arXiv:2304.04458v1 [astro-ph.EP] 10 Apr 2023

减少行星历表数值积分中的舍入误差

Maksim Subbotin、Alexander Kodukov 和 Dmitry Pavlov

俄罗斯联邦，圣彼得堡，197022，波波娃教授街5号，圣彼得堡电工大学计算机科学与技术学院。

*通讯作者。电子邮箱：dapavlov@etu.ru;

摘要

现代月球—行星历表在超过100年的观测时间跨度内进行数值积分（其中最近20年拥有非常精确的天文测量数据）。在如此长的时间跨度内，不仅有限差分近似误差会变得重要，累积的算术舍入误差也会变得重要，因为它们超过了月球、火星和水星高精度距离观测量的随机误差。解决这一问题的一种方法是使用x86处理器上可用的扩展精度算术。注意到这种方法的缺点后，我们提出了一种替代方案：在适当情况下使用双双精度算术。这样便只需使用具有普遍支持的双精度浮点原语。

关键词：数值积分、浮点运算、N体问题、行星历表。

1 引言

月球—行星历表本质上是一张表，其中包含太阳系行星和月球在时间网格上的坐标。月球—行星历表有多种用途，包括地球和太空导航、预测小行星和航天器的轨道、预测日食和月食，以及检验有关时空基本性质的科学假设。自 20 世纪 60 年代以来，星历表一直通过计算机上的数值

积分程序求解运动方程。现代月球—行星历表（按首次发布的时间排序）包括：JPL DE [1], IAA EPM [2], INPOP IMCCE [3]。此外，还有基于解析理论而非数值积分的星历表，例如 VSOP [4], ，但它们不在本文的讨论范围内。

运动方程包含许多通过天体测量观测值确定的自由参数。几十年来，随着观测值精度的提高和数学模型的改进，星历表的精度也不断提高。

通常，数学模型的精度必须达到或优于观测值的精度。在本文中，我们将重点研究截至目前拥有最精确观测值的三个天体的轨道：

- 月球——最精确的月球激光测距法截点达到 1 毫米的精度 [5, 6].
- 火星——精度最高的航天器无线电测距法截点拟合后均方根达到 55 厘米[7]。
- 水星——MESSENGER 航天器无线电测距法截点的拟合后均方根达到 60 厘米[2]。

本文所要处理的问题是运动方程数值积分过程中累积的算术舍入误差，因此这些误差会影响星历表的精度。

2 运动方程

本文采用 EPM2021 星历表的动力学模型，不过其他星历表也有类似的动力学模型。模型中纳入了以下天体：

- 所有行星；
- 太阳和月球；
- 冥王星及其他 30 个海王星外天体；
- 277小行星；
- 用于近似小行星带的离散均匀 180 点环带；
- 用于近似柯伊伯带的离散均匀 160 点环。

有关天体选择的详细信息，请参阅 [8–10]。十六个天体（太阳、行星、冥王星、谷神星、智神星、灶神星、虹神星、班贝加）遵循写到 $1/c^2$ 项的爱因斯坦–英费尔德–霍夫曼运动方程：

$$
\begin{aligned}
\boldsymbol{a}_i = & \sum_{j\neq i} \frac{Gm_j \boldsymbol{r}_{ij}}{r_{ij}^3} \\
& + \frac{1}{c^2} \sum_{j\neq i} \frac{Gm_j \boldsymbol{r}_{ij}}{r_{ij}^3} \left[v_i^2 + 2v_j^2 - 4(\boldsymbol{v}_i \boldsymbol{v}_j) - \frac{3}{2} \frac{(\boldsymbol{r}_{ji} \boldsymbol{v}_j)^2}{r_{ji}^2} \right. \\
& \qquad \left. - 4 \sum_{k\neq i} \frac{Gm_k}{r_{ik}} - \sum_{k\neq j} \frac{Gm_k}{r_{jk}} + \frac{1}{2} (\boldsymbol{r}_{ij} \boldsymbol{a}_j) \right] \\
& + \frac{1}{c^2} \sum_{j\neq i} \frac{Gm_j}{r_{ij}^3} \left[\boldsymbol{r}_{ji} (4\boldsymbol{v}_i - 3\boldsymbol{v}_j) \right] (\boldsymbol{v}_i - \boldsymbol{v}_j) \\
& + \frac{7}{2c^2} \sum_{j\neq i} \frac{Gm_j \boldsymbol{a}_j}{r_{ij}}
\end{aligned}
\tag{1}
$$

其中， $\boldsymbol{a}$ 是加速度； $\boldsymbol{v}$ 是速度； Gm 是引力参数； $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$ 是天体 j 相对于天体 i的位置； c 是光速。

请注意，式（1）中的第一项是牛顿加速度；还要注意，式（1）的右侧也包含加速度，因此看起来像一个隐式方程。不过在实际应用中， $\boldsymbol{a}_j$ 项会被牛顿加速度替代，这对结果的影响可以忽略不计。

模型中的其他点质量仅通过牛顿力与这 16 个点质量相互作用，彼此之间则不相互作用。作出这一折衷是为了加快计算速度；在 40 年的时间跨度内，这对行星轨道的影响小于 1 厘米，对月球地心轨道的影响小于 1 毫米。

本研究省略了其他对星历动力学模型至关重要的效应，包括：太阳扁率和Lense–Thirring效应引起的加速度；由太阳、金星、火星、木星和月球产生的地球“点质量—形状”加速度；以及将月球视为具有旋转液态核的弹性体 [11] ，并用时滞微分方程 [12]描述其所受潮汐力时的运动。

这些方程以 1/16天为步长，采用13阶的 Adams–Bashforth–Moulton 多步预测器-校正器格式（PECEC模式）进行数值积分。本研究中使用的实现方式类似于 [12], 中所述的方法，但不包含“延迟”部分。预测器和校正器的公式见式（2）。

$$
\begin{aligned}
\boldsymbol{x}_{n+1}^{(\mathrm{p})} &= \boldsymbol{x}_n + h \sum_{j=0}^{k-1} \gamma_j \nabla^j \boldsymbol{f}_n \\
\boldsymbol{x}_{n+1} &= \boldsymbol{x}_{n+1}^{(\mathrm{p})} + h\gamma_k \nabla^k \boldsymbol{f}_{n+1},
\end{aligned}
\tag{2}
$$

这里， $\boldsymbol{x}_n$ 是动力系统的状态， $\boldsymbol{f}_n = \boldsymbol{f}(\boldsymbol{x}_n)$是 $\boldsymbol{x}$的时间导数， ∇ 是向后有限差分算子， h 是步长， n是步数， k 是方法阶数， γ_j 是预计算系数。

为获得 $\boldsymbol{x}_2 \ldots \boldsymbol{x}_k$，采用了 8 阶单步 Dormand–Prince格式（步长为 $h/8$）。

算术舍入误差既会出现在数值格式本身中，也会出现在 $\boldsymbol{f}$ 的计算过程中。

3 数值舍入误差的估计

3.1 数字的表示

计算机中最常见的数值数据类型是 IEEE 754 双精度——一种 64 位二进制浮点数据类型，其中小数部分占52 位。假定以这种格式表示的每个数都存在一个不超过末位单位（末位单位）一半的误差，即小数部分最后一位的值。例如，$1 + 2^{-52}$ 可以用双精度精确表示，而1与 $1 + 2^{-53}$ 之间的任何数都会向下舍入为1，$1 + 2^{-53}$ 与 $1 + 2^{-52}$ 之间的任何数都会向上舍入为 $1 + 2^{-52}$。由此可以粗略估计，双精度表示的精度约为 $-\log_{10} 2^{-53} \approx 16$ 十进制位数，因此数值表示的相对误差约为 10^{-16}。

这样的精度足以表示天文数据。例如，地月距离的最大值约为 400000 km，而最佳观测值（见第 1 节）的精度约为 1 mm；因此，在任意时刻，将月球地心位置的精度要求设为最高 1 mm 是合理的[1]；也就是说，表示 2.5×10^{-12} 的相对误差达到这一水平就足够了。

