CMST 7(2) 47-63 (2001) DOI:10.12921/cmst.2001.07.02.47-63

COMPUTATIONAL METHODS IN SCIENCE AND TECHNOLOGY 7 (2), 47-63 (2001)

USING POLYNOMIALS OF VARIABLE DEGREES FOR SOLVING THE RELATIVE *N*-KODY PROBLEM

ANDRZEJ MARCINIAK

Poznań University of Technology, Institute of Computing Science
Piotrowo 3a, 60-965 Poznań, Poland
and
Adam Mickiewicz University, Faculty of Mathematics and Computer Science
Matejki 48/49, 60-769 Poznań, Poland

Received 13 June, 2001

Abstract. A variable order method for solving the planetary type *N*-body problem, which is based on an approximation by polynomials of variable degrees, is proposed. We present an algorithm for finding such polynomials, notes on the stability and convergence of the method, and some selected numerical examples.

1. INTRODUCTION

The study of mutual positions of bodies (particles or material points) is one of the basic problems not only in celestial mechanics. In the *N*-body gravitational problem, the motion of *N* material points attracting one another in pairs is described by a system of differential equations of order $6N$ (motion in an inertial frame of references) or $6N - 6$ (relative motion) - see Section 2. As is well-known, the general solution of this system obtained by analytical methods is not available today. Therefore, numerical methods for solving the problem are used.

In order to solve the *N*-body problem we can use general numerical methods for solving the initial value problem or apply some special methods. A survey of the methods for solving the *N*-body problem is given, among others, in [1] and [9], From the point of view of the solution accuracy obtained, the most often used numerical methods are the Gragg-Bulirsch-Stoer method based on a rational approximation [3, 5], the Everhart method [4], and the Taylor-Steffensen method [15], which uses the Taylor series for the right-hand side functions occurring in the differential equations and recursive formulas for coefficients of this series. Some special methods conserving and using constants (integrals) of the motion should also be mentioned (see e. g. [6-13]).

Conventional numerical methods for solving the relative (planetary type) *N*-body problem with optimization or automatic step size correction do not seem to be the best for two reasons. Firstly, the optimization of step size depends on the 'speed' of change of the solution, which - in problems such as the problem of motion of the Solar system - leads to a determination of the optimum step size (in time) on the basis of the change of position and velocity for a planet which has the top mean motion. If the step size was chosen on the basis of the motion of a planet with small mean motion, the step size could be considerably greater. But the choice of a different step size for different planets (material points) is not sensible since the problem of motion of all planets should be solved at the same moments.

Secondly, in conventional methods the same accuracy of the solution for all planets is assumed, while in practice even the initial data have different accuracy for different planets. Therefore, it seems to be sensible to assume a different accuracy of the solution for each of its component.

In the method proposed in this paper we assume a constant step size, but different order for each component of the solution. The different orders, changed from step to step, we achieve using an approximation of each component by a polynomial of the degree which guarantees (for each moment) the accuracy given beforehand (see Section 3). It appears that for the method developed in such a way it is possible to prove some theorems on the consistency, stability, and convergence (see Section 4).

2. THE N-BODY PROBLEM

In the N-body gravitational problem, we are concerned with the motion of N mass particles of masses $m_i > 0$ ($i = 1, 2, \ldots, N$) attracting one another in pairs with force

$$G\frac{m_i m_j}{r_{ij}^2},$$

where r_{ij} is the distance between the i^{th} and j^{th} particle, and G denotes the gravitational constant. In an inertial and rectangular frame of reference the problem can be written in the for of the initial value problem as follows

$$\ddot{\xi}_{li} = -G\sum_{\substack{j=1\\ j\neq i}}^{N} m_j \frac{\xi_{li}-\xi_{lj}}{r_{ij}^3}, \quad \xi_{li}(t_0) = \xi_{li}^0, \quad \dot{\xi}_{li}(t_0) = \dot{\xi}_{li}^0, \qquad (2.1)$$

$$l = 1, 2, 3, \quad i = 1, 2, \ldots, N,$$

where

$$r_{ij} = \sqrt{\sum_{p=1}^{3}(\xi_{pi}-\xi_{pj})^2},$$

and where ξ_{li} and $\dot{\xi}_{li}$ are the l^{th} coordinate and l^{th} component of velocity of the i^{th} particle, respectively. Of course, we assume that ξ_{li}^0 and $\dot{\xi}_{li}^0$ are known at an initial moment t_0.

Since the basic problem in celestial mechanics is the study of mutual positions of bodies, we usually consider the motion of those bodies with respect to a central body of the system. Usually, it is a body with the greatest mass. For instance, in the Solar system we determine the motion of planets with respect to the Sun.

If we put the origin of a Cartesian coordinate system in the center of a particle with the mass m_N , then from (2.1) we get

$$\ddot{x}_{li} = -G\left[(m_N + m_i)\frac{x_{li}}{r_{iN}^3} + \sum_{\substack{j=1 \\ j \neq i}}^{N-1} m_j\left(\frac{x_{li} - x_{lj}}{r_{ij}^3} + \frac{x_{lj}}{r_{jN}^3}\right)\right], \tag{2.2}$$

$$x_{li}(t_0) = x_{li}^0, \quad \dot{x}_{li}(t_0) = \dot{x}_{li}^0, \quad l = 1, 2, 3, \quad i = 1, 2, \ldots, N-1,$$

where

$$r_{ij} = \sqrt{\sum_{p=1}^{3}(x_{pi} - x_{pj})^2}, \quad r_{iN} = \sqrt{\sum_{p=1}^{3} x_{pi}^2}, \quad i, j = 1, 2, \ldots, N-1, \quad i \neq j.$$

Between the coordinates ξ_{li} and x_{li} the following relation holds

$$x_{li} = \xi_{li} - \xi_{lN}, \quad l = 1, 2, 3, \quad i = 1, 2, \ldots, N-1.$$

It is common knowledge (see any handbook of celestial mechanics, e.g. [2] or [14]) that the above problem can be solved analytically only in the case $N = 2$ and in some special cases for $N = 3$. Thus, for arbitrary N we have to apply numerical methods.

