Reducing round-off errors in symmetric multistep methods

Paola Console[a], Ernst Hairer[a]

[a]*Section de Mathématiques, Université de Genève, 2-4 rue du Lièvre, CH-1211 Genève 4, Switzerland. (Paola.Console@unige.ch, Ernst.Hairer@unige.ch).*

Abstract

Certain symmetric linear multistep methods have an excellent long-time behavior when applied to second order Hamiltonian systems with or without constraints. For high accuracy computations round-off can be the dominating source of errors. This article shows how symmetric multistep methods should be implemented, so that round-off errors are minimized and propagate like a random walk.

Keywords: Symmetric linear multistep methods, constrained Hamiltonian systems, propagation of round-off error
2000 MSC: 65L06, 65G50, 65P10

1. Introduction

This article considers the numerical solution of constrained Hamiltonian systems

$$\begin{aligned} M\,\ddot{q} &= -\nabla U(q) - G(q)^{\mathsf{T}}\lambda \\ 0 &= g(q), \end{aligned} \tag{1}$$

where $q \in \mathbb{R}^d$, M is a positive definite constant matrix, $U(q)$ is a smooth real potential, $g(q) \in \mathbb{R}^m$ (with $m < d$) collects holonomic constraints, and $G(q) = g'(q)$ is the matrix of partial derivatives. Assuming that $G(q)M^{-1}G(q)^{\mathsf{T}}$ is invertible, the system (1) is a differential-algebraic equation of index 3. With the momentum $p = M\dot{q}$, the problem can be interpreted as a differential equation on the manifold

$$\mathcal{M} = \{(q,p)\,;\, g(q) = 0,\, G(q)M^{-1}p = 0\}. \tag{2}$$

Its flow is a symplectic transformation on $\mathcal{M}$, and it preserves the Hamiltonian (total energy)

$$H(q,p) = \tfrac{1}{2}\, p^{\mathsf{T}}M^{-1}p + U(q). \tag{3}$$

For a qualitative correct long-time integration of such systems the use of a geometric integrator is essential. An excellent choice is the Rattle algorithm [1], which is a symplectic, symmetric one-step method. However, due to its low order 2, it is not efficient for high accuracy computations. Symplectic partitioned Runge–Kutta methods (such as the Lobatto IIIA–IIIB pair) have arbitrarily high order, but they are implicit in the force evaluations. This article considers the use of explicit, symmetric multistep methods. With the notation $f(q) = -\nabla U(q)$ they are given by

$$\begin{aligned} \sum_{j=0}^{k} \alpha_j\, q_{n+j} &= h^2 \sum_{j=1}^{k-1} \beta_j\, M^{-1}\Big(f(q_{n+j}) - G(q_{n+j})^{\mathsf{T}}\lambda_{n+j}\Big) \\ 0 &= g(q_{n+k}). \end{aligned} \tag{4}$$

For given $q_n, \dots, q_{n+k-1}$ and $\lambda_{n+1}, \dots, \lambda_{n+k-2}$, the second relation implicitly defines λ_{n+k-1}, and the first relation gives an explicit expression for q_{n+k}. An approximation of the momentum $p = M\dot{q}$ is obtained *a posteriori* by symmetric finite differences supplemented with a projection onto $\mathcal{M}$:

$$p_n = M \frac{1}{h} \sum_{j=-l}^{l} \delta_j q_{n+j} + h\, G(q_n)^{\mathsf{T}} \mu_n, \qquad G(q_n) M^{-1} p_n = 0. \tag{5}$$

The second relation represents a linear system for μ_n, and the first relation is an explicit formula for p_n. By definition, this method yields a numerical solution on the manifold $\mathcal{M}$. It is proved in [2] (in the absence of constraints, see [3]) that under suitable assumptions on the coefficients α_j and β_j, the method can have high order, and the numerical approximations nearly conserve the Hamiltonian over very long time intervals. Its computational cost is essentially the same as that for the Rattle algorithm, which makes it to an excellent choice for high accuracy computations.

