Mon. Not. R. Astron. Soc. (1992) **259**, 505–518

On the reliability of gravitational *N*-body integrations

Gerald D. Quinlan[1] and Scott Tremaine[1,2]

[1] *Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St George Street, Toronto, Canada M5S 1A 7*
[2] *California Institute of Technology, Pasadena, CA 91125, USA*

Accepted 1992 May 26. Received 1992 May 5

ABSTRACT

In a self-gravitating system of point particles such as a spherical star cluster, small disturbances to an orbit grow exponentially on a time-scale comparable with the crossing time. The results of *N*-body integrations are therefore extremely sensitive to numerical errors: in practice it is almost impossible to follow orbits of individual particles accurately for more than a few crossing times. We demonstrate that numerical orbits in the gravitational *N*-body problem are often shadowed by true orbits for many crossing times. This result enhances our confidence in the use of *N*-body integrations to study the evolution of stellar systems.

Key words: methods: numerical – celestial mechanics, stellar dynamics – galaxies: kinematics and dynamics.

1 INTRODUCTION

The gravitational *N*-body problem is to find the solution to the equations of motion

$$\frac{\mathrm{d}^2\boldsymbol{r}_i}{\mathrm{d}t^2} = G\sum_{j\neq i} m_j \frac{\boldsymbol{r}_j-\boldsymbol{r}_i}{|\boldsymbol{r}_j-\boldsymbol{r}_i|^3}, \tag{1}$$

where m_i and $\boldsymbol{r}_i$ are the mass and position of particle i, $i=1, \ldots, N$. These equations describe a wide range of astrophysical problems, including the collision, collapse and long-term evolution of stellar systems. For an equilibrium cluster of characteristic radius R containing N particles of mass m, the virial theorem tells us that the rms velocity is $v^2 \sim GNm/R$. A perturbed cluster approaches dynamical equilibrium on a crossing time-scale $t_{\rm cr}=2R/v$, and then continues to evolve through two-body gravitational scattering on the longer relaxation time-scale $t_{\rm r} \sim Nt_{\rm cr}/\log N$. The challenge for the *N*-body simulator is to find an accurate solution of the equations of motion over the time-scale of interest.

The results of *N*-body integrations are extremely sensitive to changes in the initial conditions and to numerical errors. The first systematic study of this problem was published by Miller (1964), who integrated small ($N \leq 32$) *N*-body systems and showed that two systems started from slightly different initial conditions diverge exponentially on a short time-scale (see also Miller 1971). Miller noted that a numerical error made on one time-step is equivalent to a change in the initial conditions for the next time-step, and he questioned whether *N*-body integrations could be used to study the evolution of star clusters over relaxation time-scales because the results would be corrupted by numerical errors on a much shorter time-scale. Miller thought that the exponential divergence was too rapid to be accounted for by two-body encounters and had to result from a collective effect, but Standish (1968) showed that the divergence rate was reduced if the denominator of equation (1) was replaced by

$$(|\boldsymbol{r}_j-\boldsymbol{r}_i|^2+h^2)^{3/2}, \tag{2}$$

with a non-zero softening length h, and therefore concluded that the divergence resulted mainly from close encounters. The sensitivity to integration errors was demonstrated dramatically by Lecar (1968), who coordinated a study in which 11 different integrations (from eight observatories) of the same 25-body problem were compared. Starting from identical initial conditions the integrations proceeded for 2.5 crossing times, with an accuracy sufficient to conserve the total energy to one part in 10^4. At the end of the integrations, quantities such as the half-mass radius, the moment of inertia and the rms velocity differed among the integrations by as much as 100 per cent. Hayli (1970) repeated the experiment and showed that, even with the same program and same initial data, different computers gave significantly different results after a few crossing times, presumably because of the different numerical precisions of the computer.

The extreme sensitivity of the *N*-body problem to small changes in the initial conditions is a property shared by all chaotic dynamical systems. The sensitivity is quantified by a Lyapunov exponent γ which describes how the phase-space separation d between two nearby orbits grows exponentially with time (a more precise definition is given in Section 3.1):

$$d(t)/d(0) \sim \mathrm{e}^{\gamma t}. \tag{3}$$

Recent work on the growth of errors in *N*-body integrations has largely been directed towards determining the relation

between the Lyapunov exponent γ and the dynamical and relaxation time-scales. This question was first addressed by Gurzadyan & Savvidy (1986), who predicted that the Lyapunov exponent should vary with the number of particles as $\gamma^{-1} \sim N^{1/3} t_{\rm cr}$. Heggie (1988) argued that the Lyapunov exponent should vary as $\gamma^{-1} \sim t_{\rm cr}$, not $\gamma^{-1} \sim N^{1/3} t_{\rm cr}$, and recent numerical work appears consistent with Heggie's prediction (see, e.g., Kandrup & Smith 1991; Heggie 1991; Goodman, Heggie & Hut 1992). The Lyapunov exponent might have an additional logarithmic dependence on the number of particles ($\gamma^{-1} \sim t_{\rm cr}/\log N$ according to J. Goodman, as quoted in Heggie 1991), but this is difficult to detect in numerical experiments because the range of N is limited by practical considerations. Gurzadyan & Savvidy suggested that the exponential divergence of nearby trajectories was a new form of global relaxation occurring on a time-scale much shorter (for large N) than the standard relaxation time-scale $t_{\rm r}$, but this suggestion is incorrect, as Heggie (1991) has emphasized. It is not appropriate to compare the Lyapunov time-scale γ^{-1} with the relaxation time-scale $t_{\rm r}$ because the exponential growth of the phase-space separation $d(t)$ in equation (3) occurs only in the linear regime where the separation is small; once the separation becomes large (of order the interparticle separation or less) its growth slows down and equation (3) does not imply that a stellar system will relax on a time-scale shorter than $t_{\rm r}$.

The sensitivity of the N-body problem to small errors is an inherent property of the physical problem, not simply a numerical artefact (numerical techniques such as regularization can increase the accuracy of an integration but do not eliminate the chaotic nature of the problem). Because of this sensitivity there is no obvious way to estimate how accurately an N-body integration should be done. None of the usual tests for the accuracy of numerical integrations of regular orbits is satisfactory. Comparing the results with exact solutions is not possible because exact solutions exist for only a few artificial configurations. Comparing the results after more than a few crossing times with those of independent calculations (using different time-steps, programs, computers, etc.) almost always shows large differences, as was found by Lecar (1968), because of the exponential error growth. Reversing the velocities at the end of an integration and integrating backwards usually fails to recover the initial conditions, for the same reason (unless the integration method is explicitly reversible, in which case the test is meaningless). The most commonly used test is the conservation of integrals such as the total energy, but errors in the positions and velocities can be large even though the energy is conserved to high accuracy, and hence the meaning of this test is not clear.

Because of these difficulties, the results of N-body integrations are sometimes viewed with suspicion (Miller 1964). Most N-body practitioners understand that it is not possible to follow accurately the positions and velocities of the individual particles over many crossing times, and say that integration methods should be designed to be 'faithful' or reliable in some other way, such as yielding results that resemble (in some statistical sense) those that would have been obtained by advancing in time exactly a whole neighbourhood around the initial conditions (Press 1986), but little work has been done to check that N-body integrations satisfy such requirements. Smith (1977) did numerical integrations of varying accuracy for 16-body systems and found no significant differences in the statistical quantities describing the overall characteristics of the systems, provided that energy conservation was not grossly violated. Heggie (1991) did experiments on a somewhat larger system ($N = 100$) and reached a similar conclusion, but warned his readers that 'It may be that numerical simulations of adequate accuracy give consistent results only because they are all equally inaccurate ...' Even if we are interested in only statistical results, we still need an objective criterion to decide how accurate the integration has to be; otherwise it is tempting to use larger step-sizes and cruder integration algorithms until something is obviously wrong with the final state. The problem can be summed up by the question, to paraphrase Heggie (1991): 'How badly are we allowed to integrate?'

The difficulties described above are common to all numerical investigations of chaotic systems. The best justification of such investigations that we know of relies on the existence of shadow orbits. A shadow orbit is a true (error-free) orbit of a dynamical system that remains close to a noisy numerical orbit for a long time. A numerical orbit can be considered reliable if a shadow orbit exists, because the numerical orbit then closely follows some true orbit of the system. Mathematicians have known for many years that shadow orbits exist for a restrictive class of chaotic dynamical systems known as hyperbolic systems, but it has been shown only recently (Hammel, Yorke & Grebogi 1988, hereafter HYG) that shadowing can also work for the more common non-hyperbolic systems. In this paper we study the existence of shadow orbits for the gravitational N-body problem. In Section 2 we discuss shadow orbits in more detail and describe our method for determining whether a numerical orbit can be shadowed by a true orbit. We apply this test in Section 3 to single-particle orbits in an $N = 100$ Plummer model to demonstrate that typical numerical N-body orbits can be shadowed for times much longer than γ^{-1}. In the final section we discuss the results and their implications. Some technical details of our numerical method and some illustrative examples of shadow orbits are presented in the appendices.

2 SHADOW ORBITS

2.1 The shadowing of noisy orbits by true orbits

We start with a definition of a shadow orbit (following HYG). Consider a dynamical system whose evolution is determined by a discrete map $\boldsymbol{f}$ that advances the phase-space coordinates $\boldsymbol{p}$ from one time-step to the next: $\boldsymbol{p}_{n+1} = \boldsymbol{f}(\boldsymbol{p}_n)$. (There is no loss of generality in restricting our attention to discrete maps, because a continuous dynamical system described by a set of ordinary differential equations can always be integrated over a time $\Delta t = t_{n+1} - t_n$ to generate a map that advances the phase-space coordinates by time Δt. The formalism can easily be generalized to allow for variable step-sizes Δt_n, in which case the map $\boldsymbol{f}_n$ will depend on the step n.) A *true orbit* of the map is a set of points $\{\boldsymbol{p}_n\}_{n=0}^{N}$ that satisfies $\boldsymbol{p}_{n+1} = \boldsymbol{f}(\boldsymbol{p}_n)$ for $n = 0, \ldots, N-1$. Numerical calculations with floating-point arithmetic cannot produce true orbits because of the inevitable round-off and truncation errors that occur when iterating the map. We call a set of points $\{\boldsymbol{x}_n\}_{n=0}^{N}$ a ε-*pseudo-orbit* (or *noisy orbit*) of the map $\boldsymbol{f}$ if $|\boldsymbol{x}_{n+1} - \boldsymbol{f}(\boldsymbol{x}_n)| \le \varepsilon$

for $n=0, \ldots, N-1$. We say that the true orbit $\{\boldsymbol{p}_n\}_{n=0}^{N}$ δ-*shadows* the noisy orbit $\{\boldsymbol{x}_n\}_{n=0}^{N}$ if $|\boldsymbol{p}_n-\boldsymbol{x}_n|\leq\delta$ for all $n=0, \ldots, N$.

