

EFFICIENT INTEGRATION OF HIGHLY ECCENTRIC ORBITS BY SCALING METHODS APPLIED TO KUSTAANHEIMO-STIEFEL REGULARIZATION

TOSHIO FUKUSHIMA

National Astronomical Observatory, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan; Toshio.Fukushima@nao.ac.jp

Received 2004 June 23; accepted 2004 August 26

ABSTRACT

We apply our single scaling method to the numerical integration of perturbed two-body problems regularized by the Kustaanheimo-Stiefel (K-S) transformation. The scaling is done by multiplying a single scaling factor with the four-dimensional position and velocity vectors of an associated harmonic oscillator in order to maintain the Kepler energy relation in terms of the K-S variables. As with the so-called energy rectification of Aarseth, the extra cost for the scaling is negligible, since the integration of the Kepler energy itself is already incorporated in the original K-S formulation. On the other hand, the single scaling method can be applied at every integration step without facing numerical instabilities. For unperturbed cases, the single scaling applied at every step gives a better result than either the original K-S formulation, the energy rectification applied at every apocenter, or the single scaling method applied at every apocenter. For the perturbed cases, however, the single scaling method applied at every apocenter provides the best performance for all perturbation types, whether the main source of error is truncation or round-off.

Key words: celestial mechanics — methods: numerical

1. INTRODUCTION

Recently we published a series of new methods to efficiently integrate perturbed two-body problems. They are classified into three categories: (1) the four methods of manifold correction, consisting of the single scaling method (Fukushima 2003a),¹ the dual scaling method (Fukushima 2003b),² the rotation method (Fukushima 2003c),³ and the linear transformation method (Fukushima 2004a),⁴ (2) three simplifications of the linear transformation method, consisting of two first-stage simplifications using nine variables per body (Fukushima 2004b),⁵ two second-stage simplifications using seven variables per body, and a final simplification using six variables per body, which we call the orbital longitude method (Fukushima 2004c);⁶ and (3) a couple of modifications of the orbital longitude method using the true and the antifocal orbital longitude, respectively, enhanced by a technique to reduce the accumulation of round-off error in the angle variables (Fukushima 2004d, 2004e).⁷ Table 1 provides a summary.

Among these methods, the last two exhibit the best cost performance, as we showed in Papers VII and VIII. For unperturbed cases, they produce only periodic errors at the machine-epsilon level if a sufficiently high order integrator is used with a sufficiently small step size (see Fig. 1 of Paper VIII). We experimentally confirmed that this is true even in highly eccentric cases. For perturbed orbits, the errors of the orbital longitude methods first grow in proportion to the square root of time for some period of time, the length of which depends on the magnitude of the perturbation. After that, the errors grow more rapidly (see Fig. 8 of Paper VII for the true-longitude method and Fig. 12 of Paper VIII for the antifocal longitude method).

The overall errors of the two orbital longitude methods are the lowest among the methods of manifold correction for various types of perturbations. In any sense, this feature is of great advantage when compared with the usual manner of error growth achieved by existing methods, a quadratic or linear⁸ increase with respect to time from the beginning of the integration whether perturbations are present or not.

Unfortunately, the cost of ensuring the desirable properties of the orbital longitude methods increases rapidly when the eccentricity is large. In other words, if we choose a large step size, the precision of the two improved orbital longitude methods rapidly degrades when the eccentricity increases. Such demerits are eminent in the study of the long-term orbital evolution of the periodic comets, some peculiar asteroids, and a few natural and artificial satellites with highly eccentric orbits. Typical examples are Halley's comet ($q = 0.577$ AU, $e = 0.967$, $P = 76.0$ yr), Icarus ($a = 1.078$ AU, $e = 0.827$, $P = 326$ days) and Hidalgo ($a = 5.746$ AU, $e = 0.661$, $P = 13.8$ yr), Nereid ($a = 222 R_{\text{Neptune}}$, $e = 0.75$, $P = 360$ days), and the artificial satellite *HALCA* (perigee altitude 560 km, $e = 0.600$, $I = 31.3^\circ$, $P = 6.3$ hr) of VSOP, the Japanese space VLBI program (Hirabayashi et al. 2000). This deterioration occurs not only in the magnitude of the error itself, but also in the rate of its growth.

Consider Figure 1, which shows the errors in the mean longitude at epoch, L_0 , of a highly eccentric ($e = 0.5$) orbit obtained using various methods of manifold correction. In preparing this figure, we (1) adopted the implicit Adams method in PECE (predict, evaluate, correct, evaluate) mode as the integrator, (2) fixed the step size such that one orbital period is covered by 90 steps,⁹ (3) prepared the starting tables using Gragg's extrapolation method, (4) measured the errors by comparing with reference solutions that we obtained using the same integrator and the same model parameters but with half

¹ Hereafter Paper I.

² Hereafter Paper II.

³ Hereafter Paper III.

⁴ Hereafter Paper IV.

⁵ Hereafter Paper V.

⁶ Hereafter Paper VI.

⁷ Hereafter Papers VII and VIII, respectively.

⁸ Only a limited number of integrators are known to produce linearly growing errors: the symplectic integrators and the symmetric multistep methods.

⁹ This is to make the cost of integration almost the same.

TABLE 1
METHODS OF MANIFOLD CORRECTION FOR ORBIT INTEGRATIONS

Method	Variables	Relations	Correction	Conserved Elements	Reference
Standard	6: $\mathbf{x}, \mathbf{v}$	None	None	None	
Single scaling	7: $\mathbf{x}, \mathbf{v}, K$	K	$(\mathbf{x}, \mathbf{v}) \rightarrow (s\mathbf{x}, s\mathbf{v})$	a	Paper I
Dual scaling	10: $\mathbf{x}, \mathbf{v}, K, \mathbf{P}$	K, F	$(\mathbf{x}, \mathbf{v}) \rightarrow (s_X\mathbf{x}, s_Y\mathbf{v})$	a, e^*, ω^*	Paper II
Single scaling + rotation	10: $\mathbf{x}, \mathbf{v}, K, \mathbf{L}$	K, x_L, v_L	$(\mathbf{x}, \mathbf{v}) \rightarrow (s\mathbf{R}\mathbf{x}, s\mathbf{R}\mathbf{v})$	a, I, Ω	Paper III
Dual scaling + rotation	13: $\mathbf{x}, \mathbf{v}, K, \mathbf{L}, \mathbf{P}$	K, F, x_L, v_L	$(\mathbf{x}, \mathbf{v}) \rightarrow (s_X\mathbf{R}\mathbf{x}, s_Y\mathbf{R}\mathbf{v})$	$a, e^*, I, \Omega, \omega^*$	Paper III
Linear transformation	13: $\mathbf{x}, \mathbf{v}, K, \mathbf{L}, \mathbf{P}$	$K, \mathbf{P}, x_L, v_L$	$(\mathbf{x}, \mathbf{v}) \rightarrow (s_X\mathbf{R}\mathbf{x}, s_Y(\mathbf{R}\mathbf{v} - \alpha\mathbf{R}\mathbf{x}))$	a, e, I, Ω, ω	Paper IV
First-stage simplification:					
First	9: $\mathbf{x}, \mathbf{L}, \mathbf{P}$	r, x_L	$\mathbf{x} \rightarrow s(\mathbf{x} - \beta\mathbf{L})$	a, e, I, Ω, ω	Paper V
Second	9: $\mathbf{n}, \mathbf{L}, \mathbf{P}$	$ \mathbf{n} , n_L$	$\mathbf{n} \rightarrow s(\mathbf{n} - \beta\mathbf{L})$	a, e, I, Ω, ω	Paper V
Second-stage simplification:					
First	7: $x_A, x_B, \mathbf{L}, P_A, P_B$	r	$(x_A, x_B) \rightarrow (sx_A, sx_B)$	a, e, I, Ω, ω	Paper VI
Second	7: $n_A, n_B, \mathbf{L}, P_A, P_B$	$ \mathbf{n} $	$(n_A, n_B) \rightarrow (sn_A, sn_B)$	a, e, I, Ω, ω	Paper VI
Original true-longitude	6: $g, \mathbf{L}, P_A, P_B$	None	None	a, e, I, Ω, ω	Paper VI
True-longitude	6: $g, \mathbf{L}, P_A, P_B$	None	$g \rightarrow \text{mod}(g, 2\pi)$	$a, e, I, \Omega, \omega, L_0$	Paper VII
Antifocal longitude	6: $w, \mathbf{L}, P_A, P_B$	None	$w \rightarrow \text{mod}(w, 2\pi)$	$a, e, I, \Omega, \omega, L_0$	Paper VIII
K-S regularization:					
Standard	10: $\mathbf{u}, \mathbf{u}', t, h_K$	None	None	None	Kustaanheimo & Stiefel 1965
E - J scaled	10: $\mathbf{u}, \mathbf{u}', t, h_K$	$\mu, C_1 C_2 = 1$	$(\mathbf{u}, \mathbf{u}') \rightarrow (C_1\mathbf{u}, C_2\mathbf{u}')$	a	Aarseth 2003
Single-scaled	10: $\mathbf{u}, \mathbf{u}', t, h_K$	μ	$(\mathbf{u}, \mathbf{u}', t) \rightarrow (\sigma\mathbf{u}, \sigma\mathbf{u}', kT + t^*)$	a	This work

