Интегратор Гаусса—Эверхарта*

В. А. Авдюшев
НИИ прикладной математики и механики
Томского государственного университета, Россия
e-mail: sch@niipmm.tsu.ru

Представлена новая версия интегратора Эверхарта для решения обыкновенных дифференциальных уравнений первого порядка. Обсуждаются особенности в компьютерной реализации интегратора, исследуются его возможности при решении задач небесной механики.

Ключевые слова: численное интегрирование, обыкновенные дифференциальные уравнения, интегратор Эверхарта, неявные методы Рунге—Кутты, численное моделирование орбит.

Введение

В 1973 г. Э. Эверхарт [1] предложил интегратор, специально разработанный для численного исследования орбит, и продемонстрировал его высокую эффективность в задачах кометной динамики. По-видимому, обнаружив в дальнейшем принадлежность данного интегратора к семейству интеграторов бутчеровского типа, Эверхарт акцентировал внимание на оригинально реализованный им алгоритм интегрирования и обобщил его для численного решения любых обыкновенных дифференциальных уравнений первого и второго порядка [2, 3], тем самым расширив область применения предложенного интегратора, который, тем не менее, остается одним из самых популярных именно в решении задач небесной механики.

Интегратор Эверхарта (RA15) основан на видоизмененных формулах неявных коллокационных методов Рунге—Кутты бутчеровского типа, поэтому он наследует все их замечательные свойства [4]. Более того, именно благодаря оригинальному представлению вычислительной схемы интегратор Эверхарта с точки зрения численного интегрирования имеет ряд следующих преимуществ:

1) алгоритм интегрирования универсален для любого порядка;

2) интегратор имеет простой критерий для выбора шага интегрирования;

3) в интеграторе реализован достаточно точный предиктор решения, что позволяет выполнять численное интегрирование всего с двумя итерациями на шаге.

Несмотря на это, программный код Эверхарта RA15 [3] (впрочем, как и любая его модификация типа RADAU_27), на наш взгляд, довольно существенно ограничивает возможности интегратора и поэтому нуждается в дополнительной редакции.

*Работа выполнена при финансовой поддержке РФФИ (грант № 08-02-00359) и Министерства образования РФ (грант № РНП.2.1.1/2629).

Среди главных недостатков в программной реализации интегратора можно назвать следующие:

1) трудночитаемый и громоздкий код;

2) много констант, связанных с порядком интегратора, что затрудняет обобщение кода на другие порядки;

3) интегратор реализован только для определенных порядков, причем для нечетных с разбиением Гаусса—Радо, хотя известно, что неявные коллокационные методы Рунге—Кутты, построенные на симметричных разбиениях Гаусса—Лобатто и Гаусса–Лежандра, обладают геометрическими свойствами [5];

4) алгоритм выбора шага для уравнений первого порядка используется такой же, как и для уравнений второго порядка, поэтому шаг при интегрировании уравнений первого порядка выбирается неверно;

5) стартовый шаг интегрирования в режиме переменного шага выбирается независимо от дифференциальных уравнений, поэтому не всегда оптимально;

6) ограничения на величину выбираемого переменного шага, по-видимому, заданы в интеграторе просто из эмпирических соображений и, кроме того, не зависят от порядка интегратора.

В настоящей работе представлен новый код интегратора Эверхарта GAUSS_15, который позволяет устранить перечисленные выше недостатки:

1) путем использования возможностей Фортран 90 программный код сокращен почти в 2 раза;

2) устранены все константы, связанные с порядком метода (оставлены лишь константы узловых значений на шаге);

3) код позволяет получать решение 2–15 порядка точности (хотя при необходимости код без изменений можно обобщить на любой другой порядок: для этого нужно лишь получить соответствующие узловые значения);

4) исправлен алгоритм выбора переменного шага;

5) стартовый шаг выбирается по оценке интегрирующей схемы второго порядка с учетом поведения правых частей уравнений;

6) накладываются ограничения на выбираемый шаг в соответствии с порядком интегратора.

Кроме того, предложенный интегратор имеет новые возможности:

1) интегрирование на шаге до полной сходимости итерационного процесса;

2) запоминание величины предпоследнего шага после выполнения процедуры интегрирования при многократном использовании программного кода в режиме переменного шага;

3) быстрый выбор стартового шага, требуемый лишь для первого обращения к интегратору (при повторном обращении применяется запоминаемый шаг предыдущего обращения).

В работе представлена общая теория интегратора Эверхарта с внесенными автором коррективами, предлагается новый программный код интегратора, а также показаны результаты тестирования интегратора на примере дифференциальных уравнений задачи двух тел. В дальнейшем ввиду того что интегратор использует гауссовы разбиения, будем называть его интегратором Гаусса—Эверхарта (хотя обычно такие интеграторы называются гауссовыми [5]).

1. Основные формулы

Предположим, что на шаге h мы решаем задачу

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \quad \mathbf{x}_0 = \mathbf{x}(t_0). \tag{1}$$

Здесь t — независимая переменная, $\mathbf{x}$ — интегрируемые переменные, $\mathbf{f}$ — заданная вектор–функция t и $\mathbf{x}$. Введем переменную $\tau = (t - t_0)/h$ и представим правую часть уравнений (1) в виде полинома степени k:

$$\mathbf{x}' = \mathbf{x}'_\tau/h = \mathbf{f} = \mathbf{f}_0 + \mathbf{A}_1\tau + \mathbf{A}_2\tau^2 + \mathbf{A}_3\tau^3 + \ldots + \mathbf{A}_k\tau^k, \tag{2}$$

где коэффициенты $\mathbf{A}_i$ пока не определены. Интегрируя (2) по τ, получаем решение

$$\mathbf{x} = \mathbf{x}_0 + h\left(\mathbf{f}_0\tau + \frac{1}{2}\mathbf{A}_1\tau^2 + \frac{1}{3}\mathbf{A}_2\tau^3 + \frac{1}{4}\mathbf{A}_3\tau^4 + \ldots + \frac{1}{k+1}\mathbf{A}_k\tau^{k+1}\right). \tag{3}$$

Перепишем (2) в виде интерполяционного многочлена Ньютона на сетке $\tau_0, \tau_1, \ldots, \tau_k$ ($\tau_0 = 0$):

$$\mathbf{f} = \mathbf{f}_0 + \boldsymbol{\alpha}_1\tau + \boldsymbol{\alpha}_2\tau(\tau - \tau_1) + \boldsymbol{\alpha}_3\tau(\tau - \tau_1)(\tau - \tau_2) + \ldots + \boldsymbol{\alpha}_k\tau(\tau - \tau_1)\ldots(\tau - \tau_{k-1}). \tag{4}$$

Из соотношений

$$\begin{aligned}
&\mathbf{f}_1 = \mathbf{f}_0 + \boldsymbol{\alpha}_1\tau_1,\\
&\mathbf{f}_2 = \mathbf{f}_0 + \boldsymbol{\alpha}_1\tau_2 + \boldsymbol{\alpha}_2\tau_2(\tau_2 - \tau_1),\\
&\mathbf{f}_3 = \mathbf{f}_0 + \boldsymbol{\alpha}_1\tau_3 + \boldsymbol{\alpha}_2\tau_3(\tau_3 - \tau_1) + \boldsymbol{\alpha}_3\tau_3(\tau_3 - \tau_1)(\tau_3 - \tau_2),\\
&\ldots
\end{aligned} \tag{5}$$

