Is the solar system stable? and Can we use chaos to make measurements?

Jack Wisdom
Massachusetts Institute of Technology, USA

Abstract. This talk addresses two separate questions: "Is the solar system stable?" and "Can we use chaos to make better measurements?" In the first part, a review is presented of the numerical experiments which indicate that the motion of Pluto, and indeed the whole solar system, is chaotic. The time scale for the exponential divergence of nearby trajectories is remarkably short compared to the age of the solar system. In the second part, numerical experiments are presented which indicate that the exponential sensitivity of trajectories to changes in initial conditions and parameters cannot be used to exponentially constrain initial conditions and parameters from trajectory measurements. It does appear though that parameters are better constrained by measurements of chaotic trajectories than might naively be expected.

Introduction

First, it is useful to remind ourselves of the reality of chaos, and just how much fun it is. I have a nice demonstration to show you, of a double pendulum (Fig. 1). The double pendulum is one of the simplest dynamical systems one can build after the pendulum: one pendulum supported at the end of another pendulum, constrained to move in a plane. This simple system exhibits outrageously complicated behavior. How could anyone watch the double pendulum and continue to assume that all solutions of Newton's equations could be developed in quasiperiodic perturbation expansions? Did hundreds of years really go by without anyone looking at a dynamical system in action? The double pendulum can also be used to illustrate the divided phase-space characteristic of Hamiltonian systems (even though there is friction in the physical pendulum): trajectories with the same energy can be either chaotic or regular depending on the initial condition. Given that such a simple system as the double pendulum exhibits such complicated motion, it is hard to understand why it has taken so long for the importance of chaotic behavior to be realized. Chaos is not an irrelevant mathematical curiosity.

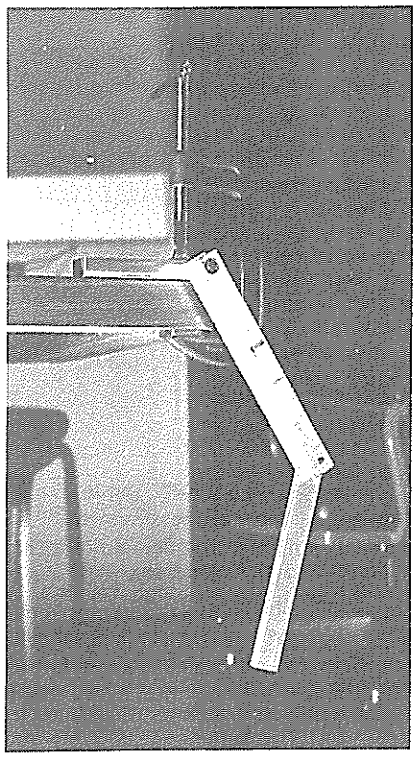

Fig. 1. The double pendulum provides a nice demonstration of chaotic behavior and the divided phase space in simple physical systems.

Question 1: Is the solar system stable?

Surely this is one of the oldest questions in modern science. As soon as Newton's equations of motion are written down one has to wonder about the long-term consequences. It has generally been assumed that our solar system is quasiperiodic and consequently stable, and that it is only a matter of time until a mathematical proof of this fact is given. Tremendous progress towards that goal has been made. Arnold (1963) has proven that solar systems are quasiperiodic in large measure provided that the masses, inclinations, and eccentricities of the planets are sufficiently small. On the other hand, we know that dynamical systems generally display chaotic behavior as well as regular behavior, and the solar system is, after all, just another dynamical system. The question of its stability should be approached with an open mind.

What we really would like to know is whether our solar system is on a chaotic or quasiperiodic trajectory. Since the physical experiment runs

too slowly for us to deci
of the stability of the so
numerical experiments i
(Sussman and Wisdom,

1.1. Previous integration

Numerical integrations
amount of computer tim
time scales. A direct cal
follow each planet arou
interesting on time scale

This time scale separ
two-body problem. The
frequency, the orbital fr
freedom after eliminatio
by planetary perturbatio
orbital nodes precess. T
orbital frequency multip
that of the sun. The pre
of years to a couple of m
precession time scale s
coupling between the pl

We have performed
(Applegate *et al.*, 1985)
centuries (Fig. 2), the I
puter designed specifica
one-third the speed of a
jector. It consists of ten
because there are nine
capabilities: it can add,
integrate Newton's equa
bits on the Cray; it has

In our first calculatio
planets (Sun, Jupiter,
Plutos) forward and ba
is longer than the clas
Cohen, Hubbard, and C
orders of magnitude.

One result from that
tion theory (Bretagnon,
merically resolving the

too slowly for us to decide the matter, we have approached the question of the stability of the solar system through numerical experiments. Our numerical experiments indicate that, in fact, the solar system is chaotic (Sussman and Wisdom, 1988).

1.1. Previous integration

Numerical integrations of the solar system take an extraordinarily large amount of computer time. This is because there is a tremendous range of time scales. A direct calculation must take steps that are small enough to follow each planet around the sun, yet the motion of the planets is only interesting on time scales of millions of years.

This time scale separation is a consequence of the degeneracy of the two-body problem. The unperturbed two-body problem has only a single frequency, the orbital frequency, even though there are three degrees of freedom after elimination of the center of mass. The degeneracy is broken by planetary perturbations. The largest effect is to make the perihelia and orbital nodes precess. The frequency of these motions is of order the orbital frequency multiplied by the mass ratio of the perturbing planet to that of the sun. The precession time scales range from tens of thousands of years to a couple of million years. Only on a time scale longer than this precession time scale should we expect to find interesting dynamical coupling between the planets.

We have performed our numerical integrations on the Digital Orrery (Applegate *et al.*, 1985). Named after the orreries of the 18th and 19th centuries (Fig. 2), the Digital Orrery (Fig. 3) is a special purpose computer designed specifically for solar system dynamics. It runs at about one-third the speed of a Cray 1, but is smaller than the viewgraph projector. It consists of ten computers which run in parallel. There are ten because there are nine planets plus the sun. Each computer has limited capabilities: it can add, multiply, and take $-3/2$ powers—just enough to integrate Newton's equations. The mantissa has 56 bits compared to 48 bits on the Cray; it has turned out that this extra precision was crucial.

In our first calculation (Applegate *et al.*, 1985) we integrated the outer planets (Sun, Jupiter, Saturn, Uranus, Neptune, and several massless Plutos) forward and backward in time for about 100 million years. This is longer than the classic million year ($\pm 500\,000$ yr) integration of Cohen, Hubbard, and Oesterwinter (1973) by a factor of more than two orders of magnitude.

One result from that integration was that the best analytic perturbation theory (Bretagnon, 1982) for the solar system was inadequate. Numerically resolving the observed motions of the massive planets into a

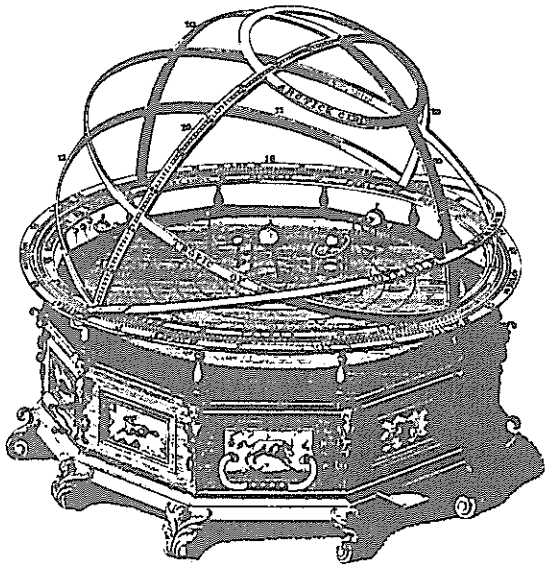

Fig. 2. The mechanical orreries are nice symbols of the apparent clockwork predictability of the motions of the planets.

quasiperiodic series, we found that the spectrum of Jupiter contained terms which were larger than all but seven of the 200 terms listed in

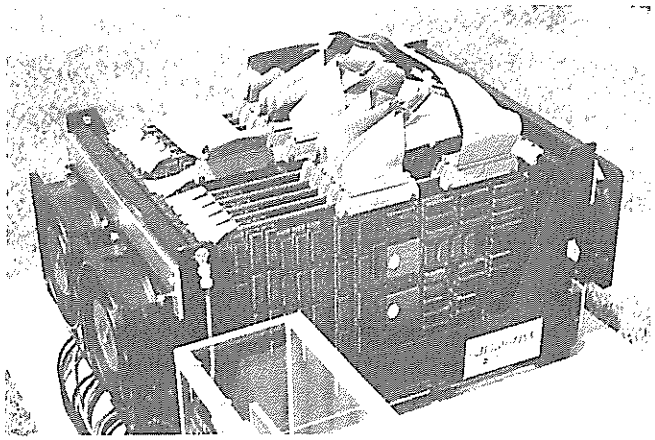

Fig. 3. The "Digital Orrery" is a computer which was specifically designed to investigate the dynamics of the solar system.