NOTE.—Listed are some key features of the methods of manifold correction to integrate perturbed orbits in rectangular coordinates, as well as the standard method to directly integrate in real rectangular coordinates and the K-S regularization (Stiefel & Scheifele 1971). The numbers in the second column are the number of dependent variables per celestial body. The variables listed in the second column include the position vector $\mathbf{x}$, the velocity vector $\mathbf{v}$, the Kepler energy K , the Laplace integral vector $\mathbf{P}$, the orbital angular momentum vector $\mathbf{L}$, and the unit position vector $\mathbf{n} \equiv \mathbf{x}/r$; x_A and x_B are two independent components of $\mathbf{x}$ in the orbital plane, n_A and n_B are the corresponding components of $\mathbf{n}$, and P_A and P_B are the corresponding components of $\mathbf{P}$; g is a true orbital longitude, w is a similar orbital longitude for the antifocal anomaly, $\mathbf{u}$ and $\mathbf{u}'$ are the four-dimensional position and velocity vectors of an associated harmonic oscillator, t is the real time, and $h_K = -K$ is the negative Kepler energy. In the third column, we indicate the relations to be maintained: “ K ” for $K = (v^2/2) - (\mu/r)$, “ F ” for $|\mathbf{v} \times \mathbf{L}| = |\mathbf{P} + \mu\mathbf{n}|$, “ $\mathbf{P}$ ” for $\mathbf{v} \times \mathbf{L} = \mathbf{P} + \mu\mathbf{n}$, “ x_L ” for $\mathbf{x} \cdot \mathbf{L} = 0$, “ v_L ” for $\mathbf{v} \cdot \mathbf{L} = 0$, “ n_L ” for $\mathbf{n} \cdot \mathbf{L} = 0$, “ r ” for $r = L^2/(\mu - \mathbf{P} \cdot \mathbf{x}/r)$ or $r = L^2/[\mu - (P_A x_A + P_B x_B)/(x_A^2 + x_B^2)^{1/2}]$, “ $|\mathbf{n}|$ ” for $|\mathbf{n}| = 1$ or $(n_A^2 + n_B^2)^{1/2} = 1$, and “ μ ” for $2(\mathbf{u}')^2 + h_K \mathbf{u}^2 = \mu$. In the fourth column, listing the type of manifold correction, s is the single scaling factor, s_X and s_Y are the dual scaling factors, $\mathbf{R}$ is a rotation matrix to make $\mathbf{x}$ and $\mathbf{v}$ perpendicular to $\mathbf{L}$, α is a factor to maintain the angle between $\mathbf{x}$ and $\mathbf{v}$ properly, β is a factor for orthogonalization, $\text{mod}(x, y)$ is a function to take the modulus of x with respect to y , C_1 and C_2 are the scaling factors of the energy–angular momentum (E - J) scaling for the K-S regularization (Aarseth 2003), and σ is the single scaling factor for the K-S regularization. The second-to-last column shows the orbital elements that are conserved roughly (with asterisks) or completely (without asterisks) for the Keplerian orbits. Here we use the phrase “completely” in the sense of being at the machine-epsilon level. Note that the error in the mean longitude at epoch, L_0 , reduces completely if a sufficiently high order integrator is used with a sufficiently small step size.

the step size, and (5) set the orders of the implicit Adams method as the highest among those that led to no numerical instabilities.¹⁰ Even the antifocal longitude method, which produces the lowest errors among the methods of manifold correction in rectangular coordinates, provides a poor result, although its rate of error growth is relatively low, being linear with respect to time in the long run. This weakness of the orbital longitude methods against increases in eccentricity is more clearly shown in Figure 2. This figure illustrates the eccentricity dependence of the longitude errors in a Keplerian orbit after a sufficiently long integration time, 32,768 orbital periods, for the same integration conditions as in Figure 1. Under the condition that the number of steps per orbital period is fixed, the errors of the antifocal longitude method grow exponentially with the eccentricity. This is a well-known property common to almost all integration methods that employ the policy of fixing the step size in physical time.

Of course, a common recipe for dealing with such highly eccentric orbits is the introduction of two-body regularizations (Aarseth 2003), the most famous of which is the Kustaanheimo-Stiefel (K-S) regularization (Kustaanheimo & Stiefel 1965). A concise summary of the K-S regularization is given by Funato et al. (1996), while details are provided in a textbook by one of the inventors (Stiefel & Scheifele 1971). See Table 1 and Figures 1 and 2 again—apparently, the integration of regularized

orbits is superior to the antifocal longitude method in highly eccentric cases, say, when $e > 0.5$ for a step size of 90 steps per period. Such superiority of the K-S regularization in precision orbital integrations has been reported many times. For example, Palmer et al. (1998) present a comparison of various combinations of the K-S regularization and different integrators for practical problems in the orbital motion of artificial satellites.

Unfortunately, the errors of the K-S regularization will exceed those of the orbital longitude methods over the long run, since its error growth is quadratic. Yet, there remains a chance for improvement. As summarized by Stiefel & Scheifele (1971), the K-S regularization transforms the nonlinear Kepler problem in real three-dimensional space into a set of linear problems for a harmonic oscillator in a virtual four-dimensional space. This change in the form of the equation of motion provides some benefit in its numerical integration. In fact, Funato et al. (1996) presented an enhancement of K-S regularization by requiring its integration to be time-symmetric. Further, we ourselves reported that numerical integration of K-S regularized orbits using the special second-order symmetric linear multistep methods designed for special second-order ordinary differential equations avoids the so-called step-size resonance, the only known drawback of these integrators (Arakida & Fukushima 2000, 2001).

