Perturbed Kepler problem in quaternionic form

1995 J. Phys. A: Math. Gen. 28 6245

(http://iopscience.iop.org/0305-4470/28/21/027)

J. Phys. A: Math. Gen. **28** (1995) 6245–6252. Printed in the UK

Perturbed Kepler problem in quaternionic form

Jan Vrbik†

Department of Mathematics, Brock University, St Catharines, Ontario, Canada, L2S 3A1

Received 5 June 1995, in final form 4 August 1995

Abstract. A new technique for solving the perturbed Kepler problem is presented. As its approach is essentially algebraic, it can be easily computerized and carried out to any perturbation order. It also eliminates the notorious difficulty of small divisors, and is well suited to deal with commensurable orbital periods, as demonstrated by its explanation of Kirkwood gaps.

1. Introduction

It has been known for quite some time [8] that transforming the Kepler problem to a linear and regular form (the Kustaanheimo–Stiefel equation) has several advantages over the traditional formulation. The new approach utilizes quaternion algebra, and solves the problem in terms of a quaternion-valued function, which is, in a very symbolic sense, a square root of the satellite's location $\mathbf{r}$. Since quaternions possess an extra (redundant) dimension, a 'gauge' has to be designed to make this solution unique (thus, a gauge is simply a choice of a one-dimensional constraint, restricting each potential solution). Until recently, only numerical aspects of the new technique had been fully exploited, using a gauge appropriate for that purpose. The objective of this paper is to advance yet another line of development, which builds an analytical solution to the Kustaanheimo–Stiefel equation based on a novel gauge.

The technique is a special case of a wider class of possibilities of linearizing Kepler's equation [7], and has been chosen as a basis of this paper because it lends itself easily to a traditional perturbation treatment.

2. Linearized Kepler equation

It has been shown [10], and then further expounded [11], that the Kepler problem under a small perturbing force $\mathbf{f}$, namely

$$\ddot{\mathbf{r}} + \mu\frac{\mathbf{r}}{r^3} = \varepsilon\mathbf{f} \tag{1}$$

is equivalent to the following quaternion-algebra equation

$$2\mathfrak{U}'' - [2\mathfrak{U}'\overline{\mathfrak{U}}' - 4(A^2+B^2)]\frac{\mathfrak{U}}{r} + 2\mathfrak{k}\mathfrak{U}'\frac{\Gamma}{r} + \mathfrak{k}\mathfrak{U}\left(\frac{\Gamma}{r}\right)' - \left(2\mathfrak{U}' + \mathfrak{k}\mathfrak{U}\frac{\Gamma}{r}\right)\frac{A'A+B'B}{A^2+B^2}$$
$$= -\varepsilon\mathfrak{U}\mathbf{r}\mathbf{f}H^2. \tag{2}$$

† E-mail: jvrbik@abacus.ac.brocku.ca

Proof's summary. (The bullets below explain our notation.)

$$\dot{\mathbf{r}} = \left(2\overline{\mathfrak{U}}\,\mathfrak{k}\mathfrak{U} + \frac{\Gamma}{Hr}\right)^{\cdot} = 2\overline{\mathfrak{U}}\mathfrak{k}\mathfrak{U}^{\cdot} - \frac{\Gamma}{rH}$$

premultiplied by

$$H\mathfrak{k}\mathfrak{U} \Rightarrow H\mathfrak{k}\mathfrak{U}\dot{\mathbf{r}} = -2\mathfrak{U}' - \frac{\mathfrak{k}\mathfrak{U}\Gamma}{r}.$$

Apply

$$rH\frac{\mathrm{d}}{\mathrm{d}t} \Rightarrow rH(H\mathfrak{k}\mathfrak{U}\dot{\mathbf{r}})^{\cdot} = -2\mathfrak{U}'' - \left(\frac{\mathfrak{k}\mathfrak{U}\Gamma}{r}\right)'.$$