Fig. 4. The power spectrum
common logarithm of the p

Bretagnon's solution
perturbation theory wa

The motion of Plut
crosses the orbit of Nep
are in an orbital resona
orbital period of Neptu
Pluto. With the systen

Fig. 5. The power spectrum
perturbation th ories. Lines
which are th rder in t
planetary m

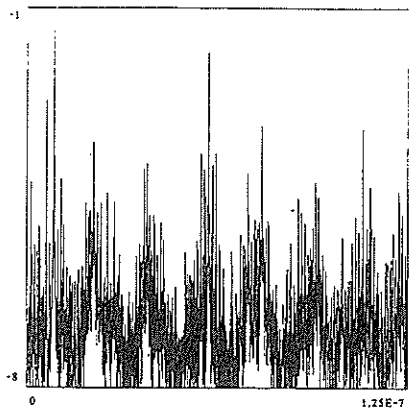

Fig. 4. The power spectrum of a variable related to the eccentricity of Jupiter's orbit. The common logarithm of the power is plotted vs frequency (in cycles per day).

Bretagnon's solution (see Figs. 4 and 5). The problem was that the perturbation theory was not carried to high enough order.

The motion of Pluto is particularly complicated. The orbit of Pluto crosses the orbit of Neptune. This is only possible because the two planets are in an orbital resonance (Cohen and Hubbard, 1965): three times the orbital period of Neptune is approximately two times the orbital period of Pluto. With the system in this resonance close encounters do not occur

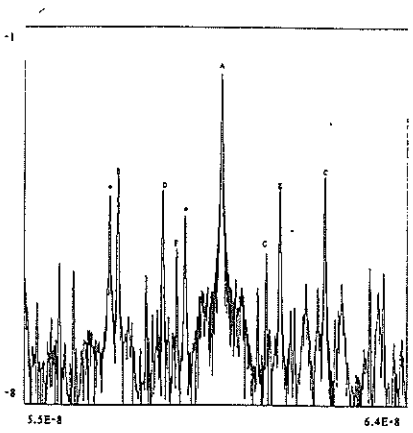

Fig. 5. The power spectrum of Jupiter was not adequately represented by the best analytic perturbation theories. Lines marked D and E are not recovered in perturbation theories which are third order in the eccentricities and inclinations and second order in the planetary masses.

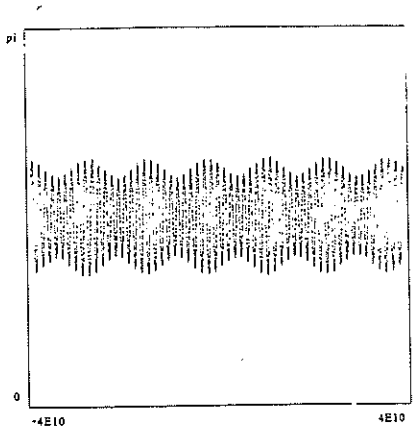

Fig. 6. The argument of perihelion of Pluto had a strong long-period modulation, with a period of 34 million years.

even though the orbits cross. Integrations by Williams and Benson (1973) showed that Pluto was involved in yet another resonance: Pluto's perihelion (the longitude at which the planet is closest to the sun) and its ascending node (the longitude at which the orbit plane crosses the plane perpendicular to the angular momentum of the solar system) are locked together. The regression of the perihelion and the ascending node have precisely the same periods. The difference between the two angles, which is called the argument of perihelion, oscillates about $\pi/2$ with a period of about 3.8 million years.

Several new features in the motion of Pluto were revealed by our calculation. We found a surprisingly large number of strong, long-period variations. The argument of perihelion showed a strong modulation with a period of 34 million years (Fig. 6). The variable $h = e \sin \varpi$, where e is the orbital eccentricity and ϖ is the longitude of perihelion, had a strong component with a period near 137 million years (Fig. 7). (This variable plays an important role in analytic theories.) This frequency may be associated with a near resonance between one of the fundamental frequencies associated with Pluto and one of the fundamental frequencies of the system of massive planets. The inclination of Pluto showed some evidence of a period longer than our integration, or perhaps even a secular decline (Fig. 8). The most suspicious bit of evidence was the very noisy power spectrum of the resonance variable associated with the basic 3:2 commensurability of the orbital periods (Fig. 9). The power spectrum of a quasiperiodic trajectory should have no more independent frequencies than the number of degrees of freedom. Noisiness of power

Fig. 7. There was a very lo
to Pluto's eccentricity.

spectra or the presen
associated with chaotic
spectrum is difficult to

The proper thing to
quasiperiodic is to co
whether or not neighb
calculation of the Ly
displays $\log_{10} \gamma$ versus

Fig. 8. The inclination of P
million year integration.

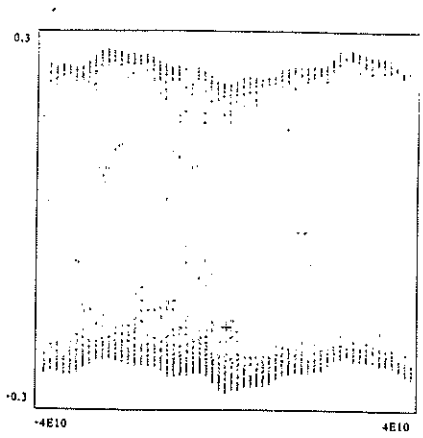

Fig. 7. There was a very long period (137 million year) component in a variable related to Pluto's eccentricity.

spectra or the presence of a broadband component have been widely associated with chaotic behavior. Unfortunately, the noisiness of a power spectrum is difficult to quantify.

The proper thing to do to determine whether a trajectory is chaotic or quasiperiodic is to compute the Lyapunov exponents, which measure whether or not neighboring trajectories diverge exponentially. Our first calculation of the Lyapunov exponent is shown in Fig. 10. The plot displays $\log_{10}\gamma$ versus $\log_{10}(t - t_0)$, where $\gamma = \ln[d(t)/d(t_0)]/(t - t_0)$,

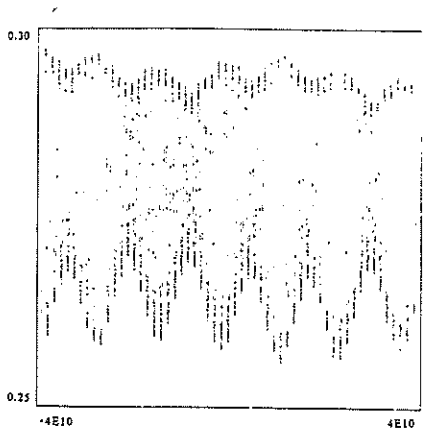

Fig. 8. The inclination of Pluto had periods longer than could be resolved by our 214 million year integration.

