JOURNAL OF COMPUTATIONAL PHYSICS 91, 490–494 (1990)

Note

A Counterexample of the Use of Energy as a Measure of Computational Accuracy

When solving a conservative dynamical system numerically, it is a common practice to equate accuracy of the energy of the numerical solution with the accuracy of the numerical solution itself. Our objective in this note is to produce a counterexample to such thinking.

Consider, then, the following elementary problem posed, but not solved, in the Feynman lectures [1]. In the XY plane, a particle of unit mass is positioned at (0.5, 0.0) and has an initial velocity of (0.0, 1.63). Its trajectory is to be determined if it is acted upon by a central, attractive force $\mathbf{F}$ whose magnitude F satisfies

$$F = 1/r^2. \tag{1}$$

The resulting conservative motion can be determined exactly by the methods of classical mechanics [4]. The motion is an ellipse whose major axis has length $2a$ and minor axis has length $2b$. One focus of the ellipse is at the origin, and one finds

$$a = \frac{10000}{13431}, \qquad b = \sqrt{\frac{664225}{1343100}} \tag{2}$$

the constant energy E of the system is

$$E = -\frac{1}{2a} = -0.67155, \tag{3}$$

the period τ of the motion is

$$\tau = \frac{2\pi}{\sqrt{(-2E)^3}} = \frac{2\pi}{(1.3431)^{3/2}}, \tag{4}$$

and parametric equations of the motion are

$$x = -\frac{6569}{26862} + \frac{10000}{13431}\cos t, \qquad y = \left(\frac{664225}{1343100}\right)^{1/2} \sin t, \qquad t \geqslant 0. \tag{5}$$

Note that the quantities a, b, E, τ given in (2)–(4) are *exact*, as is the solution by (5). The graph of (5) is the ellipse shown in Fig. 1.

0021-9991/90 $3.00

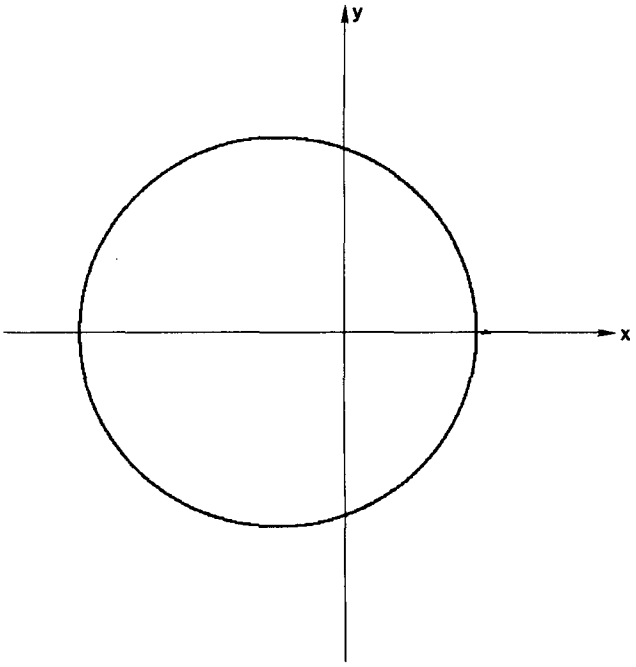

FIG. 1. The analytical orbit.

Next, let us formulate the problem dynamically and then solve it numerically. The differential equations of motion are

$$\frac{d^2x}{dt^2} = -\frac{1}{x^2+y^2}\cdot\frac{x}{(x^2+y^2)^{1/2}}; \qquad \frac{d^2y}{dt^2} = -\frac{1}{x^2+y^2}\cdot\frac{y}{(x^2+y^2)^{1/2}} \tag{6}$$

and the initial conditions are

$$x(0)=0.5, \qquad y(0)=0.0, \qquad v_x(0)=0.0, \qquad v_y(0)=1.63. \tag{7}$$

The central potential $\phi(r)$ associated with (1) is

$$\phi(r) = -1/r. \tag{8}$$

To proceed numerically, we first rewrite (6) as an equivalent first-order system:

$$\frac{dx}{dt} = v_x \tag{9}$$

$$\frac{dy}{dt} = v_y \tag{10}$$

$$\frac{dv_x}{dt} = -\frac{1}{x^2+y^2}\cdot\frac{x}{(x^2+y^2)^{1/2}} \tag{11}$$

$$\frac{dv_y}{dt} = -\frac{1}{x^2+y^2}\cdot\frac{y}{(x^2+y^2)^{1/2}}, \tag{12}$$

which, in the usual numerical notation, is approximated by

$$\frac{x_{k+1}-x_k}{\Delta t}=\frac{v_{k+1,x}+v_{k,x}}{2} \tag{13}$$

$$\frac{y_{k+1}-y_k}{\Delta t}=\frac{v_{k+1,y}+v_{k,y}}{2} \tag{14}$$

$$\frac{v_{k+1,x}-v_{k,x}}{\Delta t}=-\frac{1}{(x_k^2+y_k^2)^{1/2}(x_{k+1}^2+y_{k+1}^2)^{1/2}} \cdot\frac{x_{k+1}+x_k}{[(x_k^2+y_k^2)^{1/2}+(x_{k+1}^2+y_{k+1}^2)^{1/2}]} \tag{15}$$

$$\frac{v_{k+1,y}-v_{k,y}}{\Delta t}=-\frac{1}{(x_k^2+y_k^2)^{1/2}(x_{k+1}^2+y_{k+1}^2)^{1/2}} \cdot\frac{y_{k+1}+y_k}{[(x_k^2+y_k^2)^{1/2}+(x_{k+1}^2+y_{k+1}^2)^{1/2}]}. \tag{16}$$

Note that as $\Delta t \to 0$, (13)–(16) converge to their counterparts in (9)–(12).