получаем конечные разности $\boldsymbol{\alpha}$

$$\begin{aligned}
&\boldsymbol{\alpha}_1 = (\mathbf{f}_1 - \mathbf{f}_0)/\tau_1,\\
&\boldsymbol{\alpha}_2 = ((\mathbf{f}_2 - \mathbf{f}_0)/\tau_2 - \boldsymbol{\alpha}_1)/(\tau_2 - \tau_1),\\
&\boldsymbol{\alpha}_3 = (((\mathbf{f}_3 - \mathbf{f}_0)/\tau_3 - \boldsymbol{\alpha}_1)/(\tau_3 - \tau_1) - \boldsymbol{\alpha}_2)/(\tau_3 - \tau_2),\\
&\ldots
\end{aligned} \tag{6}$$

В свою очередь, сравнивая (2) и (4), будем иметь

$$\begin{aligned}
&\mathbf{A}_1 = \boldsymbol{\alpha}_1 + (-\tau_1)\boldsymbol{\alpha}_2 + (\tau_1\tau_2)\boldsymbol{\alpha}_3 + \ldots + (-1)^{k-1}(\tau_1\ldots\tau_{k-1})\boldsymbol{\alpha}_k,\\
&\mathbf{A}_2 = \boldsymbol{\alpha}_2 + (-\tau_1 - \tau_2)\boldsymbol{\alpha}_3 + \ldots,\\
&\ldots\\
&\mathbf{A}_k = \boldsymbol{\alpha}_k,
\end{aligned}$$

или

$$\begin{aligned}
&\mathbf{A}_1 = c_{11}\boldsymbol{\alpha}_1 + c_{21}\boldsymbol{\alpha}_2 + \ldots + c_{k1}\boldsymbol{\alpha}_k,\\
&\mathbf{A}_2 = c_{22}\boldsymbol{\alpha}_2 + \ldots + c_{k2}\boldsymbol{\alpha}_k,\\
&\ldots\\
&\mathbf{A}_k = c_{kk}\boldsymbol{\alpha}_k.
\end{aligned} \tag{7}$$

Тогда обратный переход от $\mathbf{A}$ к $\boldsymbol{\alpha}$ можно представить как

$$\begin{aligned}
\boldsymbol{\alpha}_1 &= d_{11}\mathbf{A}_1 + d_{21}\mathbf{A}_2 + \ldots + d_{k1}\mathbf{A}_k,\\
\boldsymbol{\alpha}_2 &= d_{22}\mathbf{A}_2 + \ldots + d_{k2}\mathbf{A}_k,\\
&\ldots\\
\boldsymbol{\alpha}_k &= d_{kk}\mathbf{A}_k.
\end{aligned} \tag{8}$$

Коэффициенты c_{ij} и d_{ij} являются числами Стирлинга и вычисляются по формулам

$$\begin{aligned}
&c_{ii} = d_{ii} = 1, \quad c_{i0} = d_{i0} = 0 \quad (i > 0),\\
&c_{ij} = c_{i-1,j-1} - \tau_{i-1}c_{i-1,j}, \quad d_{ij} = d_{i-1,j-1} - \tau_j d_{i-1,j} \quad (i > j > 0).
\end{aligned} \tag{9}$$

В соответствии с (9) каждый коэффициент c_{ij} представляет собой сумму всевозможных произведений $i-j$ величин $\tau_1, \ldots, \tau_{i-1}$ со знаком $(-1)^{i-j}$, например,

$$c_{62} = \tau_1\tau_2\tau_3\tau_4 + \tau_1\tau_2\tau_3\tau_5 + \tau_1\tau_2\tau_4\tau_5 + \tau_1\tau_3\tau_4\tau_5 + \tau_2\tau_3\tau_4\tau_5.$$

Отсюда согласно теореме Виета значения $\tau_1, \ldots, \tau_{i-1}$ будут являться корнями полинома вида

$$P_{i-1} = c_{i1} + c_{i2}\tau + c_{i3}\tau^2 + \ldots + c_{ii}\tau^{i-1}. \tag{10}$$

2. Интегрирование на шаге

Величины $\boldsymbol{\alpha}$ определяются по $\mathbf{f}$, которые, в свою очередь, вычисляются по решениям $\mathbf{x}$. Согласно (3) эти решения будем задавать как

$$\begin{aligned}
\mathbf{x}_1 &= \mathbf{x}_0 + h\left(\mathbf{f}_0\tau_1 + \frac{1}{2}\mathbf{A}_1\tau_1^2 + \frac{1}{3}\mathbf{A}_2\tau_1^3 + \ldots + \frac{1}{k+1}\mathbf{A}_k\tau_1^{k+1}\right),\\
&\ldots\\
\mathbf{x}_k &= \mathbf{x}_0 + h\left(\mathbf{f}_0\tau_k + \frac{1}{2}\mathbf{A}_1\tau_k^2 + \frac{1}{3}\mathbf{A}_2\tau_k^3 + \ldots + \frac{1}{k+1}\mathbf{A}_k\tau_k^{k+1}\right).
\end{aligned} \tag{11}$$

Формулы (11) представляют собой неявные уравнения относительно $\mathbf{x}$, поэтому решаются итерационным способом.

Для получения начального приближения $\bar{\boldsymbol{\alpha}}$ на следующем шаге $\bar{h}$ используется информация о коэффициентах $\mathbf{A}$ на текущем шаге h. Безразмерная независимая переменная следующего шага будет $\bar{\tau} = (t - t_h)/\bar{h}$, где $t_h = t_0 + h$. Отсюда

$$\tau = r\bar{\tau} + 1, \tag{12}$$

где $r = \bar{h}/h$. Согласно (2)

$$\mathbf{f}_0 + \mathbf{A}_1\tau + \mathbf{A}_2\tau^2 + \mathbf{A}_3\tau^3 + \ldots + \mathbf{A}_k\tau^k = \bar{\mathbf{f}}_0 + \bar{\mathbf{A}}_1\bar{\tau} + \bar{\mathbf{A}}_2\bar{\tau}^2 + \bar{\mathbf{A}}_3\bar{\tau}^3 + \ldots + \bar{\mathbf{A}}_k\bar{\tau}^k. \tag{13}$$

Подставляя (12) в (13) и приравнивая коэффициенты при одинаковых степенях $\bar{\tau}$, получаем

$$\begin{aligned}
\bar{\mathbf{f}}_0 &= e_{00}\mathbf{f}_0 + e_{10}\mathbf{A}_1 + e_{20}\mathbf{A}_2 + e_{30}\mathbf{A}_3 + \ldots + e_{k0}\mathbf{A}_k,\\
\bar{\mathbf{A}}_1 &= r(e_{11}\mathbf{A}_1 + e_{21}\mathbf{A}_2 + e_{31}\mathbf{A}_3 + \ldots + e_{k1}\mathbf{A}_k),\\
\bar{\mathbf{A}}_2 &= r^2(e_{22}\mathbf{A}_2 + e_{32}\mathbf{A}_3 + \ldots + e_{k2}\mathbf{A}_k),\\
&\ldots\\
\bar{\mathbf{A}}_k &= r^k e_{kk}\mathbf{A}_k,
\end{aligned} \tag{14}$$

где e_{ij} — числа арифметического треугольника, вычисляемые по рекуррентным формулам

$$e_{ii} = e_{i0} = 1, \quad e_{ij} = e_{i-1,j-1} + e_{i-1,j} \quad (i > j > 0). \tag{15}$$

Далее оценка $\bar{\mathbf{A}}$ для $\bar{h}$ уточняется путем внесения поправки $\Delta\mathbf{A}$, получаемой как разность между значениями $\mathbf{A}$ после итераций и оценкой $\bar{\mathbf{A}}$ на текущем шаге h. Наконец, пользуясь соотношениями (8), будем иметь начальное приближение $\bar{\boldsymbol{\alpha}}$.