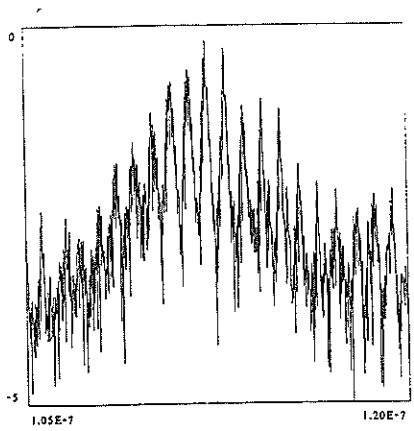

Fig. 9. Expanded views of power spectra for Pluto were suspiciously noisy.

and $d(t)$ the phase-space distance between neighboring Plutos. There was no sign that γ was leveling off to a positive Lyapunov exponent.

1.2. New integration

We were compelled by the very long periods and noisy spectra we found in our integrations to carry out a longer integration of the solar system. The motion of Pluto just seemed too complicated.

Our earlier integration was limited to ± 100 million years because of the accumulation of roundoff errors. We found a trick that allowed us to

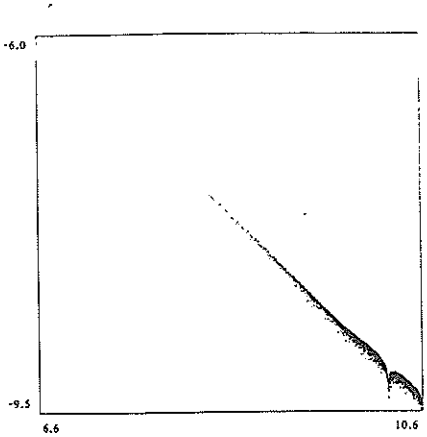

Fig. 10. The computation of the Lyapunov exponent did not show any indication of a nonzero exponent greater than $10^{-6.8}$ yr^{-1}.

Fig. 11. Relative er

significantly redu
much longer time
etary orbits the er
11). This energy
growth of the err
in energy corresp
grated yields an
The origin of this
which can estima
would be that the
grow with the sq
understand why
found in our nur
pended on the ste
positive and for o
W. Kahan, we in
which there was n
was indeed the ca
size which elimir
length of the test i
integrations were
million days versu
measured in days,
of the solar mass.
The best step size
for our new integr

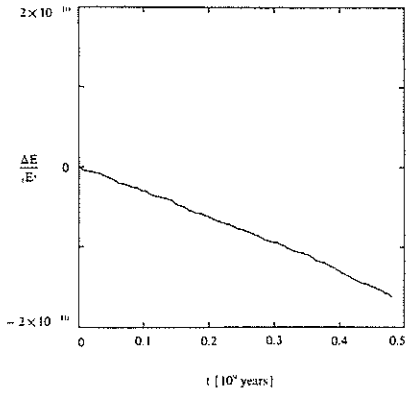

Fig. 11. Relative energy error in our new long-term integration of the outer planets.

significantly reduce our numerical error and extend our integrations to much longer times. In all direct long-term numerical integrations of planetary orbits the error energy is found to grow linearly with time (see Fig. 11). This energy error dominates all other errors; it leads to a quadratic growth of the error in all positions and longitudes since the linear error in energy corresponds to a linear error in frequency which when integrated yields an error in longitude which grows quadratically in time. The origin of this energy error is not understood; there is as yet no theory which can estimate the rate of the linear growth. The naive expectation would be that the error in energy should behave like a random walk and grow with the square root of time. However, even though we do not understand why the error grows as it does we can make use of it. We found in our numerical studies that the slope of the energy error depended on the step size; in fact, for some step sizes the energy error was positive and for others the error was negative. Following a suggestion of W. Kahan, we investigated whether there might be special step sizes for which there was no linear growth in the energy error. We found that this was indeed the case. More importantly, we found that the special step size which eliminated the linear energy error did not depend on the length of the test integrations, but rather became better defined as the test integrations were extended. Figure 12 shows the energy error after 5 million days versus step size. The units are not important, but time is measured in days, distance is in astronomical units, and mass is in units of the solar mass. Figure 13 shows the energy error after a billion days. The best step size of these test runs is 32.7 days, which is what we chose for our new integrations. We extensively studied the numerical accuracy

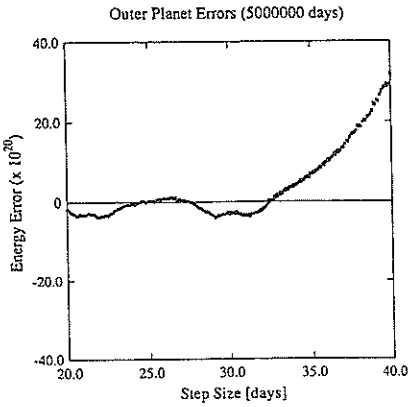

Fig. 12. The energy error after 5 million days as a function of step size.

of integrations with various step sizes. These errors were determined by integrating forward in time, then reversing the system to recover the initial state. This is a valid test since our integrator is not microscopically reversible or explicitly symplectic. Figure 14 shows the round trip errors in the position of Jupiter, which display a minimum at the special step size of 32.7 days. Thus removing the energy error actually gives a better long-term trajectory (as indicated by the round trip errors). It is curious that at the special step size the round trip errors in all the planets are comparable; this is not true for other step sizes.

Table I lists for comparison the rate of growth of energy error in

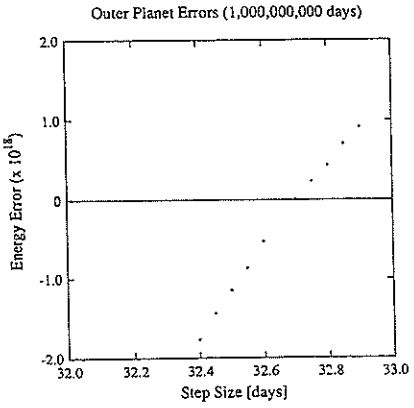

Fig. 13. The energy error after a billion days as a function of step size.

Fig. 14. The error in
days and then backwa
at the same step size

several long-term
significantly more
tions of the solar
for nearly a billio
at the end of the
longitude of Plut
months on the D
puter time.

Table I. Energy errors
integration is marked
energy error is abou
integrations.

Integration
CHO (1973)
K&N (1984)
200 MYR (1986)
LONGSTOP (1986)
LONGSTOP (1987)
This work

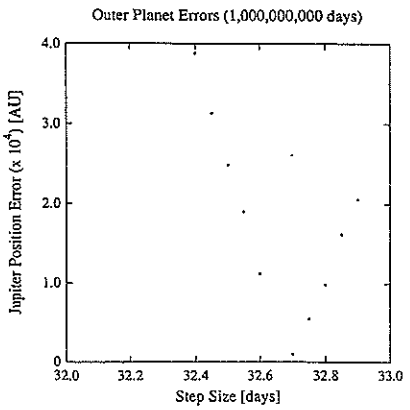

Fig. 14. The error in recovering the position of Jupiter after integrating forward a billion days and then backward a billion days vs step size. The round trip errors are a minimum at the same step size that minimizes the energy error.

several long-term integrations of the solar system. Our new integration is significantly more accurate and longer than all other long-term integrations of the solar system. We simulated the motion of the outer planets for nearly a billion years. The estimated error in the longitude of Jupiter at the end of the integration is only a few degrees, and the error in the longitude of Pluto is a few arc minutes. The calculation took about five months on the Digital Orrery; this is about 4 MegaFlop-years of computer time.

Table I. Energy errors in the various long-term integrations of the outer planets. Our first integration is marked "200 MYR (1986)." In our new integration the growth of the energy error is about three orders of magnitude smaller than all previous long-term integrations.