For a chaotic system, a noisy orbit rapidly diverges from the true orbit with the same initial conditions. Can we trust the noisy orbit to represent faithfully the dynamics of the system? The view we adopt in this paper is that we can if there is some true orbit that shadows the noisy orbit (with a suitably small shadow distance δ) over the time interval of interest. If a shadow orbit exists then the noisy orbit does follow a true orbit of the system, but one that starts from initial conditions slightly displaced from those of the noisy orbit.

We stress that the noise in the noisy orbit is assumed to influence the evolution of the orbit, i.e., the noise introduced on time-step n changes the initial conditions for the map at step $n+1$. Noise of this kind is sometimes called dynamical noise, to contrast it with observational noise (e.g., measurement errors in laboratory systems, or printing out the coordinates with less precision than the internal storage in a numerical calculation) which does not influence the subsequent evolution of the orbit. If the noise is purely observational then it is obvious that a shadow orbit exists.

As a simple example of shadowing, consider the equation of motion $x''=x$, or, in position–velocity phase space, $x'=v$, $v'=x$ (it is easier to present this example as a continuous system than as a discrete map). Suppose that we start from initial conditions $x=v=0$ at some time $t<0$ and integrate the equation exactly except for noise at time $t=0$ consisting of a discontinuous jump ε in the velocity. The noisy orbit is thus

$$x_{\rm n}(t)=\begin{cases}0 & \text{if } t<0,\\ \varepsilon\sinh(t) & \text{if } t\geq 0.\end{cases} \tag{4}$$

After the jump the noisy orbit diverges exponentially from the true orbit $x_{\rm tr}(t)=0$ starting from the same initial conditions. However, there is a shadow orbit $x_{\rm sh}(t)=\varepsilon e^{t}/2$, which is a true orbit that remains within $\varepsilon/\sqrt{2}$ (in phase space) of the noisy orbit at all times. Note that we construct the shadow orbit by combining the two independent solutions of the equation with the coefficient of the growing (or expanding) component $\exp(t)$ chosen to match the noisy orbit as $t\to\infty$ and the coefficient of the decaying (or contracting) component $\exp(-t)$ chosen to match the noisy orbit as $t\to-\infty$. This is the idea behind the numerical algorithm we use to find shadow orbits in more complex examples: match the beginning of the noisy orbit in the contracting directions and match the end in the expanding directions. It is easy to generalize this example to allow for noise occurring at more than one time. If the noisy orbit is

$$x_{\rm n}(t)=\sum_{t_j<t}\varepsilon_j\sinh(t-t_j) \tag{5}$$

then the shadow orbit is

$$x_{\rm sh}(t)=\frac{1}{2}e^{t}\sum_{j}\varepsilon_j e^{-t_j}, \tag{6}$$

and if the noise is bounded ($|\varepsilon_j|<\varepsilon$) the shadow orbit remains close to the noisy orbit for all time.

Contrast this example with the seemingly simpler case of motion in a uniform force field, $x''=g$. A noisy orbit of the form

$$x_{\rm n}(t)=\begin{cases}gt^2/2 & \text{if } t<0,\\ \varepsilon t+gt^2/2 & \text{if } t\geq 0,\end{cases} \tag{7}$$

cannot be shadowed by a true orbit for all time because any true orbit will have the form $x(t)=a+bt+gt^2/2$ and must diverge linearly from the noisy orbit in either the future or the past (or both), depending on the choice of b. Thus shadowing works better for problems like $x''=x$, where nearby orbits diverge exponentially with time, than it does for problems like $x''=g$, where nearby orbits diverge linearly.

Does shadowing always work for chaotic systems? The answer is no. After some time it may not be possible to find a true orbit that remains close to the noisy orbit. When this happens the noisy orbit is said to have separated from the true orbits, and the point where this happens is called a 'glitch'. Simple non-linear maps show examples of glitches where numerical errors can introduce new behaviour. Consider for example the (chaotic) one-dimensional logistic map $f(x)=1-2x^2$ on the interval $-1<x<1$ (Sauer & Yorke 1991). The interval $[-1, 1]$ is mapped on to itself if the map is computed exactly, but a point near $x=0$ can be mapped out of this interval if a numerical error causes $f(x)$ to be larger than 1. The noisy orbit then moves off towards $-\infty$ (if the subsequent evolution of the orbit is computed without error), and it is clear that no true orbit can shadow the noisy orbit when this happens. The occurrence of this glitch is a robust phenomenon [it occurs also for maps $f(x)=1-ax^2$ with $a\neq 2$]; Sauer & Yorke (1991) suggest that the same phenomenon occurs in higher dimensional chaotic dynamical systems, because of the folds caused by homoclinic tangencies and near-tangencies of stable and unstable manifolds. Additional examples of simple problems exhibiting glitches are presented in Appendix A.

In the example $x''=x$, shadowing worked for all time and no glitches occurred. This is always true for hyperbolic systems. The essential property of a hyperbolic system is that the tangent space at any point on a trajectory is decomposable into unstable (or expanding) and stable (or contracting) directions, with infinitesimal displacements in the unstable direction growing exponentially when followed forwards in time and displacements in the stable direction growing exponentially when followed backwards. Also, the angle between the stable and unstable directions must be uniformly bounded away from zero. If these conditions are met then it is possible to prove (Anosov 1967; Bowen 1975) that, given a shadow distance $\delta>0$, there exists an $\varepsilon>0$ such that any ε-pseudo-orbit can be δ-shadowed by a true orbit for all time.

Most non-linear dynamical systems are non-hyperbolic and the Anosov–Bowen theorem is not applicable. but it is remarkable that noisy orbits in these systems can often be shadowed by true orbits for long times – much longer than one might have guessed. As an example, consider the area-preserving map of the unit square on to itself, known as the standard map, one of the simplest non-trivial Hamiltonian systems:

$$J_{n+1}=J_n+(K/2\pi)\sin(2\pi\theta_n) \qquad (\text{mod } 1), \tag{8}$$

$$\theta_{n+1}=\theta_n+J_{n+1} \qquad (\text{mod } 1). \tag{9}$$

For the choice $K=3$ most of phase space is filled with chaotic orbits, and the separation between a noisy orbit and a true orbit typically increases by a factor of 3 with each iteration. Thus even if the noisy orbit is iterated with double-precision arithmetic (16 decimal places accuracy) it will be completely different from the true orbit starting from the same initial conditions after about 30 iterations. Yet Grebogi et al. (1990, hereafter GHYS) have shown that noisy orbits of the standard map can be shadowed by true orbits for much longer than this: in one example GHYS constructed a noisy orbit with one-step errors of order $\varepsilon = 10^{-14}$ and proved that this orbit was shadowed by a true orbit for $N = 10^7$ iterations, with shadow distance $\delta = 10^{-8}$ (their method of proof will be discussed in Section 2.3). The long shadowing time is impressive, considering the rapid rate at which nearby orbits diverge.

GHYS conjectured that, for a typical chaotic two-dimensional Hamiltonian map, noisy orbits with noise amplitude ε should be shadowed by true orbits for $N \sim 1/\sqrt{\varepsilon}$ iterations with a shadow distance $\delta \sim \sqrt{\varepsilon}$ (a similar conjecture for dissipative systems was made by HYG). The logistic map example $f(x) = 1 - 2x^2$ illustrates the motivation for the conjecture that $N \sim 1/\sqrt{\varepsilon}$ (Sauer & Yorke 1991): if the map is iterated with noise level ε the noisy orbit is in danger of being mapped outside the interval $[-1, 1]$ whenever the orbit approaches within $\sim\sqrt{\varepsilon}$ of the origin, and if this dangerous interval is sampled in proportion to its length we expect a glitch to occur on one out of every $\sim 1/\sqrt{\varepsilon}$ steps.

2.2 Numerical method for finding shadow orbits

We describe here the refinement procedure of HYG, which is the method we use to find shadow orbits in our N-body experiments. Given a noisy orbit, the procedure returns a less noisy ('refined') orbit that stays close to the original noisy orbit. The procedure can be iterated to converge on to a true orbit of the system, similar to the way in which Newton's method converges on to a root of a function. The papers of HYG and GHYS described the refinement procedure for two-dimensional systems only. We have generalized the procedure to work for Hamiltonian systems with more than two dimensions. The details of our generalization are described in Appendix B; we concentrate on the two-dimensional case here for ease of exposition.

We are given a noisy orbit $\{\boldsymbol{p}_n\}_{n=0}^N$ of a map $\boldsymbol{f}$, and our goal is to find a less noisy orbit $\{\tilde{\boldsymbol{p}}_n\}_{n=0}^N$ that remains uniformly close to the noisy orbit. We let $\boldsymbol{e}_{n+1}$ represent the one-step error:

$$\boldsymbol{e}_{n+1} = \boldsymbol{p}_{n+1} - \boldsymbol{f}(\boldsymbol{p}_n). \tag{10}$$

The refined orbit is constructed by setting

$$\tilde{\boldsymbol{p}}_n = \boldsymbol{p}_n + \boldsymbol{\Phi}_n. \tag{11}$$

The equation satisfied by $\boldsymbol{\Phi}_n$, using equations (10) and (11), is then

$$\boldsymbol{\Phi}_{n+1} = \boldsymbol{f}(\tilde{\boldsymbol{p}}_n) - \boldsymbol{e}_{n+1} - \boldsymbol{f}(\boldsymbol{p}_n), \tag{12}$$

where we have set $\tilde{\boldsymbol{p}}_{n+1} = \boldsymbol{f}(\tilde{\boldsymbol{p}}_n)$. Assuming $\boldsymbol{\Phi}_n$ to be small, we can expand $\boldsymbol{f}(\tilde{\boldsymbol{p}}_n)$ about $\boldsymbol{p}_n$ in a Taylor series to get $\boldsymbol{f}(\tilde{\boldsymbol{p}}_n) \simeq \boldsymbol{f}(\boldsymbol{p}_n) + L_n\boldsymbol{\Phi}_n$, where L_n is the linearized map.[1] Equation (12) becomes

$$\boldsymbol{\Phi}_{n+1} = L_n\boldsymbol{\Phi}_n - \boldsymbol{e}_{n+1}. \tag{13}$$

We assume that at each step the linearized map L_n has an expanding direction and a contracting direction, and we choose unit basis vectors $\boldsymbol{u}_n$ and $\boldsymbol{s}_n$ at each step to lie along these directions and to follow the linearized map, i.e.,

$$\boldsymbol{u}_{n+1} = L_n\boldsymbol{u}_n / |L_n\boldsymbol{u}_n|, \tag{14}$$

$$\boldsymbol{s}_{n+1} = L_n\boldsymbol{s}_n / |L_n\boldsymbol{s}_n|. \tag{15}$$

In practice the unstable vectors $\boldsymbol{u}_n$ are constructed by picking an arbitrary unit vector $\boldsymbol{u}_0$ and iterating equation (14) forwards, and the stable vectors are constructed by picking an arbitrary unit vector $\boldsymbol{s}_N$ and iterating equation (15) backwards. After a few steps the vectors $\boldsymbol{u}_n$ and $\boldsymbol{s}_n$ become nearly aligned with the unstable and stable directions, and become independent of the initial choices for $\boldsymbol{u}_0$ and $\boldsymbol{s}_N$.