Каждая итерация выполняется следующим образом. Сначала определяется решение $\mathbf{x}_1$, из которого по первой формуле (5) улучшается значение $\boldsymbol{\alpha}_1$. Далее определяется $\mathbf{x}_2$, по которому улучшается $\boldsymbol{\alpha}_2$, и т. д. до $\mathbf{x}_k$. Как правило, для получения достаточно хороших $\boldsymbol{\alpha}$ необходимы всего две итерации, очень редко — три.

Как только величины $\boldsymbol{\alpha}$ получены, решение на шаге h ($\tau = 1$) будет

$$\mathbf{x}_h = \mathbf{x}_0 + h\left(\mathbf{f}_0 + \frac{1}{2}\mathbf{A}_1 + \frac{1}{3}\mathbf{A}_2 + \ldots + \frac{1}{k+1}\mathbf{A}_k\right). \tag{16}$$

В начале интегрирования, на первом шаге, в качестве $\bar{\boldsymbol{\alpha}}$ выбирают нулевые значения и запускается вышеописанный итерационный процесс. Если начальный шаг достаточно большой, чтобы обеспечить заданную локальную точность, то его следует уменьшить. При оптимально выбранном шаге высокая точность $\boldsymbol{\alpha}$ достигается уже на четвертой итерации.

3. Формулы интегратора как одно из представлений неявного метода Рунге—Кутты

Пользуясь (6), (7) и (11), нетрудно показать, что решение (16) представимо в виде

$$\mathbf{x}_h = \mathbf{x}_0 + h\sum_{i=0}^{k} b_i\mathbf{k}_i, \quad \text{где} \quad \mathbf{k}_i = \mathbf{f}(t_0 + h\tau_i, \mathbf{x}_0 + h\sum_{j=0}^{k} a_{ij}\mathbf{k}_j) \quad (i = 0, \ldots, k), \tag{17}$$

а коэффициенты a_{ij} и b_i — постоянные, зависящие только от τ_i. Таким образом, интегратор Гаусса—Эверхарта фактически основан на видоизмененных формулах неявного метода Рунге—Кутты, причем эти формулы являются коллокационными [6]. Действительно, в качестве коллокационного полинома в интеграторе выступает приближенное представление решения (3), тогда как условия коллокации формально задаются соотношениями (5).

Произвольный выбор узлов τ_i дает метод порядка $k+1$. Однако, используя свойства неявных методов Рунге—Кутты, Эверхарт выбрал разбиение Гаусса—Радо, что позволило повысить точность метода до порядка $2k+1$.

4. Повышение порядка интегратора

Значения узлов $\tau_1, \ldots, \tau_k$ — свободные параметры и их можно выбирать как угодно, лишь бы они не равнялись между собой. В общем случае интегратор будет иметь порядок $k+1$, однако существуют такие узловые значения, которые позволяют значительно повысить точность решения — до порядка $2k$ либо $2k+1$.

Приближенное решение порядка $2k+1$ на сетке $\tau_1, \ldots, \tau_{2k}$ будет

$$\mathbf{x}_h = \mathbf{x}_0 + h\left(\mathbf{f}_0 + \frac{1}{2}\mathbf{A}'_1 + \frac{1}{3}\mathbf{A}'_2 + \ldots + \frac{1}{2k+1}\mathbf{A}'_{2k}\right). \tag{18}$$

Определим $\tau_1, \ldots, \tau_k$ так, чтобы решение (16) имело порядок $2k+1$. Это возможно в случае, если разность решений (16) и (18) будет равна нулю:

$$\Delta\mathbf{x}_h = h\left(\frac{\mathbf{A}'_1 - \mathbf{A}_1}{2} + \frac{\mathbf{A}'_2 - \mathbf{A}_2}{3} + \ldots + \frac{\mathbf{A}'_k - \mathbf{A}_k}{k+1} + \frac{\mathbf{A}'_{k+1}}{k+2} + \ldots + \frac{\mathbf{A}'_{2k}}{2k+1}\right) = \mathbf{0}. \tag{19}$$

Согласно (6) коэффициенты $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ в (16) и (18) совпадают. Тогда

$$\begin{aligned}
&\mathbf{A}'_1 - \mathbf{A}_1 = c_{k+1,1}\boldsymbol{\alpha}_{k+1} + \ldots + c_{2k,1}\boldsymbol{\alpha}_{2k},\\
&\mathbf{A}'_2 - \mathbf{A}_2 = c_{k+1,2}\boldsymbol{\alpha}_{k+1} + \ldots + c_{2k,2}\boldsymbol{\alpha}_{2k},\\
&\ldots\\
&\mathbf{A}'_k - \mathbf{A}_k = c_{k+1,k}\boldsymbol{\alpha}_{k+1} + \ldots + c_{2k,k}\boldsymbol{\alpha}_{2k},\\
&\mathbf{A}'_{k+1} = c_{k+1,k+1}\boldsymbol{\alpha}_{k+1} + \ldots + c_{2k,k+1}\boldsymbol{\alpha}_{2k},\\
&\mathbf{A}'_{k+2} = c_{k+2,k+2}\boldsymbol{\alpha}_{k+2} + \ldots + c_{2k,k+2}\boldsymbol{\alpha}_{2k},\\
&\ldots\\
&\mathbf{A}'_{2k} = c_{2k,2k}\boldsymbol{\alpha}_{2k}.
\end{aligned} \tag{20}$$

Используя рекуррентные соотношения (9), выразим в (20) коэффициенты $c_{k+i,j}$ через $c_{k+1,j}$:

$$\begin{aligned}
&c_{k+2,1} = -\tau_{k+1}c_{k+1,1}, \quad c_{k+2,j} = c_{k+1,j-1} - \tau_{k+1}c_{k+1,j} \quad (j>1),\\
&c_{k+3,1} = \tau_{k+2}\tau_{k+1}c_{k+1,1}, \; c_{k+3,j} = c_{k+1,j-2} + (-\tau_{k+1} - \tau_{k+2})c_{k+1,j-1} + \tau_{k+2}\tau_{k+1}c_{k+1,j} \quad (j>1),\\
&\ldots
\end{aligned} \tag{21}$$

Подставляя (21) в (20), получим

$$\begin{aligned}
&\mathbf{A}'_1 - \mathbf{A}_1 = \mathbf{B}_1 c_{k+1,1},\\
&\mathbf{A}'_2 - \mathbf{A}_2 = \mathbf{B}_1 c_{k+1,2} + \mathbf{B}_2 c_{k+1,1},\\
&\ldots\\
&\mathbf{A}'_k - \mathbf{A}_k = \mathbf{B}_1 c_{k+1,k} + \mathbf{B}_2 c_{k+1,k-1} + \ldots + \mathbf{B}_k c_{k+1,1},\\
&\mathbf{A}'_{k+1} = \mathbf{B}_1 c_{k+1,k+1} + \mathbf{B}_2 c_{k+1,k} + \ldots + \mathbf{B}_k c_{k+1,2},\\
&\mathbf{A}'_{k+2} = \mathbf{B}_2 c_{k+1,k+1} + \ldots + \mathbf{B}_k c_{k+1,3},\\
&\ldots\\
&\mathbf{A}'_{2k} = \mathbf{B}_k c_{k+1,k+1},
\end{aligned} \tag{22}$$