As we stressed in Paper II, the concept of manifold correction is not limited to a specific form of orbit integration, for example, those in rectangular coordinates that we studied extensively. Thus, it could be worthwhile to apply the idea to the K-S

¹⁰ This is to compare the best available precision for each method of integration.

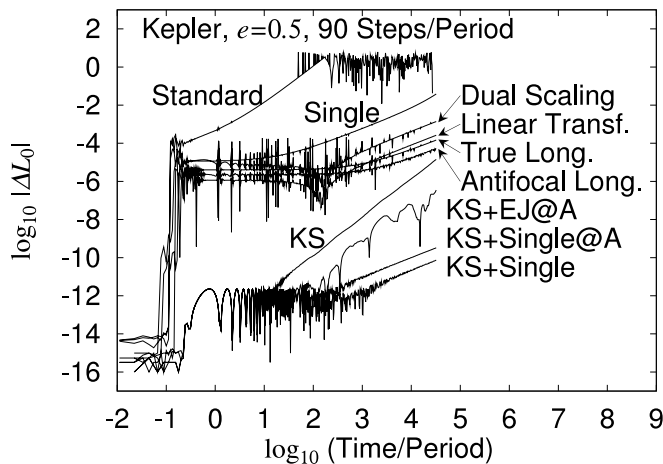


FIG. 1.—Dependence of growth of integration error on method for a highly eccentric orbit. Illustrated are the errors in the mean longitude, L_0 , of a Keplerian orbit with $e = 0.5$. The compared methods of integration are (1) the standard method, (2) the single scaling method from Paper I, (3) the dual scaling method from Paper II, (4) the linear transformation method from Paper IV, (5) the improved true-longitude method from Paper VII, (6) the antifocal longitude method from Paper VIII, (7) the original K-S regularization, (9) the K-S regularization with the E - J scaling (Aarseth 2003) applied at every apocenter, (10) the K-S regularization with the single scaling applied at every apocenter, and (11) the K-S regularization with the single scaling applied at every step. The notation “@A” means that the manifold correction is applied at the apocenter only. The adopted integrator was the implicit Adams method in PECE mode, the step size was fixed throughout the integration and chosen such that one orbital period is covered by 90 steps, the starting tables were prepared using Gragg’s extrapolation method, and the errors were measured by comparing with reference solutions obtained by the same integrator and with the same model parameters but half the step size. The order of the integrator was selected as the highest among those that led to no numerical instabilities for this step size: the 16th for the antifocal longitude method, the 14th for the linear transformation method, the 13th for the improved true-longitude method, and the 11th for the others.

regularization.¹¹ In our work on manifold correction methods, we have repeatedly seen that the most important relation to be maintained is that of the Kepler energy. In the K-S formulation, this quantity is already incorporated as an auxiliary quantity to be integrated simultaneously with the main variables, the position and velocity vectors of a four-dimensional harmonic oscillator. Then a natural question emerges: Why not apply the single or another scaling method to the K-S formulation?

S. Mikkola invented such a method of manifold correction around 1990. It was then adopted in the codes described by Aarseth (2003), where he referred to it as *energy rectification* (see § 11.3 of Aarseth 2003 for details; see also Appendix A for a concise summary). In short, energy rectification is a sort of dual scaling method, in the vocabulary of our methods of manifold correction. It applies different scaling factors to the position and velocity vectors of the associated harmonic oscillator in the K-S formulation to satisfy the Kepler energy relation and conservation of the magnitude of the orbital angular momentum. In this sense, we will call this rectification the energy–angular momentum scaling, or the E - J scaling for short, since it is a scaling method to maintain the energy and the angular momentum. Experimentally, we have learned that the E - J scaling is not suitable for application at every integration

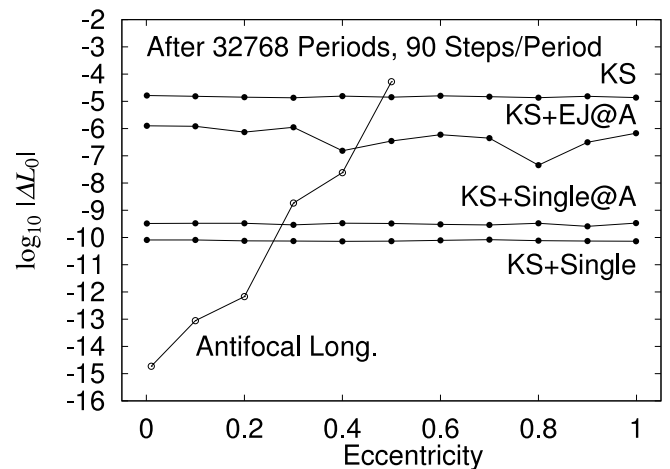


FIG. 2.—Eccentricity dependence of numerical integration errors of a Keplerian orbit. Similar to Fig. 1, but the errors after 32,768 periods are plotted as functions of eccentricity. Here we show the results of the antifocal longitude method and four kinds of K-S regularization.

step. If one attempts to do so, the integration error is greatly enhanced after one orbital period and the integration itself becomes numerically unstable sooner or later. However, as recommended by Aarseth (2003), when it is applied only at the apocenter—or, rigorously speaking, only at the nearest point to the apocenter among the discrete points on the integrated orbit¹²—the precision of the orbit integration is greatly enhanced without facing any instabilities. We confirmed that such limited use of the E - J scaling is much superior to the K-S regularization itself. Refer to Figures 1 and 2 again; for Keplerian orbits, the E - J scaling applied at the apocenter significantly reduces the magnitude of the longitude errors.

From our experience with the manifold correction methods for orbit integrations without regularizations, we expected that the simplest method of manifold correction, the single scaling method, would also be applicable to the K-S regularization without requiring any additional components to be integrated. To our surprise, the single scaling method, whether applied at every apocenter or at every step, not only reduces the magnitude of longitude errors but also changes their growth rate from quadratic to linear with respect to time in the unperturbed case (see Figs. 1 and 2 once again). This is a significant improvement, especially over the long run. However, it remains to be examined whether this superiority persists when perturbations are present.

In this paper, we report that the single scaling, especially if it is applied at every apocenter, is generally superior to the E - J scaling whether perturbations exist or not. In the following, we describe the application of the single scaling method to K-S regularized orbital motions in § 2 and present a numerical comparison with existing methods in § 3.

2. SINGLE SCALING METHOD FOR K-S REGULARIZED ORBITAL MOTION

Here we apply the single scaling method, the simplest method of manifold correction and extensively described in Paper I, to the perturbed two-body problem in the K-S regularization (Kustaanheimo & Stiefel 1965; Stiefel & Scheifele 1971). The equation of motion of a K-S regularized orbit

¹¹ There is another method of regularization, named the Burdet–Ferrándiz (B-F) regularization (Ferrándiz et al. 1992; Martín & Ferrándiz 1995). In this paper, we do not discuss the B-F regularization since its main auxiliary quantity is not the Kepler energy but the total angular momentum.

¹² A practical scheme to realize this situation is presented in Appendix C.

