REGULARIZING TIME TRANSFORMATIONS IN SYMPLECTIC AND COMPOSITE INTEGRATION

SEPPO MIKKOLA[1] and PAUL WIEGERT[2]
[1]*Turku University Observatory, Tuorla, 21500 Piikkiö, Finland*
[2]*Department of Physics and Astronomy, York University, 4700 Keele Street, Toronto, M3J 1P3, Ontario, Canada*

(Received: 25 April 2001; accepted: 5 November 2001)

Abstract. We consider numerical integration of nearly integrable Hamiltonian systems. The emphasis is on perturbed Keplerian motion, such as certain cases of the problem of two fixed centres and the restricted three-body problem. We show that the presently known methods have useful generalizations which are explicit and have a variable physical timestep which adjusts to both the central and perturbing potentials. These methods make it possible to compute accurately fairly close encounters. In some cases we suggest the use of composite (instead of symplectic) alternatives which typically seem to have equally good energy conservation properties.

Key words: time transformations, symplectic integration, three-body problem, planetary systems

1. Introduction

Symplectic integration is ideal for the study of the longterm behavior of nearly integrable dynamical systems. For planetary n-body problems the Wisdom–Holman (1991, WH hereafter) method (also Kinoshita et al., 1991) is the most popular. These methods require the use of a constant timestep in the integration and thus problems may arise due to close approaches which will not be properly resolved. Also a high eccentricity causes numerical instability in the WH-method (Rauch and Holman, 1999). Some ideas to deal with these problems have been suggested by Duncan et al. (1998), who use shorter step-size for close encounters, and by Mikkola (1997), Mikkola and Tanikawa (1999a, MT99 hereafter), and Preto and Tremaine (1999, PT99 hereafter) who utilize time transformations. In this paper we first review the use of time transformations and (starting in §3) present our new ideas.

2. Time Transformations

More than a century ago Poincare (Siegel, 1956, p. 35) introduced a technique to transform the independent variable, usually the time, in a Hamiltonian system. This was then used, for example, by Kustaanheimo and Stiefel (1965) (also Stiefel and Scheifele, 1971) for regularization of two-body motions with perturbations.

Celestial Mechanics and Dynamical Astronomy **82:** 375–390, 2002.

Later Mikkola (1997) applied the same idea in symplectic integration. Recently a new transformation, allowing time step adjustment, was found by Preto and Tremaine (PT99) and also by Mikkola and Tanikawa (MT99, 1999b). In the next two subsections we briefly review these methods.

2.1. POINCARE'S TIME TRANSFORMATION

Let $H(\mathbf{p}, \mathbf{q}, t)$ be a Hamiltonian. Take the time to be a canonical coordinate ($t = q_0$) by adding the momentum of time p_0 to the Hamiltonian. Thus we have, in the extended phase space, what Stiefel and Scheifele call the homogeneous Hamiltonian

$$H_h = H(\mathbf{p}, \mathbf{q}, q_0) + p_0, \tag{1}$$

which is a constant of motion since it does not depend on time. If we choose initially $p_0 = -H$, then numerically $H_h(t) = 0$ along the entire trajectory (since it is a constant).

Multiplication of the homogeneous Hamiltonian by the differential time transformation $\mathrm{d}t = g(\mathbf{p}, \mathbf{q})\mathrm{d}s$, where $g > 0$, leads to the new Hamiltonian

$$\Gamma = g H_h = g(\mathbf{p}, \mathbf{q})\,[H(\mathbf{p}, \mathbf{q}, q_0) + p_0]. \tag{2}$$

The time evolution of this new system is measured by the new independent variable s. If one forms the equations of motion using this Hamiltonian, drops the (zero) terms containing the factor H_h and divides all equations by $\mathrm{d}t/\mathrm{d}s = g$, the result is nothing but the usual Hamiltonian equations of motion. Thus one can conclude that the time transformed Hamiltonian (2) is equivalent to the original one displayed in (1).

Although above we used the notation q_0 for the time coordinate we will, in what follows, use the more familiar notation t. Yet for the momentum of time the symbol p_0 will be used throughout.

2.2. FUNCTIONAL TIME TRANSFORMATION

A new possibility for introducing time transformation was given by Mikkola and Tanikawa (MT99) who considered a homogeneous separable Hamiltonian $H_h = T(\mathbf{p}) + p_0 - U(\mathbf{q}, t)$ and wrote the new one in the logarithmic form

$$\Lambda = \ln(T + p_0) - \ln(U). \tag{3}$$

This can be shown to be equivalent to the original one provided one takes initially $p_0 = -(T - U)$. Here the time transformation is $\mathrm{d}t/\mathrm{d}s = \partial\Lambda/\partial p_0 = 1/(T + p_0) = 1/U$.

Preto and Tremaine (1999) noted the more general possibility that the logarithm can be replaced by any function $f(z)$ (which has a positive derivative $f'(z) > 0$)

$$\Lambda = f(T + p_0) - f(U). \tag{4}$$

In this case the time transformation is

$$\frac{\mathrm{d}t}{\mathrm{d}s} = f'(T + p_0), \tag{5}$$

which, along the correct orbit, is also $\mathrm{d}t/\mathrm{d}s = f'(U)$.