For computations close to machine accuracy, round-off errors can become more important than discretization errors. This motivates the present study of the propagation of round-off errors. This article gives hints on how the method should be implemented to reduce round-off errors and to obtain approximations for which the error in the Hamiltonian behaves like a random walk. Numerical experiments are presented in a final section.

2. Reducing round-off errors

For a straight-forward implementation of method (4)-(5), the round-off error typically increases linearly with time. This can be observed for step sizes, for which the discretization error is close to machine precision. It is known [5] that symplectic implicit Runge–Kutta methods can be implemented such that the round-off error is improved quantitatively (using compensated summation) and qualitatively. This means that it behaves like a random walk and grows like the square root of time. This section shows how the same behavior can be achieved for symmetric multistep methods.

2.1. Separation into difference equations for position and momentum

For consistent multistep methods (4) the characteristic polynomal $\rho(\zeta)$ of the coefficients α_j has a double zero $\zeta = 1$. In the limit $h \to 0$, the solution of the difference equation (4) is unbounded and this fact provokes an undesired accumulation of round-off errors. There are two possibilities to avoid this weak instability. Either one works with sums of f_j values (*summed form* of [7, Sect. 6.4-1]) or with differences of q_j values (*stabilized algorithm* of [6, Sect. III.10]). We use the second approach, because it is closer to the standard use of the Rattle algorithm.

For the difference of two consecutive q_j values we introduce momentum approximations on a staggered grid. Denoting by $\hat{\alpha}_j$ the coefficients of the polynomial $\rho(\zeta)/(\zeta - 1)$, i.e., $\alpha_k = \hat{\alpha}_{k-1}$ and $\alpha_j = \hat{\alpha}_{j-1} - \hat{\alpha}_j$ for $j = 1, \dots, k-1$, the method (4) is mathematically equivalent to

$$\begin{aligned} \sum_{j=0}^{k-1} \hat{\alpha}_j\, p_{n+j+1/2} &= h \sum_{j=1}^{k-1} \beta_j \left(f(q_{n+j}) - G(q_{n+j})^{\mathsf{T}} \lambda_{n+j} \right) \\ q_{n+k} &= q_{n+k-1} + h\, M^{-1} p_{n+k-1/2} \\ 0 &= g(q_{n+k}). \end{aligned} \tag{6}$$

Concerning the propagation of round-off errors there is a big difference to (4), because in the

limit $h \to 0$ the two difference equations (for position and momentum) have bounded solutions. The approximation of the momenta can be expressed in terms of $p_{n+j+1/2}$ as

$$p_n = \sum_{j=-l}^{l-1} \hat{\delta}_j \, p_{n+j+1/2} + h \, G(q_n)^{\mathsf{T}} \mu_n, \qquad G(q_n) M^{-1} p_n = 0, \tag{7}$$

where the coefficients $\hat{\delta}_j$ are given by $\hat{\delta}_{l-1} = \delta_l$ and $\delta_j = \hat{\delta}_{j-1} - \hat{\delta}_j$ for $j = -l+1, \ldots, l-1$. Compared to (5) this formula for p_n is less affected by round-off errors, because the difference operator approximates a function and not a derivative. This reformulation is less important than the previous one, because p_n is not used in the step by step application of the method.

2.2. Use of symmetric and rational coefficients

For symmetric methods (4) the coefficients satisfy $\alpha_{k-j} = \alpha_j$ and $\beta_{k-j} = \beta_j$, which implies that $\hat{\alpha}_{k-1-j} = -\hat{\alpha}_j$. To retain this symmetry in the implementation, it is recommended to apply the formula for $p_{n+k-1/2}$ as follows (here for $k = 4$):

$$p_{n+7/2} = p_{n+1/2} + \frac{1}{\hat{\alpha}_0}\Big(\hat{\alpha}_1 (p_{n+3/2} - p_{n+5/2}) + h\left(\beta_1 (f_{n+1} + f_{n+3}) + \beta_2 f_{n+2}\right)\Big),$$

where f_n is an abbreviation for $f(q_n) - G(q_n)^{\mathsf{T}} \lambda_n$. If the coefficients of the method are rational (the typical situation), they should be multiplied by a factor, so that all $\hat{\alpha}_j$ and β_j become integers. This avoids round-off errors in the computation of the coefficients.