(eq. [52] in § 9 of Stiefel & Scheifele 1971), is expressed in four-dimensional vector form as

$$\mathbf{u}'' + \left(\frac{h_K}{2}\right) \mathbf{u} = \mathbf{Q}, \quad t' = \mathbf{u}^2, \quad h'_K = -4 \left(\frac{\mathbf{u}' \cdot \mathbf{Q}}{\mathbf{u}^2}\right), \quad (1)$$

where primes denote differentiation with respect to the fictitious time s , $\mathbf{u}$ is the four-dimensional fictitious position vector, $\mathbf{u}' \equiv d\mathbf{u}/ds$ is the four-dimensional fictitious velocity vector, t is the real time, h_K is the negative Kepler energy, and $\mathbf{Q}$ is the perturbing force, which is a function of $\mathbf{u}$, $\mathbf{u}'$, and t . Even when $\mathbf{Q}$ is nonzero, equation (1) has an integral¹³

$$2(\mathbf{u}')^2 + h_K \mathbf{u}^2 = \mu, \quad (2)$$

where $\mu \equiv G(M + m)$ is the gravitational constant of the two-body problem. In the course of numerical integration of equation (1), this relation may not always be satisfied. In this case, we assume that the errors in the fast variables, $\mathbf{u}$ and $\mathbf{u}'$, are the cause of the observed inequality and correct both of them to satisfy the relation. Since there is a single relation to be maintained, there is 1 degree of freedom in the correction. We thus select a single scaling

$$(\mathbf{u}, \mathbf{u}') \rightarrow (\sigma \mathbf{u}, \sigma \mathbf{u}') \quad (3)$$

as the method of correction. As we studied extensively in § 2.3 of Paper I, the harmonic oscillator form of the main equation of motion strongly implies that there is a benefit to be had in taking the same scaling factor for the fictitious positions and velocities when the integration formulae applied to the position and the velocity are the same. Fortunately, equation (2) is purely quadratic with respect to $\mathbf{u}$ and $\mathbf{u}'$. Then the scaling factor is uniquely¹⁴ determined as

$$\sigma = \sqrt{\frac{\mu}{2(\mathbf{u}')^2 + h_K \mathbf{u}^2}}, \quad (4)$$

where the quantities $\mathbf{u}$, $\mathbf{u}'$, and h_K on the right-hand side are the integrated values. The scaling is feasible as long as the argument of the above square root is positive. This condition is always satisfied for elliptical and parabolic orbits, even under perturbations, since h_K is nonnegative in these cases. When $h_K < 0$, on the other hand, there remains the possibility that the argument will become negative. However, we seldom face such cases in precise integrations.¹⁵ In any event, our experience tells us that the scaling will be applicable in almost all cases, independent of the type of orbit or perturbations.

We skip the extension to n -body integrations since it is trivial, as in Papers I–VIII. See Aarseth (2003) for practical applications of K-S regularizations. We have only to determine the scaling factor for each body and apply the scaling, body by body.

Finally, let us mention a couple of practical techniques to enhance the scaling method. First, the quantity h_K remains

almost constant throughout an integration. Thus, in order to reduce the round-off error, it is better to integrate not h_K itself but its deviation from the initial value, $\Delta h_K \equiv h_K - (h_K)_0$, instead. In practice, we replace the last component of equation (1) with

$$(\Delta h_K)' = -4 \left(\frac{\mathbf{u}' \cdot \mathbf{Q}}{\mathbf{u}^2}\right) \quad (5)$$

and evaluate h_K as

$$h_K = (h_K)_0 + \Delta h_K \quad (6)$$

whenever it is needed. Secondly, the solution of t contains a secular component that grows linearly with respect to the fictitious time, s , even in the unperturbed case. In this situation, it is better to treat separately the integral and fractional parts measured in some unit of time in order to reduce the accumulation of round-off errors. We therefore decompose t as $t = kT + t^*$, where k denotes the integer part of t measured in units of T , such that the condition $|t^*| < T$ always holds. Usually we take half the nominal orbital period as the value of T . If this separation is already done before the integration, the relation can be maintained by applying the following procedure, written in C-style pseudocode, at each integration step:

```
if (t* > T) { k += 2; t* -= 2T; }
else if (t* < -T) { k -= 2; t* += 2T; };
```

(see the similar discussions in Papers VII and VIII).

Before concluding this section, we stress that the single scaling described above retains the universality of the original K-S formulation. Namely, it is applicable to all types of orbits: elliptical, parabolic, hyperbolic, and linear. This is true even under perturbations.

3. NUMERICAL EXPERIMENTS

We next examine the effects of the single scaling method applied to K-S regularized orbits and compare them with the E - J scaling method. In the following, we restrict ourselves to perturbed two-body problems. The extension to n -body cases is postponed to future work.¹⁶

Before going further, let us explain how we measure the integration errors. Among the various approaches discussed in § 3.1 of Paper I, we adopt a comparison with a reference solution that is obtained using the same integrator, the same initial conditions, and the same model parameters but with half the step size. Based on this policy and the technique for synchronizing the errors at the same real time explained in Appendix B, we define the errors of the K-S variables and Keplerian orbital elements, $X \equiv (\mathbf{u}, \mathbf{u}', h_K, a, e, I, \Omega, \omega, L_0)$, obtained from the integration with a fixed step size in the fictitious time, h , at the same real time, $t^* \equiv t(s, h/2)$, as

$$\Delta X(t^*, h) \equiv \delta X(s, h) - \frac{dX}{dt} \delta t(s, h), \quad (7)$$

where

$$\delta Y(s, h) \equiv Y(s, h) - Y(s, h/2) \quad (8)$$

denotes the equal- s difference of a solution Y at s , which is obtained by integration with the step size in the fictitious time

¹³ This is a rewriting of the Kepler energy relation in terms of the fictitious vectors.

¹⁴ The uniqueness of the solution of the equation to specify the scaling factor is different from the situation for the E - J scaling described in Appendix A.

¹⁵ Rare exceptions happen when originally hyperbolic orbits are integrated by very crude integration methods, say, with a very low-order integrator or with a very large step size, such that the integrated position and velocity become so circularized that the kinetic energy greatly reduces while h_K itself, which is integrated separately, remains negative.

¹⁶ We feel no pressing need to do long-term integrations of multiple gravitating bodies in highly eccentric orbits.

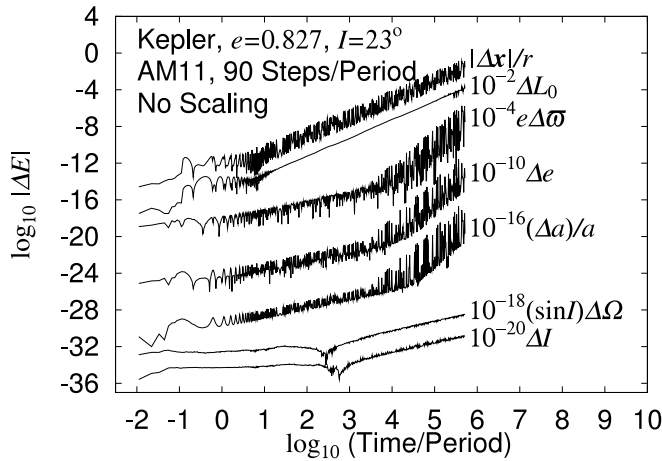


FIG. 3.—Element errors of a Keplerian orbit regularized by K-S transformation with no scaling. Similar to Fig. 1, but the errors in position and in the modified orbital elements of a highly eccentric Keplerian orbit integrated with the K-S regularization are plotted as functions of the real time on a log-log scale. The eccentricity is as high as $e = 0.827$ and the inclination is as moderate as $I = 23^\circ$. They are taken from the nominal values of Icarus in the ecliptic coordinate system. Most of the curves are offset by some factor to avoid overlap.

at h fixed. The coefficient on the right-hand side of equation (7), the t -derivative of X , is evaluated directly for Keplerian orbital elements. On the other hand, for the K-S variables, the derivatives are evaluated by way of s -derivatives as

$$\frac{dX}{dt} = \frac{X'}{t'}, \quad (9)$$

where the s -derivatives, X' and t' , are evaluated using the reference solution, $X(s, h/2)$ and $t(s, h/2)$. Note that this definition of the equal- t errors satisfies the trivial condition $\Delta t = 0$.