同样，地火距离的最大值约为 4 亿 km；若要达到约 0.5 m 的观测精度，1.25×10^{-12} 的相对误差达到这一水平就足够了。

然而，在数值积分过程中，每次涉及双精度数的算术运算都会再次将结果舍入为双精度，这种舍入误差会传播到后续计算中；因此，最终轨道会受到不断累积的数值误差影响。

3.2 双向积分

为了估算舍入误差，必须找到一种方法，将它们与数值格式固有的近似误差以及底层模型中的误差区分开来。可以通过先正向、再反向积分相同的方程来实现这一点。反向积分以正向积分的结果作为初始状态。从数学上讲，这两次积分应产生相同的结果。实际上，两次积分都存在模型误差、近似误差和数值误差——但这些误差

[1]星历中月球轨道的半长轴与后向反射器的 X坐标高度相关；因此，所述坐标的不确定度至少约为3cm [13] 。然而，当关注后向反射器的地心坐标时，要求月球星历具有更高精度仍然是有意义的；在这些坐标中，后向反射器 X 坐标与月球轨道半长轴的不确定度在很大程度上相互抵消。

前两类误差会相互抵消[2]。因此，比较两次积分所得的轨道，就可以估算累积的舍入误差。关于这种方法以往的应用，参见例如 [14]。

在现代历表中，轨道的初始参数由覆盖约 100 年时间跨度的观测值确定。历元——即这些参数所对应的日期——位于该时间跨度内的某处，随后从历元开始沿两个方向进行数值积分。为减小数值积分的总体误差，历元可以设在时间跨度的中点；但考虑到现代观测值的精度更高，历元通常会从中点向后移动。在 EPM 中，历元为 1984年10月27日（JD 2446000.5）。本文将研究 40 年积分时间跨度（单向）内舍入误差的累积情况。

使用双精度计算、步长为 1/16天时得到的月球、水星和火星位置的舍入误差如图1–3所示。在此配置下，单向积分 40 年耗时 81 秒。显然，对于这三个天体，双精度产生的数值舍入误差并未全部低于观测值的精度：月球为 25 厘米（法截点精度达到 1 毫米），水星为 2.5 米（法截点的拟合后均方根达到 60 厘米），火星为 70 厘米（法截点的拟合后均方根达到 55 厘米）。

3.3 质心

由 N 个天体组成的系统的（相对论）质心定义如下：

$$\boldsymbol{b} = \frac{\sum_i \mu_i^* \boldsymbol{r}_i}{\sum_i \mu_i^*} \tag{3}$$

其中， μ_i^* 是天体的相对论引力参数：

$$\mu_i^* = \mu_i \left(1 + \frac{1}{2c^2} \dot{r_i}^2 - \frac{1}{2c^2} \sum_{j \neq i} \frac{\mu_j}{r_{\mathrm{ij}}} \right) \tag{4}$$

对天体的初始位置和速度进行等量调整，使质心的位置和动量均为零：

$$\begin{cases} \sum_i \mu_i^* \boldsymbol{r}_i = \boldsymbol{0} \\ \sum_i \dot{\mu_i^*} \boldsymbol{r}_i + \mu_i^* \dot{\boldsymbol{r}}_i = \boldsymbol{0} \end{cases} \tag{5}$$