3. APPROXIMATING THE SOLUTION BY POLYNOMIALS

If the functions occurring in the equations of motion (2.2) fulfill the assumptions of the Weierstrass theorem (what is easy to guarantee), then on the basis of this theorem we can search for the solution of (2.2) in the class of polynomials.

Let us try to find the polynomials $w_{li}(t)$ such that

$$x_{li} \cong w_{li}(t) = \sum_{k=0}^{P_{li}} a_{lik} t^k, \quad l = 1, 2, 3, \quad i = 1, 2, \ldots, N-1, \tag{3.1}$$

where P_{li} denotes the degree of $w_{li}(t)$ and may be different for different l and i, and a_{lik} are coefficients of the polynomial $w_{li}(t)$. Both, P_{li} and a_{lik} must be determined for each l=1,2,3; i=1,2,..., N - 1, and k =1,2,..., P_{li}.

From (3.1) it follows that

$$\ddot{x}_{li}(t) \cong \ddot{w}_{li}(t) = \sum_{k=0}^{P_{li}} (k+2)(k+1)\, a_{li,k+2}\, t^k, \tag{3.2}$$

$$x_{li}(t) - x_{lj}(t) \cong w_{li}(t) - w_{lj}(t) = \sum_{k=0}^{\max(P_{li}, P_{lj})} (a_{lik} - a_{ljk}) t^k, \tag{3.3}$$

where

$$a_{li,P_{li}+1} = \ldots = a_{li,P_{lj}} = 0, \text{ if } P_{li} < P_{lj},$$

$$a_{li,P_{lj}+1} = \ldots = a_{li,P_{li}} = 0, \text{ if } P_{li} > P_{lj}.$$

Moreover, we have

$$r_{iN}^2 \cong \sum_{q=1}^{3}\left(\sum_{k=0}^{P_{qi}} a_{qik}\, t^k\right)^2 = \sum_{k=0}^{2\underset{q}{\max}(P_{qi})} b_{ik}\, t^k, \qquad (3.4)$$

where

$$b_{ik} = \sum_{q=1}^{3}\begin{cases}\displaystyle\sum_{s=0}^{k} a_{qis}\, a_{qi,k-s}, & \text{for } k \le \underset{q}{\max}(P_{qi}),\\[2ex] \displaystyle\sum_{s=k-\underset{q}{\max}(P_{qi})}^{\underset{q}{\max}(P_{qi})} a_{qis}\, a_{qi,k-s}, & \text{for } k \ge \underset{q}{\max}(P_{qi})+1,\end{cases}$$

and

$$a_{qi,P_{qi}+1} = \ldots = a_{qi,\,\underset{q}{\max}(P_{qi})} = 0, \text{ for } q = 1, 2, 3,$$

$$r_{ij}^2 \cong \sum_{q=1}^{3}\left[\sum_{k=0}^{\underset{q}{\max}(P_{qi},\,P_{qj})} (a_{qik} - a_{qjk})\, t^k\right]^2 = \sum_{k=0}^{2\underset{q}{\max}(P_{qi},\,P_{qj})} c_{ijk}\, t^k, \qquad (3.5)$$

where

$$c_{ijk} = \sum_{q=1}^{3}\begin{cases}\displaystyle\sum_{s=0}^{k} (a_{qis} - a_{qjs})(a_{qi,k-s} - a_{qj,k-s}), & \text{for } k \le \underset{q}{\max}(P_{qi}, P_{qj}),\\[2ex] \displaystyle\sum_{s=k-\underset{q}{\max}(P_{qi},\,P_{qj})}^{\underset{q}{\max}(P_{qi},\,P_{qj})} (a_{qis} - a_{qjs})(a_{qi,k-s} - a_{qj,k-s}), & \text{for } k \ge \underset{q}{\max}(P_{qi}, P_{qj})+1,\end{cases}$$

and

$$a_{q\nu,P_{q\nu}+1} = \ldots = a_{q\nu,\underset{q}{\max}(P_{qi},\,P_{qj})} = 0, \text{ for } q = 1, 2, 3 \text{ and } \nu = i,\ j.$$

We also have

$$\left(\sum_{k=0}^{2\underset{q}{\max}(P_{qi})} b_{ik}\, t^k\right)^3 = \sum_{k=0}^{6\underset{q}{\max}(P_{qi})} d_{ik}\, t^k,$$

where

$$d_{ik} = \begin{cases} \sum\limits_{s=0}^{k} b_{is} \sum\limits_{z=0}^{k-s} b_{iz} b_{i,k-s-z}, \text{ for } 2\max\limits_q(P_{qi}), \\ \sum\limits_{s=k-\max\limits_q(P_{qi})}^{2\max\limits_q(P_{qi})} b_{is} \sum\limits_{z=0}^{k-s} b_{iz} b_{i,k-s-z} + \sum\limits_{s=0}^{k-1-2\max\limits_q(P_{qi})} b_{is} \sum\limits_{z=k-s-2\max\limits_q(P_{qi})}^{2\max\limits_q(P_{qi})} b_{iz} b_{i,k-s-z}, \\ \qquad \text{for } 2\max\limits_q(P_{qi})+1 \le k \le 4\max\limits_q(P_{qi}), \\ \sum\limits_{s=k-4\max\limits_q(P_{qi})}^{2\max\limits_q(P_{qi})} b_{is} \sum\limits_{z=k-s-2\max\limits_q(P_{qi})}^{2\max\limits_q(P_{qi})} b_{iz} b_{i,k-s-z}, \text{ for } k \ge 4\max\limits_q(P_{qi})+1, \end{cases}$$