In this paper we omit the discussion of nonelliptical orbits. In such cases, whether perturbations are present or not, a long integration is not necessary, since the nonlinear effects are serious only for a limited time, in the neighborhood of close encounters with gravitating bodies (including the central one), and there are only one or two such close encounters at most for each gravitating body. Therefore the original K-S regularization is accurate enough to integrate the nonelliptical cases with sufficient precision. In fact, we have confirmed that the scaling does not significantly enhance the performance of the K-S regularization. Thus, we deal with the elliptical but highly eccentric cases in what follows.

We begin with unperturbed cases. Figure 3 illustrates the growth of the position error and the errors in six modified orbital elements for a highly eccentric Keplerian orbit integrated in the K-S regularized form. The errors in the semimajor axis, Δa , in the orbital eccentricity, Δe , and in the longitude of pericenter, $\Delta \varpi$, first grow linearly with respect to the real time t and then grow more rapidly, as a cubic function of t , after some amount of time—a few thousand periods in this case. As a result of the initial linear growth of Δa , the errors in the mean longitude, ΔL_0 , and those in the real position vector, $\Delta \mathbf{x}$, too, seem to grow quadratically with respect to t . Since Δa has started to grow in a cubic manner, sooner or later ΔL_0 will increase as a quartic function of t . The initial quadratic growth with respect to t is the same as the standard method in the nonregularized form (cf. Fig. 1). On the other hand, the errors in the two angles specifying the orbital plane, namely, those in

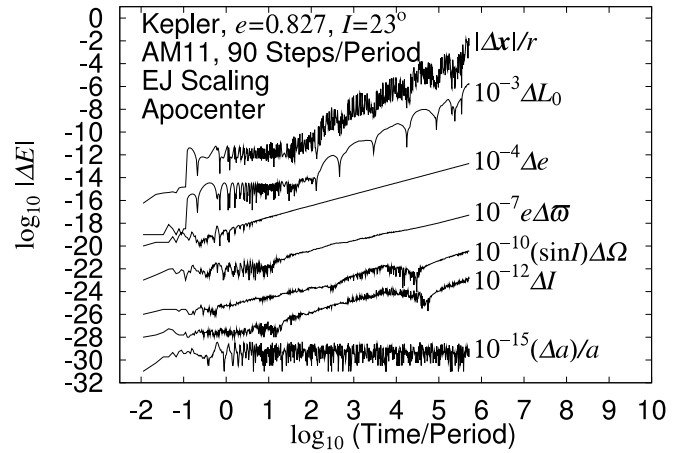


FIG. 4.—Same as Fig. 3, but with application of the E - J scaling at every apocenter.

the inclination, ΔI , and in the longitude of the ascending node, $\Delta \Omega$, are negligibly small for the first few hundred periods. Then they grow in proportion to $\sqrt{t}$. This suggests that these errors are caused by the statistical accumulation of round-off in the single summation of the random errors of the zero-mean value (Brouwer 1937). We confirmed this conjecture by examining the step-size dependence of the results integrated by fixing the integrator in a separate numerical experiment.

Figure 4 shows the same curves as in Figure 3 but with the application of the E - J scaling at every apocenter.¹⁷ This time Δa remains almost constant, at some tens of machine epsilons, throughout the integration. This is the effect of scaling to maintain the Kepler energy relation. However, ΔL_0 grows quadratically with respect to t . This seems curious. As will be shown in the next example, the reason is not that we limit the application of the scaling only to every apocenter but the nature of the scaling applied, although we have no idea about its mechanism. On the other hand, the errors in some of the elements first remain at the machine-epsilon level (ΔI and $\Delta \varpi$) or grow in proportion to the square root of time ($\Delta \Omega$) for some amount of time, 10 or so periods in this case, and then the errors grow linearly, whereas Δe grows linearly from the beginning. In any event, the magnitude of the errors in the position vector, $|\Delta \mathbf{x}|$, grows quadratically with time. This is mainly due to the quadratic growth of the longitude error. However, we note that their magnitude is significantly smaller than in the case without any scaling. Thus we have confirmed that the application of the E - J scaling enhances the precision of orbit integration.

Figure 5 depicts similar curves for when the single scaling is applied in place of the E - J scaling. In order to examine the genuine effect of the difference in the scaling prescription, we did not change the timing at which we apply the scaling—at every apocenter. Again, Δa remains finite throughout the integration, at the same low level. This time, however, ΔL_0 grows linearly with respect to t . This is the effect of the single scaling. The errors in the eccentricity vector, Δe and $\Delta \varpi$, first remain at the level of the machine epsilon for some amount of time, around a hundred periods in this case. Then they grow linearly with time. On the other hand, the errors in the direction of the

¹⁷ As we noted earlier, the application of the E - J scaling at every integration step frequently leads to numerical instabilities. Thus, we did not include it in our comparison.

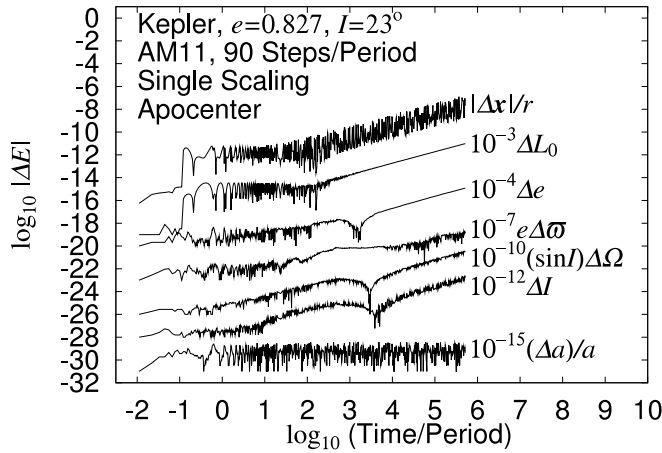


FIG. 5.—Same as Fig. 3, but with application of the single scaling at every apocenter.

angular momentum vector, ΔI and $\Delta \Omega$, grow linearly after one orbital period. The faster growth of ΔI and $\Delta \Omega$ is the result of ignoring conservation of the angular momentum direction, which is taken into account in the E - J scaling, in the rotation method in Paper III, and in the subsequent methods of manifold correction in Papers IV–VIII. In the long run, the errors in position also grow linearly. This is because there are no components whose errors grow quadratically with time.

Figure 6 plots the same curves as in Figure 5 but with the timing of the scaling changed from every apocenter to every integration step. Here Δa further decreases and remains exactly at the machine-epsilon level. This is the effect of applying as many scalings as possible. As a result, the magnitude of ΔL_0 decreases significantly while maintaining the same linear growth as in Figure 5. On the other hand, the errors in the other elements do not change very much. In conclusion, the position error in this case is smaller than in the three previous figures because the Kepler energy relation is maintained more rigorously through application of the scaling at every integration step. This situation is exactly the same as we faced with the single scaling method developed in Paper I.