从数学上讲，系统的质心在整个积分过程中应保持在原点，但近似误差和数值舍入误差会导致其发生漂移。在图4中，我们考察质心到原点的距离（随时间的函数），将其作为对

[2]在多步格式中，由于存在“预热”阶段，积分时间段两端附近的近似值略有不同，但我们稍后会看到，这种差异不会产生任何明显影响。

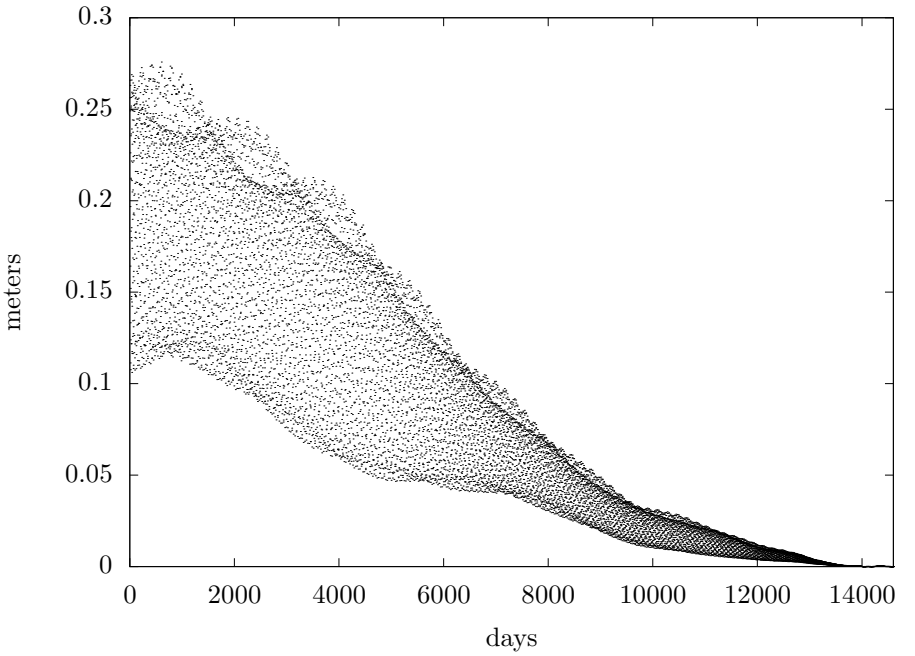

图 1 双精度下月球位置的舍入误差。下文所说的月球位置均指地心位置，而行星位置均指质心位置。

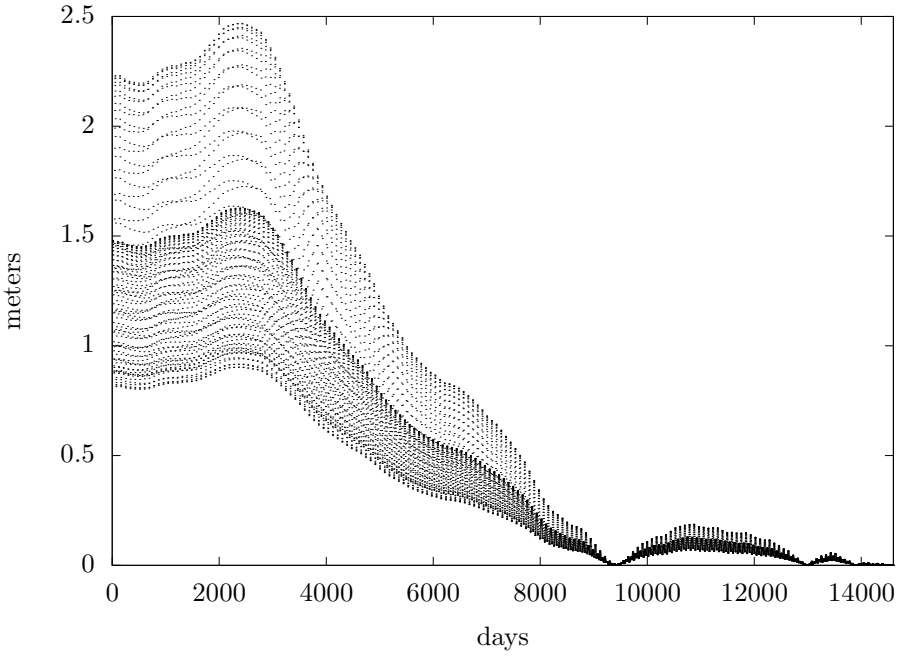

图 2 双精度下水星位置的舍入误差

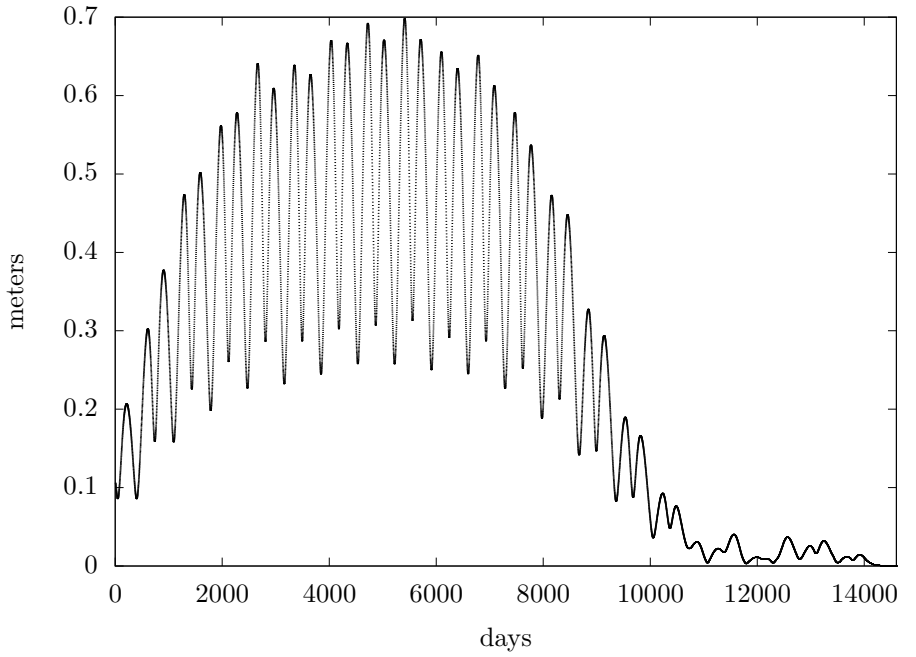

图 3 双精度下火星位置的舍入误差

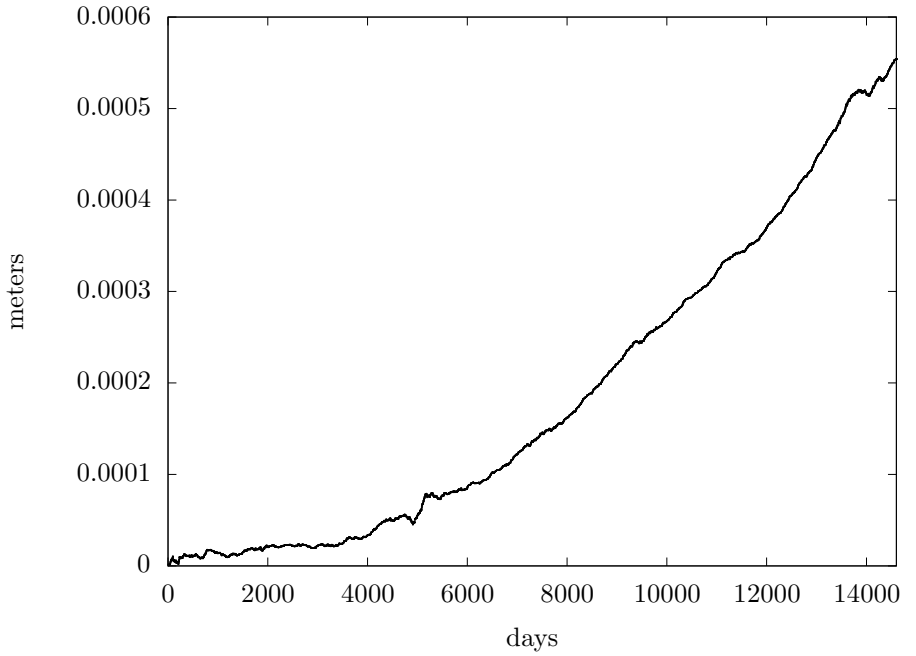

图 4 双精度下质心的漂移

EIH 方程实现的正确性（1）以及差分格式的正确性和稳定性（2）。漂移值小于毫米，证实一致性检查已通过。

4 减少舍入误差

4.1 现有方法

所有针对累积舍入误差问题的缓解措施，都或多或少地建立在使用双精度以外的数据类型之上。DE405（1997年）之前的DE星历表是在UNIVAC大型机上使用其原生的72位浮点类型、其中尾数为60位进行积分的（巧合的是，这种类型也被称为双精度）；DE405则是在VAX Alpha上使用其原生的128位浮点类型、其中尾数为112位进行积分的，即四精度 [15]。DE410（2003年）是在配备UltraSPARC处理器的SunUltra上使用普通双精度进行积分的，这导致了数值噪声问题，使轨道的精度被限制在1–2米 [16]。用于构建DE的数值积分器称为DIVA [17]。到DE414（2006年）时，DIVA升级为QIVA，其中“Q”代表四精度（并非由硬件实现，而是由Fortran编译器模拟实现）。实际上，采用的是混合模式：只有加速度中的牛顿部分以四精度计算，而其余所有摄动则以双精度计算。即便如此，混合模式所需的时间仍是双精度模式的30倍。QIVA可能至今仍被用于构建DE星历表，不过编译器的模拟功能此后可能已有改进。需要注意的是，对于这项任务而言，四精度远远超出了实际需要。

自 EPM2004 以来的所有公开版 EPM，以及自 INPOP06 以来的所有版本 INPOP，均采用 80 位 扩展精度类型（63 位尾数）进行数值积分。对于行星轨道而言，这种精度已足够，见图 5–8。然而，它也有自身的缺点：

- 扩展精度仅在 x86 架构（英特尔和 AMD 处理器）上受支持，在 ARM 架构处理器上则不受支持，而后者似乎正在获得越来越大的市场份额（Apple M1、Microsoft SQ1、AWS Graviton，以及几乎所有移动芯片）。
- 很少有编程语言提供对扩展精度的支持，即 Delphi、C/C++、D、Perl、Racket BC[3]、Swift，以及 CommonLisp和 Fortran 的一些实现。Fortran 中的扩展精度仅限于 gfortran编译器；Intel 和 Absoft Fortran 编译器不支持该功能。
- 虽然使用扩展精度数字进行的简单运算在 x86 中与使用双精度数字进行的运算一样快，但 x86 架构的 SIMD 扩展（MMX、SSE、AVX、AVX-512）不支持扩展

[3]“BC”代表“Chez 之前”，指的是被 Chez Scheme 取代之前使用的 Racket 编译器和虚拟机，但目前仍受支持。

精度数值。这一限制使其相较于双精度模式速度降低了 4 倍。（单向积分 40 年耗时 323 秒。）

4.2 双双精度算术的实现

双双精度算术是一种提高浮点数精度的软件方法，它使用成对的双精度值存储数值，并基于双精度算术执行算术运算。双双精度数被定义为两个双精度分量（较大部分和较小部分）的未求值之和：$x = x_{\text{low}} + x_{\text{high}}$，$|x_{\text{low}}| \leq 0.5 \cdot$ 末位单位(x_{high})，因此它们的有效数不会重叠。该技术能够存储有效数至少包含 $2 \cdot 53 = 106$ 位的数值。

双双精度算术不需要特殊硬件或编程语言支持（软件实现[18] 可用）。下面展示了双双精度类型的一些基本运算算法。

• Dekker求和或带误差的两个双精度数 [19]。RN(x) 表示舍入到最近值。

算法 1 Fast2Sum (a, b)，$a \geq b$

$s \leftarrow RN(a + b)$
$z \leftarrow RN(s - a)$
$t \leftarrow RN(b - z)$
return (s, t)

• Dekker 的双精度求和 [19]。Fast2Sum 算法用于累加较大的分量，然后将求和误差加到较小的分量上。

算法 2 DDSum (x_h, x_l, y_h, y_l)

如果 $|x_h| \geq |y_h|$ 则
  $(r_h, r_l) \leftarrow$ Fast2Sum(x_h, y_h)
  $s \leftarrow$ RN(RN($r_l + y_l$) + x_l)
else
  $(r_h, r_l) \leftarrow$ Fast2Sum(y_h, x_h)
  $s \leftarrow$ RN(RN($r_l + x_l$) + y_l)
结束如果
返回 Fast2Sum(r_h, s)