The goal is to find $\{\boldsymbol{\Phi}_n\}_{n=0}^N$, and hence $\{\tilde{\boldsymbol{p}}_n\}_{n=0}^N$ by equation (11), in the coordinates $\{\boldsymbol{u}_n\}_{n=0}^N$ and $\{\boldsymbol{s}_n\}_{n=0}^N$. We represent $\boldsymbol{\Phi}_n$ and $\boldsymbol{e}_n$ as

$$\boldsymbol{\Phi}_n = \alpha_n\boldsymbol{u}_n + \beta_n\boldsymbol{s}_n, \tag{16}$$

$$\boldsymbol{e}_n = \eta_n\boldsymbol{u}_n + \zeta_n\boldsymbol{s}_n. \tag{17}$$

To find $\{\alpha_n\}_{n=0}^N$ and $\{\beta_n\}_{n=0}^N$ in terms of $\{\eta_n\}_{n=0}^N$ *and* $\{\xi_n\}_{n=0}^N$, rewrite equation (13) as

$$\alpha_{n+1}\boldsymbol{u}_{n+1} + \beta_{n+1}\boldsymbol{s}_{n+1} = L_n(\alpha_n\boldsymbol{u}_n + \beta_n\boldsymbol{s}_n) - (\eta_{n+1}\boldsymbol{u}_{n+1} + \zeta_{n+1}\boldsymbol{s}_{n+1}). \tag{18}$$

The unit vectors follow the linearized map. The substitution of equations (14) and (15) in (18) yields recursive relations for $\{\alpha_n\}_{n=0}^N$ and $\{\beta_n\}_{n=0}^N$, which are made computationally stable by calculating the α_n coefficients by starting at the end-point $n = N$, and the β_n coefficients by starting at the initial point $n = 0$:

$$\alpha_N = 0, \qquad \alpha_n = (\alpha_{n+1} + \eta_{n+1})/|L_n\boldsymbol{u}_n|, \tag{19}$$

$$\beta_0 = 0, \qquad \beta_{n+1} = |L_n\boldsymbol{s}_n|\,\beta_n - \zeta_{n+1}. \tag{20}$$

Once we have the refined orbit we can iterate the procedure. When the one-step errors become sufficiently small the number of significant digits in the refined orbit tends to double on each iteration, yielding a geometric convergence analogous to that of Newton's method.

The geometric convergence of the HYG procedure on to a true orbit of the system is an important feature that is lacking in most noise-reduction methods described in the literature. These methods produce a refined orbit that is less noisy than the original, but if iterated they do not converge on to a true orbit of the system. An example is the noise-reduction method of Kostelich & Yorke (1988). In this method the refined orbit is not constrained to be a deterministic orbit of a map; instead the deviation from determinism is minimized along with the deviation from the noisy orbit in a least-squares sense. The Kostelich–Yorke method was developed for the analysis of experimental time-series where the exact dynamics might be unknown, a task for which it is certainly suitable. The HYG procedure can be used as a noise-reduction tool in this manner (Hammel 1990), but it can also be used to search for shadow orbits, whereas the simpler noise-reduction methods cannot.

[1] For a discrete map, L_n is simply the Jacobian of the map at step n; for a system of ordinary differential equations L_n represents integrating the variational equations from step n to step $n+1$.

There is no guarantee that the HYG procedure will converge when applied to any given noisy orbit (if there were, then all noisy orbits could be shadowed by true orbits). Difficulties arise when the angle between the stable and unstable directions becomes small, which can result in large correction coefficients that lie outside the domain in which the linearization of equation (12) is valid. When this happens, the errors in the refined orbit near the trouble spot are no better (and often worse) than those of the original noisy orbit, and further iterations do not help. Farmer & Sidorowich (1991) have suggested a method to overcome this difficulty. Near a trouble spot the matrix of basis vectors becomes nearly rank deficient, and Farmer & Sidorowich therefore argue that singular value decomposition (SVD) is the best way to solve the equations. Unfortunately, SVD can be prohibitively slow when used on large problems. The Farmer–Sidorowich method requires inverting an $M \times M$ matrix, where $M=(2N-1)\,d$ with N being the number of time-steps and d the dimensionality of the phase space. Doing this by SVD requires of order M^3 operations and M^2 memory locations, whereas the simpler HYG procedure requires only of order M operations and M memory locations. As a compromise, Farmer & Sidorowich use a hybrid method in which SVD is used near the trouble spots and the simpler HYG procedure is used elsewhere. We have experimented with this but have not found it to work better than the HYG procedure for finding shadow orbits. For example, if we use the hybrid method on a noisy orbit that has a trouble spot in the middle we find, after iterating several times, that the errors in the refined orbit are greatly reduced (from those of the noisy orbit) away from the trouble spot but are unchanged or only slightly reduced at the trouble spot. For noise-reduction applications, this result is better than that of the HYG procedure – it would probably amplify the errors near the trouble spot – but for shadowing applications the result is unsatisfactory because if the errors near the trouble spot remain significant then we have not found a shadow orbit. Perhaps Farmer & Sidorowich were more enthusiastic about their method than we are because the problems they tested it on had observational noise and not dynamical noise.

2.3 Rigorous results on the existence of shadow orbits

Two questions arise from the preceding discussion. First, how do we know that the HYG refinement procedure converges on to a true orbit of the system? In practice we do calculations in double-precision arithmetic, so the errors in the refined orbit cannot be reduced to less than about one part in 10^{15}. How then do we know that the refined orbit would continue to improve if we could iterate the procedure further using higher precision arithmetic? Secondly, what if the refinement procedure fails and does not reduce the errors: do we then know for certain that the noisy orbit cannot be shadowed by a true orbit? How do we know that a better method would not find a shadow orbit?

The first question is the easier one to answer. It is possible to prove rigorously (using normal floating-point arithmetic with its limited precision) that a given noisy orbit can be shadowed by a true orbit. Hammel, Yorke & Grebogi (1987) did this for one-dimensional maps (such as the logistic map) using interval arithmetic. HYG used a two-dimensional generalization of this approach to construct a sequence of small parallelograms in phase space around the points of the noisy orbit in such a way that there had to be a true orbit that passed through them. A more practical method of proof for high-dimensional systems was developed by Sauer & Yorke (1991), based on the procedure for refining noisy orbits described in the previous section. If certain quantities evaluated at the points of the noisy orbit are not too large, then the Sauer–Yorke proof shows that the iterated application of the refinement procedure results in a sequence of refined pseudo-orbits with decreasing noise level whose limit is a true orbit close to the noisy orbit.

Our experiments with simple chaotic maps suggest that in most cases one can tell from the refinement procedure alone (without constructing a rigorous proof) whether a given noisy orbit can be shadowed. We have experimented with several different maps: the standard map, a four-dimensional map consisting of two coupled standard maps, and another four-dimensional map based on a crude model for the motion of a particle through a lattice of fixed gravitationally attracting particles. We programmed the refinement procedure using the symbolic manipulation language MAPLE and kept 100 decimal places of accuracy in all calculations. When iterating the refinement procedure on noisy orbits (with ε in the range 10^{13}–10^{-8}), we invariably found one of two outcomes: either the refinement would converge geometrically, with the maximum one-step error in the refined orbit becoming $\simeq 10^{-100}$ after a small number of iterations, or the refinement would stop reducing the maximum one-step error after just one or two iterations, with subsequent iterations yielding no further improvement and with the maximum one-step error being not much smaller (and often larger) than the corresponding error in the original noisy orbit. This strongly suggests (but does not prove) that if we can take a noisy orbit with errors of magnitude 10^{-3}–10^{-8} and, using double-precision arithmetic, refine it until the one-step errors approach the machine precision (e.g., 10^{-15}), then we can be confident that the refinement was converging on to a true orbit of the system and would have made the one-step errors arbitrarily small if we had been able to continue iterating with higher precision. This is the approach we adopt in our experiments on N-body orbits described in the next section.

The second question raised above is more difficult to answer. If our refinement procedure does not succeed in producing a refined orbit with one-step errors comparable with the machine precision, does this mean that the noisy orbit cannot be shadowed by a true orbit? We do not know of a general way to answer this question, but we can give two reasons why we believe that a failure of the refinement procedure usually signifies that a shadow orbit does not exist. First, we know from the simple examples discussed in Section 2.1 and Appendix A that glitches do occur and that they are not simply a failure of one specific refinement algorithm. Secondly, the logistic map example motivates the GHYS conjecture on the frequency of glitches (Section 2.1), and the numerical results we have obtained with the HYG refinement procedure are consistent with this conjecture. It is, however, possible that some noisy orbits we examine can be shadowed for longer than we think. Hammel (1990) and Farmer & Sidorowich (1991) say that they have sometimes been able to refine noisy orbits for which the HYG procedure would fail by perturbing the orbits away from the

trouble spots. For example, in the two-dimensional case Hammel recommends perturbing the noisy orbit transversally to the line of near-tangency between the stable and unstable directions when the angle between these directions becomes too small, and then repeating the refinement procedure. We have tried this but without any success; perhaps it is because our orbits contain dynamical noise whereas those of Hammel (1990) and Farmer & Sidorowich (1991) contained observational noise.