Integration	Interval of integration	$\frac{d}{dt}\frac{E-E_0}{\lvert E_0\rvert}$ [yr^{-1}]
CHO (1973)	1 000 000 yr	2.4×10^{-16}
K&N (1984)	6 000 000 yr	5×10^{-16}
200 MYR (1986)	214 000 000 yr	1.8×10^{-16}
LONGSTOP (1986)	8 000 000 yr	
LONGSTOP (1987)	100 000 000 yr	-2.5×10^{-16}
This work	845 000 000 yr	-3.0×10^{-19}

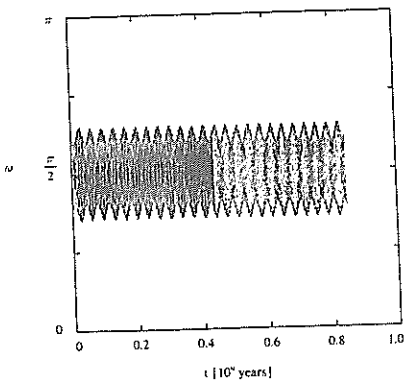

Fig. 15. The argument of perihelion of Pluto continues to display a 34 million year modulation.

The argument of perihelion of Pluto is displayed in Fig. 15. The 34 million year oscillation we previously observed (Fig. 6) is the longest period which is noticeable in the time series for the argument of perihelion. The 137 million year modulation of the quantity $h=e \sin \varpi$, is also still the longest period noticeable (Fig. 16). The inclination was not secularly declining, but there may be a weak component with a period near 600 million years (Fig. 17).

The most important indicator of chaotic behavior is the largest Lyapunov exponent. We computed the largest Lyapunov exponent in several different ways. The simplest method is to just look at the divergence of nearby trajectories. Figure 18 shows the divergence of two Plutos.

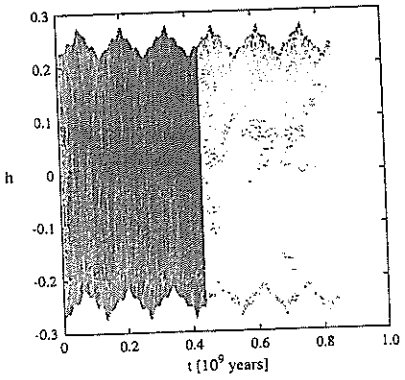

Fig. 16. The eccentricity of Pluto continues to display the 137 million year modulation.

Fig. 17. T

Already here there
The slope gives an
years. This simple
eventually separate
represent the diverg
tance between the
separate more than
basic 3:2 orbital re
are two prescriptio
calculation. One w

Fig. 18. The logarithm
growth indicates expon
simple two-trajectory n

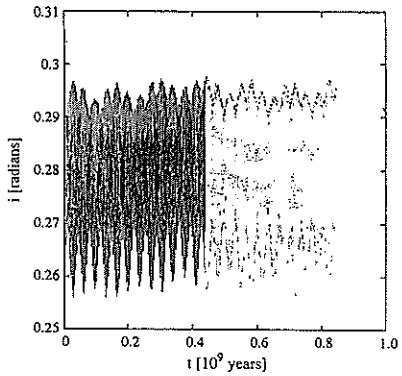

Fig. 17. The inclination of Pluto varies somewhat irregularly.

Already here there is evidence of exponential divergence of trajectories. The slope gives an exponential divergence time scale of about 20 million years. This simple method has the disadvantage that the two trajectories eventually separate so much that the two trajectories do not any longer represent the divergence of "nearby" trajectories, and ultimately the distance between the trajectories saturates. In this case two Plutos cannot separate more than about 45 AU, if the amplitude of the oscillation of the basic 3:2 orbital resonance variable does not significantly change. There are two prescriptions for improving on this simple Lyapunov exponent calculation. One way is to "renormalize" the distance between the two

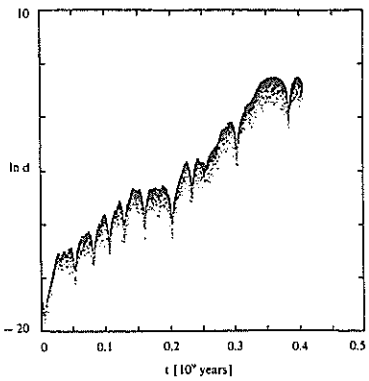

Fig. 18. The logarithm of the distance between two "Plutos" is plotted vs time. The linear growth indicates exponential divergence with a time scale of about 20 million years. The simple two-trajectory method saturates at a distance of about 45 AU.

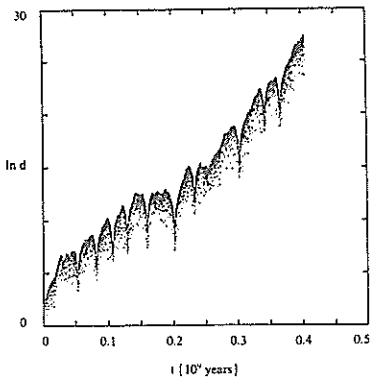

Fig. 19. The logarithm of the distance between neighboring Plutos computed with the linearized variational equations.

Fig. 20. The two n
are in excellent ag

trajectories, i.e., the test trajectory is repeatedly brought back to be near the reference trajectory preserving the direction (in phase space) between them, and keeping track of how much the test trajectory was moved. The problem with this method is knowing how often to renormalize. In a system with a small Lyapunov exponent, we found that too frequent renormalization can contaminate the result. The best method is to integrate the linearized variational equations along with the equations of motion. This method needs renormalization only if the distance grows so large that it cannot be represented in the computer—a problem not encountered here. Figure 19 shows the divergence of nearby Plutos computed using the variational method. The agreement with the two Pluto calculation is striking (Fig. 20). The two calculations differ significantly only when the two trajectory method is close to saturation. Figure 21 shows the conventional (more conservative) plot of the calculation of the Lyapunov exponent. The exponent is clearly leveling off at a positive value, in a manner typical of Lyapunov exponent calculations. The inverse of the Lyapunov exponent is about 20 million years. Thus our calculation indicates that the motion of Pluto is chaotic, with a surprisingly short divergence time scale.

ments on a sys
similar behavio
with the square
that when the s
gence begins to
gence acts as a
illustrates the c
tion during a ca
ization is repea

It is interesting to further examine the divergence of nearby trajectories by displaying the distance between Plutos on a log d versus log t plot (Fig. 22). Several different experiments are superimposed to show the envelope. The solid curve is an exponential with a 20 million year e-folding time scale. The dashed curve is a power law with distance proportional to the 3/2 power of the time. The plot shows that the initial divergence of the Plutos follows the 3/2 power law. Numerical experi-

Fig. 21. The conv
positive exponent

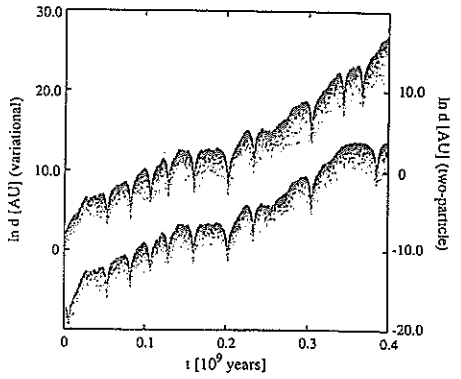

Fig. 20. The two methods for computing the divergence of trajectories give results which are in excellent agreement up to the point of saturation.

ments on a system of Plutos without planetary perturbations showed a similar behavior for the divergence. (This divergence can be contrasted with the square-law behavior observed for the round trip errors.) We see that when the slope of the exponential exceeds the power-law the divergence begins to follow the exponential. Thus the initial power-law divergence acts as a "seed" for the later exponential divergence. This plot illustrates the disaster that could result from too frequent renormalization during a calculation of a small Lyapunov exponent. If the renormalization is repeatedly carried out during the power-law phase of the di-

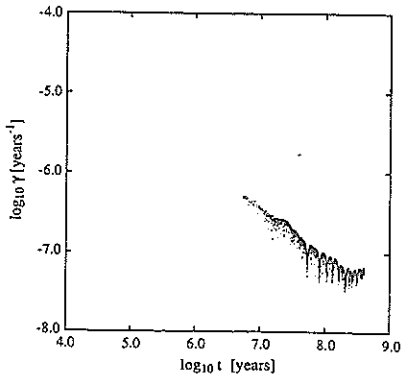

Fig. 21. The conventional representation of a Lyapunov calculation clearly indicates a positive exponent with a divergence time scale of about 20 million years.