где $\mathbf{B}$ имеют форму (7):

$$\begin{aligned}
&\mathbf{B}_1 = \boldsymbol{\alpha}_{k+1} + (-\tau_{k+1})\boldsymbol{\alpha}_{k+2} + (\tau_{k+1}\tau_{k+2})\boldsymbol{\alpha}_{k+3} + \ldots + (-1)^{k-1}(\tau_{k+1}\ldots\tau_{2k-1})\boldsymbol{\alpha}_{2k},\\
&\mathbf{B}_2 = \boldsymbol{\alpha}_{k+2} + (-\tau_{k+1} - \tau_{k+2})\boldsymbol{\alpha}_{k+3} + \ldots,\\
&\ldots\\
&\mathbf{B}_k = \boldsymbol{\alpha}_{2k}.
\end{aligned}$$

Подставляя далее (22) в (19) и уравнивая коэффициенты при одинаковых величинах $\mathbf{B}$, получим следующие уравнения для $c_{k+1,j}$:

$$\begin{aligned}
&\frac{c_{k+1,1}}{2} + \frac{c_{k+1,2}}{3} + \ldots + \frac{c_{k+1,k}}{k+1} + \frac{1}{k+2} = 0,\\
&\ldots\\
&\frac{c_{k+1,1}}{k+1} + \frac{c_{k+1,2}}{k+2} + \ldots + \frac{c_{k+1,k}}{2k} + \frac{1}{2k+1} = 0.
\end{aligned} \tag{23}$$

Решения (23) однозначно определяют значения $\tau_1, \ldots, \tau_k$, которые вычисляются из уравнения

$$c_{k+1,1} + c_{k+1,2}\tau + c_{k+1,3}\tau^2 + \ldots + c_{k+1,k}\tau^{k-1} + \tau^k = 0. \tag{24}$$

Узловые значения τ_i, определяемые из (23) и (24), задают (левое) разбиение Гаусса—Радо. Их также можно получить из уравнения

$$(\tau^{k+1}(\tau-1)^k)_\tau^{(k)} = 0. \tag{25}$$

В (25) левый корень равен нулю ($\tau_0 = 0$). Если потребуем, чтобы правый корень был равен единице ($\tau_k = 1$), то для повышения порядка численного решения получим разбиение Гаусса—Лобатто, узловые значения которого также удовлетворяют уравнению

$$(\tau^k(\tau-1)^k)_\tau^{(k-1)} = 0. \tag{26}$$

Разбиение Гаусса—Лобатто симметрично и дает численное решение порядка $2k$.

5. Выбор шага

В интеграторе Гаусса—Эверхарта контроль шага интегрирования осуществляется по величине последнего члена в (16).

Пусть $\|\mathbf{e}_{tol}\|$ — заданная точность. Потребуем, чтобы на следующем шаге выполнялось равенство

$$\frac{\bar{h}}{k+1}\|\bar{\mathbf{A}}_k\| = \|\mathbf{e}_{tol}\|.$$

Отсюда, используя последнее соотношение в (14), получим оценку

$$\bar{h} = hr = h\left(\frac{k+1}{h}\frac{\|\mathbf{e}_{tol}\|}{\|\mathbf{A}_k\|}\right)^{\frac{1}{k+1}}. \tag{27}$$

Очевидно, при разбиениях Гаусса—Радо и Гаусса—Лобатто недостаток такой оценки состоит в том, что шаг по ней выбирается как для решения порядка k, поэтому, вообще говоря, она не обеспечивает сохранение требуемой локальной точности для решения более высокого порядка.

Во избежание слишком больших (и малых) локальных ошибок на r следует наложить ограничение:

$$\frac{1}{\sigma} < r^{k+1} < \sigma. \tag{28}$$

Чтобы величина последнего члена в (16) была ограничена в пределах одного порядка, значение σ должно быть равно $\sqrt{10}$. Это следует из $\bar{h}\|\bar{\mathbf{A}}_k\| \sim r^{k+1}$.

Выполнение обоих неравенств проверяется лишь в начале интегрирования при выборе стартового шага: если (28) не выполняется, то интегрирование повторяется с новым шагом $\bar{h} = hr$ и т. д., пока не выполнится условие (28). Обычно для получения стартового шага требуется не более четырех итераций. В дальнейшем для ограничения r проверяется только правое неравенство: если оно не выполняется, то r принимает значение правого предела.

Начальное приближение стартового шага получается из оценки (27) для $k = 1$:

$$\bar{h} = \sqrt{\frac{2h\|\mathbf{e}_{tol}\|}{\|\mathbf{f}_1 - \mathbf{f}_0\|}}, \quad \mathbf{f}_1 = \mathbf{f}(t_0 + h, \mathbf{x}_0 + h\mathbf{f}_0), \tag{29}$$

где h — малая величина. Если h настолько мала, что в компьютерной арифметике $\mathbf{f}_1 = \mathbf{f}_0$, то она увеличивается в 10 раз и оценка повторяется снова.

6. Порядок и шаг интегрирования при компьютерной реализации метода

Теоретически совместное повышение порядка и уменьшение шага метода неограниченно повышает методическую точность численных результатов интегрирования. Однако при компьютерной реализации в арифметике с определенной точностью вследствие ошибок округления существуют такие значения параметров интегрирования, которые дают предельно высокую точность ввиду того, что методические ошибки становятся соизмеримыми с ошибками округления, и в этом случае не имеет смысла предпринимать какие-либо дальнейшие попытки получить более высокую точность численного интегрирования путем повышения порядка и уменьшения шага метода.

В общем случае получить оптимальные параметры интегрирования невозможно, поскольку они непосредственно зависят от специфики решаемой задачи. Впрочем, принимая во внимание, что интегратор Гаусса—Эверхарта используется, как правило, для численного моделирования орбит, найдем оценки оптимальной пары порядок—шаг интегрирования применительно к решению какой-нибудь простой задачи небесной механики, например, круговой двумерной задачи двух тел с уравнениями

$$\mathbf{r}' = \mathbf{v}, \quad \mathbf{v}' = -\frac{\mu}{|\mathbf{r}|^3}\mathbf{r}. \tag{30}$$

Здесь $\mathbf{r} = (r_1, r_2)$, $\mathbf{v} = (v_1, v_2)$ — векторы положения и скорости соответственно, μ — гравитационный параметр. Поскольку в круговом случае $|\mathbf{r}| = a = \text{const}$, будем полагать, что величина $\mu/|\mathbf{r}|^3 = n^2 = \text{const}$, т. е. решение $\mathbf{x} = (\mathbf{r}, \mathbf{v})$ описывается уравнениями гармонического осциллятора с частотой n и может быть представлено в виде

$$x_1 = r_1 = a\cos nt, \quad x_2 = r_2 = a\sin nt,$$

$$x_3 = v_1 = -an\sin nt, \quad x_4 = v_2 = an\cos nt. \tag{31}$$

Оценим методическую ошибку $|\mathbf{e}|_M$ по главному члену погрешности:

$$|\mathbf{e}|_M = \frac{a\sqrt{1+n^2}(nh)^{p+1}}{(p+1)!}, \tag{32}$$

где использована формула $|\mathbf{x}^{(p+1)}| = an^{p+1}\sqrt{1+n^2}$. Здесь p — порядок метода. Ошибку округления $|\mathbf{e}|_R$ можно оценить как

$$|\mathbf{e}|_R = \varepsilon|\mathbf{x}| = \varepsilon a\sqrt{1+n^2}, \tag{33}$$

где ε — машинный эпсилон.