• 带误差的双精度乘法（使用融合乘加指令，即 FMA）。

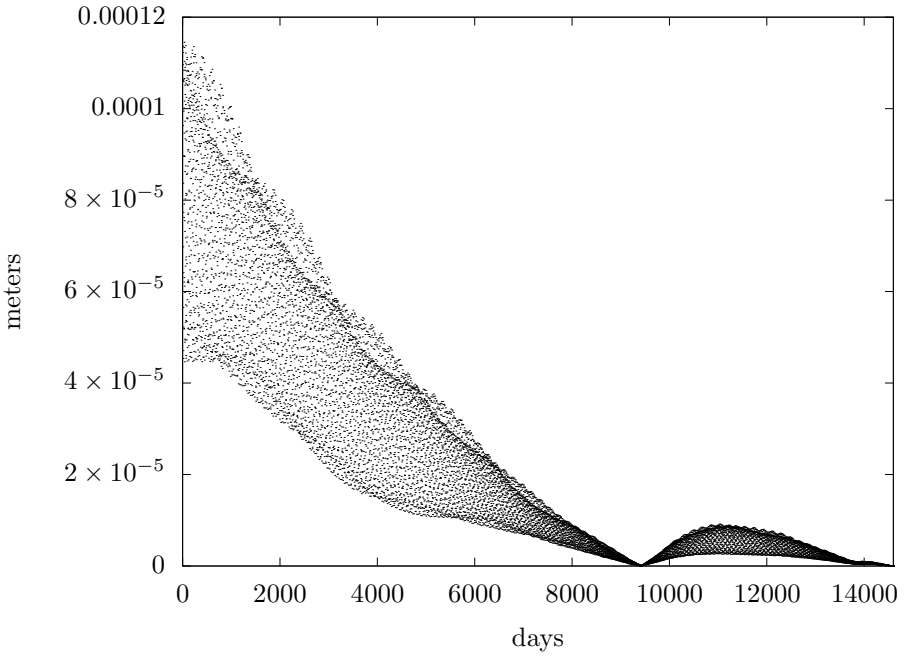

图 5 扩展精度下月球位置的舍入误差

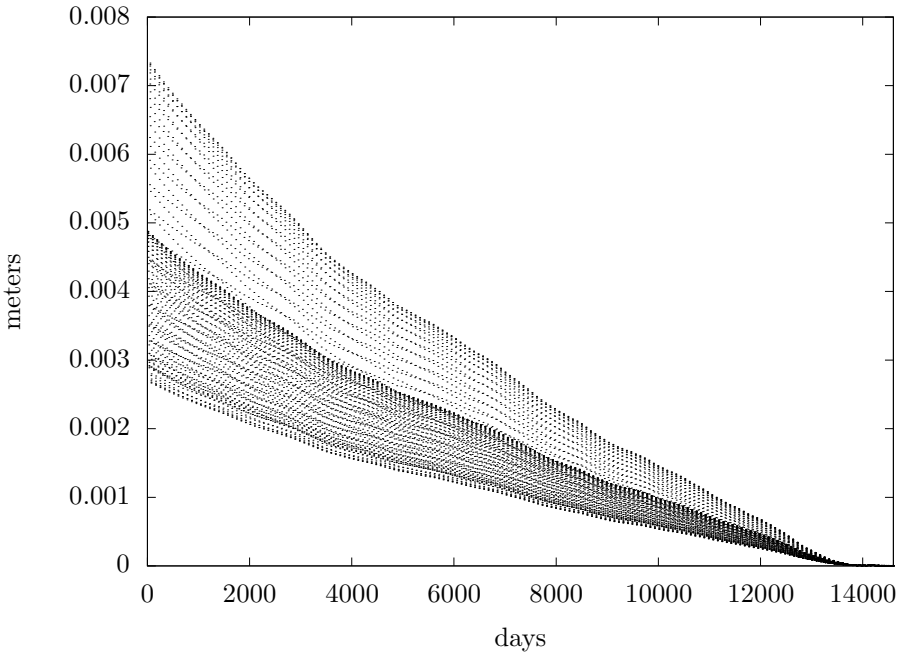

图 6 使用扩展精度时水星位置的舍入误差

减　少行星历表数值积分中的舍入误差 1　1

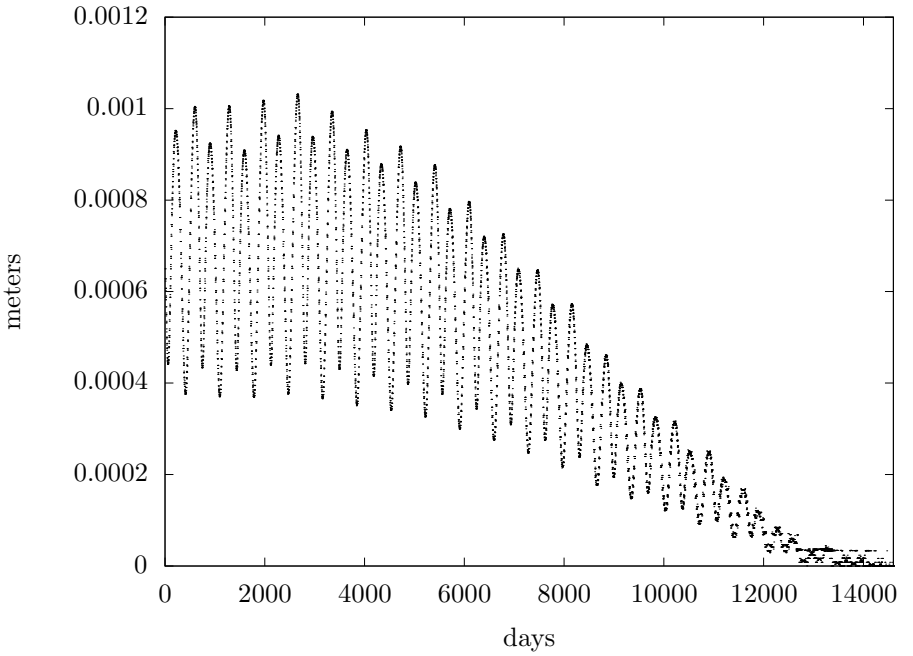

图 7 扩展精度下火星位置的舍入误差

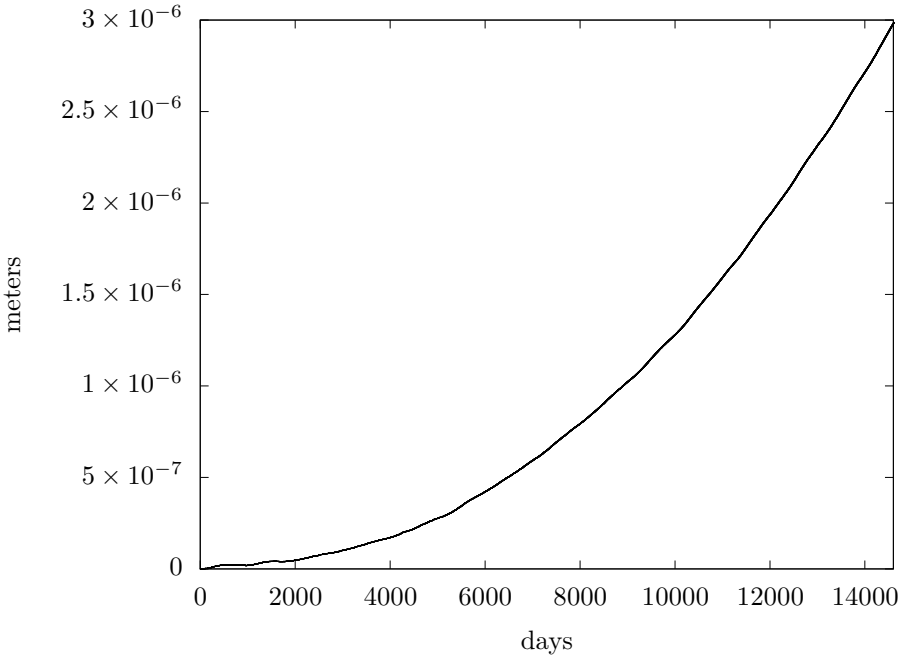

图 8 扩展精度下质心的漂移

算法 3 Product(x, y)