Expanding the left-hand side of the last equation yields:

$$rH^2\mathfrak{k}\mathfrak{U}\left(\varepsilon\mathbf{f} - \mu\frac{\mathbf{r}}{r^3}\right) + H'\mathfrak{k}\mathfrak{U}\dot{\mathbf{r}} + H\mathfrak{k}\mathfrak{U}'\dot{\mathbf{r}} = \varepsilon H^2\mathfrak{U}\mathbf{r}\mathbf{f} + 4(A^2+B^2)\frac{\mathfrak{U}}{r} + \frac{AA'+BB'}{A^2+B^2}\left(-2\mathfrak{U}' - \frac{\mathfrak{k}\mathfrak{U}\Gamma}{r}\right) + \frac{\mathfrak{k}\mathfrak{U}'}{r}(2\overline{\mathfrak{U}}'\mathfrak{k}\mathfrak{U} + \Gamma)$$

from which equation (2) easily follows. □

$\mathfrak{U}$ is a quaternion function of a modified time s (the independent variable of the equation), and can be expanded as follows

$$\mathfrak{U} = \exp(\mathfrak{k}a)\{A\exp(\mathfrak{j}b)\exp[\mathfrak{i}(s-s_0)] + B\exp(-\mathfrak{j}b)\exp[-\mathfrak{i}(s-s_0)] + \mathbb{D} + \mathfrak{k}\mathfrak{U}_{\!\not 0}\mathbb{S}\}\mathfrak{R} \qquad (3)$$

where

- $\mathfrak{i}$, $\mathfrak{j}$, $\mathfrak{k}$, are the three generators of quaternion algebra, ('fraktur' font is used for quaternion quantities, an overbar implying quaternion conjugation);
- a is an arbitrary real function of s, having no physical meaning (in our gauge we make $a \equiv 0$);
- A, B, s_0, and b are real parameters defining the orbit's major semi-axis A^2+B^2, eccentricity $2AB/(A^2+B^2)$, modified time at apocenter s_0, and perpendicular dislocation b;
- $\mathfrak{R}$ is a quaternion representation of the orbit's attitude [3], which must satisfy $\mathfrak{R}\overline{\mathfrak{R}} \equiv 1$ (together with A, B and s_0 the three parameters of $\mathfrak{R}$ represent a direct analogue of the six classical orbital elements and will be referred to as such from now on);
- $\mathbb{D}$ is a series in odd powers of $E \equiv \exp[\mathfrak{i}(s-s_0)]$ with complex (no $\mathfrak{j}$ and $\mathfrak{k}$) coefficients and zero E and E^{-1} terms;
- $\mathbb{S}$ is similarly an expansion in even powers of E, complex coefficients, zero E^{-2}, E^0 and E^2 terms, and such that $\mathbb{S} = -\mathbb{S}^*$ (the asterisk indicating complex conjugation);
- $\mathfrak{U}_0$ equals $AE + BE^{-1}$ and represents the unperturbed solution in Kepler's frame;
- $\mathbf{r} \equiv \mathfrak{k}x + \mathfrak{j}y + \mathfrak{i}z$ is a quaternion representation of the orbit's location, and equals $\overline{\mathfrak{U}}\mathfrak{k}\mathfrak{U}$, implying, for the corresponding distance, $r = \overline{\mathfrak{U}}\mathfrak{U}$;
- similarly $\mathbf{f} \equiv \mathfrak{k}f_x + \mathfrak{j}f_y + \mathfrak{i}f_z$ (note that $\mathbf{rf}$ denotes the implied quaternion multiplication);
- $\Gamma = \overline{\mathfrak{U}}\mathfrak{k}\mathfrak{U}' - \overline{\mathfrak{U}}'\mathfrak{k}\mathfrak{U}$, and is equal to zero for the unperturbed solution (the old gauge would make Γ identically equal to zero);
- $H^2 = 4(A^2+B^2)/\mu$, which also serves to define the modified time s, via $\mathrm{d}s = \mathrm{d}t/rH$.

We also introduce $\beta = B/A$ and $z = E^2$ for future convenience.

One can show that this form of $\mathfrak{U}$ is fully general [10].