Очевидно, что не имеет смысла выбирать такие шаг и порядок интегрирования, при которых методическая ошибка будет меньше ошибки округления. Из условия $|\mathbf{e}|_M = |\mathbf{e}|_R$ получим отношение между оптимальными параметрами интегрирования h и p:

$$h_\varphi^{p+1} = \varepsilon(p+1)!, \tag{34}$$

где $h_\varphi = nh$ — шаг по долготе $\varphi = nt$, соответствующий шагу h. Уравнение (34) дает нижнюю границу для шага h, в то время как для неявных методов (17) имеет место верхняя граница, задаваемая условием сходимости итерационного процесса для решения нелинейных уравнений в неявных методах [6]:

$$h < \frac{1}{L\max_i \sum_j |a_{ij}|},$$

где L — постоянная Липшица, которая задает неравенство

$$|\mathbf{f}(t,\mathbf{x}) - \mathbf{f}(t,\mathbf{y})| \le L|\mathbf{x}-\mathbf{y}| \tag{35}$$

для любых $\mathbf{x}$ и $\mathbf{y}$. Если положить $\max_i \sum_j |a_{ij}| = 1$ (в действительности максимум близок к единице), то получим следующее ограничение на шаг интегрирования: $h < 1/L$ или $h_\varphi = n/L$. Оценим постоянную Липшица L для исследуемой задачи. Рассмотрим отношение

$$\frac{|\delta\mathbf{f}|^2}{|\delta\mathbf{x}|^2} = \frac{n^4|\delta\mathbf{r}|^2 + |\delta\mathbf{v}|^2}{|\delta\mathbf{r}|^2 + |\delta\mathbf{v}|^2},$$

где $\delta\mathbf{x}$, $\delta\mathbf{r}$ и $\delta\mathbf{v}$ — всевозможные разности векторов в соответствующих переменных. Принимая $|\delta\mathbf{r}| = \xi\cos\psi$, $|\delta\mathbf{v}| = \xi\sin\psi$, где $\xi \ge 0$ и $0 \ge \psi \ge \pi/2$, будем иметь

$$\frac{|\delta\mathbf{f}|^2}{|\delta\mathbf{x}|^2} = (n^4-1)\cos^2\psi + 1.$$

Отсюда следует, что все значения отношения лежат между 1 и n^4. Следовательно, согласно (35) в качестве постоянной Липшица можно выбрать $L = \max(1, n^2)$. Тогда получим верхнюю границу шага

$$h_\varphi < \frac{n}{\max(1,n^2)} \le 1.$$

Таким образом, в лучшем случае, а именно при $n = 1$, когда верхняя граница максимальна, шаг интегрирования h_φ должен удовлетворять неравенствам

$$\varepsilon(p+1)! < h_\varphi^{p+1} < 1.$$

Очевидно, условие

$$\varepsilon(p+1)! > 1 \tag{36}$$

означает, что порядок метода завышен и использование такого метода при вычислениях в арифметике с точностью ε не логично в том смысле, что ту же точность результатов интегрирования можно получить с использованием методов более низких порядков. Оптимальные порядки p неявных методов Рунге—Кутты для различных ε, соответствующих одинарной, двойной, расширенной и четверной точности, представлены в табличном виде (следует иметь в виду, что эти порядки получены для задачи с $n = 1$, в ином случае они могут быть меньше):

ε	$1.1 \cdot 10^{-7}$	$2.2 \cdot 10^{-16}$	$1.1 \cdot 10^{-19}$	$1.9 \cdot 10^{-34}$
p	9	16	19	29

7. Фортран-код интегратора Гаусса—Эверхарта

Представленный выше фортран-код интегратора Гаусса—Эверхарта GAUSS_15 (до 15 порядка), основанного на гауссовых разбиениях Лобатто и Радо, а также Лежандра, доступен в интернете по адресу http://astro.tsu.ru/sch/Gauss_15.txt.

`GAUSS_15(X,TS,TF,STEP,ERR,N,NOR,NI,NS,NBS,NF,FUN)` — заголовок процедуры интегрирования. Опишем входные и выходные параметры интегратора, помечая их соответственно индексами I и O. Итак, $\mathtt{X}_O^I$ — интегрируемые переменные (до выполнения процедуры — начальные для значения `TS`, после — конечные для `TF`); $\mathtt{TS}^I$ и $\mathtt{TF}^I$ — начальное и конечное значения независимой переменной соответственно; $\mathtt{STEP}_O^I$ — величина стартового шага интегрирования (если $\mathtt{STEP} = 0$, то стартовый шаг выбирается по формуле (29)), после выполнения процедуры `STEP` — величина предпоследнего шага; $\mathtt{ERR}^I$ — значение $\|\mathbf{e}_{tol}\|$ для выбора переменного шага (27) (условие $\mathtt{ERR} \neq 0$ задает режим переменного, $\mathtt{ERR} = 0$ — постоянного шага); $\mathtt{N}^I$ — число уравнений; $\mathtt{NOR}^I$ — порядок интегратора p; $\mathtt{NI}^I$ — максимальное число итераций на шаге (при $\mathtt{NI} \leq 0$ итерационный процесс выполняется до сходимости); $\mathtt{NS}_O$ — число шагов интегрирования за выполнение процедуры; $\mathtt{NBS}_O$ — число шагов, на которых итерационный процесс не сходится (несходимость итерационного процесса регистрируется, когда число итераций достигает 100); $\mathtt{NF}_O$ — число обращений к процедуре правых частей; $\mathtt{FUN}^I$ — название процедуры правых частей $\mathbf{f}$.

8. Тестирование интегратора Гаусса—Эверхарта в задаче двух тел

Интегратор тестировался на дифференциальных уравнениях задачи двух тел (30). Начальные условия задачи

$$r_1 = 1 - e, \quad r_2 = 0, \quad v_1 = 0, \quad v_2 = \sqrt{(1+e)/(1-e)}$$

соответствуют однопараметрическому семейству орбит с эксцентриситетом e в качестве параметра и с единичной большой полуосью a. Интегрирование выполнялось на интервале 1000 оборотов для различных эксцентриситетов. При этом исследовались те или иные характеристики интегратора.

8.1. Круговой случай $e = 0$

Как уже было замечено, шаг в интеграторе выбирается таким образом, чтобы сохранялась величина члена порядка $k+1$. Если бы интегратор имел порядок k, то таким способом можно было бы обеспечить сохранение локальной точности на всем интервале интегрирования. Однако порядок интегратора Гаусса—Эверхарта существенно выше k. На примере круговой задачи ($e = 0$) исследовалась зависимость реальной локальной точности от задаваемой для выбора переменного шага.