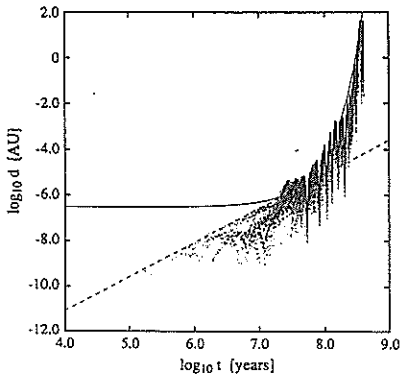

Fig. 22. A log–log plot of the distance between two Plutos vs time emphasizes the early stages of the separation. The initial divergence due to round-off error provides a seed to the exponential divergence associated with chaotic behavior.

vergence, the measured Lyapunov exponent may have no relation to the true exponent. The variational method does not have this problem.

There is a disturbing aspect to this result that the motion of Pluto is chaotic. In other examples of chaotic behavior in the solar system (see Wisdom, 1987 for a review) trajectories always look decidedly irregular when there is a positive Lyapunov exponent. That is not the case here. For example, the argument of perihelion (Fig. 16) looks remarkably regular. The inclination is a little more complicated, but not obviously irregular. The strongest evidence for irregular behavior is in the power spectra. Figure 23 shows a portion of the power spectrum of $h = e \sin \varpi$

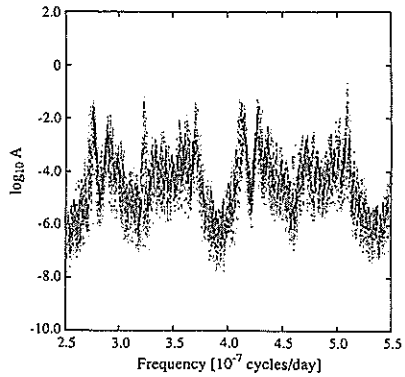

Fig. 23. This expanded view of the Pluto spectrum still has a broadband character.

Fig. 24. This expanded
riodic line spectrum.

for Pluto. Figure
power spectrum o
ferent. The spectru
line spectrum; the
that the broadband
spectrum of Neptu
consistent with our

In planetary mo
esting of the orbita
closely related to th
longitudes of the p
the motion on the
associated with the
age out and produc
gitudes are ignorabl
the average, consta
the sampled semim
orbital element of I
Pluto and Neptune
The semimajor axis
28 show expanded
Neptune shows pat
On the other hand,
of the semimajor ax

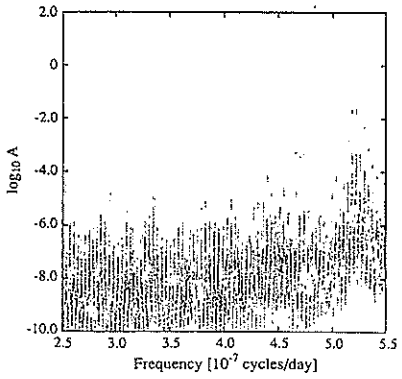

Fig. 24. This expanded view of the Neptune spectrum is still consistent with a quasiperiodic line spectrum.

for Pluto. Figure 24 shows, for comparison, a similar portion of the power spectrum of Neptune. The power spectra are qualitatively different. The spectrum of Neptune is consistent with a very complicated line spectrum; the spectrum of Pluto has a broadband character. Note that the broadband component is stronger than most of the lines in the spectrum of Neptune. The broadband component of Pluto's spectrum is consistent with our measurement of a positive Lyapunov exponent.

In planetary motion the semimajor axes are usually the most uninteresting of the orbital elements. The reason is that the semimajor axes are closely related to the momenta conjugate to the fastest moving angles, the longitudes of the planets. Because of the large separation of time scales, the motion on the orbital time scale is not very important. The effects associated with the extremely rapid variations in longitude tend to average out and produce no cumulative effect. The fact that the orbital longitudes are ignorable implies that the variables conjugate to them are, on the average, constant and hence not very interesting. Curiously, plots of the sampled semimajor axis appear more irregular than plots of any other orbital element of Pluto. Figures 25 and 26 show the semimajor axes of Pluto and Neptune (for comparison) over several hundred million years. The semimajor axis of Pluto is markedly more irregular. Figures 27 and 28 show expanded portions of these time series. The semimajor axis of Neptune shows patterns characteristic of a sampled quasiperiodic signal. On the other hand, the word "turbulent" is brought to mind by the plot of the semimajor axis of Pluto. Apparently, sampling the semimajor axis

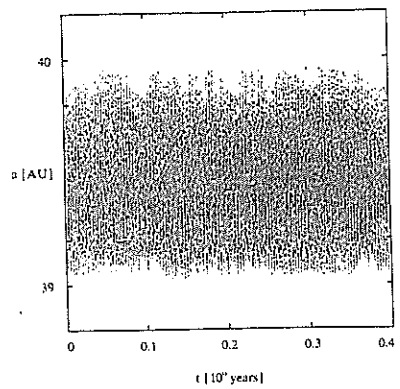

Fig. 25. The time series of the semimajor axis of Pluto has an irregular appearance.

has brought out the chaotic character of Pluto's motion. Perhaps there is a clue here to the dynamical resonance mechanism responsible for Pluto's chaotic motion?

Since we performed this numerical experiment which indicated that the motion of Pluto was chaotic an analogous experiment has been performed by J. Laskar (1989) for the whole solar system. With today's computers and today's computer time allotments it is impossible to directly integrate the whole solar system for hundreds of millions of years. Instead, Laskar has studied the whole solar system by first analytically averaging the equations of motion to yield a system of about 150 000 secular equations which he then integrated on a conventional supercomputer. Figure 29 shows the calculation of the Lyapunov exponent from

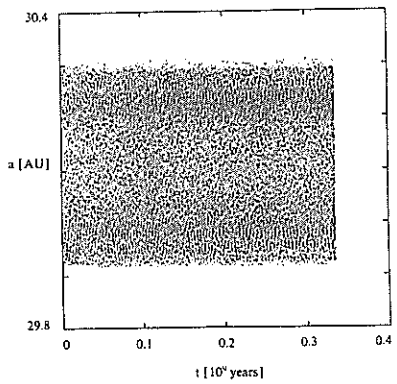

Fig. 26. The time series of the semimajor axis of Neptune has a quasiperiodic appearance.

Fig. 27. On a shorter time s
irregular appearance.

his paper. Laskar has fc
exponential divergence
long-term evolution of

One word of caution
one of a kind calculati
our results to uncertai
initial conditions of the
tiny and with the true
unlikely to me since tv

Fig. 28. On a shorter time sc
appears quasiperiodic.

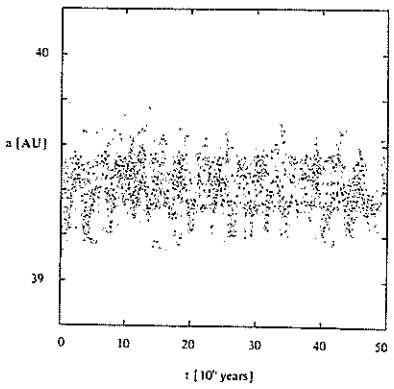

Fig. 27. On a shorter time scale the time series of the semimajor axis of Pluto still has an irregular appearance.

his paper. Laskar has found that the whole solar system is chaotic with an exponential divergence time of only about 5 million years. The detailed long-term evolution of the Earth is unpredictable!

One word of caution is in order. Both of these calculations are so far one of a kind calculations. We have not yet evaluated the sensitivity of our results to uncertainties in our knowledge of the true masses and initial conditions of the planets. It may be that the chaotic zones are very tiny and with the true masses the system is not chaotic. This seems unlikely to me since two very different kinds of calculation have been

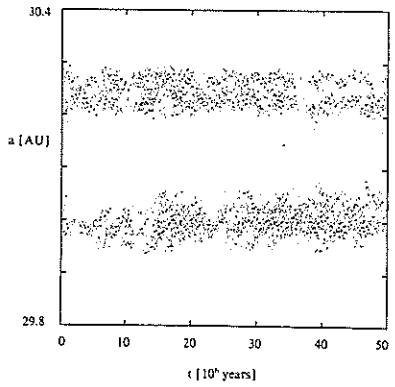

Fig. 28. On a shorter time scale the time series of the semimajor axis of Neptune still appears quasiperiodic.