$p \leftarrow x \cdot y$
$e \leftarrow$ FMA($x, y, -p$)
返回 (p, e)

- 双双精度乘法[19]。

算法 4 DDProduct (x_h, x_l, y_h, y_l)

$(c_h, c_l) \leftarrow$ 产品(x_h, y_h)
$p_1 \leftarrow$ RN($x_h \cdot y_l$)
$p_2 \leftarrow$ RN($x_l \cdot y_h$)
$c_l \leftarrow$ RN($c_l + RN(p_1 + p_2)$)
返回 Fast2Sum(c_h, c_l)

- 双双精度除法 [18]。第一次近似值为 $q_0 = x_h/y_h$。余数为 $r = x - q_0 \cdot y$。修正项为 $q_1 = r_h/b_h$。最后，Fast2Sum(q_0, q_1) 对结果重新规格化。双一乘积是双双精度数与双精度数的乘积（其写法与 DDProduct 类似）。二和用于 Fast2Sum 的参数顺序未知的情况。

算法 5 双双精度除法 (x_h, x_l, y_h, y_l)

$t_h \leftarrow$RN(x_h/y_h)
$(r_h, r_l) \leftarrow$ 双一乘积(y_h, y_l, t_h)
$(\pi_h, \pi_l) \leftarrow$ 2和($x_h, -r_h$)
$\sigma_h \leftarrow$ RN($\pi_l - r_l$)
$\sigma_l \leftarrow$ RN($\sigma_h + x_l$)
$\sigma \leftarrow$ RN($\pi_h + \sigma_l$)
$t_l \leftarrow$RN(σ/y_h)
返回 Fast2Sum(t_h, t_l)

- 双双精度平方根。QD 库 [18] 使用 Karp 技巧计算平方根 [20]。如果 s 是 $\sqrt{x}$ 的近似值（QD 使用 $\sqrt{x_h}$），那么 $\sqrt{x} \approx x \cdot s + [x - (x \cdot s)^2] \cdot s/2$。

作为实验的一部分，本工作实现了数值积分器的完整双双精度版本。结果见图 9–12。显然，数值舍入误差可以忽略不计，但代价是性能大幅下降：单向积分 40 年耗时 32 分钟，比双精度慢 24 倍。