В (16) величину $k+1$ члена $\|\mathbf{e}_k\|$ можно оценить как

$$\|\mathbf{e}_k\| \equiv \frac{h}{k+1}\|\mathbf{A}_k\| \approx \frac{h^{k+1}}{(k+1)!}\|\mathbf{x}^{(k+1)}\|, \tag{37}$$

тогда как ошибка численного решения порядка p будет

$$\|\mathbf{e}_p\| \approx \frac{h^{p+1}}{(p+1)!}\|\mathbf{x}^{(p+1)}\|. \tag{38}$$

Поскольку решение $\mathbf{x}$ имеет вид (31), то

$$\|\mathbf{x}^{(i)}\| = a\sqrt{1+n^2}n^i. \tag{39}$$

Подставляя (39) в (37) и (38), получим

$$\|\mathbf{e}_p\| \approx \frac{(a\sqrt{1+n^2})^{-\frac{p-k}{k+1}}}{(p+1)!}\left((k+1)!\|\mathbf{e}_k\|\right)^{\frac{p+1}{k+1}}.$$

В частности, для разбиения Гаусса—Радо ($p = 2k+1$)

$$\|\mathbf{e}_p\| \approx \frac{1}{a\sqrt{1+n^2}}\prod_{i=1}^{k+1}\frac{i}{k+1+i}\|\mathbf{e}_k\|^2. \tag{40}$$

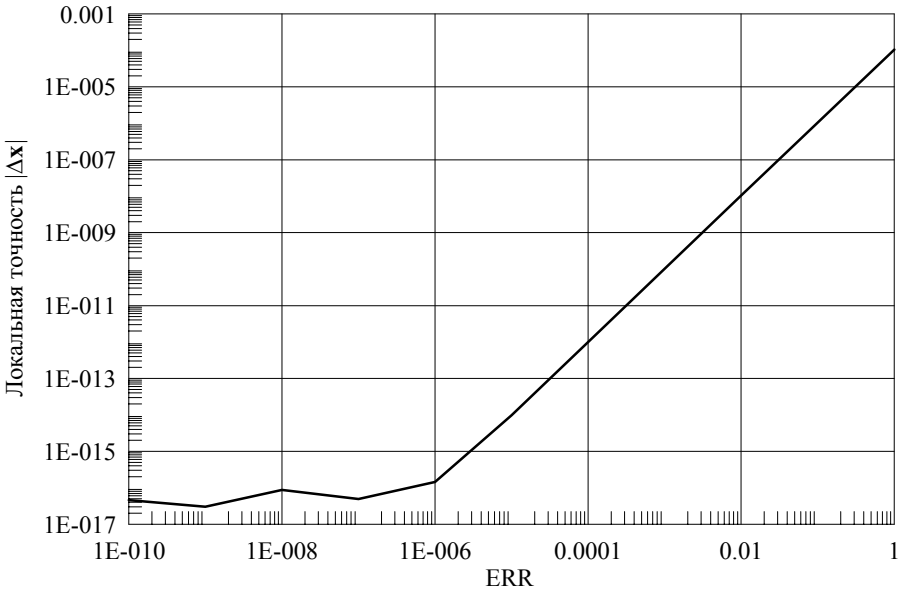

Рис. 1. Зависимость локальной точности от параметра ERR

На рис. 1 (см. с. 41) представлены оценки реальной локальной ошибки

$$\|\mathbf{e}_p\| = \|\Delta\mathbf{x}\| = \sqrt{\Delta r_1^2 + \Delta r_2^2 + \Delta v_1^2 + \Delta v_2^2}$$

в зависимости от `ERR` ($\approx \|\mathbf{e}_k\|$) для $p = 11$. Оценивание выполнялось в режиме `STEP` $= 0$ на первом шаге, подбираемом в соответствии с параметром `ERR`. Видно, что полученные результаты хорошо согласуются с (40). В частности, экспериментальная оценка при `ERR` $> 10^{-6}$ (`ERR` $< 10^{-6}$ — область доминирования вычислительных ошибок) подтверждает квадратичную зависимость $\|\mathbf{e}_p\| \propto \|\mathbf{e}_k\|^2$, отчего главным образом реальная точность численного решения существенно выше задаваемого значения параметра `ERR`.

8.2. Слабоэксцентричный случай $e = 0.1$

Для оценки оптимизации составленного фортран-кода сравнили быстродействия (в затратах процессорного времени) интегратора GAUSS_15 и широко используемого в небесной механике интегратора RADAU_27. При этом в обоих случаях выполнялись все операции, определяемые основными формулами. В то же время расхождение двух численных решений, полученных интеграторами, оказалось меньше методической ошибки на несколько порядков.

На рис. 2 показаны отношения временных затрат, которые потребовались интеграторам для получения численных решений 7, 11 и 15 порядков. Как видно, интегратор GAUSS_15 работает быстрее, однако этот выигрыш с увеличением порядка уменьшается. Впрочем, следует заметить, что в задачах со сложной функцией $\mathbf{f}$ оперативность интеграторов должна быть одинаковой, поскольку большая часть времени будет затрачиваться на процедуру ее вычисления `FUN` и тогда быстродействие главным образом определяется числом обращения к этой процедуре.

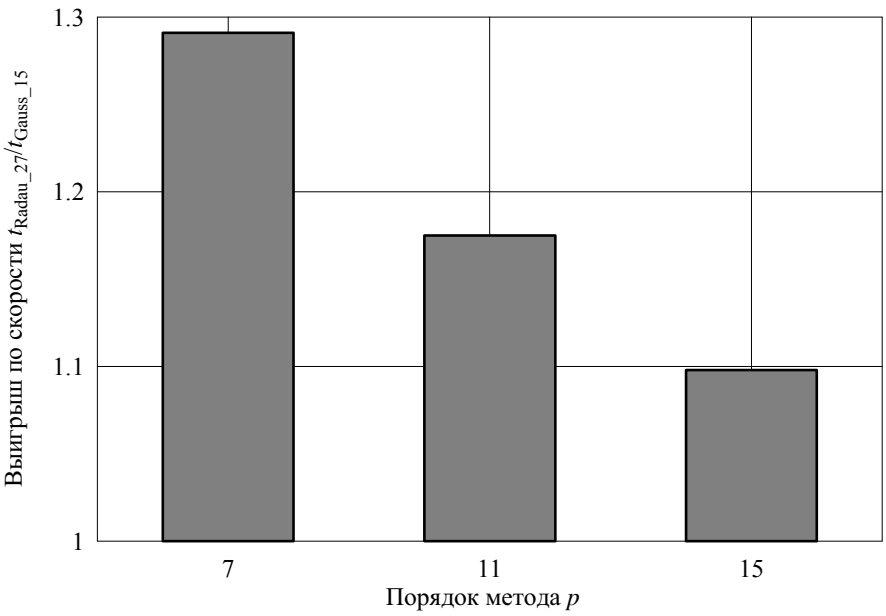

Рис. 2. Оптимальность интегратора GAUSS_15 относительно RADAU_27

Далее была оценена ошибка интегрирования $|\Delta\mathbf{r}| = \sqrt{\Delta r_1^2 + \Delta r_2^2}$ в зависимости от величины постоянного шага h для порядков 2–11. Интегрирование выполнялось с шестью итерациями (на шаге) на интервале 1000 оборотов.

В случае $k = 1$–3 представленные на рис. 3 характеристики показывают, что при достаточно больших h (глобальная) ошибка решения p-порядка, как и ожидалось, с уменьшением шага ведет себя как $|\Delta\mathbf{r}| \sim h^p$, что указывает на соответствие ошибки методу порядка p. Для $k = 4$ и 5 ситуация несколько иная. Это объясняется тем, что шести итераций недостаточно для сходимости итерационного процесса на шаге и получаемое решение не соответствует порядку интегрирующих формул. Случайное поведение характеристик ниже $|\Delta\mathbf{r}| = 10^{-6}$ обусловлено влиянием ошибок округления.