2.3. GENERALIZED LEAPFROG IN EXTENDED PHASE SPACE

The most popular symplectic integration method is the so called generalized leapfrog: One divides the Hamiltonian in two individually integrable parts

$$H = H_0 + H_1,$$

and the motion is approximated the leapfrog operator

$$\mathbf{L}(h) = \mathbf{H}_0\left(\frac{h}{2}\right)\mathbf{H}_1(h)\mathbf{H}_0\left(\frac{h}{2}\right), \tag{6}$$

where $\mathbf{H}_0(h/2)$ means advancing the system over a timestep of length $= h/2$ using H_0 as the Hamiltonian. Similarly $\mathbf{H}_1(h)$ means using H_1 only as the Hamiltonian and advancing it over a timestep of $= h$. In the WH-method H_0 is a Keplerian Hamiltonian and H_1 is the perturbation, which normally depends only on coordinates. In this case the $\mathbf{H}_0(h/2)$ step means pure Kepler motion over the timestep required, while the $\mathbf{H}_1(h)$ operation changes only the momenta by the amount $-h\,\partial H_1/\partial \mathbf{q}$. Often this is called the velocity jump.

In the extended phase space one can similarly divide the Hamiltonian in integrable parts and apply the generalized leapfrog. Mikkola (1997) used $g = r$ for the perturbed two-body problem

$$H = K - R = \frac{1}{2}\mathbf{p}^2 - \frac{M}{r} - R(t, \mathbf{r}),$$

which was time transformed into

$$\Gamma = r(K + p_0) - rR(t, \mathbf{r}). \tag{7}$$

Here the division $\Gamma_0 = r(K + p_0)$ and $\Gamma_1 = -rR(t, \mathbf{r})$ suggests itself. (Here and in what follows we mostly consider the perturbed two-body problem and, therefore, use the familiar notation $\mathbf{r}$, instead of $\mathbf{q}$, for coordinates.)

The Γ_0 step consists of calculating the motion defined by Γ_0 over a (fixed) step h_s of the new independent variable s. This motion turns out to be integrable, and in fact just the Kepler motion, but with a new effective mass

$$M_{\mathrm{eff}} = r\left(\tfrac{1}{2}\mathbf{p}^2 + p_0\right). \tag{8}$$

The motion can be explicitly calculated using standard two-body formulae (e.g. Stumpff, 1962), save that in place of the mass one must take the effective value

$M_{\rm eff}$ and the time increment is obtained by $\Delta t = \int_0^{h_s} r\,\mathrm{d}s$, which is a form of Kepler's equation.

The second part, advancement under Γ_1 (i.e. the velocity jump), simply consists of computing

$$\Delta\mathbf{p} = -h_s\frac{\partial\Gamma_1}{\partial\mathbf{r}} = h_s\frac{\partial(rR)}{\partial\mathbf{r}}; \qquad \Delta p_0 = -h_s\frac{\partial\Gamma_1}{\partial t} = h_s\frac{\partial(rR)}{\partial t}, \tag{9}$$

plus the addition of these jumps to $\mathbf{p}$ and p_0.

In the case of the functional time transformation with the Hamiltonian $\Lambda = f\left(\frac{1}{2}\mathbf{p}^2 + p_0\right) - f(M/r + R)$ the obvious division is

$$\Lambda_0 = f\left(\frac{1}{2}\mathbf{p}^2 + p_0\right); \qquad \Lambda_1 = -f\left(\frac{M}{r} + R\right) \tag{10}$$

and in this case, since Λ_0 only depends on the momenta and Λ_1 on the coordinates, the resulting algorithm is close to the standard leapfrog. For most functions $f(z)$ this algorithm is a rather rough approximation to the motion, however, the case $f(z) = \ln(z)$ is remarkable in that, if $R = 0$, it gives exact two-body orbit (MT99, PT99), except for a phase error. Thus one may expect good performance for weakly perturbed two-body orbits with this algorithm.

3. General Functional Time Transformation

Our new ideas are based on the observation that the Hamiltonian manipulation and the resulting time transformation can be generalized in several ways. Instead of dividing the Hamiltonian in the way that separates the momenta and coordinates (considered in §2.2) one may use any other division $H = H_0 - R$ and a function $f(z)$ to obtain

$$\Lambda = f(H_0 + p_0) - f(R). \tag{11}$$

Furthermore, one may use some other function $f(z, \mathbf{r}, \mathbf{p})$ and write

$$\Lambda = f(H_0 + p_0, \mathbf{r}, \mathbf{p}) - f(R, \mathbf{r}, \mathbf{p}), \tag{12}$$

which may also be shown to be equivalent to the original Hamiltonian.

The particular special case which we will consider here is a somewhat simpler function, namely

$$\Lambda = f(g(r)[K + p_0]) - f(g(r)R(t, \mathbf{r})), \tag{13}$$

where K is a Kepler Hamiltonian (but could in principle be any part of the original Hamiltonian) and R is the perturbing function. The separation into the Keplerian part and perturbation is, however, not necessarily the normal one. One may, for example, split the $1/r$ potential into two parts, placing one part into the perturbing

function. In some problems this turns out to be a useful alternative, and will be discussed in more detail later.

First we provide a simple demonstration that Λ can be used as the Hamiltonian: Following Preto and Tremaine (1999) we introduce a new time transformation function $\tilde{g}$ defined by

$$\tilde{g}(p_0, \mathbf{p}, t, \mathbf{r}) = g\frac{f(z_0) - f(z_1)}{z_0 - z_1}; \qquad z_0 = g(r)(K + p_0);$$
$$z_1 = g(r)R(t, \mathbf{r}), \tag{14}$$

where we see a difference quotient $(f(z_0) - f(z_1))/(z_0 - z_1)$ which reduces to the derivative $f'(z)$ in case $z_0 \to z_1 \to z$. Thus the fact that in the exact orbit $z_0 = z_1$ does not mean any singularity of $\tilde{g}$. When we use the time transformation function $\tilde{g}$ by multiplying the homogeneous Hamiltonian $K + p_0 - R$ with it we get $\tilde{g}(p_0, \mathbf{p}, t, \mathbf{r})(K + p_0 - R) = \Lambda$. Thus, the fact that Λ can be used as the Hamiltonian is proved in terms of our earlier arguments.