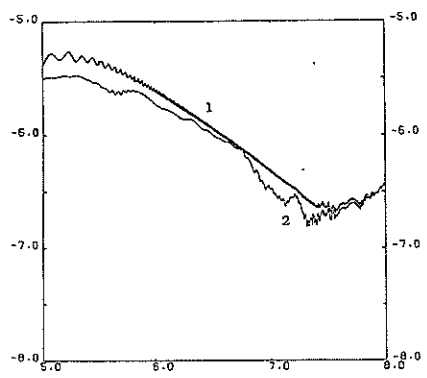

Fig. 29. Calculation of the Lyapunov exponent for the whole solar system indicates exponential divergence with a time scale of only 5 million years.

performed and both gave positive Lyapunov exponents. Still the sensitivity to parameters needs to be addressed.

It is difficult to evaluate the implications of these results. The chaotic character of the solar system opens many possibilities. The high eccentricity and inclination of Pluto's orbit are puzzling since we understand that planets form from a nebula which collapses to a disk; all the planets should more or less be in the same plane. However, I have shown (Wisdom, 1987) that asteroids placed in the chaotic zones associated with orbital resonances with Jupiter often evolve to high inclination and eccentricity. Thus Pluto may have formed in an orbit near the plane of the other planets, and evolved through the chaotic zone to its current high eccentricity and high inclination orbit. This seems a very natural suggestion. Of course, it could also be the case that the chaotic zone in which Pluto moves is very tiny with only exceedingly small holes into other parts of the phase space. In this case, Pluto may always have had much the same orbit. In an extreme view, the chaotic character of Pluto's motion opens the possibility that Pluto was initially in an orbit radically different from the orbit in which it is found today. Chaos enlarges the possibilities rather than constrains them.

The chaotic character of the dynamics of the solar system also precludes any rigorous statement about its future evolution (at the moment). The fact that the solar system is 4.6 billion years old does not demand any conclusion regarding its future stability. Recall my curious results regarding the motion of asteroids in the 3:1 orbital resonance with Jupiter (Wisdom, 1982, 1983). Figure 30 shows the eccentricity of a chaotic asteroid trajectory versus time. For over a hundred thousand

Fig. 30. Eccentricity
display long periods o
During the regular pe

years the eccentri
and then suddenl
behavior is that th
it breaks loose, an
chaotic zone in w
1985). Until we
Pluto and the res
the possibility, sa
exhibit similar wi
say whether the E
years from now.

Question 2: Ca

The hallmark of
The exponential

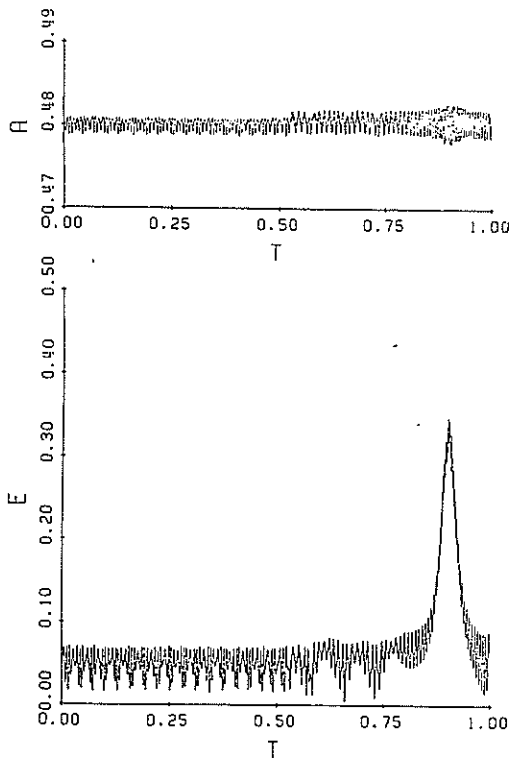

Fig. 30. Eccentricity vs time of a chaotic asteroidlike trajectory. These trajectories often display long periods of apparently regular motion, followed by bursts of irregular motion. During the regular periods the trajectory is stuck to an island in the phase space.

years the eccentricity looks quasiperiodic, then it becomes more irregular, and then suddenly it begins shooting to large values. The reason for this behavior is that the trajectory is originally stuck to an island, from which it breaks loose, and eventually it finds its way into a different region of the chaotic zone in which large jumps in eccentricity take place (Wisdom, 1985). Until we understand the structure of the chaotic zone in which Pluto and the rest of the solar system move, it is not possible to exclude the possibility, say, that the orbit of the Earth will suddenly begin to exhibit similar wild excursions in eccentricity. It is not even possible to say whether the Earth will still be a member of the solar system a billion years from now.

Question 2: Can we use chaos to make measurements?

The hallmark of chaos is the sensitive dependence on initial conditions. The exponential divergence of chaotic trajectories precludes long-term

prediction given limited knowledge of the state of the system. Any error in our knowledge of either the state of the system or the system parameters will exponentially explode into complete uncertainty in the state of the system. Is it possible, though, to turn this sensitivity to our advantage? Rather than trying to predict the future from limited knowledge, can we use the sensitivity of a system undergoing chaotic motion to deduce the properties of the system? Can we measure the behavior of a system and deduce, exponentially well, properties of the system which were required to produce the observed behavior. Several years ago (Wisdom, 1987) I suggested that it might be possible to use noisy observations of the brightness of Hyperion, the chaotically tumbling satellite of Saturn (Wisdom *et al.*, 1984; Klavetter, 1989), to determine exponentially well both the state of rotation of Hyperion and the ratios of its principal moments of inertia. Letting the dream get out of hand, it even seemed conceivable that if extensive observations of this otherwise uninteresting 400 km satellite could be made, an arbitrarily large amount of information could be deduced about the Saturn system, without a spacecraft going there. And why restrict oneself to observations of objects which are difficult to observe? Surely the behavior of the double pendulum depends outrageously sensitively on the parameters of the system such as the lengths and masses of the individual pendula, and more importantly on the value of the gravitational acceleration g. Can we make a chaotic double pendulum to measure g, or a chaotic Cavendish balance to measure the gravitational constant G? Unfortunately, it appears there is a sort of "classical uncertainty principle," which prevents us from using the extreme sensitivity of chaotic behavior to make exponentially accurate measurements.

There has recently been considerable interest in the closely related problem of "noise reduction" (Farmer and Sidorowich, 1987, 1988; Kostelich and Yorke, 1989). Given the dynamical system (with its parameters), it certainly is possible to use noisy measurements to exponentially constrain the trajectory of the system, away from the end points of the time series. In the work reported here, I used neither of the techniques for noise reduction mentioned above. Rather, I use the first part of the shadowing algorithm as described by Hammel *et al.* (1989). By the way, the results of Hammel *et al.*, are especially nice, if anyone is unfamiliar with them. They prove, by "computer proof," that for particular trajectories of particular maps, there are true trajectories of the map which stay close to the numerically computed trajectory for an astonishingly long time. Typical results are that for trajectories computed with 64 bit arithmetic true trajectories shadow computed trajectories within a relative accuracy of 10^{-8} for 10^8 iterations. Who could ask for more of a numerical trajec-

tory? The first p
jectory a more c
the roundoff err
was mentioned l
works pretty we
algorithm does
fails. The algori
result. Despite
essence of the p
ments, as we sh

In order to fo
set up a highly i
simple map, the

The map is used
I make measure
each time step.
being measured
measure all the
cations of embe
the dynamical s
is unknown. Thi
the dynamical s
ation, to what e
exponentially de

Figure 31 illu
dynamical syste
The common lc
the true orbit, a
the shadow orbi
number. We see
observations on
except at the en
machine is not
could be recove
tially with the l

Now what h
noise with a slig
trates the result
10^{-12}. In each

tory? The first part of the proof is to compute from the computed trajectory a more consistent pseudotrajectory which is less contaminated by the roundoff error. That this algorithm could be used for noise reduction was mentioned by Sidorowich and Farmer (1988). I have found that it works pretty well, though it is not ideal for a number of reasons. The algorithm does not work for regular orbits. The algorithm sometimes fails. The algorithm does not give a natural estimate of the error in the result. Despite these shortcomings the shadow algorithm captures the essence of the problem with parameter estimation from noisy measurements, as we shall shortly see.