and

$$\left(\sum_{k=0}^{2\max\limits_q(P_{qi}, P_{qj})} c_{ijk} t^k \right)^3 = \sum_{k=0}^{6\max\limits_q(P_{qi}, P_{qj})} e_{ijk} t^k,$$

where e_{iik} are calculated in the same way as d_{ik} (after substitutions $c_{ij\mu}$ for $b_{i\mu}$ and $\max\limits_q(P_{qi}, P_{qj})$ for $\max\limits_q(P_{qi})$).

Further, we have

$$\frac{1}{\sqrt{\sum\limits_{k=0}^{6\max\limits_q(P_{qi})} d_{ik} t^k}} = \frac{1}{\sqrt{d_{i0}}} \sum_{k=0}^{\infty} f_{ik} t^k,$$

where

$$f_{i0} = 1,$$
$$f_{i1} = -\frac{d_{i1}}{2d_{i0}},$$
$$f_{i2} = \frac{3d_{i1}^2}{8d_{i0}^2} - \frac{d_{i2}}{2d_{i0}},$$
$$f_{i3} = -\frac{5d_{i1}^3}{16d_{i0}^3} + \frac{3d_{i1}d_{i2}}{4d_{i0}^2} - \frac{d_{i3}}{2d_{i0}},$$

....

and

$$\frac{1}{\sqrt{\sum\limits_{k=0}^{6\max\limits_{q}(P_{qi},P_{qj})} e_{ijk}t^k}}=\frac{1}{\sqrt{e_{ij0}}}\sum_{k=0}^{\infty} g_{ijk}t^k,$$

where g_{ijk} can be found in a similar way to f_{ik} (after substitution $e_{ij\mu}$ for $d_{i\mu}$).

If we insert all of the above relations into the equations (2.2), we obtain

$$\begin{aligned}\sum_{k=0}^{P_{li}-2}(k+2)(k+1)a_{li,k+2}t^k=-G\Bigg(&(m_N+m_i)\frac{1}{\sqrt{d_{i0}}}\left(\sum_{k=0}^{P_{li}}a_{lik}t^k\right)\left(\sum_{k=0}^{\infty}f_{ik}t^k\right)\\&+\sum_{\substack{j=1\\ j\neq i}}^{N-1}m_j\left\{\left[\sum_{k=0}^{\max(P_{li},P_{lj})}(a_{lik}-a_{ljk})t^k\right]\frac{1}{\sqrt{e_{ij0}}}\left(\sum_{k=0}^{\infty}g_{ijk}t^k\right)\right.\\&\left.+\left(\sum_{k=0}^{P_{lj}}a_{ljk}t^k\right)\frac{1}{\sqrt{d_{j0}}}\left(\sum_{k=0}^{\infty}f_{jk}t^k\right)\right\}\Bigg),\end{aligned}\tag{3.6}$$

$$l=1,2,3,\quad i=1,2,\dots,N-1.$$

Since

$$\frac{1}{\sqrt{d_{i0}}}\left(\sum_{k=0}^{P_{li}}a_{lik}t^k\right)\left(\sum_{k=0}^{\infty}f_{ik}t^k\right)=\sum_{k=0}^{\infty}h_{lik}t^k,$$

where

$$h_{lik}=\begin{cases}\dfrac{1}{\sqrt{d_{i0}}}\sum\limits_{s=0}^{k}a_{lis}f_{i,k-s}, & \text{for } k\le P_{li},\\[2ex] \dfrac{1}{\sqrt{d_{i0}}}\sum\limits_{s=0}^{P_{li}}a_{lis}f_{i,k-s}, & \text{for } k\ge P_{li}+1,\end{cases}$$

and

$$\begin{aligned}\sum_{\substack{j=1\\ j\neq i}}^{N-1}m_j&\left\{\frac{1}{\sqrt{e_{ij0}}}\left[\sum_{k=0}^{\max(P_{li},P_{lj})}(a_{lik}-a_{ljk})t^k\right]\left(\sum_{k=0}^{\infty}g_{ijk}t^k\right)\right.\\&\left.+\frac{1}{\sqrt{d_{j0}}}\left(\sum_{k=0}^{P_{lj}}a_{ljk}t^k\right)\left(\sum_{k=0}^{\infty}f_{jk}t^k\right)\right\}=\sum_{\substack{j=1\\ j\neq i}}^{N-1}m_j\left(\sum_{k=0}^{\infty}u_{lijk}t^k+\sum_{k=0}^{\infty}h_{ljk}t^k\right),\end{aligned}$$

where u_{lijk}, should be calculated in the same was as h_{lik} (after substitutions e_{ij0} for d_{i0}, $a_{li\mu}$ - $a_{lj\mu}$ for $a_{li\mu}$, $g_{lj\mu}$ for $f_{i\mu}$, and $\max(P_{li}, P_{lj})$ for P_{li}), we can rewrite the equations (3.6) in the form

$$\sum_{k=0}^{P_{li}-2}(k+2)(k+1)a_{li,k+2}t^k = -G\sum_{k=0}^{\infty}\left[(m_N + m_i)h_{lik} + \sum_{\substack{j=1\\ j\neq i}}^{N-1} m_j(u_{lijk} + h_{ljk})\right]t^k,$$