Let us now divide Λ into the two parts

$$\Lambda_0 = f(g(r)[K + p_0]); \qquad \Lambda_1 = -f(g(r)R(t, \mathbf{r})), \tag{15}$$

and consider the motion under the Hamiltonian Λ_0. Writing $\Gamma_0 = g(r)(K + p_0)$ we have the equation of motion for the time

$$\frac{\mathrm{d}t}{\mathrm{d}s} = f'(\Gamma_0)g(r). \tag{16}$$

From this, or by considering Hamilton's equation of motion, it is easy to see that the only difference to what we would obtain, for any of the dependent variables, using Λ_0 versus the basic Poincare-transformed Hamiltonian Γ_0, is just the additional factor $f'(\Gamma_0)$. On the other hand Γ_0 is constant over the integration step under consideration, thus the solution for the Λ_0-problem is just the same as that for Γ_0 save for the constant scaling factor in the (physical) timestep corresponding to the constant steplength in the new independent variable. Here, it is also important to realize that $\Gamma_0 \approx g(r)R(t, \mathbf{r})$ and thus we may consider the time transformation to be

$$\frac{\mathrm{d}t}{\mathrm{d}s} \approx f'(\, g(r)R(t, \mathbf{r})\,)g(r). \tag{17}$$

This is a crucial result, as it will allow the physical timestep to adjust in response both to the central potential and the perturbation. However, the details depend on our choice of the function f and R (how much, if any, of the $1/r$ potential has been moved to the perturbing function).

The various consequences of this are most simply demonstrated by considering some specific examples, such as the problem of two fixed centres and the restricted three-body problem.

3.1. TWO FIXED CENTRES

The problem of two fixed centres has the Hamiltonian

$$H = \frac{1}{2}\mathbf{p}^2 - \frac{m_1}{r_1} - \frac{m_2}{r_2}, \tag{18}$$

where $\mathbf{r}_k = \mathbf{r} - \mathbf{x}_k$ and $\mathbf{x}_1$ and $\mathbf{x}_2$ are the positions of the two attracting centres. We consider now especially the case when $m_2 \ll m_1$ which has some formal resemblance to the problem of motions in a planetary system. Let us write $M = m_1$, $m = m_2$, and write

$$H = \frac{1}{2}\mathbf{p}^2 - \frac{M}{r} - \frac{m}{\Delta}, \tag{19}$$

where we have taken $\mathbf{x}_1 = \mathbf{0}$ and thus $\Delta = |\mathbf{x}_1 - \mathbf{r}|$. At first we use $g(r) = 1$ and thus take for the functional Hamiltonian the form

$$\Lambda = f\big(\underbrace{\tfrac{1}{2}\mathbf{p}^2 - (M - \tilde{m})/r + p_0}_{\Gamma_0}\big) - f(\underbrace{\tilde{m}/r + m/\Delta}_{R}), \tag{20}$$

where $\tilde{m}$ is an adjustable constant, which splits the Kepler potential into two parts, one of which we consider part of Γ_0, the other part of R. For the special case $f(z) = \ln(z)$, the time transformation is

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{1}{\Gamma_0} \approx \frac{1}{R} = \frac{1}{(\tilde{m}/r + m/\Delta)} = \frac{r\Delta}{\tilde{m}\Delta + mr}, \tag{21}$$

which ensures that the physical timestep becomes small as either r or Δ go to zero. If, on the other hand we choose $g(r) = r$ and $f(z) = \ln(z)$ we obtain

$$\Lambda = \ln\big(\underbrace{r\big[\tfrac{1}{2}\mathbf{p}^2 - (M - \tilde{m})/r + p_0\big]}_{\Gamma_0}\big) - \ln\left(r\left[\frac{\tilde{m}}{r} + \frac{m}{\Delta}\right]\right), \tag{22}$$

which leads to

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{r}{\Gamma_0} \approx \frac{r}{(rR)} = \frac{r\Delta}{\tilde{m}\Delta + mr}. \tag{23}$$

Somewhat surprisingly we note that the time transformation looks the same. However, in the discrete algorithm this is not the case exactly and there is a significant difference between the two formulations. While carrying out the Λ_0 step in the first case (21), the physical time advances by the amount

$$\Delta t = \frac{h_s}{\Gamma_0}; \qquad \Gamma_0 = K + p_0,$$

which means that, to advance the motion, we must solve the Kepler's equation

$$\Delta t = r_0 X c_1(\beta X^2) + \mathbf{r}_0 \cdot \mathbf{v}_0 X^2 c_2(\beta X^2) + M_{\mathrm{eff}} X^3 c_3(\beta X^2), \tag{24}$$

for X, in terms of which the Kepler motion can be advanced by known formulae. Here $\mathbf{r}_0$, $\mathbf{v}_0$, are the values of coordinates and velocities at the beginning of the step, $\beta = 2M_{\text{eff}}/r_0 - v_0^2$ and c_k are Stumpff-functions. In the second case (22) it is the quantity $X(= \int_0^{\Delta t} \mathrm{d}t/r)$ that advances by the amount

$$X = \frac{h_s}{\Gamma_0}; \qquad \Gamma_0 = r(K + p_0),$$

and the Kepler's equation gives the corresponding amount of time increment by explicit evaluation of the expression (24).