Main points of proof. $\mathfrak{U}$ is built in a Fourier-like manner as a series of *odd* (positive and negative) powers of $\exp(\mathrm{i}s)$, each post-(or, equivalently, pre)-multiplied by a quaternion coefficient $\mathfrak{C}_n$ (n being the corresponding power) with, potentially, its own slow time dependence (eventually, 'slow' will acquire two distinct meanings: either the time derivative is of the $\mathrm{O}(\varepsilon)$ type or the coefficient is oscillatory with a period which is at least twice as long as that of the perturbed satellite). This will result in $\mathbf{r}$ being a general series in *even* powers of $\exp(\mathrm{i}s)$ with 'vector' coefficients (an ordinary Fourier series). $\mathfrak{C}_1$ and $\mathfrak{C}_{-1}$ are then rewritten in an equivalent form of equation (3), to parametrize them in a physically meaningful manner (six of them to closely resemble the classical orbital elements). This parametrization is one-to-one, therefore a, and the corresponding gauge, are well defined. Note that the algebra of matching the Fourier coefficients of both sides of equation (2) will be greatly simplified by expanding all expressions in terms of an auxiliary variable $z \equiv \exp[2\mathrm{i}(s - s_0)]$, with the quaternion coefficients redefined accordingly. □

Equation (2) is usually solved in the Kepler frame, i.e. post-multiplied by $\overline{\mathfrak{R}}$, with $\mathfrak{R}'\overline{\mathfrak{R}} = \mathbf{Z}/2$ ($\mathbf{Z}$, as all our vectors, will be expressed in its quaternion representation, analogous to that of $\mathbf{r}$ and $\mathbf{f}$; it represents the orbit's slow rotation).

The iterative solution of the next section will consider the above 'parameters' (in the case of the six orbital parameters, their time derivatives) as functions of s, expanded as a power series in ε (with no absolute term). The coefficients of these ε-expansions will explicitly (carrying the s-dependence) depend on: the orbital elements, external parameters, and small (between $-\frac{1}{2}$ and $\frac{1}{2}$) powers of z. Subsequent integration of the resulting set of differential equations is thus necessary. On the other hand, b and the individual parameters of both $\mathbb{D}$ and $\mathbb{S}$ are found directly, in the form of similar ε-expansions.

3. Iterative solution

The solution to equation (2) can be built in an iterative manner by substituting the current, ε^N-accurate solution (starting with $\mathfrak{U}_0$), into equation (2) and collecting all ε^{N+1}-proportional terms on the right-hand side of the equation (let us call the result $\mathfrak{T}_{[N+1]}$). To find the ε^{N+1}-component of A', B', s_0', $\mathbf{Z}$, b, $\mathbb{D}$ and $\mathbb{S}$, we decouple $\mathfrak{F}_{[N+1]}$ into two complex expressions by pre-multiplying it first by $\overline{\mathfrak{U}_0}/2r_0(1+\beta z)$ and then by $-\overline{\mathfrak{U}_0}\mathfrak{k}/r_0$ (where $r_0 = \overline{\mathfrak{U}_0}\mathfrak{U}_0$), keeping the complex part of each result only (let us call the results $\mathbb{Q}_{[N+1]}$ and $\mathbb{W}_{[N+1]}$ respectively). Note [10] that $\mathbb{W}_{[N+1]} = -\mathbb{W}^*_{[N+1]}$, reflecting, and effectively removing, the extra quaternion-algebra dimension. Each $\mathbb{Q}_{[N+1]}$ and $\mathbb{W}_{[N+1]}$ can be considered a complex function of z, analytic in a region containing the unit circle, and can be expanded in the corresponding Laurent series (this is normally performed algebraically by expanding each function with respect to all 'small' parameters: β of the perturbed planet, the eccentricity of a perturbing planet, the ratio of the two planets' axes, etc). We denote the coefficient of z^n in each of these expansions by Q_n and W_n, respectively. When the perturbing function $\mathbf{f}$ has no explicit time dependence of its parameters (note that $\mathbf{f}$ is ordinarily a function of $\mathbf{r}$, and possibly its time derivative, which introduces what we call an implicit time dependence), both Q_n and W_n are constant.