In conclusion, the scaling methods applied to the K-S regularization for unperturbed orbits perform excellently, in the following order: (1) the single scaling applied at every step, (2) the single scaling applied at every apocenter, (3) the E - J

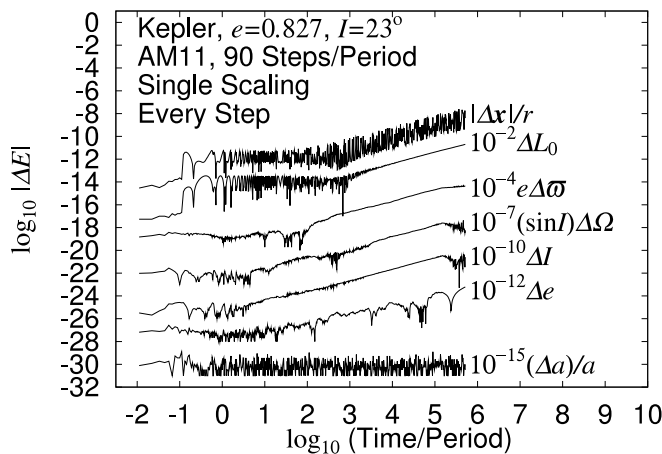


FIG. 6.—Same as Fig. 3, but with application of the single scaling at every step.

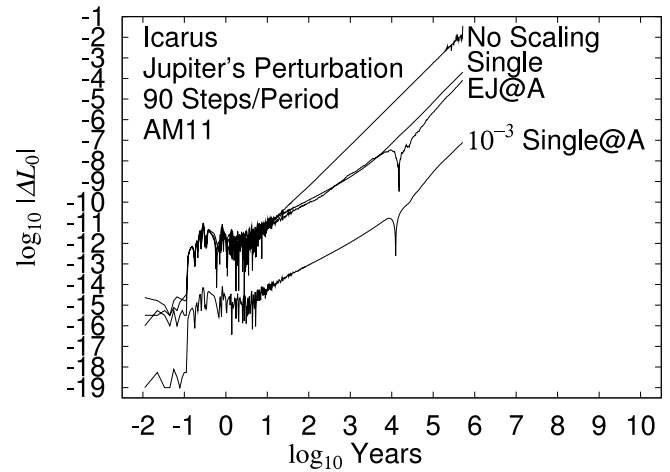


FIG. 7.—Effect of the scaling on the K-S regularized orbital motion under a third body's perturbations. Compared are the longitude errors of a model Icarus under Jupiter's perturbation. The orbit of Jupiter was given as a fixed Keplerian orbit. The step size in the fictitious time was fixed so as to cover one nominal orbital period in 90 steps. This corresponds to an averaged step size of 3.6 days in real time. The curve for the single scaling applied at every apocenter is mostly the same as that for the E - J scaling at every apocenter, so the former is offset to show the similarity clearly.

scaling applied at every apocenter, and (4) no application of scaling. This fact has already manifested in Figures 1 and 2.

Let us move on to the perturbed orbits. First of all, we confirmed that the situation observed in case of unperturbed orbits essentially persists when the perturbation is sufficiently weak. Thus, we mainly examine the case of moderate and strong perturbations in the following.

Figure 7 illustrates the case of a typical perturbation, Icarus under Jupiter's third-body influence. In preparing the figure, we set the initial conditions for Icarus and Jupiter as those at J2000.0. Also, the orbit of Jupiter was fixed as its osculating one at J2000.0. The conditions of the integration are the same as in Figure 1, namely, the 11th-order implicit Adams method in PECE mode with the step size set to cover one nominal orbital period in 90 steps. Thus the averaged step size is as large as 3.6 days. In the case of K-S regularization without scaling, the errors begin to grow quadratically with time after the first close encounter with Jupiter. On the other hand, with the three scaling methods, the errors grow linearly for the first few thousand years. Then they grow quadratically. The result of the single scaling method applied at every step is a little worse than both the E - J scaling applied at every apocenter and the single scaling method applied at every apocenter, while the latter two give almost the same result. In any case, the scaling suppresses the magnitude of the errors over the long run by two digits or so.

The scaling is effective even for dissipative perturbations, where the Kepler energy secularly decreases. Figure 8 shows a similar comparison of integration errors for *HALCA* under the perturbations of air drag only. The eccentricity is high, at 0.6, and the perigee altitude is as low as 560 km. The other conditions of the integration are basically the same as in the previous figures. Thus the averaged step size is 2.1 minutes. The errors of the K-S regularization without scaling increase quadratically with time. Those with scaling first grow linearly for some amount of time, about 1 month for the E - J scaling applied at every apocenter, around a few decades for the single scaling applied at every step, and more than a thousand years for the single scaling method applied at every apocenter. Then the

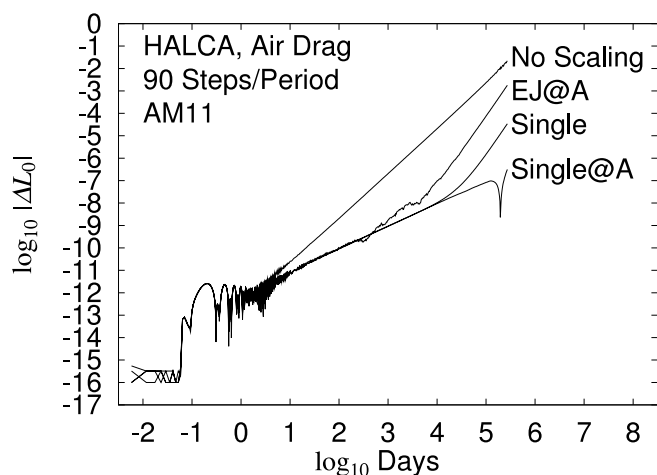


FIG. 8.—Same as Fig. 7, but integrated is the orbit of *HALCA*, an artificial Earth satellite with $e = 0.600$, under perturbations due to a model air drag only. The perigee altitude was as low as 560 km. The step size was chosen such that 90 steps cover the nominal orbital period, 6.3 hr. In this example, the single scaling applied at every apocenter gives the best result.

errors grow more rapidly, about as a cubic function of time. In any case, the effect of scaling is so eminent that the gain in precision after 7 years, the project lifetime of *HALCA*, amounts to three to four digits.

Of course, there is a sort of perturbation for which the scaling seems to be ineffective. Specifically, we present a case in which the error component of quadratic growth is so small for the original K-S regularization that the application of scaling does not succeed in further reduction of the errors. Figure 9 shows the integration errors for *HALCA* under the J_2 perturbation of Earth. The eccentricity is as high as 0.6 while the inclination is as moderate as 31° . The J_2 perturbation is so strong that we halved the step size to obtain precise results. In the figure, we scaled the results of the integrations to show the similarity of the error growth. Note that all the errors increase almost linearly with time for the first few thousand periods, say, a few years. This is an excellent property of the K-S regularization itself.