We stress that the quantities Γ_0 are different in the two formulations by a factor of r, but the really significant point is that in the first case we must solve Kepler's equation (Δt is given) but in the second case we only compute the increment of the physical time using Kepler's equation (X is given).

A difficult question here is how to choose the adjustable constant $\tilde{m}$. Numerical experiments suggest that the best value is $\tilde{m} = m$, but we have not found any obvious reason for this. One should, however, note that if we set $\tilde{m} = 0$ and take $f(z) = \ln(z)$, then we will not have the $\mathrm{d}t \propto r$ scaling for small central body distance. On the other hand, if we use some other function, like $f'(z) = 1/\sqrt{1 + (z/m)^2}$, the said scaling remains.

One should note that the two-body motion defined by the Hamiltonian $\Lambda_0 = f\left(r\left[\frac{1}{2}\mathbf{p}^2 - (M - \tilde{m})/r + p_0\right]\right)$ is independent of the value of $\tilde{m}$ save for a scaling in the timestep. This is because, as we explained before, the motion is Keplerian with the effective mass $M_{\text{eff}} = r\left(\frac{1}{2}\mathbf{p}^2 + p_0\right)$ and this does not depend in any way on the value of $\tilde{m}$. Thus $\tilde{m}$ only affects the lengths of the timesteps.

Of course, the leapfrog must also advance the motion due to the remainder R, whichever formulation we use. However, this only consists of the velocity jump

$$\Delta\mathbf{p} = h_s f'(gR)\frac{\partial(gR)}{\partial\mathbf{r}}; \qquad \Delta p_0 = h_s f'(gR)\frac{\partial(gR)}{\partial t}. \tag{25}$$

In the considered case, however, p_0 will remain constant since R does not contain the time.

3.2. RESTRICTED THREE-BODY PROBLEM

3.2.1. *Heliocentric coordinates*

One of the possible forms of the restricted three-body problem Hamiltonian reads

$$H = \frac{1}{2}\mathbf{p}^2 - \frac{M}{r} - m\left(\frac{1}{\Delta} - \frac{\mathbf{r}\cdot\mathbf{r}_1}{r_1^3}\right), \tag{26}$$

where $\mathbf{r}_1$ is the position vector of the small mass m with respect to the big mass M. Similarly to the two fixed centres problem we can write the functional Hamiltonian $\Lambda = \Lambda_0 + \Lambda_1$, where

$$\Lambda_0 = f\left(r\left[\frac{1}{2}\mathbf{p}^2 - \frac{(M-\tilde{m})}{r} + p_0\right]\right) \tag{27}$$

$$\Lambda_1 = -f\left(r\left[\frac{\tilde{m}}{r} + \frac{1}{\Delta} - \frac{\mathbf{r}\cdot\mathbf{r}_1}{r_1^3}\right]\right). \tag{28}$$

Again one could take $f(z) = \ln(z)$ and get results that are very similar to those in the case of the two fixed centres. However, we note that in the problem of two fixed centres (with $f(z) = \ln(z)$) we have $\mathrm{d}t/\mathrm{d}s \approx 1/R$, and R is greater than zero. This may not be true in the present case owing to the indirect term in the potential which makes $R < 0$ at large distance. The method breaks down at this point unless we restrict ourselves to orbits with small enough r. Another possibility is to choose something else for $f(z)$ or rather, for its derivative $f'(z)$ in order that $f'(rR)$ remains positive. A reasonable looking alternative in this case is

$$f'(z) = \frac{1}{\left(1 + z/m + \sqrt{1 + (z/m)^2}\right)}, \tag{29}$$

which reduces for large z to $f'(z) \sim m/z$ and the method then behaves, in this limit, in a way similar to the $f(z) = \ln(z)$ alternative. In the case of large negative z we have simply $f \sim 1$, and the time transformation then approaches the simple $\mathrm{d}t/\mathrm{d}s \approx r$ law. However, at large distances the physical timestep would depend on the angle between $\mathbf{r}$ and $\mathbf{r}_1$, which does not seem reasonable. Also the physical timestep should not really be too big when the distance is large. This is because the perturbation is periodic and the timestep should remain a small fraction of that period. An alternative is to consider the system in the centre of mass frame, as in this case the indirect term in the potential does not appear.

3.2.2. *Inertial coordinates*

If the Hamiltonian is written in an inertial coordinate system, such as the centre-of-mass system, we have

$$H = \frac{1}{2}\mathbf{p}^2 - \frac{M}{r_1} - \frac{m}{r_2}, \tag{30}$$

where r_1 and r_2 are distances measured from M and m, respectively. In this case we may take

$$\Lambda_0 = f\left(r\left[\frac{1}{2}\mathbf{p}^2 - \frac{\tilde{m}}{r} + p_0\right]\right), \tag{31}$$

$$\Lambda_1 = -f\left(r\left[\frac{M}{r_1} + \frac{m}{r_2} - \frac{\tilde{m}}{r}\right]\right), \tag{32}$$

with arbitrary $\tilde{m}$ (note the different meaning from our earlier usage of the symbol $\tilde{m}$). Again the Λ_0 part gives always the same two-body motion independent of $\tilde{m}$.

An obvious possibility would be $\tilde{m} = 0$, but then the proximity of the small mass m becomes important in the time transformation only when $r_2 < mr_1/M$ which is too close for small m. The most reasonable choice seems to be $\tilde{m} \approx M$, but then we have similar difficulties (perturbing function may become zero or and/or negative) as in the M-centric formulation (26). Also very eccentric orbits are troublesome in this formulation because near the big primary the perturbations grow very large due to the choice of the reference point where there is actually no attracting body at all.