Typically though, $\mathbf{f}$ contains parameters with their own time variation which, to simplify matters, will be considered harmonic (since any periodic function can be decomposed into a Fourier series, the subsequent formulae cover all periodic perturbations). To be specific, we assume that each Q_n can be written as $q_n \exp[\mathrm{i}\alpha(s - s_0)]$, and each W_n with a positive index n as $w_n \exp[\mathrm{i}\alpha(s - s_0)]$. Thus, $\frac{2}{\alpha}$ is simply the period of the 'external' perturbation, relative to the period of the perturbed planet. Due to the $W_n = -W_n^*$ relationship, the W's with a negative index constitute a 'mirror image' of the positive-index coefficients, and carry no additional information. Due to the same relationship, the coefficient of W_0 will need to have two terms, $w_0 \exp[i\alpha(s - s_0)] - w_0^* \exp[-i\alpha(s - s_0)]$. In the general case, each coefficient can actually consist of a sum of terms of the above type, all with distinct values of α. It is sufficient to present a solution assuming a single value of α only, as the extension to the general case is trivial (the problem is, at this level, linear).

Furthermore, since the solution we are constructing does not differentiate between the implicit (r dependent) and explicit ('external') time dependence, we may combine these together and *have the integer part of* $\frac{\alpha}{2}$ *contribute the corresponding extra power of* z. Thus, α can always be reduced to a value from the interval $[-1, 1)$ for a positive index m, from $(-1, 1]$ for m negative, and from $(-1, 1)$ when $m = 0$ (to preserve the symmetry of the W's). This eliminates any possibility of experiencing the 'small divisor' syndrome! One can then find the following set of formulae for the ε^{N+1} component of each of the basic parameters of our solution (3).

$$Z_1 = \sum_{n=2}^{\infty} \frac{4\alpha(-\beta)^{n-1}\,\mathrm{Im}(W_n)}{[(2n+\alpha)^2 - 4](2n+\alpha)} - \left(\frac{2}{4+\alpha} + \frac{\beta^2}{1-\beta^2}\frac{2}{2+\alpha}\right)\mathrm{Im}(W_1) - \frac{\beta}{1-\beta^2}\frac{4}{4-\alpha^2}\,\mathrm{Im}(W_0) \tag{4a}$$

$$\frac{1+\beta^2}{1-\beta^2} Z_2 = \sum_{n=2}^{\infty} \frac{4\alpha(-\beta)^{n-1}\,\mathrm{Re}(W_n)}{[(2n+\alpha)^2 - 4](2n+\alpha)} - \left(\frac{2}{4+\alpha} + \frac{\beta^2}{1-\beta^2}\frac{2}{2+\alpha}\right)\mathrm{Re}(W_1) + \frac{\beta}{1-\beta^2}\frac{2\alpha}{4-\alpha^2}\,\mathrm{Re}(W_0) \tag{4b}$$

$$\begin{aligned}\beta Z_3 = {}& \beta\frac{1-\alpha+\beta^2}{(2-\alpha)^2}\,\mathrm{Re}(Q_{-2}) + \left(\frac{1+\beta^2}{2-\alpha} + \frac{1-\beta^2}{(2-\alpha)^2} - \frac{\beta^2}{2+\alpha}\right)\mathrm{Re}(Q_{-1}) \\ & + \beta\left(\frac{1}{2-\alpha} - \frac{1+\beta^2}{2+\alpha} + \frac{1-\beta^2}{(2+\alpha)^2}\right)\mathrm{Re}(Q_0) \\ & - \left(\frac{1-\beta^2}{(2+\alpha)^2} + \frac{\beta^2}{2+\alpha}\right)\mathrm{Re}(Q_1)\end{aligned} \tag{4c}$$