$$l = 1, 2, 3, \quad i = 1, 2, \ldots, N-1.$$

Hence

$$\sum_{k=0}^{P_{li}-2}(k+2)(k+1)a_{li,k+2}t^k = -G\sum_{k=0}^{P_{li}-2}\left[(m_N + m_i)h_{lik} + \sum_{\substack{j=1\\ j\neq i}}^{N-1} m_j(u_{lijk} + h_{ljk})\right]t^k + O(t^{P_{li}-1}), \tag{3.7}$$

$$l = 1, 2, 3, \quad i = 1, 2, \ldots, N-1.$$

if the functions $h_{li,P_{li}-1}$, $u_{lij,P_{li}-1}$ and $h_{lj,P_{li}-1}$ are bounded. From (3.7) it follows that — excepting terms $O(t^{P_{li}-1})$ — for each $k = 0, 1, \ldots, P_{li} - 2$ we have

$$a_{li,k+2} = -\frac{G}{(k+2)(k+1)}\left[(m_N + m_i)h_{lik} + \sum_{\substack{j=1\\ j\neq i}}^{N-1} m_j(u_{lijk} + h_{ljk})\right], \tag{3.8}$$

$$l = 1, 2, 3, \quad i = 1, 2, \ldots, N-1.$$

Let us note that the functions h_{lik}, u_{lijk} and h_{ljk} on the right-hand side of (3.8) do not contain the coefficient $a_{li,k+2}$. Moreover, taking into account the previous formulas, we get (for each $k = 0, 1, \ldots, P_{li} - 2$)

$$h_{lik} = \frac{1}{\sqrt{d_{i0}}}\sum_{s=0}^{k} a_{lis} f_{i,k-s}, \tag{3.9}$$

$$u_{lijk} = \frac{1}{\sqrt{e_{ij0}}}\sum_{s=0}^{k}(a_{lis} - a_{ljs}) g_{ij,k-s}, \tag{3.10}$$

where

$$f_{ik} = f_{ik}(d_{i0}, d_{i1}, \ldots, d_{ik}), \tag{3.11}$$

$$g_{ik} = g_{ik}(e_{ij0}, e_{ij1}, \ldots, e_{ijk}), \tag{3.12}$$

and where

$$d_{ik} = \sum_{s=0}^{k} b_{is} \sum_{z=0}^{k-s} b_{iz} b_{i,k-s-z},$$

$$e_{ijk} = \sum_{s=0}^{k} c_{ijs} \sum_{z=0}^{k-s} c_{ijz} c_{ij,k-s-z},$$

$$b_{ik} = \sum_{q=1}^{3} \sum_{s=0}^{k} a_{qis} a_{qi,k-s},$$

$$c_{ijk} = \sum_{q=1}^{3} \sum_{s=0}^{k} (a_{qis} - a_{qjs})(a_{qi,k-s} - a_{qj,k-s}).$$

The functions f_{ik} and g_{ik} are given by somewhat complicated formulas, but there is a way to simplify them. First, let us introduce a *multiple sum* symbol.

Definition.

$$\begin{aligned} &{}^{(0)}\sum_{s_0=p}^{s_1} \alpha_{s_0} = \alpha_{s_1-p+1}, \\ &{}^{(k)}\sum_{s_0=p}^{s_{k+1}} \alpha_{s_0} = \sum_{s_k=p}^{s_{k+1}} \left({}^{(k-1)}\sum_{s_0=p}^{s_k} \alpha_{s_0} \right) \alpha_{s_{k+1}+p+1-s_k}. \end{aligned} \tag{3.13}$$

The above recursive definition makes a rule that the 0-based multiple sum is a single element, and the k-based multiple sum one can obtain as a regular sum of products of the $(k - 1)$-based multiple sum and an element. Using this definition we can significantly simplify some notations, for instance

$$\sum_{s=2}^{k} \left[\sum_{p=2}^{s} \left(\sum_{l=2}^{p} \alpha_{l-1} \alpha_{p-1-l} \right) \alpha_{s-1-p} \right] \alpha_{k-1-s} =^{(3)} \sum_{s_0=2}^{s_4=k} \alpha_{s_0}.$$

Using (3.13), we can write the formulas that determine f_{ik} and g_{ijk} (see (3.11) and (3.12)) in the form

$$\begin{aligned} &f_{i0} = 1, \\ &f_{ik} = \sum_{n=1}^{k} \left[\binom{-\frac{1}{2}}{n} \frac{1}{d_{i0}^{n}} {}^{(n-1)}\sum_{s_0=0}^{k-n} d_{i,s_0} \right], \quad k = 1, 2, \ldots, \end{aligned} \tag{3.14}$$

$$\begin{aligned} &g_{ij0} = 1, \\ &g_{ijk} = \sum_{n=1}^{k} \left[\binom{-\frac{1}{2}}{n} \frac{1}{e_{ij0}^{n}} {}^{(n-1)}\sum_{s_0=0}^{k-n} e_{ij,s_0} \right], \quad k = 1, 2, \ldots, \end{aligned} \tag{3.15}$$

The above formulas may also be written as follows

$$f_{i0} = 1,$$
$$f_{ik} = \sum_{n=1}^{k} \binom{-\frac{1}{2}}{n} \alpha_{ikn}, \quad k = 1, 2, \ldots, \tag{3.16}$$

where

$$\alpha_{ik1} = \frac{d_{ik}}{d_{i0}},$$
$$\alpha_{ikn} = \frac{1}{d_{i0}} \sum_{p=n-1}^{k-1} \alpha_{ip,n-1} d_{i,k-p}, \quad n = 2, 3, \ldots, k, \tag{3.17}$$

and

$$g_{ij0} = 1,$$
$$g_{ijk} = \sum_{n=1}^{k} \binom{-\frac{1}{2}}{n} \beta_{ijkn}, \quad k = 1, 2, \ldots, \tag{3.16}$$

where

$$\beta_{ijk1} = \frac{e_{ijk}}{e_{ij0}},$$
$$\beta_{ikn} = \frac{1}{e_{ij0}} \sum_{p=n-1}^{k-1} \beta_{ijp,n-1} e_{ij,k-p}, \quad n = 2, 3, \ldots, k, \tag{3.17}$$