3 SHADOWING OF CHAOTIC ORBITS IN AN *N*-BODY SYSTEM

3.1 Numerical experiments with a discrete Plummer model

Searching for shadow orbits for large N-body integrations with all N particles moving simultaneously would require enormous computational resources. We have therefore examined a simpler problem in which a single particle moves through a cluster of N fixed particles. For the fixed cluster we chose 100 equal-mass particles at random from a spherical Plummer model, whose density varies with radius as $\rho(r)\sim(1+r^2/r_p^2)^{-5/2}$ with $r_p=3\pi/16$ (using the standardized units of Heggie & Mathieu 1986). A projection of the cluster on to the x–y plane is shown in Fig. 1. If the particles in this cluster were moving in virial equilibrium the central velocity dispersion would be $v_{\rm rms}(r=0)=\sqrt{8/3\pi}$. Our experiments consist of taking noisy orbits that start at the cluster centre with velocity $v_{\rm rms}(0)$ in a random direction and testing them to see for how long they can be shadowed by true orbits and how large the distance between the noisy and shadow orbits is.

We have verified that the orbits in this discrete Plummer model are chaotic. We measured the maximum Lyapunov

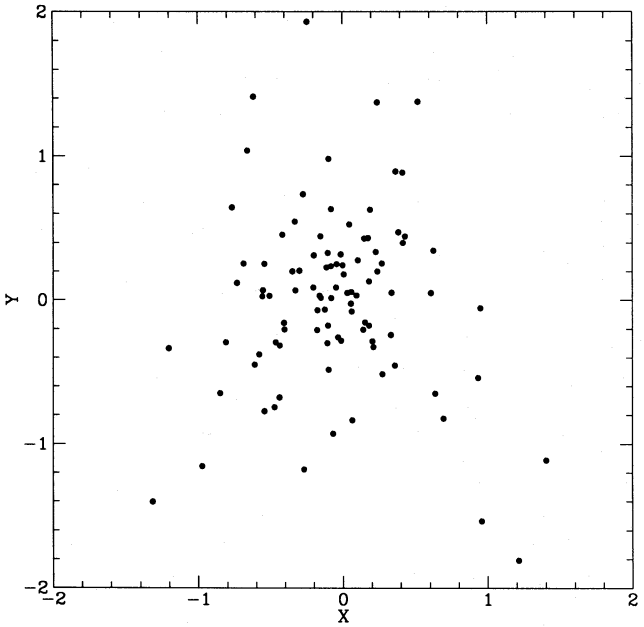

Figure 1. Projection of a 100-particle Plummer model on to the x–y plane. All physical quantities in this paper are measured in the standardized units of Heggie & Mathieu (1986), where $G=1$, $Nm=1$, and the gravitational potential energy of the cluster is $-1/2$.

exponent by the standard practice of integrating the variational equations. We digress for a moment to discuss the definition of the maximum Lyapunov exponent and some difficulties that result when this definition is applied to an N-body system.

If the equation of motion is ($\boldsymbol{z}$ denotes the phase-space coordinates)

$$\frac{d\boldsymbol{z}}{dt}=\boldsymbol{F}(\boldsymbol{z}) \tag{21}$$

then an infinitesimal displacement $\delta\boldsymbol{z}$ evolves according to the variational equation

$$\frac{d\delta z_i}{dt}=\frac{\partial F_i}{\partial z_j}\,\delta z_j, \tag{22}$$

where the partial derivatives on the right-hand side are evaluated on the unperturbed orbit. For a chaotic orbit $|\delta\boldsymbol{z}|$ grows exponentially with time and the maximum Lyapunov exponent γ is defined by

$$\gamma=\lim_{t\to\infty}\gamma_t=\lim_{t\to\infty}\frac{1}{t}\log\frac{|\delta\boldsymbol{z}(t)|}{|\delta\boldsymbol{z}(0)|}. \tag{23}$$

This gives the maximum Lyapunov exponent for almost all choices of the initial displacement $\delta\boldsymbol{z}(0)$; it can fail for special choices of $\delta\boldsymbol{z}(0)$ but these choices form a set of measure zero. In the non-linear dynamics literature, definition (23) is usually applied to dissipative systems with a chaotic attractor on which the limit is independent of an orbit's initial conditions. The Lyapunov exponent γ is then called a 'global' exponent, whereas γ_t is called a 'local' exponent because its value will vary with position on the attractor.[2] The local exponents are often of more interest than the global exponent, however, because the distribution of local exponents on the attractor (for some relevant time t) shows the variations to be expected in the growth rate of small perturbations (Abarbanel, Brown & Kennel 1991).

In the N-body problem there is even more reason for preferring the local exponent γ_t over the global exponent γ. In some N-body problems it is not obvious that the limit in definition (23) exists, and in other problems where we know the limit exists it is not what we are interested in. An example of the latter is the three-body problem of celestial mechanics known as the Pythagorean problem, which starts with three particles at rest at the vertices of a Pythagorean triangle. The problem was studied numerically by Szebehely & Peters (1967), who followed the motion (in which numerous close encounters occur) until $t=69$ (in their units) when two of the three particles form a permanent binary and the third is ejected to infinity. The motion prior to $t=69$ certainly looks

[2]The local exponent γ_t defined by equation (23) will also vary with the initial displacement $\delta\boldsymbol{z}(0)$, unlike the limit γ which is independent of $\delta\boldsymbol{z}(0)$. The proper way to define the local maximum Lyapunov exponent is to integrate the variational equations for a set of initial displacement vectors that span the tangent space and to define γ_t using the maximum eigenvalue of the resulting linearized mapping, but the value of t in our experiments is large enough that the answer found in this way would not differ by much from that found by our simpler definition using one randomly chosen initial displacement $\delta\boldsymbol{z}(0)$.

chaotic; Dejonghe & Hut (1986) have shown that small displacements to the orbits grow exponentially, with the total amplification of a typical displacement from $t=0$ to 69 being approximately 10^{10}. But the final state of the system is not chaotic – a stable binary with a third particle moving off to infinity – and if we apply definition (23) we get $\gamma=0$. This is not of much use; we are interested in the exponential growth rate of small displacements prior to $t=69$, even if this is only a transient phase of the motion.

When we use the term 'Lyapunov exponent' in this paper, we therefore have in mind a local exponent γ_t measured over some relevant time-scale t during which small displacements are growing exponentially, although we usually drop the subscript t and write simply γ. Fig. 2 shows the exponents we measured in this way for orbits in our discrete Plummer model, plotted versus the softening length h used in the force law (see equation 2). For these single-particle orbits there is no difficulty in taking the limit $t\to\infty$ in definition (23), because there is no possibility of binary formation or escape to infinity. We have integrated for as long as 150 time units and have found that the error bars in Fig. 2 decrease as $t^{-\nu}$ with $\nu\simeq 0.5$, with no systematic change in the exponents as t increases (there is no reason why ν has to be 0.5; in the problems studied by Abarbanel et al. ν was found to lie in the range 0.5–1.0). The figure shows that γ levels off to a constant $\simeq 2.2\pm 0.2$ at small h and tends towards zero as h increases. Goodman et al. (1992) predict that the transition between these extremes occurs at $h\sim R/\sqrt{N}$, with R being the radius of the cluster. This prediction cannot be tested with our cluster because the number of particles in the cluster core is too small for there to be much of a difference between N and $\sqrt{N}$, but it is consistent with experiments we have done on larger clusters [we have tested single-particle orbits in Plummer models with $N=10^2$, 10^3 and 10^4, and found $\gamma\simeq 2$ at small h in all cases, consistent with Heggie's (1988) prediction that the exponential divergence in an N-body system occurs on a crossing time-scale]. Goodman et al. (1992) measured the Lyapunov exponent of an $N=256$ Plummer model and found $\gamma_t\simeq 3.7\pm 0.3$ at $t=10$. It is not surprising that they found a slightly higher exponent than we do because they were following the motion of all N particles, not just one particle.

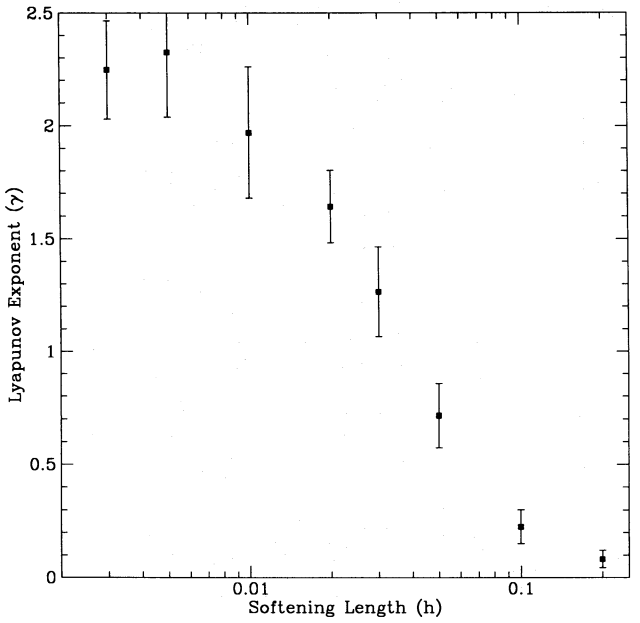

Figure 2. Lyapunov exponent γ as a function of the softening length h for single-particle orbits in the discrete Plummer model of Fig. 1. Each point and error bar shows the mean and standard deviation of the exponents measured from 20 orbits starting from the cluster centre with randomly chosen initial velocities [but with fixed speed $v=v_{\rm rms}(0)$], and followed for 18 time units.

The integrations described above were done by the Bulirsch–Stoer method. To generate noisy single-particle orbits for our shadowing experiments we wanted a numerical method similar to those used in real N-body integrations, so we chose the N-body routine by Aarseth listed in appendix 4B of Binney & Tremaine (1987), modified to keep the positions of all but one of the particles fixed. We refer to this routine as NBODY0, as it is a simplified version of the widely used NBODY1 routine. NBODY0 contains an accuracy parameter η that determines the time-step Δt for a given particle through the relation $\Delta t=(\eta F/\ddot{F})^{1/2}$, where F and $\ddot{F}$ are the force and the second time derivative of the force acting on the particle. The recommended value is $\eta=0.03$, which usually conserves the total energy of an N-body system to better than one part in 10^4 over one crossing time. The truncation errors in position and velocity after one time-step vary as $(\Delta t)^7$ and $(\Delta t)^6$ respectively, and hence the mean one-step phase-space error varies as $(\Delta t)^6$. (The routine NBODY0 does not estimate the error made on each step; there can be a variation of a factor of 10 or more than this mean, which will be important later for our interpretation of the shadowing results.)

