., New York. In addition to running his own debt management consulting service, he is currently a faculty member of the Graduate School of Business Administration at Fordham University, a trustee of the Financial Management Association, and a member of the Queen's Alumni Fund Committee.

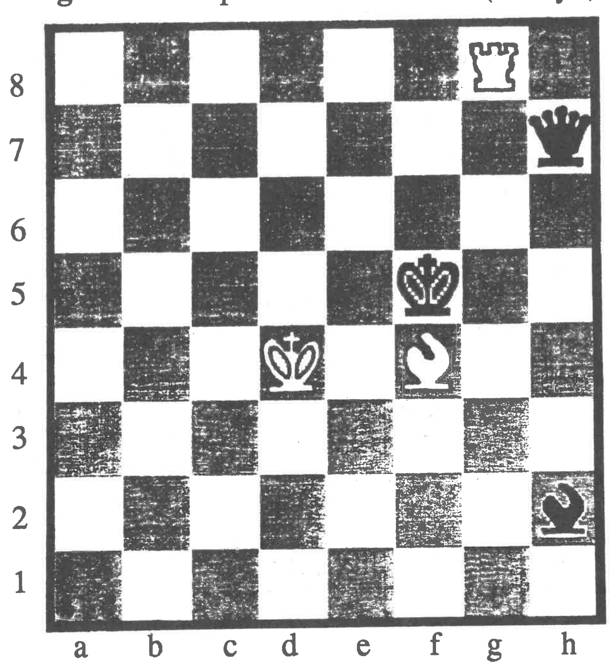
Although the game of chess has many mathematical elements, these tend to be more pronounced in composed problems. Unlike their close relatives, endgames ("white to play and win"), problems ("white to play and mate in two moves") are void of any practical value (whatever practical value playing chess may have). The goal of composers is to express artistic ideas and to discover new structures.

Although I have always enjoyed the challenge of solving problems, I became interested in composing only relatively recently. I have been specializing in problems employing relatively few pieces. A problem with at most 7 pieces is called a *miniature*, a problem with at most 12 pieces is a *Meredith*; most modern problems express complicated themes that require more than 12 pieces.

Several computer programs for solving problems are commercially available. The more sophisticated of these do not rely on brute force methods, but instead exploit the existence of a limited number of terminal positions, which may further be reduced by symmetry considerations. These programs are used by composers for validation, i.e. to check that the intended solution exists and that it is unique.

I have been particulary interested in so-called *helpmates*, where the two sides *cooperate* to create a terminal position in which white checkmates black. The inventor of helpmates was the 19th century American problemist Sam Loyd. His original composition (Fig.1) appeared in the Chess Monthly in 1860. In the solution, white builds a *battery* by placing the bishop in front of the rook, while the black king cooperates by marching to the line of the rook. The mate results when white's bishop fires the battery, producing a discovered check.

Figure 1. Helpmate in 3 moves (S.Loyd, Chess Monthly, 1860)



WHITE	BLACK
1 . ——	Kf6
2 . Ra8	Kg7
3 . Bb8	Kh8
4 . Be5#	

Incidentally, Loyd is also believed to be the inventor of the puzzle with a 4 by 4 square with 15 numbers to be rearranged sequentially. I recall discussing this puzzle with respect to the notion of parity in our linear algebra course.

Most modern helpmates attempt to express a common theme several times. Often a problem is in fact a *series* of problems obtained by *twinning*, i.e. by changing the original position slightly.

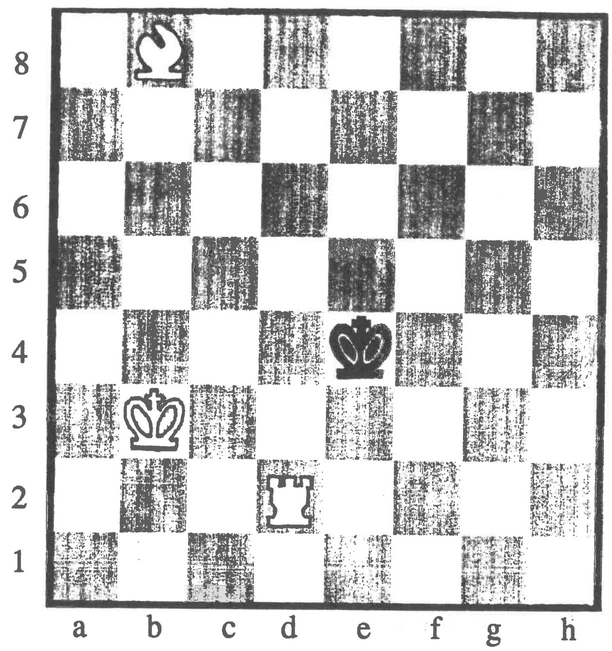
Since 1978 a jury of international experts has been awarding a prize for the best problem published during the year that employs at most four pieces, of which two are normally kings. Because of the limited material, it is difficult to express profound ideas; a mathematical analogy might be the proof of a theorem based upon two well-chosen axioms.

The winning composition in 1989 in this "Four Men Only" tourney was a joint work by Pal Benko and me. This helpmate in 3 1/2 moves had appeared in the July 1989 issue of the British publication called The Problemist, where it also won fourth prize in the helpmate category. Benko, by the way, is unique among chessplayers in being both an international grandmaster and a first-rate composer of both endgames and problems.

The problem (Fig.2) is an extension of Loyd's original idea. Each solution consists of the building and the firing of a battery. In the original position the rook discovers the bishop, in the second position the bishop discovers the rook, and in the final position the king discovers the rook. The interested reader may wish to show that this effect cannot be achieved by the king discovering the bishop.

If you would like to learn more about chess problems, please write to me (25 East 9th Street, New York, NY 10003)

Figure 2. Helpmate in 3 1/2 moves (P.Benko and A.Kalotay, The Problemist, July 1989)



a) Diagram	*b) Bb8→a7*	*c) Kb3→c2*
1. Re2+ Kd3	1. Rh2 Kd3	1. Rd1 Kf3
2. Re3+ Kd2	2. Rh1 Kd2	2. Kd2 Kg2
3. Bf4 Kc1	3. Bg1 Kc1	3. Ke1 Kh1
4. Re1#	4. Be3#	4. Kf2#

Comment by judge for Problemist: "Three different battery formations with only four pieces! Good twinning, antiduals."

Comment by "Four Men Only" Jury: "Three time battery with discovered mate: rook-bishop, bishop-rook, king-rook!"

QUEEN'S MATHEMATICAL COMMUNICATOR

SUMMER 1991

THANKS to several of our readers who sent donations to help keep the Communicator going. If you would like to help please send your cheque to the address below, payable to the Communicator, Queen's University.

EDITOR: BILL WOODSIDE

Address all correspondence, news, problems and solutions to:

Queen's Mathematical Communicator
Department of Mathematics and Statistics
Queen's University
Kingston, Ontario
K7L 3N6