If we fix i, then the quantities a_{ikn} form an upper triangular matrix with elements calculated — according to (3.17) — on the basis of the following scheme

$$\begin{tikzcd}
\alpha_{i11} \arrow[dr] \arrow[ddr] \arrow[dddr] & & & \\
\alpha_{i21} \arrow[dr] \arrow[ddr] & \alpha_{i22} \arrow[dr] \arrow[ddr] & & \\
\alpha_{i31} \arrow[dr] & \alpha_{i32} \arrow[dr] & \alpha_{i33} \arrow[dr] & \\
\alpha_{i41} & \alpha_{i42} & \alpha_{i43} & \alpha_{i44}
\end{tikzcd}$$

. . . .

The quantities β_{ijkn}, given by (3.19), are evaluated on the basis of the same scheme.

An approximation to the solution of (2.2) at the moment $t_{\nu+1} = t_0 + (\nu+1)h$, where h is a given step size, we determine from the formulas

$$x_{li}^{\nu+1} = \sum_{k=0}^{p_{li}} a_{lik} h^k, \quad \dot{x}_{li}^{\nu+1} = v_{li}^{\nu+1} = \sum_{k=1}^{p_{li}} k\, a_{lik} h^{k-1},$$
$$\nu = 0, 1, \ldots, \quad l = 1, 2, 3, \quad i = 1, 2, \ldots, N-1, \tag{3.20}$$

where $a_{li0} = x_{li}^{\nu}$, $a_{li1} = v_{li}^{\nu} = \dot{x}_{li}^{\nu}$, and the coefficients $a_{lik} = a_{lik}(t_0 + \nu h) = a_{lik}(x^{\nu}, v^{\nu})$ (k=2,3,..., P_{li}) are calculated from (3.8)-(3.10).

Now, let us try to find P_{li} (for each l and i). Let P_{li} denote such a degree of polynomial that

$$\left| a_{li,P_{li}} h^{P_{li}} \right| < \varepsilon_{li}, \quad \left| a_{li,P_{li}} h^{P_{li}-1} \right| < \varepsilon_{li},$$
$$\left| P_{li} a_{li,P_{li}} h^{P_{li}-1} \right| < \varepsilon_{li}, \quad \left| (P_{li}-1) a_{li,P_{li}} h^{P_{li}-2} \right| < \varepsilon_{li}, \tag{3.21}$$

The above conditions mean that the summations in (3.2) should be finished if an addition of consecutive elements does not cause a change in the result greater than $2\varepsilon_{li}$, where ε_{li} is a given accuracy for each l and i.

Let us note that in (3.21) it is necessary to take into account two consecutive elements of the sums which occur in (3.20), since according to the analytical theory of the relative N-body problem (see e.g. [2], [14] or [17]) there exist some simple case in which series expansions of x_{li} $=x_{li}(t)$ and $\dot{x}_{li} = \dot{x}_{li}(t)$ contain even or odd powers of h only. An example of such a case is the circular motion in the two-body problem.

An application of the criteria (3.21) for finding P_{li} need an existence of constants $K_{li} > 0$ such that for each $k_{li} \geq K_{li}$, one or both of the following inequalities are fulfilled:

$$\left| a_{li,k_{li+2}} h^{k_{li}+2} \right| < \left| a_{li,k_{li}} h^{k_{li}} \right| \tag{3.22}$$

or

$$\left| a_{li,k_{li+2}} h^{k_{li}+1} \right| < \left| a_{li,k_{li}} h^{k_{li}} \right|. \tag{3.23}$$

The first inequality — (3.22) — means that starting from a certain odd (even) element of the series, all further odd (even) elements are decreasing, and the second inequality — (3.23) — means that staring from a certain element all further elements are decreasing. Of, course, neither the condition (3.22) nor (3.23) follow from the convergency of the Taylor series for $x_{li} = x_{li}$ (t) and $\dot{x}_{li} = \dot{x}_{li}(t)$, which are the solution of (2.2). Taking into consideration the cases mentioned earlier (about odd and even powers of *h*, we can eliminate the inequality (3.23). Moreover, if we accept the condition (3.22) as an assumptions, we can prove [11]

Theorem 1. *If the inequality* (3.22) *holds for each* $k_{li} \geq 1$ (*l*= 1,2,3; *i = l,2,...,N* -1), *then in the method* (3.20), *in which the coefficients* a_{lik} *are calculated from* (3.8) - (3.10) *and* (3.16) - (3.19), *the step size h should fulfill the relation*

$$|h| \dot{<} \min_{l,i}\left(h_i^{(1)}, h_{li}^{(2)}\right), \tag{3.24}$$

where the dot means an approximate inequality, and

$$h_i^{(1)} = \sqrt{\frac{2\rho_{iN}^3}{G(m_N + m_i)}},$$

$$h_{li}^{(2)} = \sqrt{\frac{6\rho_{iN}^3}{G(m_N + m_i)\left(\frac{3\left|a_{li0}\right| \sum_{p=1}^{3} \left|a_{pi0}\, a_{pi1}\right|}{\left|a_{li1}\right| \rho_{iN}^2} + 1\right)}},$$

$$\rho_{iN} = \sqrt{\sum_{p=1}^{3} a_{pi0}^2},$$

but if for some l and i we have $a_{li1} = 0$, then the adequate value of $h_{li}^{(2)}$ should not be taken into account in (3.24).