For given Δt and $k=0, 1, 2, \ldots$, system (13)–(16) is an implicit, nonlinear system of four algebraic equations for the four unknowns $x_{k+1}, y_{k+1}, v_{k+1,x}, v_{k+1,y}$ in terms of the four knowns $x_k, y_k, v_{k,x}, v_{k,y}$. For $\Delta t=0.5$, the system was solved in

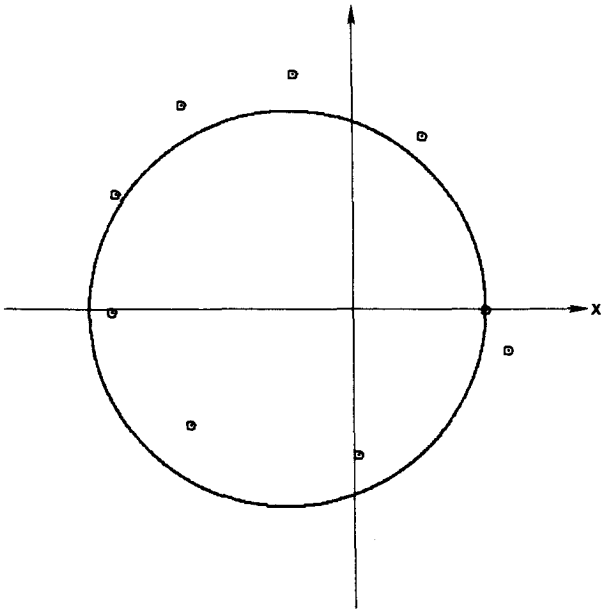

FIG. 2. The first numerical orbit.

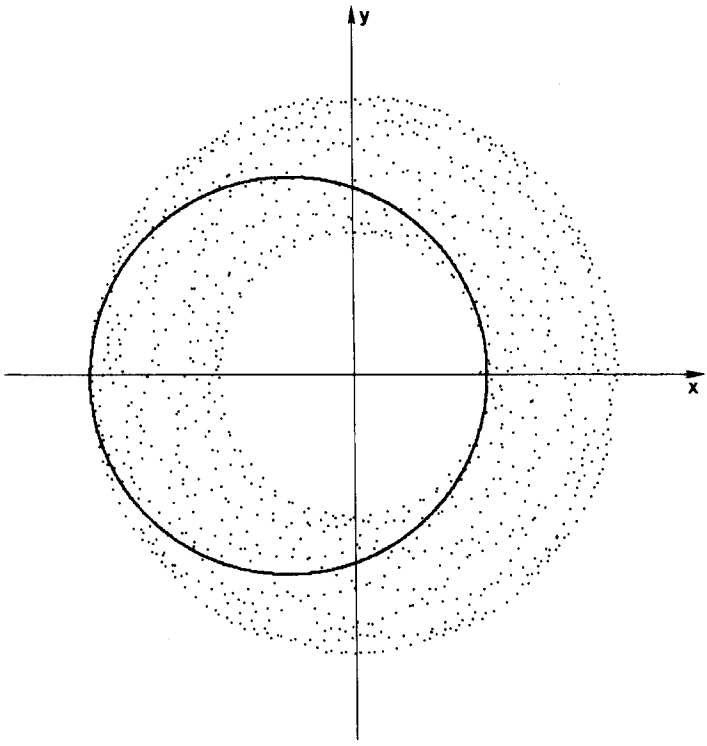

FIG. 3. The first 100 numerical orbits.

double precision on a VAX 8700 for each k by Newton's method with a convergence tolerance $\varepsilon = 10^{-10}$. The first orbit of the numerical solution is shown in Fig. 2, superimposed on the exact solution. The numerical solutions, up to and including $k = 1000$, consisted of more than 100 orbits, all of which have been plotted in Fig. 3, again superimposed on the exact solution. At each time t_k, the energy E_k was $E_k = -0.67155$, which coincides with the exact energy E given in (3). The results shown in Fig. 2 and 3 then confirm that accuracy in E_k is not equivalent to accuracy of the numerical calculations. In addition, Fig. 3 is consistent with the continuous result which states that for constant negative energy, orbits are bound by an annular region [3]. In Fig. 3, the annular region is $0.5 \leqslant r \leqslant a$.

Note that the fixed time step in the above example was chosen to ensure a relative large truncation error, while the fixed convergence tolerance was chosen to ensure a minimum roundoff error. Indeed, one can produce more striking effects simply by increasing Δt and calculating as above. The reason is that the energy E_k defined by (21)–(24) is *independent* of Δt and is *always* the same as E. This result follows because it is a special case of a more general energy invariance theorem [2].

REFERENCES

1. R. P. FEYNMAN, R. B. LEIGHTON, AND M. SANDS, *The Feynman Lectures on Physics* (Addison–Wesley, Reading, MA, 1963), p. 9–7.
2. D. GREENSPAN, *Arithmetic Applied Mathematics* (Pergamon, Oxford, 1980), p. 9.
3. L. D. LANDAU AND E. M. LIPSHITZ, *Mechanics* (Pergamon, Oxford, 1976), p. 33.
4. J. L. SYNGE AND B. A. GRIFFITH, *Principles of Mechanics* (McGraw–Hill, New York, 1942), p. 177.

RECEIVED: July 24, 1989; REVISED: November 15, 1989

DONALD GREENSPAN
Department of Mathematics
University of Texas at Arlington
Arlington, Texas 76019