$$\begin{aligned}-s_0' + \frac{Z_3}{2} = {}& \frac{\beta^4}{(1+\beta^2)}\frac{\mathrm{Re}(Q_{-2})}{(\alpha-2)^2} + \beta\left(\frac{1+\alpha}{4-\alpha^2} - \frac{\beta^2}{1+\beta^2}\frac{6-\alpha}{(2+\alpha)(2-\alpha)^2}\right)\mathrm{Re}(Q_{-1}) \\ & + \left(\frac{-1+\alpha(1+\beta^2)}{4-\alpha^2} - \frac{\beta^4}{1+\beta^2}\frac{6+\alpha}{(2+\alpha)^2(2-\alpha)}\right)\mathrm{Re}(Q_0) \\ & + \beta\left(\frac{-1}{2+\alpha} + \frac{\beta^2}{1+\beta^2}\frac{1}{(2+\alpha)^2}\right)\mathrm{Re}(Q_1)\end{aligned} \tag{4d}$$

$$\begin{aligned}\frac{A'}{A} = {}& \frac{\beta^2}{(2-\alpha)^2}\,\mathrm{Im}(Q_{-2}) - \beta\left(\frac{1}{(2-\alpha)^2} + \frac{\alpha}{4-\alpha^2}\right)\mathrm{Im}(Q_{-1}) \\ & + \left(\frac{4-\alpha}{4-\alpha^2} + \frac{\beta^2(3+\alpha)}{(2+\alpha)^2}\right)\mathrm{Im}(Q_0) + \frac{\beta(1+\alpha)}{(2+\alpha)^2}\,\mathrm{Im}(Q_1)\end{aligned} \tag{4e}$$

$$\begin{aligned}\frac{B'}{A} = {}& -\frac{\beta(1-\alpha)}{(2-\alpha)^2}\,\mathrm{Im}(Q_{-2}) - \left(\frac{3-\alpha}{(2-\alpha)^2} + \frac{\beta^2(4+\alpha)}{4-\alpha^2}\right)\mathrm{Im}(Q_{-1}) \\ & + \beta\left(\frac{1}{(2+\alpha)^2} - \frac{\alpha}{4-\alpha^2}\right)\mathrm{Im}(Q_0) - \frac{1}{(2+\alpha)^2}\,\mathrm{Im}(Q_1)\end{aligned} \tag{4f}$$

$$(1+\beta^2)b = \sum_{n=2}^{\infty}(-\beta)^n \frac{4\,\mathrm{Im}(W_n)}{[(2n+\alpha)^2-4](2n+\alpha)}$$
$$-\beta\left(\frac{1}{4+\alpha}-\frac{1}{2+\alpha}\right)\frac{\mathrm{Im}(W_1)}{2}+\frac{(1-\beta^2)}{4-\alpha^2}\frac{\mathrm{Im}(W_0)}{2} \qquad (4g)$$

$$\frac{\mathbb{D}}{A} = \sum_{\substack{n(\mathrm{odd})=-\infty \\ n\neq -1,1}}^{\infty} \left\{ \frac{-\beta(n+\alpha)\,Q_{(n-3)/2}}{(n+\alpha-1)^2(n+\alpha+1)} - \frac{(n+\alpha-2)\,Q_{(n-1)/2}}{(n+\alpha-1)^2(n+\alpha+1)} \right.$$
$$-\frac{\beta^2(n+\alpha+2)\,Q_{(n-1)/2}}{(n+\alpha-1)(n+\alpha+1)^2} - \frac{\beta(n+\alpha)\,Q_{(n+1)/2}}{(n+\alpha-1)(n+\alpha+1)^2}$$
$$+\frac{\beta^2 Q^*_{(-n-3)/2}}{(n-\alpha-1)(n-\alpha+1)^2} + \frac{2\beta Q^*_{(-n-1)/2}}{(n-\alpha-1)^2(n-\alpha+1)^2}$$
$$\left. -\frac{Q^*_{(-n+1)/2}}{(n-\alpha-1)^2(n-\alpha+1)} \right\} E^n \qquad (4h)$$

$$\mathbb{S} = \mathrm{i}\,\mathrm{Im}\left(\sum_{n=2}^{\infty} \frac{-2W_n}{[(2n+\alpha)^2-4](2n+\alpha)} \right.$$
$$\times\left\{ \sum_{m=1}^{\infty}(-\beta)^m[z^{n+m}+(-\beta)^{n-1}z^{m+1}] - \sum_{m=1}^{n-2}(-\beta)^m z^{n-m} + \frac{2n+\alpha}{2}z^n \right\}$$
$$\left. +\frac{W_1}{(2+\alpha)(4+\alpha)}\sum_{m=1}^{\infty}(-\beta)^m z^{m+1} \right). \qquad (4i)$$