4. NOTES ON THE STABILITY AND CONVERGENCE OF THE METHOD

The stability and convergence of the method presented in Section 3 may be proved on the basis of the Stetter theory about general analysis of discretization methods for ordinary differential equations [16]. In what follows we present some lemmas and theorems concerning our method. The proofs, which one can find in [11], are omitted here because of the complicated calculations involved.

Lemma 1. *If there exist the following constants*

$$0 < r = \min_{\substack{t\in[t_0,T] \\ i,j=1,2,\ldots,N \\ i\neq j}} r_{ij}(t), \quad R = \max_{\substack{t\in[t_0,T] \\ i,j=1,2,\ldots,N \\ i\neq j}} r_{ij}(t),$$
$$V = \max_{\substack{t\in[t_0,T] \\ i,j=1,2,\ldots,N \\ i\neq j}} \left|v_{ij}(t)\right|, \tag{4.1}$$

where

$$r_{ij}(t) = \sqrt{\sum_{p=1}^{3}[x_{pi}(t) - x_{pj}(t)]^2}, \quad r_{iN}(t) = \sqrt{\sum_{p=1}^{3} x_{pi}^2},$$

then there exists a constant $W = W(k) > 0$ *such that for each* $l = 1,2,3$; $i = 1, 2 \ldots N - 1$, *and* $k \geq n$ *we have*

$$\left|a_{lik}\right| \leq W,$$

where $a_{ljk} = a_{ljk}(x^v, y^v)$, $v = 0,1,\ldots, n$.

Using this lemma we can prove

Theorem 2. *If there exist the constants* (4.1), *then the method* (3.20) *is consistent with the initial value problem* (2.2).

Two next lemmas, namely

Lemma 2. *If there exist the constant* (4.1), *then for arbitrary* 1, p = 1, 2, 3; *i, q* = 1, 2,..., N - 1, *and* *we have*

$$\left|\frac{\partial a_{lik}}{\partial x_{pq}}\right| \le C(k) \ \ and \ \ \left|\frac{\partial a_{lik}}{\partial v_{pq}}\right| \le C(k),$$

where $C(k) > 0$ *denotes a constant that depends on k only.*

and

Lemma 3. *If there exist the constants* (4.1), *then for each* l = 1,2,3; i = *1,2,...,* N — 1, *and* *the following inequality holds:*

$$\left|a_{lik}(\bar{x}, \bar{y}) - a_{lik}(x, y)\right| \le C(k) \sum_{p=1}^{3} \sum_{q=1}^{N-1} \left(\left|\bar{x}_{pq} - x_{pq}\right| + \left|\bar{v}_{pq} - v_{pq}\right|\right)$$

where $C(k) > 0$ *denotes a constant.*

allow us to prove

Theorem 3. *If there exist the constants* (4.1), *then the method* (3.20) *is stable on the initial value problem* (2.2) *(in the sense of* [16, Section 1.1.4, Definition 1.1.10]) *with the stability constant*

$$S = \exp[3(N-1)(P+1)Q + P - 1],$$

where $P = \max_{\substack{l=1,2,3 \\ i=1,2,\dots,N-1}} P_{li}$, $Q = \max_{k=2,3,\dots,P} C(k)$, *and* $C(k) > 0$ *are constants from the above lemmas.*

On the basis of [16, Section 1.2.1, Theorem 1.2.3] from the theorems 2 and 3 immediately follows

Theorem 4. *If there exist the constants* (4.1), *then the method* (3.20) *is convergent on the initial value problem* (2.2).

5. NUMERICAL EXAMPLES

First, let us test our method for a problem the exact solution of which is known. Let a material point with the mass $m_1 = 1$ at the initial moment t_0 be located at $(x_1^0, x_2^0) = (1, 0)$ on the x_1x_2 plane, and let the velocity at t_0 be given by $(v_1^0, v_2^0) = (\dot{x}_1^0, \dot{x}_2^0) = (0, \alpha)$. If the material point m_1 orbits elliptically the material point m_2, in which the origin of the rectangular frame is located, then (see e.g. [2], [14] or [17])

$$x_1 = a(\cos E - e), \quad x_2 = a\sqrt{1-e^2} \sin E, \tag{5.1}$$

where e is the eccentricity, E denotes the eccentric anomaly (see further), and

$$a = \frac{c_1^2}{\mu(1-e^2)}, \quad e = \sqrt{1+\frac{2c_1^2 c_2}{\mu^2}}, \quad \mu = G(m_1+m_2),$$

and where the constants c_1, c_2 can be found from the initial conditions. Since (see e.g. [11] or [17])

$$\alpha = \sqrt{\mu(1 \pm e)},$$

then ifwe assume the '+' sign, m_2 =328900.1, and G =1.20021974563227948 x 10^{-4}, we have

$$\alpha = 6.28294194291318370 \times \sqrt{1+e}.$$

From this relation we can evaluate a, and thereby $v^0_2 = v_2\ (t_0)$, in such a way that an elliptic orbit with an eccentricity e given beforehand will be fully determined (see Table 1).