2.3. Compensated summation

The iterative application of the second equation of (6) corresponds to a sum of small quantities. This is precisely the situation where Kahan's compensated summation can significantly reduce round-off errors, see [4, Sect. VIII.5]. The effect of compensated summation is the same as if the addition would be done with higher precision arithmetic.

It turns out that an application of compensated summation to the second equation of (6) alone has only marginal effect to the propagation of round-off. One also has to reduce round-off in the computation of the first equation of (6). This is less obvious, because the recursion is not as simple. We introduce a variable e that accumulates small errors in q_{n+k}, and further variables $e_{1/2}, \ldots, e_{k-1/2}$ for accumulating small errors in k consecutive approximations $p_{n+j+1/2}$. The proposed algorithm reads then as follows (to simplify the notation, we assume $\hat{\alpha}_0 = 1$):

for $n = 0, 1, 2, \ldots$ do
  $s_1 = h\,(\sum_{j=1}^{k/2-1} \beta_j \, (f_{n+k-j} + f_{n+j}) + \beta_{k/2} \, f_{n+k/2})$
  $s_2 = -\sum_{j=1}^{k/2-1} \hat{\alpha}_j \, (p_{n+k-j-1/2} - p_{n+j+1/2})$
  $d = -\sum_{j=1}^{k/2-1} \hat{\alpha}_j \, (e_{k-j-1/2} - e_{j+1/2})$
  $a = p_{n+1/2}$
  $e_{k-1/2} = s_1 + s_2 + d + e_{1/2}$
  $p_{n+k-1/2} = a + e_{k-1/2}$
  $e_{k-1/2} = (a - p_{n+k-1/2}) + e_{k-1/2}$
  for $j = 1, \ldots, k-1$ do
    $e_{j-1/2} = e_{j+1/2}$
  end do
  $b = q_{n+k-1}$

$e = hp_{n+k-1/2} + e$
$q_{n+k} = b + e$
$e = (b - q_{n+k}) + e$
end do

The effect of this implementation of symmetric multistep methods, applied to Hamiltonian systems without constraints, will be shown in Section 3.1. Since the force evaluations and the solution of the nonlinear equation for λ_{n+k-1} are not modified, the additional overhead is negligible. For problems with constraints, it is not sufficient to apply compensated summation to the first two relations of (6). One has to improve also the computation of the solution of the algebraic equation $g(q_{n+k}) = 0$.

2.4. Solving the nonlinear equation for the Lagrange multiplier

Most parts of (6) are explicit computations, only λ_{n+k-1} is given by an implicit equation. The first relation of (6) yields $p_{n+k-1/2} = a - h\beta_{k-1}G(q_{n+k-1})^{\mathsf{T}}\lambda_{n+k-1}/\hat{\alpha}_0$. Inserted into the second relation and then into the third one gives the nonlinear equation

$$g\Big(b - M^{-1}G(q_{n+k-1})^{\mathsf{T}}\lambda\Big) = 0,$$

where $b = q_{n+k-1} + h\,M^{-1}a$, and $\lambda = h^2\beta_{k-1}\lambda_{n+k-1}/\hat{\alpha}_0$ is the vector needed for the computation of $hp_{n+k-1/2}$ and q_{n+k}. This nonlinear equation can be solved with simplified Newton iterations