It is common practice in N-body integrations to soften the force law by replacing the denominator of equation (1) by equation (2). This is sometimes done to decrease the importance of relaxation processes when an integration with small N (e.g., a few thousand) is being used to simulate a system with much larger N (e.g., the collision of two galaxies), although at other times softening is used simply as a crutch to prevent the time-step from becoming too small when the denominator of equation (1) approaches zero. Sophisticated N-body routines avoid the need for softening by regularizing the equations of motion when close encounters occur (see Aarseth 1985 and references therein), and the routine NBODY0 comes with an explicit warning that it was not designed to work well without softening, but in most of our experiments we have integrated the equations of motion (1) without softening to have a more challenging problem.

We construct a noisy orbit by starting a particle at the cluster centre with a random velocity direction and integrating for a given number of time-steps using NBODY0. The noisy orbit consists of the phase-space coordinates returned by the integrator at each time-step; the typical size of the one-step errors is $10^{-5}(\eta/0.03)^3$. We then apply the refinement procedure to the noisy orbit, noting the maximum one-step error in the refined orbit after each iteration; we stop when this maximum error has not been reduced by at least a factor of 0.9 for two successive iterations. The integrations required in the refinement procedure are done by the Bulirsch–Stoer method and are much more accurate (and time-consuming) than the original noisy integration (we use the routine of Press et al. 1986 with $EPS=10^{-13}$). If we can reduce the one-step errors in the refined orbit to 10^{-13} or better we

assume that a shadow orbit has been found, and we try again with a number of time-steps 1.5 times larger, continuing until we reach a number at which the refinement fails. We then use bisection twice to locate (to within about 10 per cent) the maximum number of time-steps for which the noisy orbit can be shadowed.

3.2 Results

We now present the results from our experiments. An example of a shadow orbit is shown in Fig. 3. The solid line is the noisy orbit of a particle starting at the cluster centre with velocity $v_{\rm rms}(0)$, integrated for 20.1 time units using the routine NBODY0 with accuracy parameter $\eta = 0.0272$ (typical of the accuracy used in large N-body integrations). The dashed line is a much more accurate Bulirsch–Stoer integration starting from the same initial conditions; this orbit soon deviates significantly from the noisy orbit and we stop plotting it after 3.25 time units. The dotted line is the shadow orbit, which starts from a position displaced from the cluster centre by about 10^{-5} and remains within a distance 0.042 (in phase space) of the noisy orbit throughout the integration. Efforts to shadow the noisy orbit for longer than shown in Fig. 3 were not successful; the glitch occurred at a close encounter with the fixed particle at the end of the noisy orbit in the figure.

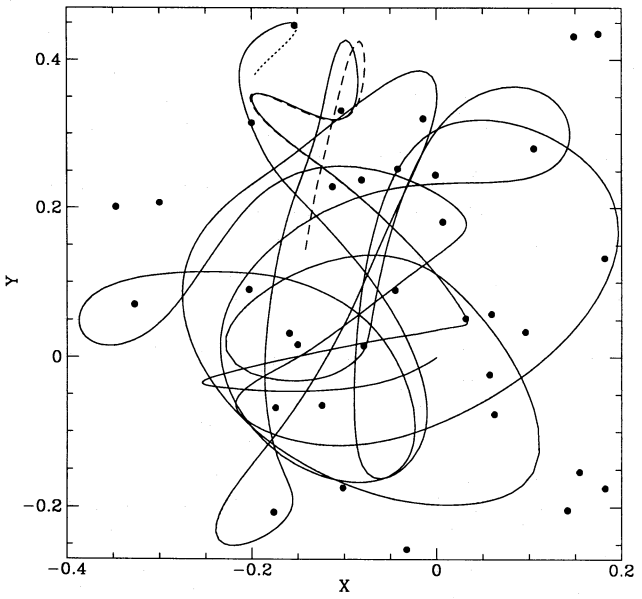

Figure 3. An example of a shadow orbit. The solid line is the projection on to the x–y plane of the noisy orbit of a particle starting at the origin of the cluster in Fig. 1 with velocity $(v_x, v_y, v_z) \simeq (-0.57, -0.46, 0.56)$ integrated for 20.1 time units with accuracy parameter $\eta = 0.027$. The filled circles are the projected positions of the fixed particles in the cluster. The dashed line is an accurate Bulirsch–Stoer integration starting from the same initial conditions, integrated for 3.25 time units. The dotted line is the shadow orbit. The phase-space separation between the shadow and noisy orbits reaches a maximum of 0.042 (resulting mostly from the separation in velocity, not position) near the point $x = 0.033$, $y = 0.051$. The shadow orbit can be seen extending from the end of the noisy orbit at the top of the figure (because the shadow orbit has been plotted for a slightly longer time interval).

The phase-space separation between the various orbits of Fig. 3 is shown in Fig. 4. We have split the phase-space separation into its position and velocity components, shown by the dotted and solid lines. The velocity separation is much more erratic and spiky than the position separation because the velocity is the derivative of the position. Fig. 4(a) shows that the separation between the noisy and accurate orbits grows exponentially with time, as expected for a chaotic system. Fig. 4(b) shows the separation between the noisy and shadow orbits. Note that the shadow distance is determined almost entirely by the separation in velocity, not position, and that at most time-steps the separation between the two orbits is considerably smaller than the maximum value of 0.042. The separation in position between the noisy and shadow orbits never gets larger than about 10^{-3} and does not grow with time.

The performance of the refinement procedure when applied to this noisy orbit is recorded in Fig. 5. The top line in the figure shows the one-step errors in the noisy orbit. Note that the errors vary by several orders of magnitude, probably a result of our using the simple NBODY0 routine without softening. The dotted line shows the errors in the refined orbit after three iterations of the refinement procedure. There are a few spikes in the dotted line, but the geometric mean error has decreased by more than 10^5. After five iterations the one-step errors in the refined orbit have been reduced to $< 10^{-14}$ everywhere (about as good as we can do with double-precision arithmetic), and they would presumably continue decreasing if we could continue iterating with higher precision. The total cpu time spent in iterating the refinement procedure five times to find the shadow orbit in Fig. 3 is about 1500 times larger than that spent in computing the original noisy orbit. One reason for this large difference is that the integrations in the refinement procedure are done much more accurately than the original noisy integration. Perhaps we could reduce the difference by work-

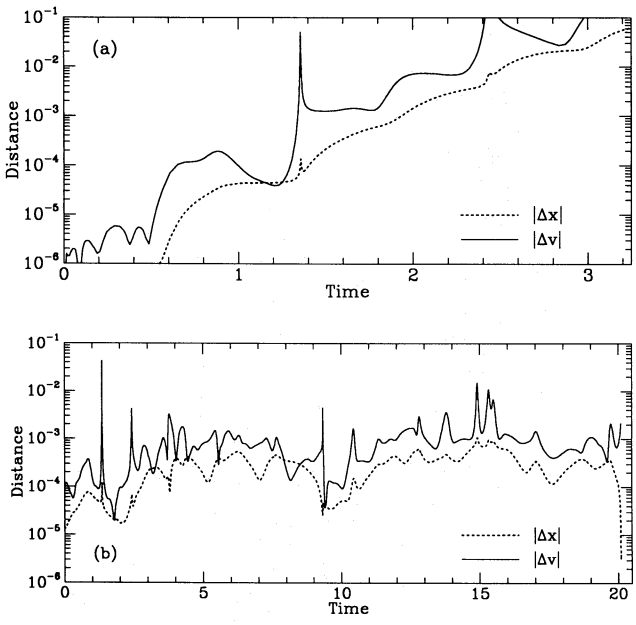

Figure 4. Phase-space separation between the orbits of Fig. 3: (a) separation between the noisy and accurate orbits; (b) separation between the noisy and shadow orbits.

ing hard to optimize the refinement calculations, but searching for shadow orbits in a system with more than a few phase-space dimensions is inevitably a time-consuming process.

A summary of the results from about 100 experiments on noisy orbits like that of Fig. 3 is shown in Fig. 6. Fig. 6(a) shows the maximum time T for which we were able to shadow each orbit as a function of the accuracy parameter η used in generating the noisy orbit. There is a clear trend of T increasing as η decreases (i.e., as the noise is reduced), although there is considerable scatter about this trend. This scatter is reduced in Fig. 6(b) where we plot instead of T the number of time-steps N, which need not be proportional to T because the step-size is reduced near close encounters. A least-squares fit to the data gives $N \sim \eta^{-1.03 \pm 0.07}$. A plot of the shadow distance δ versus η shows a large scatter, which is reduced somewhat if instead of η we use the maximum one-step error ε in the noisy orbit, as shown in Fig. 6(c). A least-squares fit gives $\delta \sim \varepsilon^{0.44 \pm 0.08}$, although there is so much scatter in the plot that the 1σ error we are quoting must be treated with some caution (a fit to the data minimizing the mean absolute deviation rather than the mean squared deviation gives $\delta \sim \varepsilon^{0.48}$).

It is difficult to say whether these results agree with the GHYS conjecture ($N \sim \varepsilon^{-1/2}$, $\delta \sim \varepsilon^{1/2}$), although there is certainly no gross inconsistency. The difficulty arises because of the large scatter in the plots, because of the limited range of η values over which we are able to do the experiments, and because of the variability in the one-step errors of the noisy orbits. The mean one-step error in the noisy orbits varies as $\langle \varepsilon_i \rangle \sim \eta^3$, but Fig. 6(d) shows that the maximum one-step error ε in the same orbits (some of which were integrated for longer than others) decreases much less rapidly with η (a least-squares fit gives $\varepsilon \sim \eta^{2.1 \pm 0.1}$) and can vary by a factor of 10 or more for a given value of η. It is not clear if we should pick $\langle \varepsilon_i \rangle$ or ε to compare with N and δ when testing the conjecture. We have plotted N versus η and δ versus ε simply because these combinations minimize the scatter in the plots. From the least-squares fit to Fig. 6(b) we find $N \sim \langle \varepsilon_i \rangle^{-0.34 \pm 0.02}$, and a similar fit to a plot of N versus ε yields $N \sim \varepsilon^{-0.36 \pm 0.04}$. These appear somewhat inconsistent with the GHYS conjecture that $N \sim \varepsilon^{-1/2}$, but the result from the fit in Fig. 6(c) is consistent with the GHYS conjecture that $\delta \sim \varepsilon^{1/2}$.