It is quite apparent that no denominator in any of these formulae can approach zero (in fact, they are all, in absolute value, greater or equal to one). This substantiates our claim of having eliminated small divisors. Note what would have happened if we did not extract the integer part of the exponent from each external $z^{\frac{\alpha}{2}}$ variation!

Also note that Z_3, and consequently s_0', may contain terms proportional to β^{-1} (in the $\beta = 0$ case Z_3 and s_0' become meaningless individually, one needs only their $\frac{Z_3}{2} - s_0'$ combination which remains finite).

The convergence of the β-related infinite summations ((4*a*), (4*b*), (4*g*) and (4*i*)) can be easily established for each specific case of the perturbing force **f**. It is not uncommon for the set of the Q_n and W_n values (let us call it Λ) to be finite, and the issue thus becoming trivial. In any case, note that β is a very convenient transformation (in terms of speeding up the convergence) of the perturbed planet's eccentricity, with values from [0, 1) (but normally quite small). Thus, for example, a bounded Λ will ensure convergence for *each* such β.

The E and z-related infinite sums of (4*h*) and (4*i*) will normally converge, as both the Q_n and the W_n sequence represent coefficients of a Fourier-like expansion, further divided by a $\propto n^3$ factor.

The issue of convergence of the iteration procedure itself (i.e. with respect to ε) will have to be left open to further investigation. A lack of the corresponding proof is the common feature of most traditional techniques which, furthermore, often yield incorrect or unreliable results beyond the first order [9]. Our technique offers an improvement in terms of the latter aspects (the driving force of its development was its practical appeal); an effort focused on its theoretical foundations will hopefully resolve the former, convergence-related issue as well.

To conclude the section, we would like to emphasize one non-trivial feature of the above process, namely the necessity of expanding all functions of time in terms of $s - s_0$ only. This is indeed the most natural choice for the 'internal' variables ($\mathbf{r}$, r, $\mathbf{r}'$, etc), but not so at all for the 'external' parameters, which are usually expressed in terms of ordinary time t. Unfortunately, the procedure requires this t to be converted to an $s-s_0$ expansion, to enable us to treat the perturbing function in a unified manner when extracting the coefficients of integer powers of z (so crucial to the procedure). Such a conversion is achieved by

$$t = \int \frac{rH}{1 - s_0'} \mathrm{d}(s - s_0) \tag{5}$$

which follows from the definition of the modified time s (the last bullet item of the previous section). The integration itself is quite trivial, as the integrand can be expanded in the manner of $\mathbb{Q}_{[N+1]}$ and $\mathbb{W}_{[N+1]}$, and easily integrated term by term. Note that it is necessary to update t after each iteration to maintain its ε^N-accuracy (the right hand side of equation (2) will supply the additional power of ε).

The attitude of the orbit can be resolved from the three components of $\mathbf{Z}$ by the method of strained coordinates [4], much in the spirit of the current technique. Some coupling of these equations to those yielding the values of s_0, A, and B (equations (4*d*), (4*e*) and (4*d*)) is to be expected (the method will easily accommodate any number of unknown functions).

4. Brief example

We demonstrate the power of the technique by explaining the formation of Kirkwood gaps in the asteroid belt (a well known and traditionally challenging problem). The details are presented only for the most conspicuous gap observed at double Jupiter's frequency; the treatment of other commensurable frequencies would be quite similar.