Since the centre-of-mass frame formulation presents its own difficulties, we conclude that the method may be most useful for orbits that remain into the neighborhood of the primaries. It may be possible to develop this method into a useful tool for studies of problems such as long term evolution of near Earth asteroids.

4. External Forces

If we consider the use of the proposed method(s) in Solar System simulations, we encounter a severe problem: the potential due to the planets, other than the one in vicinity of which the particle moves, can be so large that the time transformation does not shorten the step-size early enough and the regularizing effect is largely lost. For example, if we consider a particle moving near the Earth and take into account Jupiter as the perturber, then the potential due to Earth becomes comparable in numerical value to the part due to Jupiter only at a distance of the order of 1% of Sun's distance. This is too late and makes the method useless in this case. (Although the Hill sphere of the Earth is only about 0.01 (in our units) the trouble begins in fact at much larger distances.)

To make use of our method it seems necessary to treat the extra forces in the way suggested by Mikkola (1998) for inclusion of non-canonical forces. This means that the method is not any more symplectic, but a composite method. It remains, however, time symmetric and reversible which generally gives good behavior largely resembling that of symplectic methods. Let $\mathbf{F}$ be an extra acceleration to be added to the system represented by the Hamiltonian $\Lambda = f(g\,[K + p_0]) - f(gR)$. The only difference to the symplectic formulation is that we must add the extra terms due to $\mathbf{F}$ to the velocity jump equations. If we write $t' = gf'(gR)$, we obtain

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}s} = \frac{\partial f(gR)}{\partial \mathbf{r}} + t'\mathbf{F}, \tag{33}$$

$$\frac{\mathrm{d}p_0}{\mathrm{d}s} = \frac{\partial f(gR)}{\partial t} - \mathbf{p} \cdot (t'\mathbf{F}), \tag{34}$$

which must be solved over the s-interval of length $= h_s$ to obtain the momentum jumps. Since the coordinates t, $\mathbf{r}$ are constants in this operation, the (exact) solution is

$$\Delta\mathbf{p} = h_s \left[\frac{\partial f(gR)}{\partial \mathbf{r}} + t'\mathbf{F} \right], \tag{35}$$

$$\Delta p_0 = h_s \left[\frac{\partial f(gR)}{\partial t} - \langle \mathbf{p} \rangle \cdot (t'\mathbf{F}) \right], \tag{36}$$

where $\langle \mathbf{p} \rangle$ is the average of $\mathbf{p}$ over the step, that is, $\langle \mathbf{p} \rangle = \mathbf{p}(0) + \Delta\mathbf{p}/2$. Our numerical experiments show that this method indeed works well for most orbits. The only found exception, somewhat surprisingly, turned out to be the tadpole Trojan orbit where a weak secular error occurs.

5. Stable Two-body Advancement

We wrote experimental codes to test our ideas in the two simple cases of the two fixed centres and the restricted three-body problem. The two fixed centres problem is a good test case for its simplicity and also because it is easy to produce trajectories which develop high eccentricities, up to $e \to 1$.

Our first code used the Gaussian f, g formulation to advance the motion in the Kepler Hamiltonian part; however, we found that there is some numerical instability for $e \to 1$. This was not the one found by Rauch and Holman (1999) but something that exists even in the unperturbed case. The instability is not serious but clearly noticeable. This could be avoided by using the Kustaanheimo–Stiefel (1965) transformation (also Stiefel and Scheifele, 1971), but experiments also showed that the exact Kepler leapfrog (MT99, PT99), which is easy to program, performs well in this respect. We thus turned to use this alternative and describe the procedure here, including the necessary modifications to correct for the phase error of the basic method. One should note that this exact Kepler leapfrog is just another way of writing the solution for the two-body problem.

5.1. THE EXACT KEPLER LEAPFROG

The Kepler motion Hamiltonian in extended phase space is $K = \mathbf{p}^2/2 - M/r + p_0$, while the corresponding logarithmic form

$$\Lambda_K = \ln\left(\frac{\mathbf{p}^2}{2} + p_0\right) + \ln(r) - \ln(M), \tag{37}$$

gives the time-transformed equations of motion

$$\begin{aligned} \frac{\mathrm{d}t}{\mathrm{d}s} &= \frac{1}{(\mathbf{p}^2/2 + p_0)}, \qquad \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \frac{\mathbf{p}}{(\mathbf{p}^2/2 + p_0)}, \\ \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}s} &= -\frac{\mathbf{r}}{r^2}, \qquad \frac{\mathrm{d}p_0}{\mathrm{d}s} = 0, \end{aligned} \tag{38}$$

which can be solved exactly using the leapfrog algorithm, except for a time error. However, Preto and Tremaine (PT99) obtained an expression for the error of the time as well as for the precise meaning of the leapfrog step-size. After a minor rewriting of the PT99 expressions we have the exact Kepler leapfrog algorithm: We start with the values $\mathbf{p}_a, \mathbf{r}_a, t_a$ at the beginning of a step of length h_s, and calculate the steplength for the leapfrog by

$$h = \frac{m}{\sqrt{p_0/2}} \tan\left(\frac{h_s}{m}\sqrt{\frac{p_0}{2}}\right), \tag{39}$$

where

$$m = r_a\left(\frac{\mathbf{p}_a^2}{2} + p_0\right), \tag{40}$$

is an effective mass. (This is used instead of the true mass M, because mass does not appear explicitly in the equations of motion and thus it enters the formulation only via the value of p_0.) We proceed with