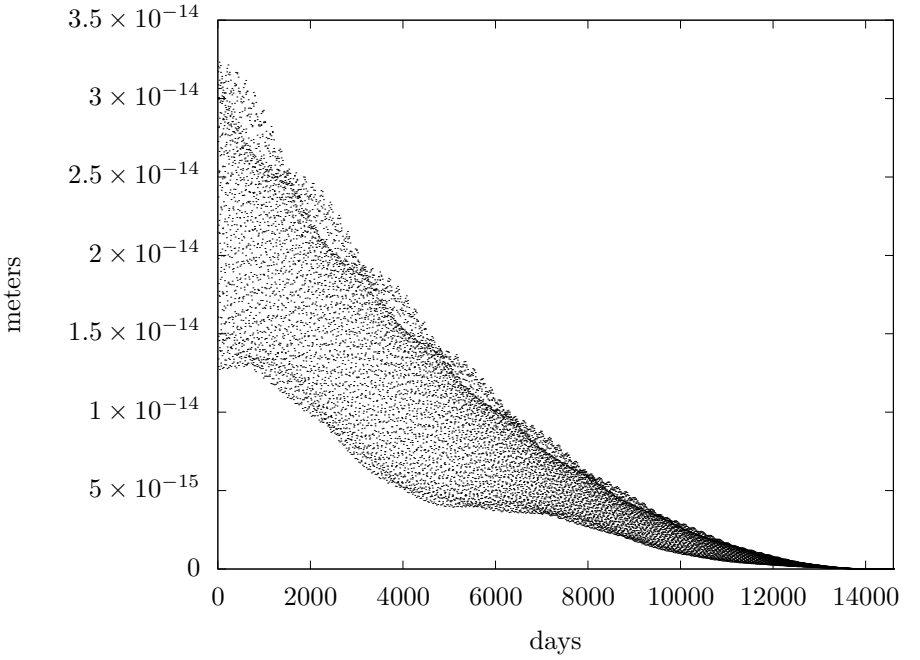

图 9 双双精度下月球位置的舍入误差

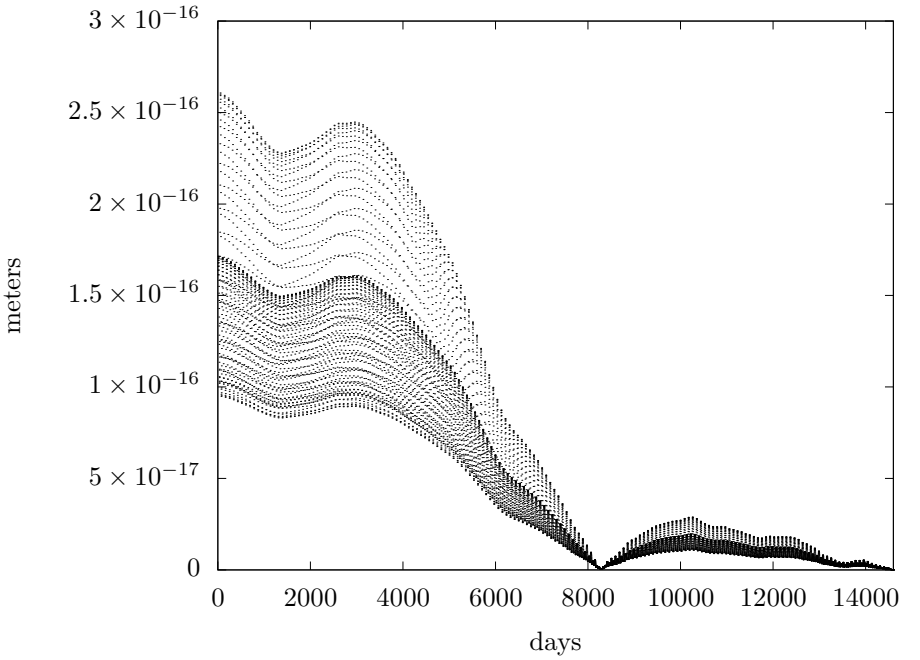

图 10 双双精度下水星位置的舍入误差

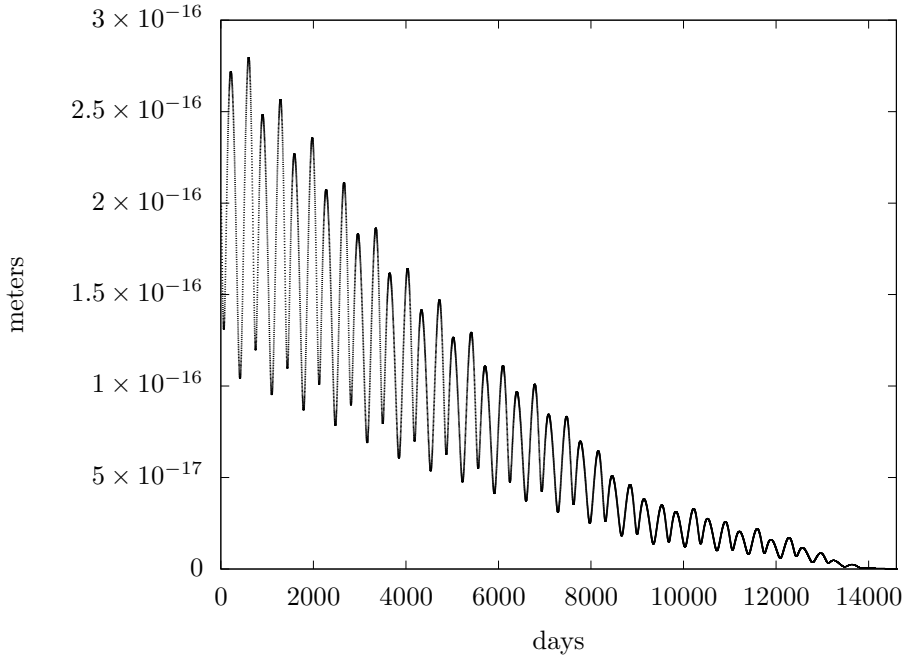

图 11 双双精度下火星位置的舍入误差

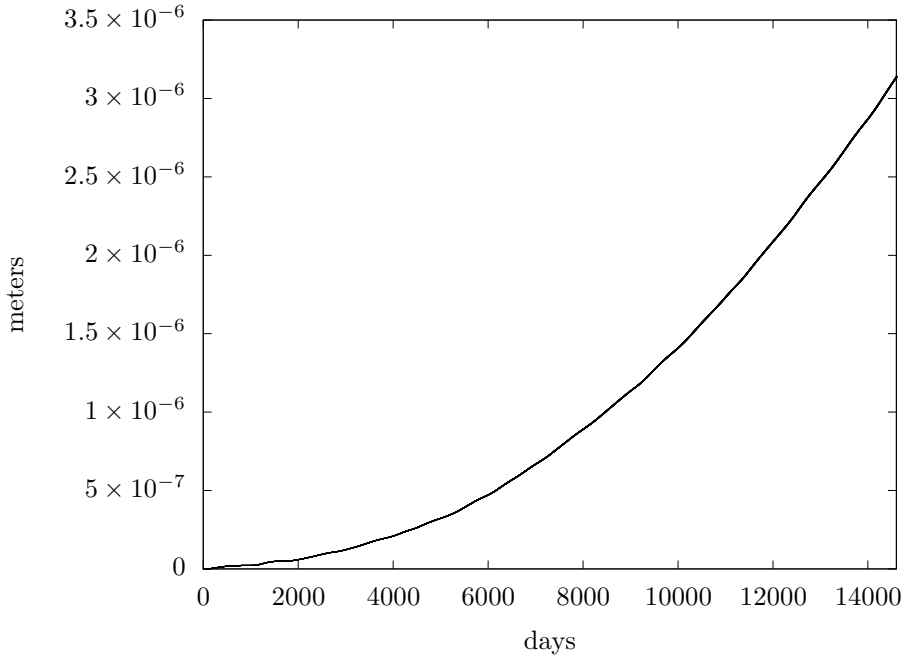

图 12 双双精度下质心的漂移

减　少行星历表数值积分中的舍入误差 15

4.3 双双精度类型的选择性使用

通过反复试验，得出了以下方案，以平衡计算的精度与性能：

- 系统的状态 $\boldsymbol{x}$ 及其时间导数 $\boldsymbol{f}$ 均以双双精度存储，PECEC方案中的中间值也采用双双精度存储。
- 有限差分 ∇^i 以双精度存储。然而，系数 γ_i 以双双精度存储。
- 在 $\boldsymbol{f}(\boldsymbol{x})$中，即在公式（1）中， $\boldsymbol{r}_{ij}$ 向量和 r_{ij} 距离以双精度存储。
- 在方程（1）的相对论部分中， $\boldsymbol{v}_i$ 和 $\boldsymbol{a}_i$ 均向下舍入为双精度，并且所有计算均以双精度进行。
- 常数 $1/c^2$ 和引力参数 Gm_i 也以双精度存储。

4.4 地月系统的特殊处理

记 $\boldsymbol{r}_{\mathrm{EM}} = \boldsymbol{r}_{\mathrm{M}} - \boldsymbol{r}_{\mathrm{E}}$ 为地月矢量。式 $\boldsymbol{r}_i$ （1）中的所有量都应当参照同一个惯性参考系（在我们的情况下为质心天球参考系，或 BCRS）。然而，直接计算 $\boldsymbol{r}_{\mathrm{M}} - \boldsymbol{r}_{\mathrm{E}}$ 将得到一个比被减数和减数小约 400 倍的值，从而由于有效位的“灾难性消去”导致精度损失。对于 $\boldsymbol{r}_{\mathrm{EM}}$ 而言，这尤其不可取，因为月球轨道对数值精度有很高的要求。在三类星历表（DE、EPM、INPOP）中，月球轨道都是在地心坐标系中积分的，因此 $\boldsymbol{r}_{\mathrm{EM}}$ 会进入 $\boldsymbol{x}$ ，而质心的 $\boldsymbol{r}_{\mathrm{M}}$ 则会在计算 $\boldsymbol{f}$期间被暂时恢复。

同样，我们旨在避免计算地心月球的牛顿加速度时发生灾难性消去。对于地球和月球之间的相互牛顿引力，这相当简单。在式（6）中，不涉及相近数值相减。

$$\boldsymbol{a}_{\mathrm{EM}}^{(\mathrm{E,M})} = \boldsymbol{a}_{\mathrm{M}}^{(\mathrm{E})} - \boldsymbol{a}_{\mathrm{E}}^{(\mathrm{M})} = -\frac{Gm_{\mathrm{M}} + Gm_{\mathrm{E}}}{r_{\mathrm{EM}}^3}\boldsymbol{r}_{\mathrm{EM}} \tag{6}$$

对于其他天体施加的月球地心加速度，我们可以应用泰勒展开。对于天体 A，我们旨在避免直接相减由该天体引起的月球 $\boldsymbol{a}^{(\mathrm{A})}{}_{\mathrm{M}}$ 和地球 $\boldsymbol{a}^{(\mathrm{A})}{}_{\mathrm{E}}$ 的牛顿加速度：

$$\begin{aligned}
\boldsymbol{a}_{\mathrm{EM}}^{(\mathrm{A})} &= \boldsymbol{a}_{\mathrm{M}}^{(\mathrm{A})} - \boldsymbol{a}_{\mathrm{E}}^{(\mathrm{A})} = Gm_{\mathrm{A}}\left(\frac{\boldsymbol{r}_{\mathrm{MA}}}{r_{\mathrm{MA}}^3} - \frac{\boldsymbol{r}_{\mathrm{EA}}}{r_{\mathrm{EA}}^3}\right) \\
&= \frac{Gm_{\mathrm{A}}}{r_{\mathrm{MA}}^3}\left[\boldsymbol{r}_{\mathrm{MA}} - \boldsymbol{r}_{\mathrm{EA}}\left(\frac{\left(\boldsymbol{r}_{\mathrm{EA}} - \boldsymbol{r}_{\mathrm{EM}}\right)^2}{r_{\mathrm{EA}}^2}\right)^{3/2}\right] \\
&= \frac{Gm_{\mathrm{A}}}{r_{\mathrm{MA}}^3}\left[\boldsymbol{r}_{\mathrm{MA}} - \boldsymbol{r}_{\mathrm{EA}}\left(1 - 2\frac{\boldsymbol{r}_{\mathrm{EA}}\boldsymbol{r}_{\mathrm{EM}}}{r_{\mathrm{EA}}^2} + \frac{r_{\mathrm{EM}}^2}{r_{\mathrm{EA}}^2}\right)^{3/2}\right]
\end{aligned} \tag{7}$$

精度	double	扩展了	混合	双精度-双精度
月球的最大误差	0.28 米	0.1 毫米	1.2 mm	$3.2 \cdot 10^{-14}$ 米
水星的最大误差	2.47 米	7 mm	8.4 毫米	$2.6 \cdot 10^{-16}$ 米
火星的最大误差	0.7 米	1 mm	0.44 毫米	$2.8 \cdot 10^{-16}$ 米
质心的最大漂移	0.5 毫米	2.9 μ米	2.6 μ米	3.14 μm
积分时间	81 s	323 秒	246 s	32 分钟
g比值				