Следует заметить, что несмотря на крутой скат характеристики нечетного порядка $2k + 1$ в сравнении с характеристикой четного порядка $2k$ (для $k = 1$–3) определенная точность для четного порядка достигается при большем шаге, т. е. быстрее. Это связано с тем, что при симметричном разбиении Гаусса—Лобатто интегратор Гаусса—Эверхарта обладает геометрическими свойствами и методическая ошибка вдоль независимой переменной t эволюционирует медленнее, чем при разбиении Гаусса—Радо.

На рис. 4 показано поведение ошибки с изменением t для четных и нечетных порядков ($k = 3$–5). Приведенные результаты были получены в режиме $\texttt{NI} \leq 0$, т. е. до полной сходимости итерационного процесса. При этом на шаге требовалось от 11 до 16 итераций. Как следует из рисунка, при постоянном шаге ($h = 2\pi/16$) ошибка интегрирования для четных порядков ведет себя линейно, тогда как для нечетных порядков — квадратически. Даже несмотря на то что интегратор порядка $2k+1$ в начале интегрирования дает решение точнее интегратора порядка $2k$, в конце интегрирования вследствие разного роста ошибок первый уже уступает второму в точности почти на два порядка. Очевидно, с увеличением интервала интегрирования это преимущество для каждого интегратора с симметричным разбиением будет возрастать. К сожалению, такие свойства интеграторов четных порядков имеют место только при использовании постоянного

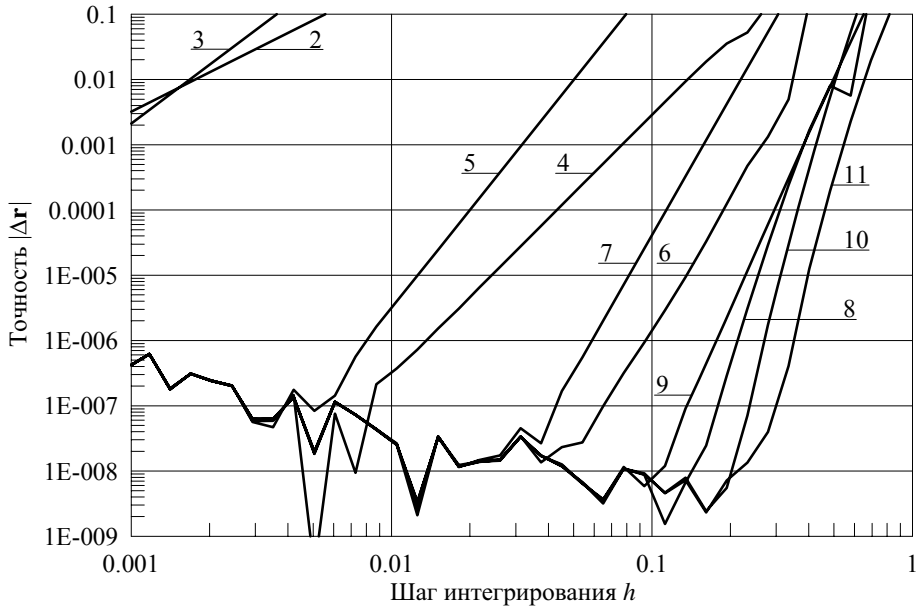

Рис. 3. Зависимость точности от величины шага интегрирования

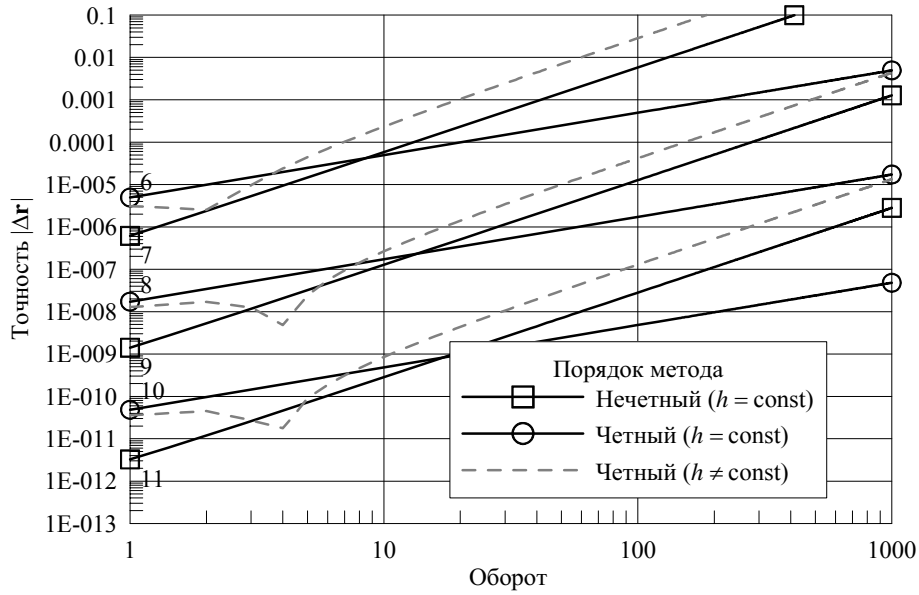

Рис. 4. Поведение ошибки для решений четных и нечетных порядков

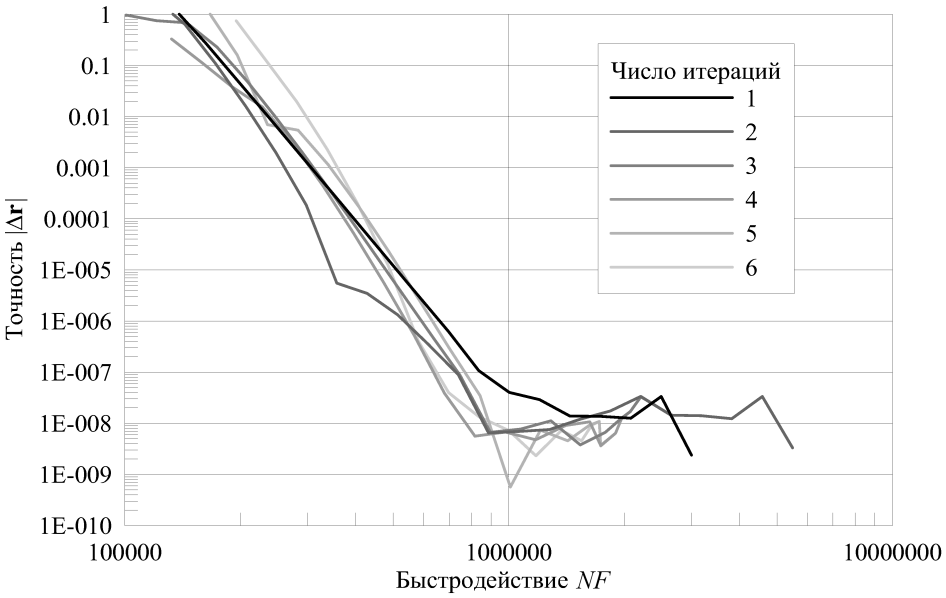

Рис. 5. Характеристики точность—быстродействие в зависимости от числа итераций на шаге

шага [5]. При автоматическом выборе шага ошибка так же, как и для интеграторов нечетных порядков, ведет себя квадратичным образом.