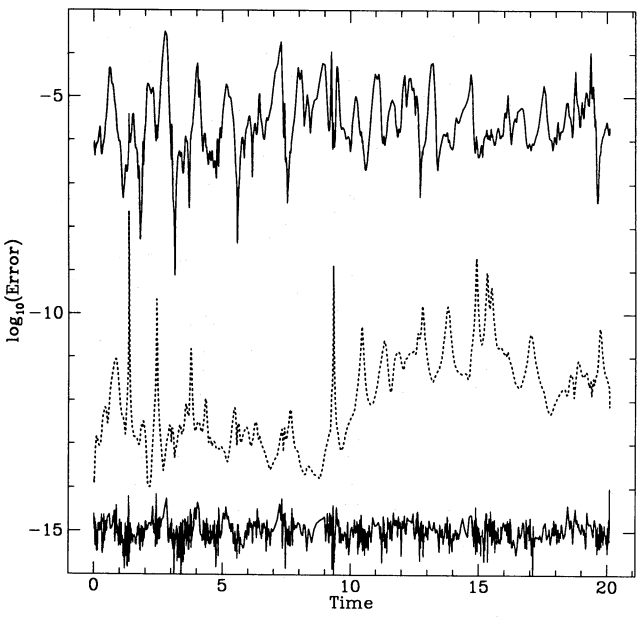

Figure 5. Progress of the refinement procedure when applied to the noisy orbit of Fig. 3. The solid line at the top shows the one-step phase-space errors in the noisy orbit. The dotted line shows the one-step errors in the refined orbit after three iterations of the refinement procedure; the solid line at the bottom shows the same errors after five iterations.

In nearly all of the noisy orbits examined in Fig. 6 the glitch that determined the maximum shadow time T occurred at a close encounter with one of the fixed particles, as it did in the noisy orbit of Fig. 3. The fraction of the total number of time-steps in a noisy orbit spent near close encounters is larger than the fraction of the total time, because the step-size is reduced near close encounters, but this alone is not sufficient to explain the glitch locations: the typical distance from a glitch location to the nearest fixed particle is smaller (by at least a factor of 10) than what would be expected if glitches occurred at randomly chosen time-steps. To study the importance of close encounters, we tried some experiments with softened force laws. Fig. 7 shows the results from about 30 such experiments, all done with a softening length $h = 0.03$ (which reduces the Lyapunov exponent by less than a factor of 2; see Fig. 2).

The experiments on the softened orbits differ from those on the unsoftened orbits in several ways. Fig. 7(a) and (b) show that the softened orbits can be shadowed for considerably longer than can the unsoftened orbits of Fig. 6. Glitches still occur in the softened orbits, but the correlation between glitch locations and close encounters is not nearly as strong as it is with the unsoftened orbits (the typical distance from a glitch location to the nearest fixed particle is about 3 times smaller than what would be expected if the glitches occurred at randomly chosen time-steps). There is much less scatter in the plots in Fig. 7 than there is in the unsoftened case, perhaps because there is no longer such a large variation in the individual step-sizes. A least-squares fit to the plot in Fig. 7(b) gives $N \sim \eta^{-1.5 \pm 0.1}$, yielding $N \sim \langle \varepsilon_i \rangle^{-0.5}$, in much better agreement with the GHYS conjecture than was the corresponding result for the unsoftened orbits (a fit to N versus ε gives $N \sim \varepsilon^{-0.65 \pm 0.08}$, but this plot again shows much more scatter than the plot of N versus η). The fit in Fig. 7(b) shows that at $\eta = 0.01$ the typical noisy orbit can be shadowed for 5000 time-steps, which is about 5 times larger than the corresponding number for unsoftened orbits in Fig. 6(b). Fig. 7(c) shows that the shadow distance for the softened orbits varies as $\delta \sim \varepsilon^{0.40 \pm 0.06}$, consistent with the GHYS conjecture that $\delta \sim \varepsilon^{1/2}$.

4 DISCUSSION

We have shown that noisy single-particle orbits in an $N = 100$ Plummer model can usually be shadowed by true orbits for times much longer than the time-scale (γ^{-1}) for the exponential growth of small errors. The more accurate a noisy orbit is, the longer it can be shadowed and the smaller the distance between the shadow and noisy orbits; our results are in reasonable agreement with the GHYS conjecture that the maximum number of time-steps N an orbit can be shadowed

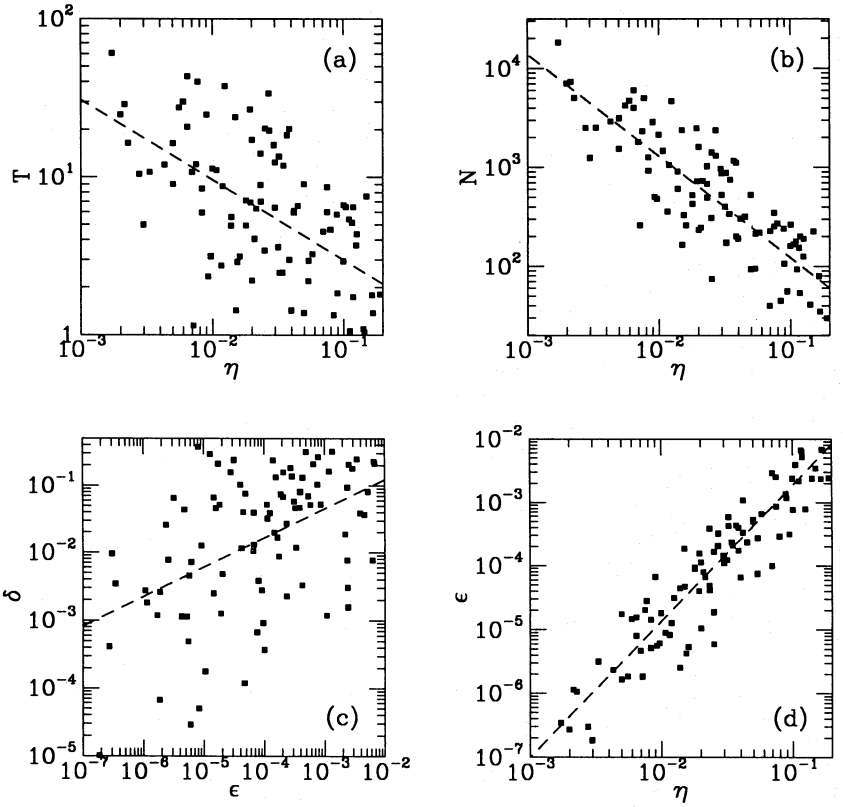

Figure 6. Results from attempts to shadow noisy orbits in the unsoftened Plummer model of Fig. 1 (each point represents a different orbit): (a) maximum time T for which the orbit could be shadowed versus the accuracy parameter η used to generate the orbit; (b) maximum number of time-steps N for which orbit could be shadowed; (c) shadow distance δ versus the maximum one-step phase-space error ε in the noisy orbit; (d) ε versus η. The dashed lines are least-squares fits to the data.

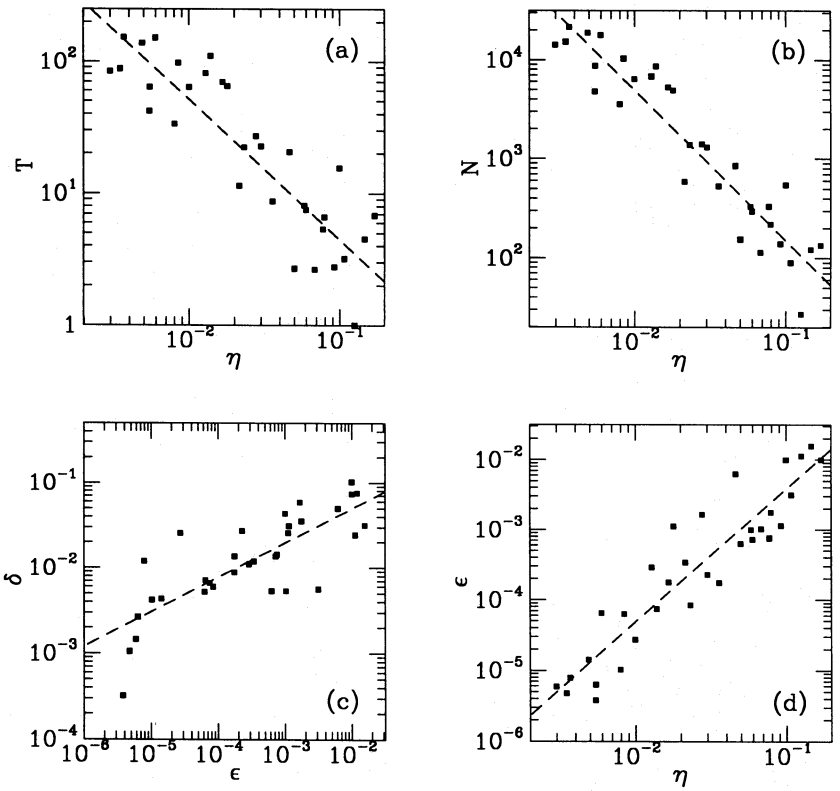

Figure 7. Results from attempts to shadow noisy orbits in the softened ($h = 0.03$) Plummer model of Fig. 1 (cf. Fig. 6).

for and the shadow distance δ vary with the noise level ε as $N \sim \varepsilon^{-1/2}$ and $\delta \sim \varepsilon^{1/2}$. We have shown that the single-particle orbits can be shadowed for even longer times if the force law is softened. The noisy orbits we tested were all generated by the same integration method (NBODY0), but we expect that similar results would be found with other integration methods.

We have tried to shadow N-body integrations with more than one moving particle, but we have not been able to gather meaningful statistics as we did for the single-particle orbits. We tried some experiments in which we picked two particles at random from the core of the cluster in Fig. 1 (giving them initial velocities chosen from the appropriate distribution for a Plummer model) and followed their motion (including both their self-interaction and their interaction with the fixed particles) using NBODY0. We were able to refine these noisy orbits to machine-level accuracy, sometimes over considerable time intervals, but it was a tedious process. We expected the work required to be about four times greater than it was for the single-particle orbits – because the computations in the refinement procedure increase as the square of the number of phase-space coordinates – but in practice it was often several times worse. More iterations of the refinement procedure were often required to reduce the errors to machine-level accuracy than were required for the single-particle orbits, for reasons that we do not fully understand.

It is difficult to extrapolate our results to predict how shadowing would work in large N-body systems. On the one hand, our comparison of softened and unsoftened single-particle orbits showed the former to be much easier to shadow. This suggests that noisy orbits in systems with large N might be easy to shadow, because increasing the number of particles decreases the importance of close encounters, as does softening the force law. On the other hand, it is possible that shadowing could become more difficult as the number of phase-space dimensions d of the system increases, i.e., it could be that as d increases the noise level has to be smaller for a noisy orbit of a given length to be shadowable. In any event, it is clear from the computing requirements alone that testing for the existence of shadow orbits cannot be used as a practical algorithm for monitoring the reliability of large N-body integrations.