$$\mathbf{r}_{1/2} = \mathbf{r}_a + \frac{h}{2}\frac{\mathbf{p}_a}{(\mathbf{p}_a^2/2 + p_0)}, \tag{41}$$

$$\mathbf{p}_b = \mathbf{p}_a - h\frac{\mathbf{r}_{1/2}}{r_{1/2}^2}, \tag{42}$$

$$\mathbf{r}_b = \mathbf{r}_{1/2} + \frac{h}{2}\frac{\mathbf{p}_b}{(\mathbf{p}_b^2/2 + p_0)}, \tag{43}$$

$$t_b = t_a + \frac{h}{2}\left(\frac{1}{(\mathbf{p}_a^2/2 + p_0)} + \frac{1}{(\mathbf{p}_b^2/2 + p_0)}\right) + \tau(h, m, P_t), \tag{44}$$

to obtain the values $\mathbf{p}_b, \mathbf{r}_b, t_b$ at the end of the step. Here the time correction is $\tau = -h^3/(12m^2) + O(h^5)$, or precisely

$$\tau = -\frac{h_s^3}{4m^2}\frac{\tan(\sqrt{\zeta}) - \sqrt{\zeta}}{\sqrt{\zeta}^3}, \tag{45}$$

where

$$\zeta = \frac{h_s^2}{m^2}\frac{p_0}{2}. \tag{46}$$

In passing we mention that the expression for h can also be written $h = (m/\sqrt{p_0/2})\tan(u/2)$, where $u = E_b - E_a$ is the increment of the eccentric anomaly. Often it is convenient to use, instead of $s = m\int^t \mathrm{d}t/r$ the variable $X = \int^t \mathrm{d}t/r = s/m$ (because this is usually used in the f, g formulation). In terms of this we thus write for the leapfrog step-size parameter

$$h = \frac{m}{\sqrt{p_0/2}} \tan\left(X\sqrt{\frac{p_0}{2}}\right) = mX\mathrm{tg}_1\left(\frac{X^2 p_0}{2}\right), \tag{47}$$

where the latter form is formally independent of the orbit type (i.e. it does not contain the square root of p_0, so it is a real expression for unbound orbits too, which have $p_0 < 0$). The function $\mathrm{tg}_1(\zeta)$ is

$$\mathrm{tg}_1(\zeta) = \frac{\tan(\sqrt{\zeta})}{\sqrt{\zeta}} = 1 + \frac{\zeta}{3} + \frac{2\zeta^2}{15} + \frac{17\zeta^3}{315} + \cdots . \tag{48}$$

For this, as well as for the function

$$\mathrm{tg}_3(\zeta) = \frac{\tan(\sqrt{\zeta}) - \sqrt{\zeta})}{\sqrt{\zeta}^3} = \frac{1}{3} + \frac{2\zeta}{15} + \frac{17\zeta^2}{315} + \frac{62\zeta^3}{2835} + \cdots \tag{49}$$

needed in the expression (45) for τ, one can produce useful Pade approximants for easy evaluation (the notation tg_3 comes from $\tan(x) = x + x^3\mathrm{tg}_3(x^2)$). Using the Stumpff c-functions it is possible to express the tg-functions as

$$\mathrm{tg}_1(\zeta) = \frac{c_1(\zeta)}{c_0(\zeta)}; \qquad \mathrm{tg}_3(\zeta) = \frac{(c_2(\zeta) - c_3(\zeta))}{c_0(\zeta)}, \tag{50}$$

which is another convenient way to evaluate these functions.

Finally, we mention that for the unperturbed Kepler problem the effective mass equals the physical mass, that is, we have $m = M$, but in case of perturbations this is not true any more since p_0 is then determined by the total Hamiltonian.

5.2. TOY CODE EXPERIMENTS

We present here some results obtained using toy codes written according to the presented theory. Our toy system consists of a 'Sun' of mass = 1, an 'Earth' of mass $= 3.0 \times 10^{-6}$ in a circular orbit and in some cases a 'Jupiter' of mass $= 1.0 \times 10^{-3}$ which was replaced by its tidal force only.

In general the heliocentric method is clearly best for orbits with small pericenter distance, while for 'near Earth asteroids' all the regularized methods worked well with no significant difference in accuracy.

While many experiments were conducted, we present here only one in detail and tabulate results from several experiments to compare the present method with the original WH. First we illustrate results obtained with the toy model of a barycentric formulation plus the 'Jupiter' tidal field added as an external force. The Figure 1 illustrates the type of motion resulting from initial values: $\mathbf{r} = (1.02, 0, 0)$; $\mathbf{v} = (0, .99, 0.0001)$ and obtained with an initial timestep of = 0.005 years, initial phase of the planet = 1 radian and for the function f we used $f'(z) = 1/\sqrt{1 + (z/m)^2}$. The semi-major axis a, eccentricity e, and inclination i vary irregularly (but obey the approximate conservation of the Jacobian constant). The variation limits in this orbit for the orbital elements were found to be approximately $0.95 < a < 1.1$, $0 < e < 0.1$, $0 < i < 0.4°$. The total period of integration was 2000 years. In Figure 2 a particularly close encounter with the Earth is illustrated. In this case the