Finally, we examine robustness against round-off. Figure 10 shows the integration errors for Icarus under the Sun's general

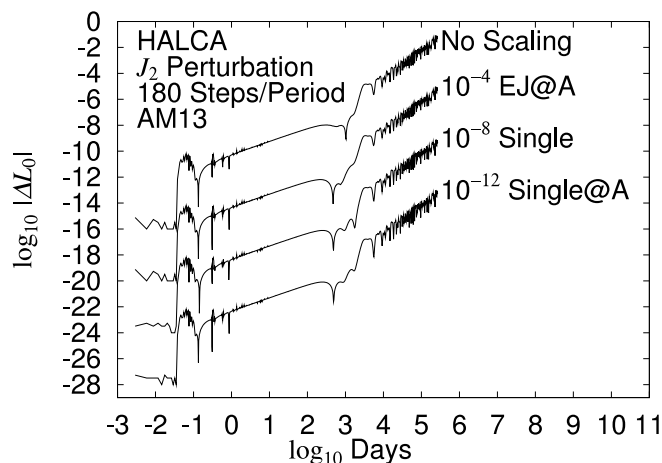


FIG. 9.—Same as Fig. 8, but under the J_2 perturbations of Earth only. This time the step size was halved. The inclination of *HALCA* is as moderate as $I = 31^\circ$. In order to show the similarity of the curves clearly, we offset the results that employ scaling. There are no practical differences among the results from the four methods compared here.

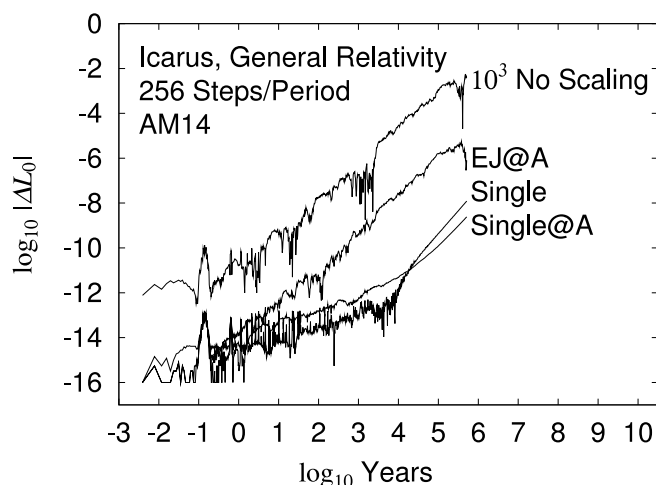


FIG. 10.—Same as Fig. 7, but under the Sun's general relativistic perturbation. This time the step size was set so small as to cover one orbital period by 256 steps. The order of the implicit Adams method was the 14th, the highest among the orders that led to no numerical instabilities for the given step size. As a result, the main error source is not truncation but round-off. Again the superiority of the single scaling method applied at every apocenter is confirmed.

relativistic perturbations using a very small step size to cover one nominal orbital period with 256 steps. The order of the implicit Adams methods was chosen as the 14th, the highest order that led to no numerical instabilities for the given step sizes. In the cases of no scaling and the E - J scaling applied at every apocenter, the errors grow in proportion to the $3/2$ power of time. This is the same as in the integrations in real rectangular coordinates, as Brouwer (1937) predicted. On the other hand, the errors for the single scaling, whether applied at every step or at every apocenter, increase in proportion to the square root of time for the first few tens of thousands of years and then increase quadratically with respect to time. This is similar to the situation for the true and antifocal longitude methods in Papers VII and VIII. In any event, the observed growth rates, $3/2$ or $1/2$, indicate that the errors are mainly due to round-off. As for the magnitude, the result of the single scaling applied at every apocenter gives significantly smaller errors than the case of application at every step. This can be well understood by considering the difference in the number of scaling operations, which significantly suppresses truncation errors but may introduce extra rounding off.

4. CONCLUSION

To K-S regularized orbital motions we have applied the idea of single scaling to maintain the Kepler energy relation consistently. The scaling is executed by applying the same multiplicative factor to all the positions and velocities of the four-dimensional harmonic oscillator associated with the orbital motion. This is different from the E - J rectification (Aarseth 2003), which multiplies two different factors with the position and velocity separately. The scaling factor is uniquely and simply determined from the integrated positions and velocities of the harmonic oscillator, as well as the integrated value of the Kepler energy, in order to exactly maintain the Kepler energy relation throughout the entire orbit integration. For unperturbed orbits, the single scaling not only reduces the magnitude of the errors but also changes the manner of error growth from quadratic to linear with respect to the real time. As for the times at which to apply the scaling, the policy of doing so at every

integration step gives better results than limiting the application to every apocenter. For perturbed orbits, the single scaling applied at every apocenter provides the best performance for third-body perturbations, dissipative perturbations, and general relativistic perturbations. For the oblateness perturbations, however, the scaling does not alter the initial linear growth of errors, which is already realized by adopting the K-S regularization. The superiority of the single scaling method applied at every apocenter is unchanged when the major error source is round-off. The cost of scaling is negligibly small, since the integration of the Kepler energy is already incorporated in the original scheme of K-S regularization and the extra computational time to judge the time of apocenter is relatively small. Also, the scaling does not destroy the universality of the original K-S formulation. In conclusion, the application of single scaling loses almost nothing in the sense of computational cost and gains a significant increase in precision for many types of perturbations, whether the errors are caused by truncation or round-off. Therefore, we recommend the application of the single scaling method whenever the K-S regularization is used.

We thank the referee for the suggestion to apply the single scaling method at every apocenter. We also appreciate other valuable suggestions to improve the quality and readability of the paper.

APPENDIX A

ENERGY RECTIFICATION

Let us summarize the so-called energy rectification for the Kustaanheimo-Stiefel regularization, which was first proposed by S. Mikkola in the early 1990s and then gradually introduced in the n -body simulation codes developed by S. Aarseth and coworkers (see § 11.2 of Aarseth 2003 for details).

The motivation behind the rectification is roughly the same as the single scaling method described in § 2 up to the point at which one specifies the functional form of the scaling. In the energy rectification, the dual scaling

$$(\mathbf{u}, \mathbf{u}') \rightarrow (C_1 \mathbf{u}, C_2 \mathbf{u}') \quad (\text{A1})$$

was selected as a method of correction. In addition to the first integral, equation (2), another condition was posed to uniquely specify the two scaling factors:

$$C_1 C_2 = 1. \quad (\text{A2})$$

This condition means that the scaling does not alter the magnitude of the orbital angular momentum, which is proportional to $|\mathbf{u}||\mathbf{u}'|$ in unperturbed orbits. With the help of this second condition, equation (2) is rewritten into a quadratic equation for $z \equiv C_2^2$,

$$Tz^2 - \mu z + U = 0, \quad (\text{A3})$$

where

$$T \equiv 2(\mathbf{u}')^2, \quad U \equiv h_K \mathbf{u}^2. \quad (\text{A4})$$

When the discriminant

$$D \equiv \mu^2 - 4TU \quad (\text{A5})$$

is nonnegative, the quadratic equation has two real roots,

$$z = \frac{\mu \pm \sqrt{D}}{2T}. \quad (\text{A6})$$

Since the scaling factor must be real, the solution z must be nonnegative. Then the solution is uniquely determined when $\sqrt{D} > \mu$, which corresponds to the hyperbolic case. Otherwise, there remain two possibilities. To specify the solution uniquely, a rule was adopted such that the more realistic solution must be the one that is closer to unity, the solution in case of no integration error. Then the solution is finally specified as

$$C_2 = \sqrt{\frac{\mu + \sqrt{D}}{2T}} \quad \text{when } 2T \geq \mu \text{ or } h_K < 0, \quad (\text{A7})$$

$$C_2 = \sqrt{\frac{\mu - \sqrt{D}}{2T}} \quad \text{otherwise.} \quad (\text{A8})$$

Of course, C_1 is computed from the C_2 thus obtained as

$$C_1 = 1/C_2. \quad (\text{A9})$$

On the other hand, when $D < 0$, a different condition,

$$C_2 = 1, \quad (\text{A10})$$

was imposed. This means that only the fictitious position vector is modified. In this case, C_1 is uniquely determined as

$$C_1 = \sqrt{\frac{\mu - T}{U}}. \quad (\text{A11})$$

As Aarseth (2003) stressed, this exceptional case rarely happens.