For simplicity we assume that Jupiter has a simple circular motion (it appears that its eccentricity would contribute only minor corrections to our solution), coplanar with the asteroid's orbit, whose initial semi-major axis ($A^2 + B^2$, in our notation) will be our unit of length (the Jupiter orbit's radius is thus $2^{2/3}$). The perturbing force will have the usual form of

$$\varepsilon\mu \left(\frac{\mathbf{R} - \mathbf{r}}{|\mathbf{R} - \mathbf{r}|^3} - \frac{\mathbf{R}}{R^3} \right) \tag{6}$$

where $\mathbf{R}$ is the Jupiter's location, and ε is its mass, relative to the mass of the Sun. Note that the time dependence of $\mathbf{R}$ needs to be expressed in terms of the asteroid's modified time s, as follows

$$\mathbf{R} = 2^{3/2}\mathfrak{k} \exp \left(\mathrm{i} \left\{ \frac{\beta}{1+\beta^2} \sin(2s - 2s_0) + (s - s_0) + s_0 + \int \left[(A^2 + B^2)^{3/2} - 1 \right] \mathrm{d}s \right\} \right). \tag{7}$$

Applying the procedure of the previous section at its crudest approximation (i.e. to the first order in ε, keeping the leading terms of the β expansion only, and assuming that $A \simeq 1$ and $B \simeq 0$ in computing the coefficients), results in

$$(\psi - s_0)' = -c\frac{\varepsilon}{2\beta} \cos(2\psi - 2s_0 - 2X) \tag{8a}$$

$$\beta' = -c\varepsilon \sin(2\psi - 2s_0 - 2X) \tag{8b}$$

$$[(A^2 + B^2)^{3/2}]' = 12c\varepsilon\beta \sin(2\psi - 2s_0 - 2X) \tag{8c}$$

$$X' = (A^2 + B^2)^{3/2} - 1 \tag{8d}$$

with

$$c = \frac{9}{4}A^6 F\left(\frac{1}{2}, \frac{5}{2}; 3; \frac{A^4}{2^{4/3}}\right) + \frac{5}{16}\frac{A^{10}}{2^{4/3}} F\left(\frac{3}{2}, \frac{7}{2}; 4; \frac{A^4}{2^{4/3}}\right) \simeq 3.00 \tag{9}$$

where ψ is the direction of the asteroid's apohelium, and the independent variable is the asteroid's modified time s (in its slightly irregular time scale π represents the length of the asteroid's cycle). The first of these equations is the result of (4*c*) and (4*d*), the next two equations follow easily from (4*e*) and (4*f*), the last equation is a consequence of

$$X \equiv \int [(A^2 + B^2)^{3/2} - 1]\, \mathrm{d}s. \tag{10}$$

We should mention that these equations apply not only to the case of the asteroid's motion as it is exactly commensurable with that of Jupiter's, but they allow any initial value of $A^2 + B^2$ as long as it is reasonably close to 1.

Solving equations (8*a*–*d*) (either numerically, or by the technique of strained coordinates [4]) reveals that all dependent variables oscillate with relatively large amplitudes around a steady mean in the case of both $(A^2 + B^2)^{3/2}$ and β, around a linear trend in the case of X and $\psi - s_0$. Without any extra perturbations this motion would continue indefinitely. But there are always other small disturbing forces acting on the asteroid (interestingly enough, even the truncation error of a numerical integration will serve this purpose) which affect most dramatically, even though extremely slowly, the mean values of $(A^2 + B^2)^{3/2}$ and β. After a correspondingly long time, the resulting slow drift will eventually bring the same two variables close to yet another possible (and trivial) solution of our differential equations, namely the one which makes $\sin(\psi - s_0 - X)$ identically equal to zero. At that point the oscillations cease and both the semi-major axis and eccentricity of the asteroid 'freeze' at their corresponding stationary values. These can be obtained from

$$(A^2 + B^2)^{3/2} = 1 \pm c\frac{\varepsilon}{2\beta} \tag{11a}$$

$$(A^2 + B^2)^{3/2} + 6\beta^2 = (A_0^2 + B_0^2)^{3/2} + 6\beta_0^2 \tag{11b}$$

where the zero subscript implies initial values. For $(A_0^2 + B_0^2)^{3/2} \leqslant 1.042$ the corresponding stationary solution is always smaller than 0.985. When $(A_0^2 + B_0^2)^{3/2} > 1.042$, the ultimate value of $(A^2 + B^2)^{3/2}$ will be the closest of three available stationary solutions, which is always bigger than 1.033. This creates a 3.2%-wide gap in observed values of $(A^2 + B^2)$. The actual gap is slightly narrower than our solution, but this may be due to some of the asteroids not having had enough time to abandon the oscillatory stage of their motion (this can be easily established by observing their eccentricity behaviour).