表 1 不同数据类型下进行 40 年数值积分的比较结果。

记 $x = 2\frac{\boldsymbol{r}_{\mathrm{EA}}\boldsymbol{r}_{\mathrm{EM}}}{r_{\mathrm{EA}}^2} - {}^{r^2}{}_{\mathrm{EM}}\overline{r_{\mathrm{EA}}^2}$ 并应用 $(1-x)^{3/2}$ 的泰勒展开，可得

$$\boldsymbol{a}_{\mathrm{EM}}^{(\mathrm{A})} \approx \frac{Gm_{\mathrm{A}}}{r_{\mathrm{MA}}^3}\left[-\mathbf{r}_{\mathrm{EM}} - \mathbf{r}_{\mathrm{EA}}\left(-\frac{3x}{2} + \frac{3x^2}{8} + \frac{x^3}{16} + \frac{3x^4}{128}\right)\right] \tag{8}$$

其中不存在相近数值相减的情况。此外，研究发现，双精度足以用于方程（8），泰勒展开需要四项，但不需要更多。

4.5 结果

采用混合精度方法（第 4.3 节）并对地月系统进行特殊处理（第 4.4 节）得到的最终结果如图 13–16 所示。其精度与采用扩展精度得到的结果相当。单向积分 40 年耗时 246 秒。

5 结论

结果汇总见表 1。显然，本研究采用的“混合”模式积分所产生的数值舍入误差足够小，可满足现代星历天文学的要求，同时无需使用除 64 位双精度以外的任何基本数据类型。

“混合”模式仅比双精度模式慢3倍，比扩展精度模式快1.3倍。后者可以解释为，扩展精度无法在 x86 硬件上利用 SIMD 扩展。

必须指出的是，此类测量的结果通常会受到所使用方程的影响（小行星和其他点质量的数量、形状效应等）。形状效应对于月球的轨道和自转运动尤其重要。此外，$\boldsymbol{x}$ 相对于模型参数的导数通常也会与动力学方程一同积分。所有这些额外的复杂性在本工作中均被忽略；然而，原则上这不太可能改变结果。月球自转方程可以很容易地在双双精度中实现，而不会对整体性能产生显著影响；加速度中的形状效应足够小，可以用双精度计算；对于这些导数，也不需要高精度。