It is also difficult to extrapolate our results to small N-body systems, but for a different reason. We remarked in Section 2.1 that shadowing works better the more chaotic a system is, and in fact is guaranteed to work for hyperbolic systems which are in some sense the most chaotic systems possible. The problem with small N-body systems is that they might not be chaotic enough. Shadowing clearly won't work well for the two-body problem, which is regular. The three-body problem sometimes can be highly chaotic (as in the Pythagorean problem discussed in Section 3.1), but at other times can be close to regular (if, for example, one of the bodies is well separated from the other two), and we expect that shadowing would work well for some configurations but not for others.

Our work on shadow orbits strengthens our belief in the validity of results from N-body integrations, but does not clear such integrations of all suspicion. The existence of shadow orbits does answer an important question of principle: it shows that meaningful results can be obtained from N-body integrations despite the chaotic nature of the problem. Shadowing also suggests why numerical integrations might correctly reproduce statistical properties of the evolution even though the integrations themselves are incorrect: if a numerical orbit can be shadowed then it is following some true orbit of the system, one that starts from initial conditions that are slightly different from those of the numerical orbit. We have not proved that shadow orbits are typical of true orbits chosen at random, but it would be surprising if they were not; the proof of shadowing given by GHYS shows that if a noisy orbit can be shadowed by one true orbit over a finite time interval then it can be shadowed by an infinite number of true orbits (although they are all squeezed into a small region of phase space), so shadow orbits are not unique. Shadowing motivates our belief in statistical results derived from N-body integrations, but does not obviate the need for independent studies to check the validity of these results. Perhaps the requirement that numerical orbits be shadowed by true orbits is too stringent for large N-body integrations, and some less restrictive requirements will suffice to ensure reliable statistical results.

ACKNOWLEDGMENTS

We thank Norm Murray, John Sidorowich and Simon White for helpful discussions. Financial support was provided by an operating grant from the Natural Sciences and Engineering Research Council of Canada, by NASA grant NAGW-1448, and by NSF grant AST-8913664. ST thanks the California Institute of Technology, where part of this work was done, for their hospitality and for the award of a Sherman Fairchild Scholarship. GDQ was supported in part by a Jeffrey L. Bishop Fellowship.

REFERENCES

Aarseth S. J., 1985, in Brackbill J. U., Cohen B. I., eds, Multiple Time Scales. Academic Press, Orlando, p. 377
Abarbanel H. D. I., Brown R., Kennel M. B., 1991, J. Nonlin. Sci., 1, 175
Abramowitz M., Stegun I. A., 1965, Handbook of Mathematical Functions. Dover, New York
Anosov D. V., 1967, Proc. Steklov Inst. Math., 90, 1
Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton, NJ
Bowen R., 1975, N. Diff. Equations, 18, 333
Dejonghe H., Hut P., 1986, in Hut P., McMillan S. L. W., eds, The Use of Supercomputers in Stellar Dynamics. Springer-Verlag, Berlin, p. 212
Farmer J. D., Sidorowich J. J., 1991, Physica D, 47, 373
Goodman J., Heggie D. C., Hut P., 1992, ApJ, submitted
Grebogi C., Hammel S. M., Yorke J. A., Sauer T., 1990, Phys. Rev. Lett., 65, 1527 (GHYS)
Gurzadyan V. G., Savvidy G. K., 1986, A&A, 160, 203
Hammel S. M., 1990, Phys. Lett. A, 148, 421
Hammel S. M., Yorke J. A., Grebogi C., 1987, J. Complex., 3, 136
Hammel S. M., Yorke J. A., Grebogi C., 1988, Bull. Am. math. Soc., 19, 465 (HYG)
Hayli A., 1970, A&A, 7, 249
Heggie D. C., 1988, in Roy A. E., ed., Long-Term Dynamical Behaviour of Natural and Artificial N-body Systems. Kluwer, Dordrecht, p. 329
Heggie D. C., 1991, in Roy A. E., ed., Predictability, Stability and Chaos in N-Body Dynamical Systems. Plenum Press, New York, p. 47

Heggie D. C., Mathieu R. D., 1986, in Hut P., McMillan S. L. W., eds, The Use of Supercomputers in Stellar Dynamics. Springer-Verlag, Berlin, p. 233
Kandrup H. E., Smith H., 1991, ApJ, 374, 255
Kostelich E. J., Yorke J. A., 1988, Phys. Rev. A, 38, 1649
Lecar M., 1968, Bull. Astron., 3, 91
Miller R. H., 1964, ApJ, 140, 250
Miller R. H., 1971, J. Comp. Phys., 8, 449
Press, W. H., 1986, in Hut P., McMillan S. L. W., eds, The Use of Supercomputers in Stellar Dynamics. Springer-Verlag, Berlin, p. 184
Press W. H., Flannery B. P., Teukolsky S. A., Vetterling W. T., 1986, Numerical Recipes. Cambridge Univ. Press, Cambridge
Sauer T., Yorke J. A., 1991, Nonlinearity, 4, 961
Smith H., 1977, A&A, 61, 305
Standish E. M., 1968, PhD thesis, Yale University
Szebehely V., Peters C. F., 1967, AJ, 72, 876

APPENDIX A: ILLUSTRATIVE EXAMPLES OF SHADOWING

A1 The parabolic cylinder differential equation

The differential equations that describe most physical systems are not hyperbolic, even though adjacent orbits may diverge exponentially if they are followed for long enough times. An instructive example is the equation of motion

$$x' = v, \qquad v' = x'' = (\tfrac{1}{4}t^2 + a)x, \tag{A1}$$

where a is a constant. Adjacent solutions to this differential equation diverge exponentially except when $a < 0$ and $t \lesssim 2(-a)^{1/2}$. The true solutions are parabolic cylinder functions $U(a, t)$, $V(a, t)$ (Abramowitz & Stegun 1965). If the initial conditions are taken to be $x = v = 0$ at some $t < 0$, and the noise consists of a discontinuous jump ε in velocity at $t = t_0 > 0$, then the noisy orbit is

$$x_n(t) = \begin{cases} 0 & \text{if } t < t_0, \\ \varepsilon(\pi/2)^{1/2}[U(a, t_0)\,V(a, t) - V(a, t_0)\,U(a, t)] & \text{if } t > t_0. \end{cases} \tag{A2}$$

As $t \to \infty$, $U(a, t) \to t^{-a-(1/2)} \exp(-\frac{1}{4}t^2)$ and $V(a, t) \to (2/\pi)^{1/2}\, t^{a-1/2} \exp(\frac{1}{4}t^2)$. Thus an orbit

$$x_{sh}(t) = \varepsilon(\pi/2)^{1/2}[U(a, t_0)\,V(a, t) + \alpha\,U(a, t)] \tag{A3}$$

will shadow the noisy orbit as $t \to \infty$, whatever the constant α may be. We now choose α so that the noisy orbit is also shadowed as $t \to -\infty$; to do so it is convenient to rewrite equation (A2) using the connection formulae

$$U(a, -t) = b_{uu}\,U(a, t) + b_{uv}\,V(a, t),$$

$$V(a, -t) = b_{vu}\,U(a, t) + b_{vv}\,V(a, t),$$

where $b_{vv} = -b_{uu} = \sin \pi a$, $b_{uv} = \pi/\Gamma(\frac{1}{2} + a)$, $b_{vu} = \Gamma(\frac{1}{2} + a)\cos^2(\pi a)/\pi$, and Γ denotes the gamma function. Setting $t = -\tau$ in equation (A3), we obtain

$$x_{sh}(-\tau) = \varepsilon(\pi/2)^{1/2}\{[U(a, t_0)\,b_{vv} + \alpha b_{uv}]\,V(a, \tau) + [U(a, t_0)\,b_{vu} + \alpha b_{uu}]\,U(a, \tau)\}. \tag{A4}$$

If x_{sh} is to shadow the noisy orbit as $t \to -\infty$ or $\tau \to \infty$, the coefficient of $V(a, \tau)$ must vanish, which requires $\alpha = -U(a, t_0)\,b_{vv}/b_{uv}$ or

$$x_{sh}(t) = \varepsilon(\pi/2)^{1/2}\,U(a, t_0)\left[V(a, t) - \frac{\sin \pi a}{\pi}\,\Gamma(\tfrac{1}{2} + a)\,U(a, t)\right]. \tag{A5}$$

For most values of the parameter a, the shadow orbit $x_{sh}(t)$ remains close to the noisy orbit $x_n(t)$, with shadow distance $O[t^{-a-1/2} \exp(-\frac{1}{4}t^2)]$ as $t \to \pm\infty$. However, for isolated values $a = a_k$, where $a_k \equiv -\frac{1}{2} - k$, $k = 0, 1, 2, \ldots$, the coefficient $\Gamma(\frac{1}{2} + a)$ diverges and there is no shadow orbit that remains close to the noisy orbit. Near $a = a_k$ the shadow distance is proportional to $|a - a_k|^{-1}$.

The absence of shadow orbits at isolated parameter values has a simple interpretation. The two linearly independent solutions of (A1) either grow or decay exponentially fast as $t \to \infty$. If the noise is localized near t_0, then a true orbit will shadow the noisy orbit as $t \to \infty$ if and only if the coefficient of its exponentially growing component is the same as that of the noisy orbit. Since the differential equation is time-reversal invariant, the same considerations apply as $t \to -\infty$. Thus, to find a shadow orbit over the interval $-\infty < t < \infty$, we must match the coefficients of the exponentially growing parts of the shadow orbit and noisy orbit both as $t \to \infty$ and as $t \to -\infty$. In general this can only be done if the solutions of the differential equations that grow as $t \to \pm\infty$ are linearly independent. An equivalent requirement is that there be no solution of the differential equation that decays to zero both as $t \to \infty$ and as $t \to -\infty$.

We now remark that equation (A1) is the Schrödinger equation for a harmonic oscillator potential. In this context, a solution that decays to zero as $t \to \pm\infty$ is simply a bound eigenstate. There is an absence of shadow orbits at the locations $a = a_k = -\frac{1}{2} - k$ because these are the energy eigenvalues of the bound states of the harmonic oscillator.