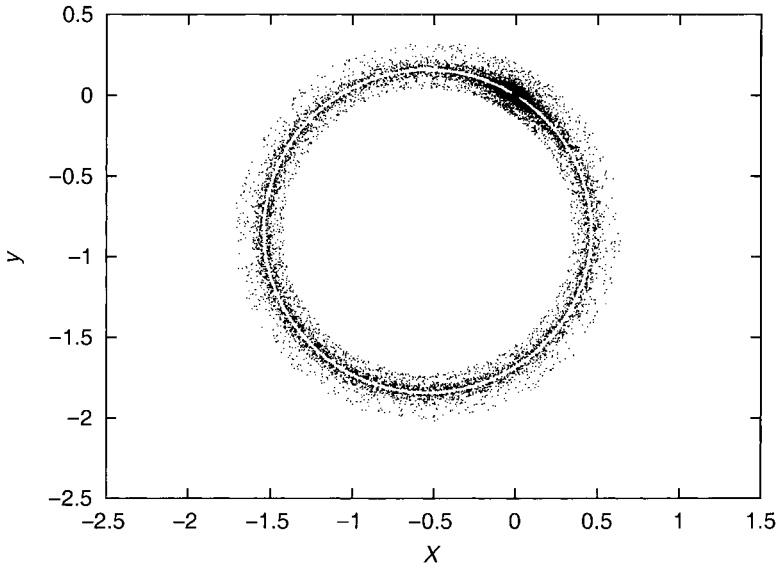

Figure 1. Snapshots of the system. The density of plotted points signify the frequency of force evaluations at any position. Complicated orbital motion and close approaches to the 'Earth' are evident.

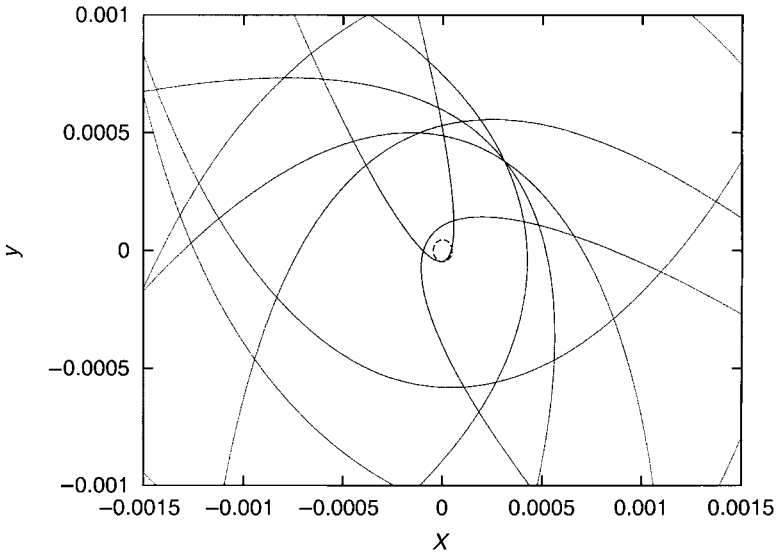

Figure 2. Orbital motion very near the Earth (the small circle represents the Earth). We see a close encounter that would actually have meant an impact on Earth. The figure was constructed by plotting a dot after every integration step. Thus we see the high density of force evaluations (which is due to the time transformation) in this critical region. The minimum timestep (in physical time) was just 7 s.

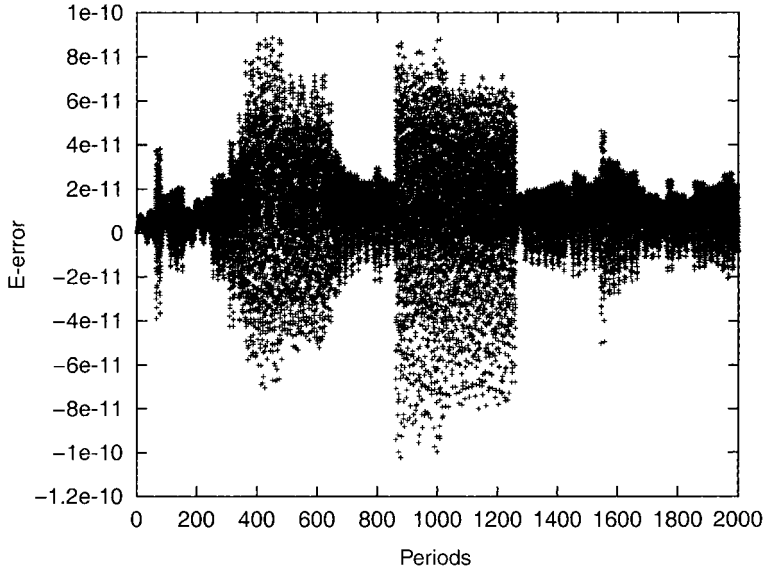

Figure 3. A measure of the energy error $= r\Delta(K + p_0 - R)$ as function of time (in years). The somewhat irregular pattern is a consequence of the frequent irregular changes in orbital elements of the particle.

particle would have actually impacted on Earth. The shortest physical timestep during the closest approach was just 7 s which is to be compared with the initial one of $\approx$1.8 days ($\sim$0.005 periods). Despite this kind of extremely close approaches our method conserves energy surprisingly well. This is illustrated in Figure 3. We see that the level of the error changes abruptly several times. This is due to the changes of the orbital elements of the particle, yet the error keeps fluctuating around zero without permanent changes of the level. Not surprisingly, the original WH-method fails in this case, even with much smaller stepsize.

To compare the behavior of the new- and the WH-method we carried out a series of experiments, the results of which are summarized in Table 1. The model used was the circular restricted three-body problem without any additional forces. The particles started on the x-axis, initial orbit was circular and in the same plane as that of the Earth. Thus only the semi-major axis was varied. The Earth had an initial phase of 1 radians as in the previous experiment. The initial stepsize in the integrations using the time-transformed method was adjusted such that the total number of force evaluations was comparable to that in the WH-computations. These numbers vary in the table since it is not possible to know in advance the average stepsize. The results clearly demonstrate the increased accuracy due to the use of the time-transformed version. Visual comparison of plots of the actual trajectories, with results obtained using a much shorter stepsize, revealed that in all these cases the orbits computed with the new method were significantly closer to the correct ones than the WH trajectories. These results clearly demonstrate that constant timesteps cannot be used for the close approaches considered here.