In order to focus on the matters of principle rather than gory details I set up a highly idealized set of numerical experiments. I chose to study a simple map, the standard map:

$$\left.\begin{aligned} y' &= y + K \sin x \\ x' &= x + y' \end{aligned}\right\} \bmod 2\pi. \tag{1}$$

The map is used to generate a time series of the trajectory (x_i, y_i). Then I make measurements of the trajectory by "observing" both x and y at each time step. Observations are performed by adding to each variable being measured a Gaussian random number with standard deviation σ. I measure all the state variables in order to avoid the unnecessary complications of embedding. In performing the noise reduction I assume that the dynamical system, the standard map, is known, but the parameter K is unknown. This avoids problems of having to build an approximation to the dynamical system from the data. Given this intentionally ideal situation, to what extent is it possible to determine the parameter K? Is it exponentially determined by the observations?

Figure 31 illustrates the now familiar result of noise reduction with the dynamical system, including the parameter, known. In this case $K = 1$. The common logarithm of the difference between the observations and the true orbit, as well as the common logarithm of the difference between the shadow orbit and the true orbit, are plotted as a function of iteration number. We see that the shadow orbit, which is computed from the noisy observations only, recovers the true orbit to about machine precision except at the end points of the time series. Of course, the precision of the machine is not a fundamental limitation; in principle a true trajectory could be recovered with an accuracy which continues to grow exponentially with the length of the time series.

Now what happens as we use the shadow algorithm to remove the noise with a slightly incorrect value of the parameter K? Figure 32 illustrates the result for three different values of $\delta K = K - K_{\text{true}}$: 0, 10^{-8}, and 10^{-12}. In each case the shadow algorithm finds a shadow orbit, and the

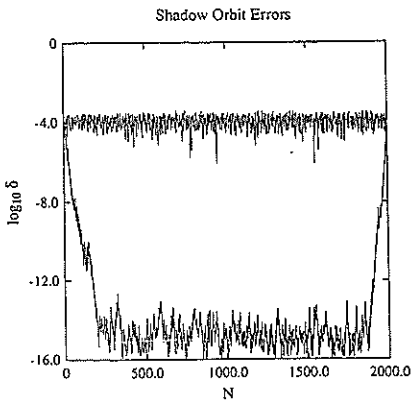

Fig. 31. Demonstration of "noise reduction" for the standard map using the shadowing algorithm. The time series of the logarithm of the errors before and after noise elimination are shown.

error in recovering the true orbit is roughly of order the error in the parameter K. We can see already that if this is generally true, then it will be difficult to determine parameters. The differences between the shadow orbits and the true orbit are far below the errors in the measurements. The measurements, which are all we know in general, apparently cannot distinguish which one is correct.

Perhaps we should try anyway. An objective measure of goodness of fit is

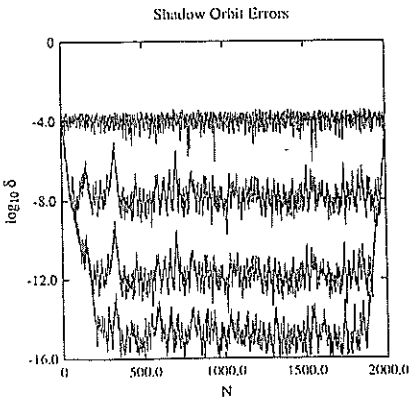

Fig. 32. The logarithm of the errors in the shadow orbit using slightly incorrect values of the parameter.

Fig. 33. Illustration of t
orbits form a single par

$$\chi^2=\overline{2}$$

where $(\overline{x}_i, \overline{y}_i)$ is
orbit. In the sum I
time series, excludi
number of points in
fitting K by minimi

We can estimate
the number of obs
these shadow orbit
shadow orbits seem
mine the shadow or
then use that shac
shadow algorithm a
true orbit is still rec
form a simple one
Fig. 33, for the ith
measured position
the position of the
trajectory varies as
shadow position fro
linearity can be see
$\delta K = 10^{-12}$ are ve
for small δK is a si
tories form a single

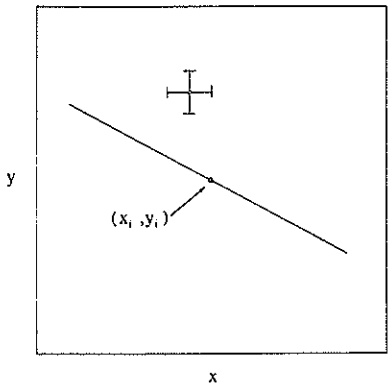

Fig. 33. Illustration of the behavior of the shadow orbits at the ith point. The shadow orbits form a single parameter family of orbits.

$$\chi^2=\frac{1}{2N\sigma^2}\sum_{i=0}^{N-1}[(\overline{x}_i-\tilde{x}_i)^2+(\overline{y}_i-\tilde{y}_i)^2], \tag{2}$$

where $(\overline{x}_i, \overline{y}_i)$ is the observational data, and $(\tilde{x}_i, \tilde{y}_i)$ is the shadow orbit. In the sum I only included those points in the interior part of the time series, excluding the exponential transients at the ends. N is the number of points included in the sum. We can try to determine the best fitting K by minimizing the χ^2 of the shadow orbit as a function of K.

We can estimate how the error in the best fitting K should depend on the number of observations. There are some interesting properties of these shadow orbits which will help us to do this. The first is that the shadow orbits seem to be uniquely determined. For instance, if I determine the shadow orbit with a slightly wrong value of the parameter, and then use that shadow orbit as the initial observational data for the shadow algorithm again, this time with the correct value of K, then the true orbit is still recovered. Away from the end points, the shadow orbits form a simple one parameter family of trajectories. This is illustrated in Fig. 33, for the ith iteration. The point with the error bars indicates the measured position for this iteration. The dot labeled (x_i, y_i) represents the position of the true trajectory. The curve represents how the shadow trajectory varies as a function of δK, for this iteration. The distance of the shadow position from the true position is linearly proportional to δK. The linearity can be seen in Fig. 32, where the curves for $\delta K = 10^{-8}$ and $\delta K = 10^{-12}$ are vertically displaced copies of one another. The linearity for small δK is a simple consequence of the fact that the shadow trajectories form a single parameter family. From the picture of the shadow

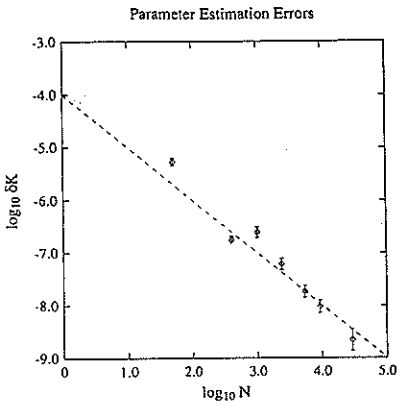

Fig. 34. Parameter estimation errors vs the number of points in the segment of the trajectory inside the exponential transients. The errors decrease as $1/N$ rather than $1/\sqrt{N}$ as was naively expected.

trajectory given in Fig. 33 we can estimate how the best fitting δK will vary with the number of points in the sum. Each measurement provides additional information as to where along the curve the true trajectory lies. The error in δK should thus be proportional to $1/\sqrt{N}$, which is not very spectacular.