Table 1. Initial velocities and periods for the given eccentricities

e	$v_2(t_0)$	period
0.00	6.28294194291318370	1.00003873412624436
0.05	6.43809967544139349	1.08001904469491341
0.10	6.58960510226671025	1.17125931415944000
0.20	6.88261805925777273	1.39759661852445071
0.30	7.16365600063424312	1.70753557924319781
0.50	7.69500092183402779	2.82853668139951299
0.70	8.19195404519764599	6.08604192288728225

In the test two-body problem considered the 'relative error' has been determine as follows

$$\varepsilon = \frac{1}{2}\left(\frac{\left\|x-\bar{x}\right\|}{\left\|\bar{x}\right\|} + \frac{\left\|v-\bar{v}\right\|}{\left\|\bar{v}\right\|}\right),$$

where $\bar{x}$ and $\bar{v} = \frac{d\bar{x}}{dt}$ denote the exact solution obtained from (5.1), and $\left\|x\right\| = \left|x_1\right| + \left|x_2\right|$.

Applying the method from Section 3 to the orbits from Tablel, we get (after adequate periods) the relative errors presented in Table 2. In all calculations the accuracy 10^{-15} has been assumed for each component of the solution, i.e. ε_1 for x_1 and v_1, ε_2 for x_2 and v_2, and

$\varepsilon_1 = \varepsilon_2 = 10^{-15}$.

In our method for each component of the solution the appropriate degree of polynomial is chosen on the basis of the accuracy given beforehand. Influences of these accuracies on the relative errors are given in the next table (Table 3), and in Table 4 we present the achieved degrees ofpolynomials for different eccentricities. Let us note that higher degrees are obtained for greater eccentricities, what corresponds to a decrease ofstep size in automatic step size correction methods. Finally, in Table 5 we present some results obtained using our method for long time integrations.

Table 2. Relative errors ($\varepsilon_1 = \varepsilon_2 = 10^{-15}$)

e	step size	mean degree of polynomials	relative error after the period ($\times 10^{16}$)
0.00	period/10 = 0.100003873412624436	18	0.2
0.05	period/10 = 0.108001904469491341	25	1.6
0.10	period/10 = 0.117125931415944000	28	6.7
0.20	period/20 = 0.0698798309262225355	22	5.0
0.30	period/20 = 0.0853767789621598905	24	0.6
0.50	period/40 = 0.0707134170349878248	20	22.9
0.70	period/90 = 0.0676226880320809140	16	183.2

Notes: 1) Mean degree of polynomials = mean order of solution obtained by our method
2) Period = period of orbiting

Table 3. Relative errors depending on given accuracies (e = 0.1, h = period/10)

$\varepsilon_1 = \varepsilon_2$	mean degree of polynomials	relative error after the period
10^{-18}	33	0.3×10^{-16}
10^{-16}	30	0.4×10^{-16}
10^{-14}	27	$49.5 \times 10^{-16} \cong 0.5 \times 10^{-14}$
10^{-12}	23	$8962.2 \times 10^{-16} \cong 0.9 \times 10^{-12}$

Table 4. Degrees of polynomials for solution coordinates x_1, x_2 ($\varepsilon_1 = \varepsilon_2 = 10^{-15}$, $h = 0.1$)

e	degrees of polynomials				
	$t = 0.2$	$t = 0.4$	$t = 0.6$	$t = 0.8$	$t = 1.0$
0.00	18, 18	18, 18	18, 17	18, 18	18, 18
0.05	26, 26	24, 23	22, 21	23, 22	25, 25
0.10	29, 29	26, 25	22, 22	22, 23	26, 26
0.20	34, 34	27, 27	21, 23	20, 19	23, 23

Table 5. Relative errors and computational times for 100 x period ($\varepsilon_1 = \varepsilon_2 = 10^{-15}$)

e	step size	degrees of polynomials at 100 × period	relative error at 100 × period (×10^{16})	relative time of computations
0.00	0.100003873412624436	18, 18	10.25	1.00
0.05	0.108001904469491341	27, 27	150.53	2.61
0.10	0.117125931415944000	32, 33	655.47	3.86

Note: "Relative time of computations" means that the time for *e* = 0.00 has been taken as a unit

I have compared the method presented in Section 3 with a number of well-known numerical methods. For the two-body problem considered, the relative errors obtained in three selected methods are presented in Table 6, while in Figure 1 we show a comparison of computational time for these methods (the computational time for the Taylor-Steffensen method with *e* = 0.0 has been taken as a unit).

Table 6. Relative errors in selected conventional methods

e	relative errors after the period (×10^{16})		
	TS (order)	GBS	EV (order)
0.00	0.39 (18)	14.99	0.77 (13)
0.05	8.56 (25)	20.99	3.14 (12)
0.10	14.74 (28)	25.76	3.92 (13)
0.20	36.38 (22)	23.82	1.17 (12)
0.30	5.68 (24)	34.37	5.64 (12)
0.50	259.80 (20)	115.37	5.83 (12)

TS - the Taylor-Steffensen method [15] with an automatic step size correction
GBS - the Gragg-Bulirsch-Stoer method [5],
EV - Everhart's method [4]

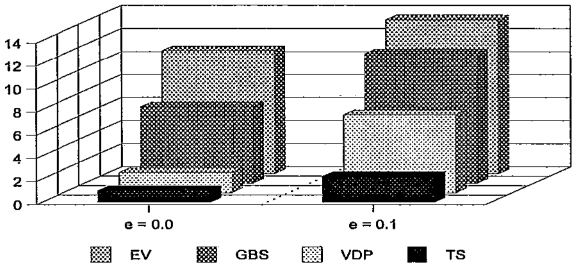

Fig 1. Computational times (VDP - the method of variable degree polynomials)

From the results presented, it follows that only accuracies obtained by the method of Everhart can be compared with those obtained by our method. On the other hand, the method of variable degree polynomials is more efficient (with respect of computational times) than that of Everhart, and greater values of e only confirms this conclusion. From the point of view of efficiency, the Taylor-Steffensen method seems to be the best.