One has to conclude that the traditional explanations of Kirkwood gaps [5, 6], all based on the oscillating (transient) mode of the solution, appear incorrect. The inadequacy of these attempts seems to have been recognized [9, 2]; the current research usually concentrates on chaotic behaviour related to the gap regions [1], still failing to fully and properly explain the main phenomenon.

Equations (8*a*–*d*) are interesting from a mathematical point of view as well. They result in two solution classes, one in which $\psi - s_0 - X$ oscillates around a linear trend, the other in which the same variable oscillates around a solution to $\sin[2(\psi - s_0 - X)] = 0$. In their unperturbed form, all these solutions are oscillatory and stable, without being 'attracted' to the stationary values of equations (11*a*–*b*). The situation is dramatically different when the equations acquire a small dissipative term. A stationary solution then seems to be the ultimate fate of practically all initial conditions (this is true mainly of systematic perturbations, those generated randomly will sometimes lead to a continual drift in $A^2 + B^2$—the asteroid being 'kicked out' of its orbit). Once a stationary solution is reached, all oscillations cease. An exact mathematical analysis of these phenomena is beyond the scope of this paper.

5. Conclusion

A new technique for solving an age old problem of a perturbed satellite motion has been presented in detail. Its main advantage is its conceptual simplicity (the basic principle is to expand each side of equation (2) in powers of z, and match the corresponding coefficients) and the consequent ease of computerization. This also results in eliminating the problem of small divisors, and enables us to deal with perturbations having commensurable frequencies. On the negative side, our solution is obtained in a series form only, but this does not appear to be a major limitation.

The procedure has been successfully tested (up to the fourth order) under a variety of situations (lunar problem [11], perihelion precession, oblateness perturbations, etc). The range of its potential applications seems to include all traditionally treated problems of this kind. Furthermore, the technique is capable of dealing with them in a truly unified manner, requiring no special adaptation to the particular-case circumstances.

To underline the improvement achieved over some previous work, we point out that the original presentation [10] dealt with conservative forces only (no 'external' time dependence) and presented the technique in its 'double iterative' form (to build the solution, one had to iterate in terms of β, in addition to ε). The subsequent article [11] removed the conservative-force limitation and introduced an external time dependence of $\mathbf{f}$, but it considered its variation to be 'slow', and dealt with it by introducing yet another iteration parameter (our α), thus making the procedure 'triply iterative'! The current paper eliminates the need for the two extra iteration levels by presenting an explicit (in terms of β and α) set of formulae for the solution to be built, iteratively, in terms of powers of ε only.

References

[1] Ferraz-Mello S 1994 *Astron. J.* **108** 2330
[2] Hartmann W K 1993 *Moons and Planets* (Belmont: Wadsworth Publishing Company)
[3] Hestenes D 1986 *New Foundations of Classical Mechanics* (Dordrecht: Kluwer)
[4] Nayfeh A H 1973 *Perturbation Methods* (New York: Wiley)
[5] Peale S J 1976 *Ann. Rev. Astron. Astrophys.* **14** 215
[6] Roy A E 1988 *Orbital Motion* (Bristol: Adam Hilger)
[7] Souriau J M 1982 *Proc. IUTAM ISIMM Symp. on Modern Developments in Analytical Mechanics* (Torino)
[8] Stiefel E L and Scheifele G 1971 *Linear and Regular Celestial Mechanics* (Berlin: Springer)
[9] Taff L G 1985 *Celestial Mechanics, A Computational Guide for the Practitioner* (New York: Wiley)
[10] Vrbik J 1994 *Can. J. Phys.* **72** 141
[11] Vrbik J 1995 *Can. J. Phys.* **73** 193