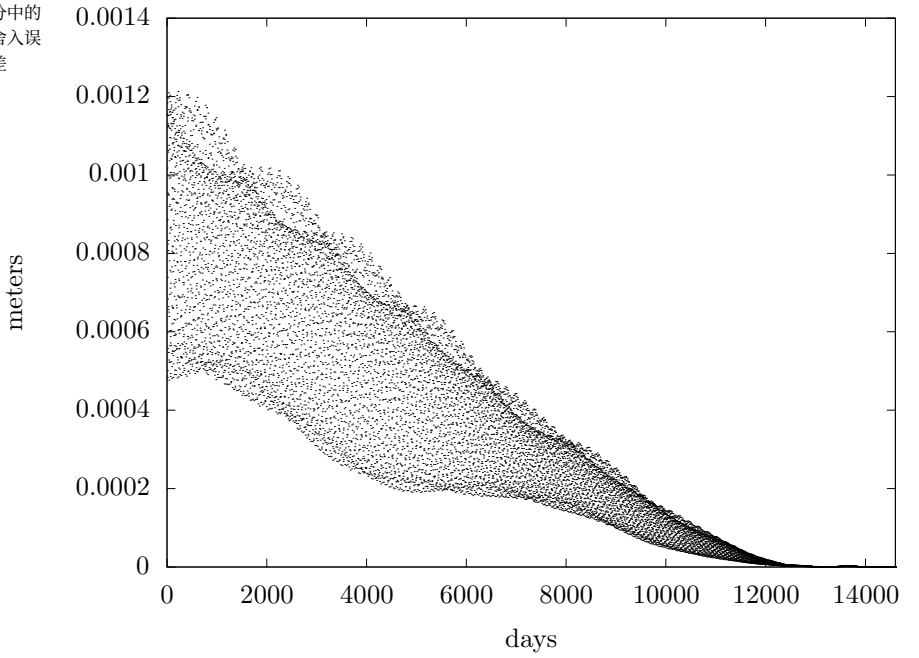

图 13 混合精度下月球位置的舍入误差

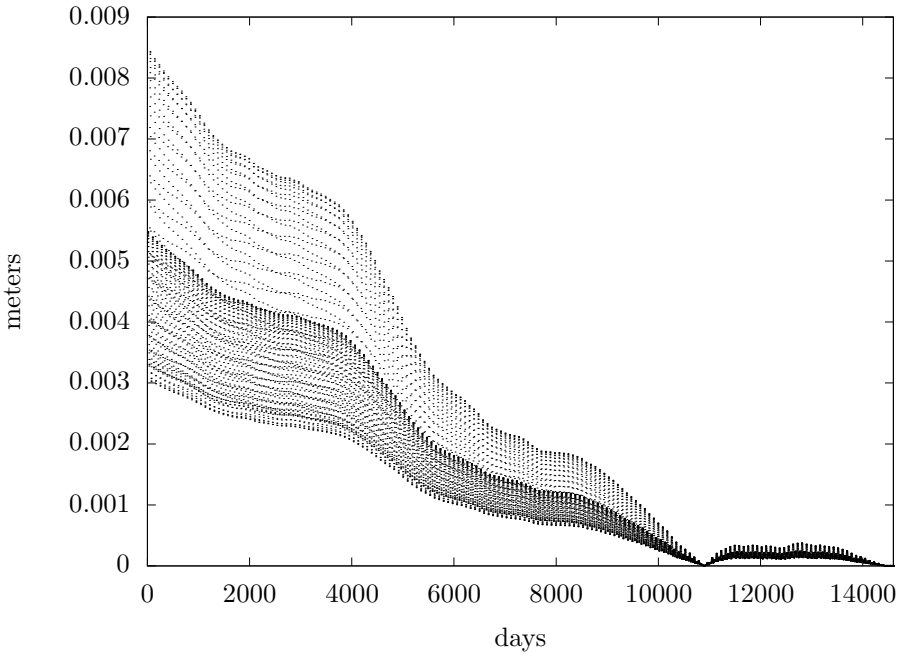

图 14 使用混合精度时水星位置的舍入误差

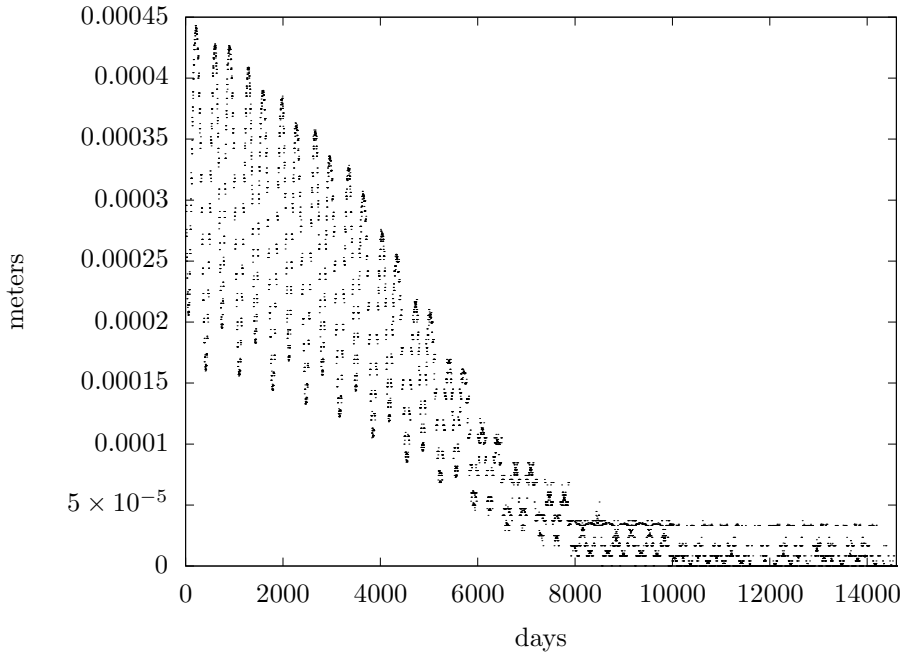

图 15 混合精度下火星位置的舍入误差

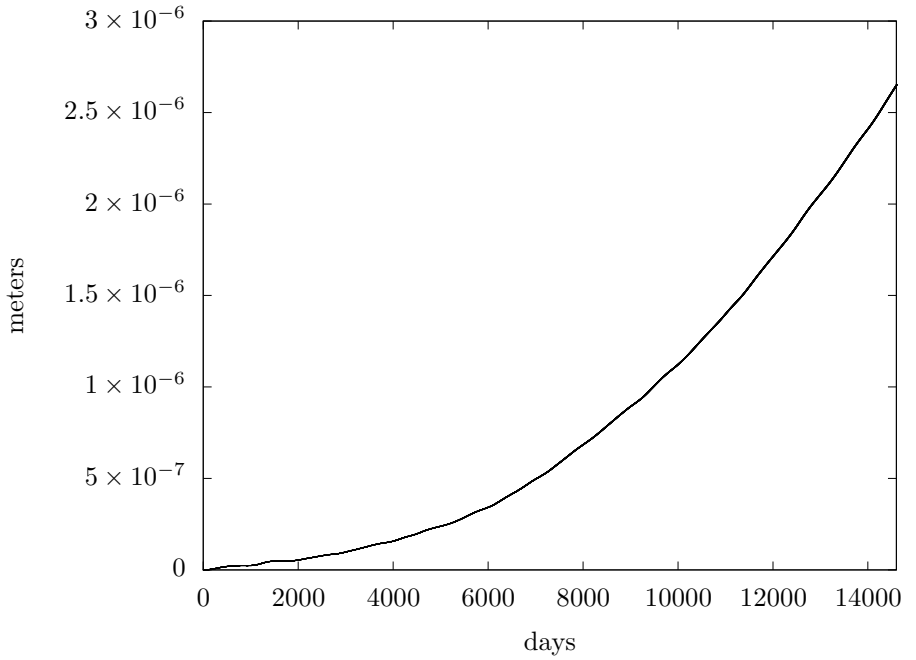

图 16 混合精度下质心的漂移

减少 行星历表数值积分中的舍入误差 19

数据可用性

本研究中用于数值积分的程序代码和数据可在线获取：
https://github.com/S1ckick/Nbodies。

参考文献

[1] Park, R.S., Folkner, W.M., Williams, J.G., Boggs, D.H.：《JPL行星和月球星历表DE440和DE441》。《天文学杂志》**161**（3），105（2021年）。
https://doi.org/10.3847/1538-3881/abd414

[2]Pitjeva, E.、Pavlov, D.、Aksim, D.、Kan, M.：行星和月球星历表 EPM2021 及其对太阳系研究的重要意义。《国际天文学联合会会议录》**15**（S364），220–225（2022年）。https://doi.org/10.1017/S1743921321001447[3] Fienga, A.、Deram, P.、Di Ruscio, A.、Viswanathan, V.、Camargo, J.I.B.、Bernus, L.、Gastineau, M.、Laskar, J.：INPOP21a 行星星历表。《巴黎天文台科学与技术笔记》´天体力学´，S110，巴黎天文台（2021年）。
https://www.imcce.fr/content/medias/recherche/equipes/asd/inpop/inpop21a.pdf
[4] Simon, J.-L.、Francou, G.、Fienga, A.、Manche, H.：新的解析行星理论 VSOP2013 和 TOP2013。《天文学与天体物理学》**557**，49（2013年）。
https://doi.org/10.1051/0004-6361/201321843

[5] Murphy, T.W., Adelberger, E.G., Battat, J.B.R., Hoyle, C.D., Johnson, N.H., McMillan,R.J., Stubbs,C.W., Swanson,H.E.：APOLLO：毫米级月球激光测距。经典与量子引力 **29**（18），184005（2012）。
https://doi.org/10.1088/0264-9381/29/18/184005

[6]Courde, C., Torre, J. M., Samain,E., Martinot-Lagarde, G., Aimar, M., Albanese, D.,Exertier, P., Fienga,A., Mariey, H., Metris,G.,Viot,H.,Viswanathan, V.：格拉斯激光站的红外月球激光测距。《天文学与天体物理学》**602**，90（2017）。
https://doi.org/10.1051/0004-6361/201628590

[7] Kuchynka, P., Folkner, W.M., Konopliv, A.S.：火星测距测量中的台站特定误差。行星际网络进展报告 42-190，NASA JPL（2012）。
https://ipnpr.jpl.nasa.gov/progressreport/42-190/190C.pdf

[8] Pitjeva, E.V.,Pitjev, N.P.：从行星和航天器的运动中确定主小行星带和柯伊伯带的质量。《天文学快报》**44**（8- 9），554–566（2018）。
https://doi.org/10.1134/s1063773718090050[9] Pitjeva,E.V., Pitjev, N.P.：柯伊伯带的质量。《天体力学》

与《天体力学与动力天文学》**130**(9), 57（2018）。
https://doi.org/10.1007/s10569-018-9853-5

[10] Kan, M.、Pavlov, D.：利用红外数据在EPM星历表中对主小行星带及部分小行星的质量进行动力学估算。载于：多信使研究时代的天文学。VAK-2021会议论文集，第93–95页。莫斯科国立大学国家天文研究所、俄罗斯科学院应用天文学研究所，莫斯科（2022年）。https://www.vak2021.ru/wp-content/uploads/2022/03/VAK2021proceedings.pdf

[11] Pavlov，D.、Williams，J.、Suvorkin，V.：利用 GRAIL 和国际地球自转与参考系服务推荐模型，根据月球激光测距观测确定月球轨道与自转运动参数。《天体力学与动力天文学》**126**(1)，61–88（2016）。
https://doi.org/10.1007/s10569-016-9712-1

[12] Aksim,D.、Pavlov, D.：关于亚当斯–巴什福思–莫尔顿方法在时滞微分方程数值积分中的扩展及其在月球轨道中的应用。《计算机科学中的数学》**14**，103–109（2020）。https://doi.org/10.1007/s11786-019-00447-y

[13] Pavlov,D.：月球激光测距在实现地球、月球和星历参考系中的作用。《大地测量学杂志》**94**（1）（2019）。https://doi.org/10.1007/s00190-019-01333-y

[14]Fienga, A., Manche, H., Laskar, J., Gastineau, M.：INPOP06：一种新的数值行星星历表。《天文学与天体物理学》**477**(1), 315–327 (2008)。
https://doi.org/10.1051/0004-6361:20066607[15] Standish E. M., Newhall X. X., Williams J. G., Yeomans D. K.：载于：Sei- delmann P. K.（编）《太阳、月球和行星的轨道星历表》，第279–374页。University Science Books，弗吉尼亚州赫恩登（1992）。
https://ssd.jpl.nasa.gov/ftp/eph/planets/ioms/ExplSupplChap8.pdf[16] Standish, E.M.：JPL行星星历表DE410。喷气推进实验室内部备忘录312.N–03–009，加州理工学院（2003）。
https://ssd.jpl.nasa.gov/ftp/eph/planets/ioms/de410.iom.pdf[17] Krogh, F.T.：一位亚当斯方法研究者使用龙格–库塔法。第395节。计算备忘录554，加州理工学院（1997）。
https://trs.jpl.nasa.gov/bitstream/handle/2014/22564/97-1071.pdf[18] Hida, Y., Li, S., Bailey, D.：双双精度和四双精度运算库。技术报告，劳伦斯伯克利国家实验室（2008）。
https://www.davidhbailey.com/dhbpapers/qd.pdf

[19] Muller, J.-M., Brisebarre,N., Dinechin, F., Jeannerod, C.-P., Lefevre, V., Melquiond, G., Revol,N., Stehle, D., Torres, S.：《浮点运算手册》（2010）。https://doi.org/10.1007/978-0-8176-4705-6

[20] Karp, A.H., Markstein, P.：高精度除法和平方根。ACM 数学软件汇刊 **23**（4），561–589（1997）。https://doi.org/10.1145/279232.279237