Curiously, this $1/\sqrt{N}$ expectation is not correct. Figure 34 shows the results of a number of numerical experiments. For each N several runs were made with different random numbers, the best fitting δK were determined by the procedure outlined above for each of these runs. The mean and standard deviation of the mean of the common logarithms of these δK's are plotted versus the logarithm of N. The dotted curve shows the expected $1/\sqrt{N}$ behavior. The dashed curve is simply σ/N. Apparently, this provides an excellent fit to the data.

Where does the extra factor of $1/\sqrt{N}$ come from? Let us continue the calculation of χ^2. It is convenient to introduce vector notation, $\mathbf{x}_i = (x_i, y_i)$ for the true orbit, $\tilde{\mathbf{x}}_i$ for the shadow orbit, and $\overline{\mathbf{x}}_i$ for the observations. Because of the linear dependence of the shadow orbit on δK, for small δK, we can write

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i + \delta K \mathbf{C}_i. \tag{3}$$

The observational data can be represented:

$$\overline{\mathbf{x}}_i = \mathbf{x}_i + \mathbf{n}_i, \tag{4}$$

where $\mathbf{n}_i$ are two-d
generator with sta

Solving for the δK

Now if all the $\mathbf{C}_i$ l
nitude of δK scale
dently the average
tory. This curious

for each value of j
resulting plot of
irregular way. Ap
difficult to shadow
the irregular close
The longer the traj
tered.

There seems to
though the trajec
initial conditions
trajectories are al
wrong system par
data. The chaos h
uncertainty in all
exponentially accu
orbit. Given K, th
the trajectory, the
can deduce from t
them.

This particular
sensitive depender
used to determine
though that chaos

where $\mathbf{n}_i$ are two-dimensional samples from a Gaussian random number generator with standard deviation σ. The sum for χ^2 becomes

$$\chi^2=\frac{1}{2N\sigma^2}\sum_{i=0}^{N-1}\|\bar{\mathbf{x}}_i-\tilde{\mathbf{x}}_i\|^2 \tag{5}$$

$$=\frac{1}{2N\sigma^2}\sum_{i=0}^{N-1}\|\mathbf{n}_i-\delta K\mathbf{C}_i\|^2. \tag{6}$$

Solving for the δK which gives a minimum χ^2 we find

$$\delta K=\frac{\Sigma_{i=0}^{N-1}\mathbf{n}_i\cdot\mathbf{C}_i}{\Sigma_{i=0}^{N-1}\mathbf{C}_i\cdot\mathbf{C}_i}. \tag{7}$$

Now if all the $\mathbf{C}_i$ have comparable lengths, say, $\|\mathbf{C}_i\|\approx C$, then the magnitude of δK scales as $\sqrt{N}/NC$, i.e., proportional to $1/\sqrt{N}$, again. Evidently the average length of the $\mathbf{C}_i$ grows with the length of the trajectory. This curious suggestion was tested by computing the quantity:

$$D=\frac{\Sigma_{i=0}^{j}\|\mathbf{C}_i\|}{\Sigma_{i=0}^{j}\mathbf{C}_i\cdot\mathbf{C}_i} \tag{8}$$

for each value of j up to the full length of the computed trajectory. The resulting plot of D versus j did indeed fall as $1/\sqrt{N}$, but in a very irregular way. Apparently, some segments of the trajectory are more difficult to shadow than others, and the $1/\sqrt{N}$ behavior of D results from the irregular close encounters of the trajectory with these difficult spots. The longer the trajectory, the more likely a really difficult spot is encountered.

There seems to be a sort of "classical uncertainty principle." Even though the trajectories are very sensitive, and almost any variation of initial conditions or parameters lead to exponential divergence, chaotic trajectories are also very flexible, so flexible in fact that even with the wrong system parameters there are chaotic trajectories that can fit the data. The chaos has reduced the dimension of the unknowns from an uncertainty in all of x_i, y_i, and K, to just an uncertainty in K. There is an exponentially accurate constraint between any given value of K and the orbit. Given K, the true trajectory can be determined; given any point on the trajectory, the parameter K can be determined. Unfortunately, all we can deduce from the data is the extremely accurate relationship between them.

This particular set of numerical experiments indicates then that the sensitive dependence of chaotic motion on system parameters cannot be used to determine system parameters exponentially well. It does appear though that chaos, in a somewhat mysterious way, gives an extra im-

provement in the determination of system parameters by one extra factor of $1/\sqrt{N}$ beyond the naive estimate. Perhaps even this level of improvement could be useful in a physical experiment.

I think it would be worthwhile to carry out similar experiments on continuous systems, and higher dimensional systems.

Acknowledgments

The work reported in this paper was carried out in collaboration with G. J. Sussman, and was supported in part by a NSF Presidential Young Investigator Award and by the NASA Planetary Geology and Geophysics program.

References

Applegate, J., Douglas, M., Gürsel, Y., Hunter, P., Seitz, C., and Sussman, G. J., "A Digital Orrery," IEEE Trans. Comput. (Sept. 1985).

Applegate, J. F., Douglas, M. R., Gürsel, Y., Sussman, G. J., and Wisdom, J., "The Outer Solar System for 200 Million Years," Astron. J. **92**, 176 (1985).

Arnold, V. I., "Small Denominators and Problems of Stability of Motion in Classical and Celestial Mechanics," Russian Math. Surv. **18**, 85 (1963).

Bretagnon, P., "Theéorie du mouvement de l'ensemble des planetes. Solution VSOP82," Astron. Astrophys. **114**, 278 (1982).

Cohen, C. J. and Hubbard, E. C., "Libration of the close approaches of Pluto to Neptune," Astron. J. **70**, 10 (1965).

Cohen, C. J., Hubbard, E. C., and Oesterwinter, C., "Elements of the outer planets for one million years," *Astronomical Papers of the American Ephemeris* **XXII**, I (1973).

Farmer, J. D., and Sidorowich, J. J., "Predicting Chaotic Time Series," Phys. Rev. Lett. **59**(8), 845 (1987).

Farmer, J. D., and Sidorowich, J. J., "Exploiting Chaos to Predict the Future and Reduce Noise," in *Evolution, Learning, and Cognition*, edited by Y. C. Lee (World Scientific, Singapore, 1988), p. 277.

Hammel, S. M., Yorke, J. A., and Grebogi, C., "Numerical Orbits of Chaotic Processes Represent True Orbits," Bull. Am. Math. Soc. **19**, 465 (1988).

Kostelich, E. J., and Yorke, J. A., "The Analysis of Experimental Data Using Time-Delay Embedding Methods," preprint (1989).

Klavetter, J. J., "Rotation of Hyperion. I. Observations," Astron. J. **97**, 570 (1989).

Laskar, J., "A Numerical Experiment on the Chaotic Behaviour of the Solar System," Nature **338**, 237 (1989).

Sussman, G. J., and Wisdom, J., "Numerical Evidence that the Motion of Pluto is Chaotic," Science **241**, 433 (1988).

Williams, J. G., and Benson, G. S., "Resonances in the Neptune–Pluto System," Astron. J. **76**(2), 167 (1971).

Wisdom, J., "The origin of the Kirkwood gaps: A mapping for asteroidal motion near the 3/1 commensurability," Astron. J. **87**, 577 (1982).

Wisdom, J., "Chaotic
(1983).

Wisdom, J., "A perturl
63, 272 (1985).

Wisdom, J., "Urey Priz
(1987).

Wisdom, J., Peale, S. J.
137 (1984).

Wisdom, J., "Chaotic behavior and the origin of the 3/1 Kirkwood gap," Icarus **56**, 51 (1983).

Wisdom, J., "A perturbative treatment of motion near the 3/1 commensurability," Icarus **63**, 272 (1985).

Wisdom, J., "Urey Prize Lecture: Chaotic Dynamics in the Solar System," Icarus **72**, 241 (1987).

Wisdom, J., Peale, S. J., and Mignard, F., "The chaotic rotation of Hyperion," Icarus **58**, 137 (1984).