Далее был исследован вопрос о соотношении точности и быстродействия при увеличении числа итераций на шаге. Путем вариации `ERR` оценивались точность и быстродействие интегратора 11 порядка с переменным шагом на интервале 1000 оборотов для `NI` = 1–6. Результаты приведены на рис. 5. Видно, что уже на второй итерации можно получить довольно хорошее решение, хотя соответствующая характеристика заметно отклоняется от степенной зависимости. Поэтому для уверенного результата необходимо использовать, по крайней мере, три итерации на шаге. Следует также заметить, что увеличение числа итераций не существенно понижает эффективность интегрирования.

8.3. Сильноэксцентричный случай $e = 0.9$

Очевидно, сильноэксцентричные орбиты необходимо интегрировать с переменным шагом. Поведение выбираемого шага на одном обороте для $e = 0.9$ представлено на рис. 6, где также приведены функции, пропорциональные $|\mathbf{r}|$, $|\mathbf{r}|^{3/2}$, $|\mathbf{r}|^2$ и $|\mathbf{v}|^{-1}$. Как видно, изменение шага очень близко к поведению $|\mathbf{r}|^{3/2}$. Это означает, что шаг переменной величины по t соответствует почти постоянному шагу по так называемой эллиптической аномалии.

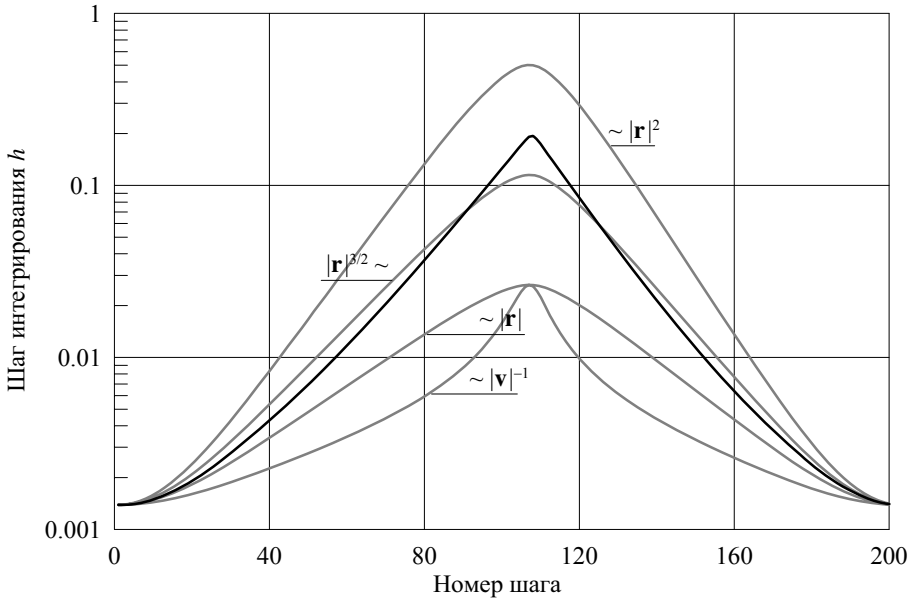

Рис. 6. Выбор шага интегрирования в случае сильновытянутой орбиты

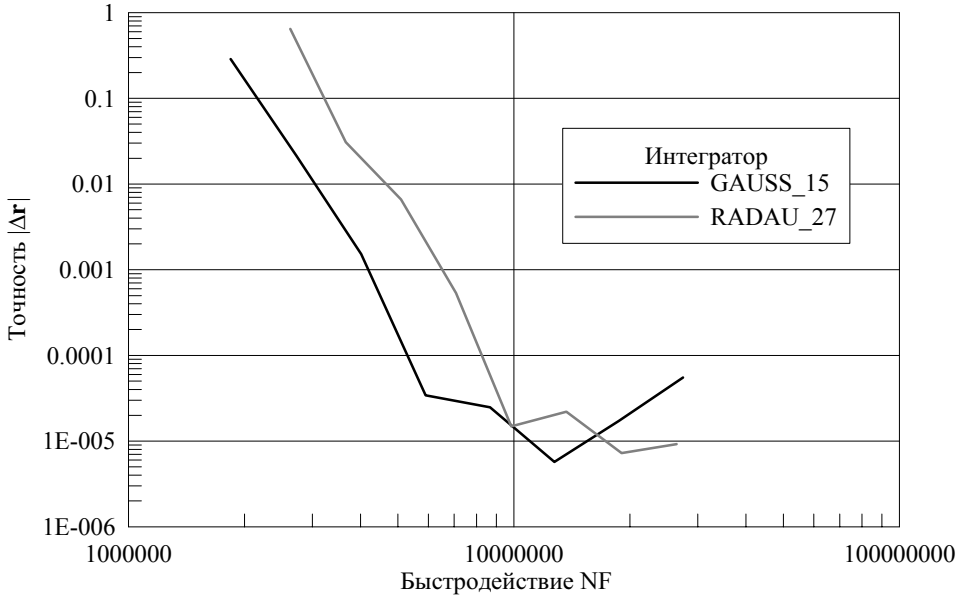

Рис. 7. Характеристики точность—быстродействие для интеграторов GAUSS_15 и RADAU_27 в экстраэксцентричном случае

8.4. Экстраэксцентричный случай $e = 0.999$

Наконец, был исследован алгоритм выбора шага при долгосрочном интегрировании очень вытянутой орбиты с эксцентриситетом $e = 0.999$. Интегрирование выполнялось на интервале 1000 оборотов двумя интеграторами: GAUSS_15 и RADAU_27. Результаты показали (рис. 7), что RADAU_27 заметно менее эффективен, нежели GAUSS_15. Так, при одинаковом быстродействии первый интегратор дает решение с точностью ниже на порядок и более. Это связано с тем, что в RADAU_27 заложен неверный алгоритм выбора шага для интегрирования систем с уравнениями первого порядка, а именно, алгоритм не точно соответствует формуле (27): вместо степени $1/(k+1)$ в RADAU_27 используется $1/(k+2)$. Очевидно, эта ошибка легко устраняется.

Таким образом, в работе представлена новая версия интегратора Гаусса—Эверхарта GAUSS_15. На примере плоской задачи двух тел при различных начальных условиях экспериментально показано, что предложенная программная реализация интегратора значительно эффективнее ранее используемой, а также предоставляет больше возможностей для высокоточного и оперативного численного интегрирования.

Список литературы

[1] Everhart E. A new method for integrating orbits // Bull. Amer. Astronom. Soc. 1973. Vol. 5. P. 389.

[2] Everhart E. Implicit single sequence methods for integrating orbits // Celest. Mech. 1974. Vol. 10. P. 35–55.

[3] Everhart E. An efficient integrator that uses Gauss—Radau spacings // Dynamics of Comets: Their Origin and Evolution. Proc. of IAU Colloq. 83. Italy, 1984 / Eds. A. Carusi and G.B. Valsecchi. Dordrecht: Reidel, Astrophys. and Space Science Library Rame, 1985. Vol. 115. P. 185–202.

[4] Butcher J.C. Implicit Runge—Kutta processes // Math. Comput. 1964. Vol. 18. P. 50–64.

[5] Hairer E., Lubich C., Wanner G. Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations / Springer Series in Comput. Math. Springer, 2002. 536 p.

[6] Hairer E., Norsett S.P., Wanner G. Solving Ordinary Differential Equations. Nonstiff Problems / Springer Series in Comput. Math. Springer, 1993. 544 p.

Поступила в редакцию 27 августа 2009 г.,
с доработки — 26 сентября 2009 г.