A2 A non-linear example

We now examine a non-linear but soluble example

$$x' = v, \qquad v' = x'' = x + bx^2\,\delta(t), \tag{A6}$$

where b is a constant. The singular delta function is somewhat unrealistic, since physical systems are generally continuous, but leads to simple algebra. The initial solution is taken to be $x = v = \exp(t)$ for $t < 0$, and the noise consists of a discontinuous jump ε in velocity at $t = t_0 > 0$. The noisy orbit is then

$$x_n(t) = \begin{cases} \exp(t) & \text{if } t < 0, \\ \cosh(t) + (1+b)\sinh(t) & \text{if } 0 < t < t_0, \\ \cosh t(1 - \varepsilon \sinh t_0) + \sinh t(1 + b + \varepsilon \cosh t_0) & \text{if } t > t_0. \end{cases} \tag{A7}$$

An arbitrary true orbit can be written in the form

$$x_{sh}(t) = \begin{cases} \alpha \cosh t + \beta \sinh t & \text{if } t < 0, \\ \alpha \cosh t + (\beta + b\alpha^2)\sinh t & \text{if } t > 0. \end{cases} \tag{A8}$$

If this true orbit is to shadow the noisy orbit (A7) as $t \to -\infty$ we require $\alpha = \beta$; to shadow the noisy orbit as $t \to \infty$ we need $\alpha + \beta + b\alpha^2 = 2 + b - \varepsilon \sinh t_0 + \varepsilon \cosh t_0$. Setting $\alpha = \beta = 1 + \delta$ we obtain the condition for a shadow orbit:

$$b\delta^2 + 2(b+1)\delta - \varepsilon \exp(-t_0) = 0. \tag{A9}$$

If the noise is small, $\varepsilon \ll 1$, then for most values of the parameter b a shadow orbit can be found, with $\delta \simeq \frac{1}{2}\varepsilon \exp(-t_0)/(b+1)$. If the parameter b is near -1, however, a shadow orbit may not exist: the explicit condition for the absence of a shadow orbit is $|b+1| < \varepsilon^{1/2} \exp(-t_0/2)$ for $0 < \varepsilon \ll 1$. The failure of shadowing is quite abrupt: just outside this interval in the parameter b, the shadow distance is only approximately $|\delta| = |b+1| \approx \varepsilon^{1/2} \exp(-t_0/2)$. Note that if the noise amplitude is $\varepsilon > 0$ the range of parameters over which a glitch occurs is $O(\varepsilon^{1/2})$, and the maximum shadow distance outside the glitch is also $O(\varepsilon^{1/2})$. Thus, as the noise is increased, shadowing fails and a glitch appears suddenly, even though the shadow distance is still small compared to unity.

APPENDIX B: THE MULTIDIMENSIONAL REFINEMENT PROCEDURE

We describe here our generalization of the HYG refinement procedure to work in an arbitrary (even) number of dimensions. The variables $\boldsymbol{e}_n$, $\tilde{\boldsymbol{p}}_n$ and $\boldsymbol{\Phi}_n$ have the same meaning as before (see equations 10–13). The construction of the stable and unstable unit vectors is a bit more complicated than in the two-dimensional case. We let $2D$ be the number of dimensions, and we assume that at each step the linearized map L has D expanding directions and D contracting directions. This is true for Hamiltonian systems, for which the expanding and contracting directions come in pairs (although there can be directions that are neither expanding nor contracting). For more general dynamical systems, however, the number of expanding and contracting directions need not be the same, and the procedure described below would have to be modified. The multidimensional generalizations of equations (16) and (17) are

$$\boldsymbol{\Phi}_n = \sum_{j=1}^{D} (\alpha_n^j \boldsymbol{u}_n^j + \beta_n^j \boldsymbol{s}_n^j), \tag{B1}$$

$$\boldsymbol{e}_n = \sum_{j=1}^{D} (\eta_n^j \boldsymbol{u}_n^j + \zeta_n^j \boldsymbol{s}_n^j). \tag{B2}$$

We first describe the construction of the unstable unit vectors and the solution for the α coefficients. The calculations are simplified if the unstable vectors form an orthonormal set at each step. We therefore start with an arbitrary orthonormal set of D vectors at step 0, and advance them forwards one step at a time by the linearized map, using the Gram–Schmidt process at each step to ensure that they remain orthonormal. Thus if $\boldsymbol{u}_n^1, \ldots, \boldsymbol{u}_n^D$ are the unstable vectors at step n, the vectors at step $n+1$ are found by

$$\boldsymbol{u}_{n+1}^1 = \frac{L_n \boldsymbol{u}_n^1}{|L_n \boldsymbol{u}_n^1|}, \tag{B3}$$

$$\boldsymbol{u}_{n+1}^2 = \frac{(\mathbf{1} - \boldsymbol{u}_{n+1}^1 \boldsymbol{u}_{n+1}^1) \cdot L_n \boldsymbol{u}_n^2}{|(\mathbf{1} - \boldsymbol{u}_{n+1}^1 \boldsymbol{u}_{n+1}^1) \cdot L_n \boldsymbol{u}_n^2|}, \tag{B4}$$

$$\boldsymbol{u}_{n+1}^{j}=\frac{(\mathbf{1}-\boldsymbol{u}_{n+1}^{1}\boldsymbol{u}_{n+1}^{1}-\ldots-\boldsymbol{u}_{n+1}^{j-1}\boldsymbol{u}_{n+1}^{j-1})\cdot L_n\boldsymbol{u}_n^j}{|(\mathbf{1}-\boldsymbol{u}_{n+1}^{1}\boldsymbol{u}_{n+1}^{1}-\ldots-\boldsymbol{u}_{n+1}^{j-1}\boldsymbol{u}_{n+1}^{j-1})\cdot L_n\boldsymbol{u}_n^j|}, \qquad j=3,\ldots,D. \tag{B5}$$

After a few steps we find that $\boldsymbol{u}_n^1$ is nearly aligned along the most unstable direction, $\boldsymbol{u}_n^2$ is nearly aligned along the second most unstable direction, and so on. The equation for the α coefficients is

$$\sum_{j=1}^{D}(\alpha_{n+1}^{j}+\eta_{n+1}^{j})\mathbf{u}_{n+1}^{j}=\sum_{j=1}^{D}\alpha_n^jL_n\boldsymbol{u}_n^j. \tag{B6}$$

If we project out the component along $\boldsymbol{u}_{n+1}^i$, we find

$$\alpha_{n+1}^{i}+\eta_{n+1}^{i}=\sum_{j=i}^{D}\alpha_n^j\boldsymbol{u}_{n+1}^{i}\cdot L_n\boldsymbol{u}_n^j, \tag{B7}$$

which is made computationally stable by starting at step N and working backwards, using

$$\alpha_n^i=\frac{1}{\boldsymbol{u}_{n+1}^{i}\cdot L_n\boldsymbol{u}_n^i}\left(\alpha_{n+1}^{i}+\eta_{n+1}^{i}-\sum_{j>i}\alpha_n^j\boldsymbol{u}_{n+1}^{i}\cdot L_n\boldsymbol{u}_n^j\right), \qquad \alpha_N^i=0. \tag{B8}$$

We first use (B8) to solve for $\{\alpha_n^D\}_{n=0}^N$ which does not require knowledge of the other α coefficients, then we solve for $\{\alpha_n^{D-1}\}_{n=0}^N$, and so on.

The construction of the stable vectors and the solution for the β coefficients must be done in the opposite direction. We start with an arbitrary set of D orthonormal vectors at step N, and advance them backwards one step at a time by the linearized map, using the Gram–Schmidt process at each step to ensure that they remain orthonormal.[3] Thus if $\boldsymbol{s}_{n+1}^1,\ldots,\boldsymbol{s}_{n+1}^D$ are the stable vectors at step $n+1$, the vectors at step n are found by[4]

$$\boldsymbol{s}_n^1=\frac{L_n^{-1}\boldsymbol{s}_{n+1}^1}{|L_n^{-1}\boldsymbol{s}_{n+1}^1|}, \tag{B9}$$

$$\boldsymbol{s}_n^2=\frac{(\mathbf{1}-\boldsymbol{s}_n^1\boldsymbol{s}_n^1)\cdot L_n^{-1}\boldsymbol{s}_{n+1}^2}{|(\mathbf{1}-\boldsymbol{s}_n^1\boldsymbol{s}_n^1)\cdot L_n^{-1}\boldsymbol{s}_{n+1}^2|}, \tag{B10}$$

$$\boldsymbol{s}_n^j=\frac{(\mathbf{1}-\boldsymbol{s}_n^1\boldsymbol{s}_n^1-\ldots-\boldsymbol{s}_n^{j-1}\boldsymbol{s}_n^{j-1})\cdot L_n^{-1}\boldsymbol{s}_{n+1}^j}{|(\mathbf{1}-\boldsymbol{s}_n^1\boldsymbol{s}_n^1-\ldots-\boldsymbol{s}_n^{j-1}\boldsymbol{s}_n^{j-1})\cdot L_n^{-1}\boldsymbol{s}_{n+1}^j|}, \qquad j=3,\ldots,D. \tag{B11}$$

After a few steps we find that $\boldsymbol{s}_n^1$ is nearly aligned along the most stable direction, $\boldsymbol{s}_n^2$ is nearly aligned along the second most stable direction, and so on.

The equation for the β coefficients can be written as

$$\sum_{j=1}^{D}(\beta_{n+1}^{j}+\zeta_{n+1}^{j})L_n^{-1}\boldsymbol{s}_{n+1}^j=\sum_{j=1}^{D}\beta_n^j\boldsymbol{s}_n^j. \tag{B12}$$

If we project out the component along $\boldsymbol{s}_n^i$ we find

$$\beta_n^i=\sum_{j=i}^{D}(\beta_{n+1}^{j}+\zeta_{n+1}^{j})\,\boldsymbol{s}_n^i\cdot L_n^{-1}\boldsymbol{s}_{n+1}^j. \tag{B13}$$

We solve for the β coefficients in the same way as we did for the α coefficients, except that this time we start at step 0 by setting

$$\beta_0^i=0 \tag{B14}$$

and iterate forwards. We first use (B13) to solve for $\{\beta_n^D\}_{n=0}^N$, (which does not require knowledge of the other β coefficients), then we solve for $\{\beta_n^{D-1}\}_{n=0}^N$, and so on.

[3]Although the stable vectors and the unstable vectors are both guaranteed to form orthonormal sets at each step, there is no guarantee that these two sets of vectors will span the two-dimensional space.

[4]For a discrete map L_n^{-1} is simply the inverse of the Jacobian L_n of the map at step n; for a map derived from a set of differential equations L_n^{-1} represents integrating the variational equations backwards from step $n+1$ to step n, so that $L_nL_n^{-1}$ is the identity operator.