APPENDIX B

EVALUATION OF ERRORS AT THE SAME REAL TIME

Let us derive the expression for errors at the same real time, equation (7) in the main text. Assume that $Y(s, h)$ represents a K-S variable Y at the fictitious time s , integrated by fixing the step size in the fictitious time at h . Then we approximate its error at the same fictitious time by the difference from the reference solution,

$$\delta Y(s, h) \equiv Y(s, h) - Y(s, h/2). \quad (\text{B1})$$

This approximation is precise because the errors of the reference solution, $Y(s, h/2)$, are negligibly small when compared with those of $Y(s, h)$, since we generally use integrators of very high order, say, the ninth to 15th. Similarly, we define the errors at the same real time, $t^* \equiv t(s, h/2)$, as

$$\Delta Y(t^*, h) \equiv Y(s^*, h) - Y(s, h/2), \quad (\text{B2})$$

where s^* is the fictitious time corrected so as to satisfy the synchronization condition

$$\Delta t(t^*, h) \equiv t(s^*, h) - t(s, h/2) = 0. \quad (\text{B3})$$

Expanding the first term in the synchronization condition as

$$t(s^*, h) = t(s, h) + t'(s^* - s) + \cdots, \quad (\text{B4})$$

where the prime indicates differentiation with respect to s , we approximately solve equation (B3) with respect to $\Delta s \equiv s^* - s$ as

$$\Delta s \approx -\delta t(s, h)/t'. \quad (\text{B5})$$

Then we finally obtain the expression of errors in $X \equiv (\mathbf{u}, \mathbf{u}', h_K, a, e, I, \Omega, \omega, L_0)$ at the same real time as

$$\begin{aligned} \Delta X(t^*, h) &= [X(s, h) - X(s, h/2)] + [X(s^*, h) - X(s, h)] \\ &\approx \delta X(s, h) + X' \Delta s \approx \delta X(s, h) - \frac{X'}{t'} \delta t(s, h) \\ &= \delta X(s, h) - \frac{dX}{dt} \delta t(s, h), \end{aligned} \quad (\text{B6})$$

which is what we aimed to prove. The correction is necessary because what is meaningful is the precise integration not in the virtual K-S spacetime but in the real world. In fact, we experimentally confirmed that the smoothness of the errors in L_0 shown in the various figures of the present paper is achieved by applying this correction term.

APPENDIX C

JUDGMENT OF APOCENTER

Let us describe a couple of ways to determine whether an individual integration step is locally the closest point to the apocenter. The first technique is rigorous. Assume that at a certain integration step in the fictitious time, $s = s_n$, we know the position vectors of the associated harmonic oscillator at three fictitious times: the previous step $s = s_{n-1}$, the current step $s = s_n$, and the next step $s = s_{n+1}$. Denote the three position vectors corresponding to these times by attaching the indices as $\mathbf{u}_{n-1}$, $\mathbf{u}_n$, and $\mathbf{u}_{n+1}$, respectively. Then the condition that the

current step is the closest to the apocenter is simply expressed as

$$r_n \geq r_{n-1}, \quad r_n \leq r_{n+1}, \quad (\text{C1})$$

where the real radii are obtained from the fictitious vectors as $r_n \equiv \mathbf{u}_n^2$, etc.

In the practical course of numerical integration, it is easy to find the current and previous values of the fictitious position vector, $\mathbf{u}_{n-1}$ and $\mathbf{u}_n$. In order to obtain a sufficiently precise estimate of that at the next step, $\mathbf{u}_{n+1}$, without spending too much computational time, we assume that the three position vectors $\mathbf{u}_{n-1}$, $\mathbf{u}_n$, and $\mathbf{u}_{n+1}$, are on the orbit of a harmonic oscillator osculated at $s = s_n$. Then, with help of Chebyshev polynomials, the following formula is easily derived:

$$\mathbf{u}_{n+1} = 2c\mathbf{u}_n - \mathbf{u}_{n-1}. \quad (\text{C2})$$

Here the coefficient c is expanded in terms of the step size, h , as

$$c = 1 - \left(\frac{h_K}{2}\right)h^2 + \left(\frac{h_K^2}{96}\right)h^4 + O(h^6), \quad (\text{C3})$$

where h_K is that evaluated at $s = s_n$. Note that this computation is universal, namely, valid independent of the sign of the Kepler energy h_K .

The second method is simpler and therefore faster, but it lacks the exactness of the first. Assume that only the positions and velocities at the previous and the current step are available. Then the condition that the time of apocenter is between the previous and the present time steps is written as

$$r'_{n-1} \geq 0, \quad r'_n \leq 0, \quad (\text{C4})$$

where the s -derivative of r is evaluated as

$$r' = 2\mathbf{u} \cdot \mathbf{u}'. \quad (\text{C5})$$

Note that the above condition does *not* mean that the current step is the closest to the time of apocenter. The condition only ensures that *either* the previous step or the current one is the closest to the timing of apocenter. Therefore the chance that the current step is the closest is 50-50. However, if the step size is sufficiently small, the effect of this approximation will be small.

We numerically confirmed that either of the above devices is sufficiently accurate and fairly fast. The difference in computational time is negligibly small.

REFERENCES

- Aarseth, S. J. 2003, *Gravitational N-Body Simulations* (New York: Cambridge Univ. Press)
- Arakida, H., & Fukushima, T. 2000, *AJ*, 120, 3333
- . 2001, *AJ*, 121, 1764
- Brouwer, D. 1937, *AJ*, 46, 149 (erratum 47, 84 [1938])
- Ferrándiz, J. M., Sansaturio, M. E., & Pojman, J. R. 1992, *Celest. Mech. Dyn. Astron.*, 53, 347
- Fukushima, T. 2003a, *AJ*, 126, 1097 (Paper I)
- . 2003b, *AJ*, 126, 2567 (Paper II)
- . 2003c, *AJ*, 126, 3138 (Paper III)
- . 2004a, *AJ*, 127, 3638 (Paper IV)
- . 2004b, *AJ*, 128, 920 (Paper V)
- Fukushima, T. 2004c, *AJ*, 128, 1336 (Paper VI)
- . 2004d, *Celest. Mech. Dyn. Astron.*, submitted (Paper VII)
- . 2004e, *AJ*, 128, 1455 (Paper VIII)
- Funato, Y., Hut, P., McMillan, S., & Makino, J. 1996, *AJ*, 112, 1697
- Hirabayashi, H., et al. 2000, *PASJ*, 52, 955
- Kustaanheimo, P., & Stiefel, E. L. 1965, *J. Reine Angew. Math.*, 218, 204
- Martín, P., & Ferrándiz, J. M. 1995, *Celest. Mech. Dyn. Astron.*, 63, 29
- Palmer, P. L., Aarseth, S. J., Mikkola, S., & Hashida, Y. 1998, *J. Astronaut. Sci.*, 46, 329
- Stiefel, E. L., & Scheifele, G. 1971, *Linear and Regular Celestial Mechanics* (New York: Springer)