TABLE I

Results of integrations using the WH-method (top) and the new method (bottom)

a_0	N_f	$\langle dt \rangle$	err	$\Delta_{\min}$	type
WH					
0.975	26737	0.047	5.3E-08	1.7E-02	N
0.980	26737	0.047	3.0E-02	8.4E-04	N
0.985	26737	0.047	4.9E-06	3.1E-03	N
0.990	26737	0.047	9.4E-11	6.9E-02	HS
0.995	26737	0.047	8.0E-13	0.22	HS
1.000	26737	0.047	1.7E-14	0.96	T
1.005	26737	0.047	7.7E-13	0.22	HS
1.010	26737	0.047	8.4E-11	7.1E-02	HS
1.015	26737	0.047	9.6E-06	2.5E-03	N
1.020	26737	0.047	3.1E-02	8.0E-04	N
1.025	26737	0.047	7.0E-07	4.9E-03	N
New					
0.975	28830	0.044	4.5E-11	6.1E-04	N
0.980	27280	0.046	9.7E-11	8.8E-05	N
0.985	31850	0.039	5.7E-11	1.1E-03	N
0.990	24150	0.052	1.0E-12	6.9E-02	HS
0.995	22410	0.056	2.0E-12	0.22	HS
1.000	19670	0.064	1.1E-13	0.96	T
1.005	19800	0.063	2.0E-12	0.22	HS
1.010	24270	0.052	1.1E-12	7.1E-02	HS
1.015	33010	0.038	4.4E-11	9.9E-04	N
1.020	26330	0.048	3.1E-11	3.4E-04	N
1.025	24410	0.051	2.1E-10	8.7E-03	N

The displayed data are: the initial semi-major axis (a_0), total number of force evaluations (N_f), average timestep $\langle dt \rangle$, maximum error measure $\text{err} = |r\Delta(K + p_0 - R)|_{\max}$, minimum distance $\Delta_{\min}$ (at force evaluations) and type of orbit (N = NEA, HS = horseshoe, T = tadpole). The period of integration was 200 years.

6. Conclusions

Various explicit, symplectic (or at least composite time-reversible) methods suitable for integrating the few-body problem in the context of the Solar System have been discussed. We have demonstrated that algorithms, with constant stepsize in some new independent variable, can be made to adjust the physical timestep to both the central and any perturbing potential. This is an important advancement when considering systems in which close approaches are important (e.g. near-Earth asteroids, Jupiter family comets, etc). These improvements are achieved through the

use of time transformations, by splitting the Hamiltonian in novel ways, and/or by considering perturbers as non-canonical 'external' forces. As a result, the physical time step can vary in size over a range of several orders of magnitude in response to the potential in which the test body is moving. Numerical experiments with these improved algorithms show significant increases in precision over standard W-H methods during close encounters, without increases in computing time.

Acknowledgements

This work was partially supported by the Academy of Finland. The paper was written while one of the authors (P. Wiegert) was a guest at the Tuorla Observatory in Finland and this institution's generosity is gratefully acknowledged.

References

Duncan, J. M., Levison, H. F. and Lee, M. H.: 1998,'A multiple timestep symplectic algorithm for integrating close encounters', *Astron. J.* **116**, 2067–2077.

Kinoshita, H., Yoshida, H. and Nakai, H.: 1991, 'Symplectic integrators and their application in dynamical astronomy', *Celest. Mech. & Dyn. Astr.* **50**, 59–71.

Kustaanheimo, P. and Stiefel, E.: 1965, 'Perturbation theory of Kepler motion based on spinor regularization', *J. Reine Angew. Math.* **218**, 204–219.

Mikkola, S.: 1997, 'Practical symplectic methods with time transformation for the few-body problem', *Celest. Mech. & Dyn. Astr.* **67**, 145–165.

Mikkola, S.: 1998, 'Non-canonical perturbations in symplectic-integration', *Celest. Mech. & Dyn. Astr.* **68**, 249–255.

Mikkola, S. and Tanikawa, K.: 1999a [MT99], 'Explicit symplectic algorithms for time-transformed Hamiltonians', *Celest. Mech. & Dyn. Astr.* **74**, 287–295.

Mikkola, S. and Tanikawa, K.: 1999b, 'Algorithmic regularization of the few-body problem', *Mon. Not. R. Astr. Soc.* **310**, 745–749.

Preto, M. and Tremaine, S.: 1999 [PT99], 'A class of symplectic integrators with adaptive timestep for separable Hamiltonian systems', *Astron. J.* **118**, 2532–2541.

Rauch, K. P. and Holman, M.: 1999, 'Dynamical chaos in the Wisdom–Holman integrator: origins and solution', *Astron. J.* **117**, 1087–1102.

Siegel. C. L.: 1956, *Vorlesungen über Himmelsmechanik*, Springer, Berlin, Göttingen, Heidelberg.

Stiefel, E. L. and Scheifele, G.: 1971, *Linear and Regular Celestial Mechanics*, Springer.

Stumpff, K.: 1962, *Himmelsmechanik*, Band I, VEB Deutscher Verlag der Wissenschaften, Berlin.

Wisdom, J. and Holman, M.: 1991, 'Symplectic maps for the N-body problem', *Astron. J.* **102**, 1520–1538.