The method of variable degree polynomials is especially efficient for small eccentricities and in problems with the number of material points $N > 2$ where for each such a point (and even for each coordinate and each component of velocity) we can assume a different accuracy. The motion of the Solar system is an example of such a problem. Applying Theorem 1 it is possible to evaluate the maximum step size for this problem, which depends on the planets considered (see Table 7). From the point of view of method accuracy we do not recommend step sizes greater than half of the values given in Table 7.

Table 7. Maximum step size in the problem of motion of the Solar system (evaluated from the initial data at 1950.0 - the beginning of the year 1950)

planets considered	maximum step size (in years)								
Mercury	0.015								
Venus	↓	0.139							
Earth+Moon	↓	↓	0.228						
Mars	↓	↓	↓	0.472					
Jupiter	↓	↓	↓	↓	2.86				
Saturn	↓	↓	↓	↓	↓	6.37			
Uranus	↓	↓	↓	↓	↓	↓	19.6		
Neptune	↓	↓	↓	↓	↓	↓	↓	36.9	
Pluto	↓	↓	↓	↓	↓	↓	↓	↓	62.6

As for $N = 2$, we performed a number of tests for $N > 2$ and compared numerous well-known conventional methods with ours. As it turned out, only Taylor-Steffensen method with an automatic step size correction was comparable from the point of view of efficiency. As an example we present some results for the problem of motion of giant planets of the Solar system (Jupiter, Saturn, Uranus and Neptune). We have solved this problem for 500 years using step size $h = 0.5$ year, and equal (10^{-12}) and different (from 10^{-12} for Jupiter to 10^{-9} for Neptune) assumed accuracies in components of the solution. It turned out that the different accuracies did not cause the solution to change significantly and they enabled to save about 5% of CPU time (see Figure 2, where the computational time for the Taylor-Steffensen method of order 9 has been taken as a unit). It should be noted that the mean degrees of polynomials in the method with equal accuracies were equal from 13 to 14 for Jupiter to 11 for Neptune, while in the method with different accuracies assumed - from 13 to 14 for Jupiter to 8 - 9 for Neptune.

Finally, let us add one remark. Any decreasing of assumed accuracies must be carried out with great care. A lesser accuracy we can assume only for a material point with a small mass or very distant from other points, i. e. for a material point whose gravitational influence on other points is relatively small.

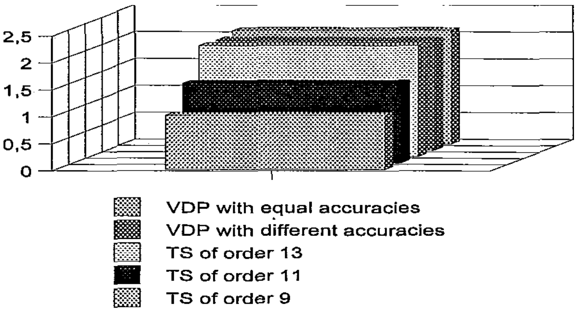

Fig 2. CPU times for the test five-body problem

References

[1] Bordovicyna.T. V., *Modern Numerical Methods in Celestial Mechanics Problems* (in Russian), Izdatel'stvo 'Nauka', Moscow 1984.

[2] Brouwer, D., Clemence, G. M., *Methods ofCelestial Mechanics,* Academic Press, New York 1961.

[3] Bulirsch, R., Stoer, J., Numerical Treatment of Ordinary Differential Equations by Extrapolation Methods, *Numer. Math.* 8 (1966), 1-13.

[4] Everhart, E., Implicit Single Sequence Methods for Integrating Orbits, *Celestial Mech.* 10 (1974), 35-55.

[5] Gragg, W. B., On Extrapolation Algorithms of Ordinary Initial Value Problems, *J. SIAM Numer Anal.* 2 (1965), 384-403.

[6] Greenspan, D., Discrete Newtonian Gravitation and the *N*-body Problem, *Utilitas Math.* 2(1972), 105-126.

[7] Greenspan, D., Conservative Numerical Methods for $\ddot{x} =f(x)$, *J. Comput. Phys.* 56 (1984), 28-41.

[8] Marciniak, A., Energy Conserving, Arbitrary Order Numerical Solutions of the *N*-body Problem, *Mumer. Math.* 45 (1984), 207-218.

[9] Marciniak, A., *Numerical Solutions ofthe N-body Problem,* D. Reidel Publishing Co., Dordrecht 1985.

[10] Marciniak, A., Arbitrary Order Numerical Solutions Conserving the Jacobi Constant in the Motion Nearby the Equilibrium Points, *Celestial Mech.* 40 (1987), 95-110.

[11] Marciniak, A., *Selected Numerical Methodsfor Solving the N-body Problem* (in Polish), Thesis No. 213, Poznań Univ. ofTechnology, Poznań 1989.

[12] Marciniak, A., Arbitrary Order Numerical Methods Conserving Integrals for Solving Dynamic Equations, *Computers Math. Applic.* 28 (1994), 33-43.

[13] Nacozy, P. E., The Use of Integrals in Numerical Integration of *N*-body Problem, *Astrophys. Space Sci.* 14 (1971), 40-41.

[14] Pollard, H., *Mathematical Introduction to Celestial Mechanics,* Prentice Hall, New Jersey 1966.

[15] Steffensen, K. F., On the Problem ofThree Bodies in the Plane, *Kong. Danske Videnskab. Seskob., Mat. Fys. Medd.* 13 (1957).

[16] Stetter.H. *L, Analysis ofDiscretizationMethodsforOrdinaiy Differential Equations, Spńngev-Veńag, Berlin* 1973.

[17] Wierzbiński, S., *Celestial Mechanics* (in Polish), PWN, Warsaw 1973.