$$\lambda^{(i+1)} = \lambda^{(i)} + \Delta\lambda^{(i)}, \qquad G(q_{n+k-1})M^{-1}G(q_{n+k-1})^{\mathsf{T}}\Delta\lambda^{(i)} = g\Big(b - M^{-1}G(q_{n+k-1})^{\mathsf{T}}\lambda^{(i)}\Big)$$

starting with $\lambda^{(0)}$ that is obtained from information of the previous step. The stopping criterion has to be chosen carefully. From the implementation of implicit Runge–Kutta methods [5] we know that a criterion like $\|\Delta\lambda^{(i)}\| \le tol$ with tol close to the round-off unit can lead to a linear growth in the energy error. This also happens with symmetric multistep methods (see Section 3.2). We therefore follow the idea of [5] and apply *iteration until convergence*. This means that we iterate until either $\Delta\lambda^{(i)} = 0$ or $\|\Delta\lambda^{(i)}\| \ge \|\Delta\lambda^{(i-1)}\|$ which indicates that the increments of the iteration start to oscillate due to round-off. For up-dating the vectors $hp_{n+k-1/2}$ and q_{n+k} we use the approximation $\lambda^{(i)}$.

2.5. Accurate evaluation of the constraint functions

Implementing the ideas of the previous sections does not give satisfactory results for constrained Hamiltonian systems. The reason is that for the computation of $\Delta\lambda^{(i)}$ one has to evaluate $g(q)$ for an argument such that the result is close to zero. This provokes cancellation of significant digits. As a remedy we use the fact that from the algorithm of Section 2.3 we obtain not only the approximation q_{n+k}, but also its accumulated error e. The idea is to compute $g(q_{n+k} + e)$ with higher precision as follows:

- either, convert q_{n+k} (only in the subroutine for $g(q)$) into quadruple precision, add the small quantity e, evaluate $g(q_{n+k}+e)$ in quadruple precision, and return the result in double precision;
- or, exploit the form of the constraint to get higher precision. For the important special case of a quadratic constraint $g(q) = q_1^2 + q_2^2 - 1$, one can approximate $q_1 \approx k_1/k$, $q_2 = k_2/k$ by rational numbers, compute the errors $d_i = (kq_i - k_i) + ke_i$, and evaluate the constraint as

$$g(q+e) = \frac{1}{k^2}\Big((k_1^2 + k_2^2 - k^2) + 2\,(k_1d_1 + k_2d_2) + d_1^2 + d_2^2\Big)$$

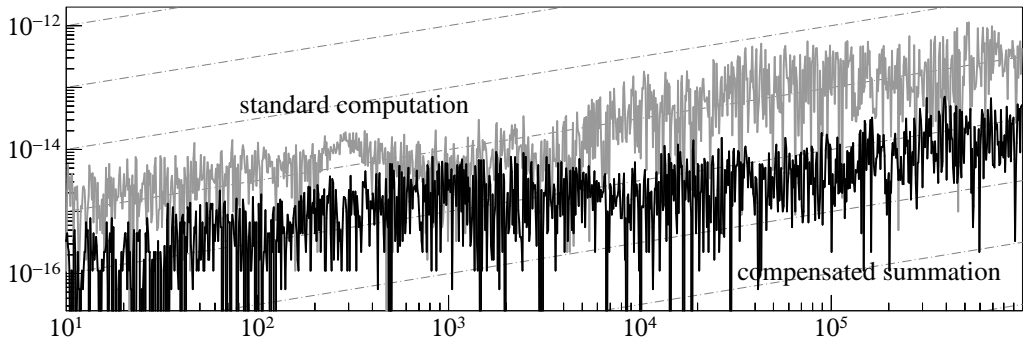

Figure 1: (Mathematical pendulum). Error in the Hamiltonian of the 8th order symmetric multistep method SY8 as function of time without (grey) and with (black) compensated summation; step size $h = 0.01$.

In this way, the dangerous subtraction is done with integers and thus without introducing round-off errors.

We shall present in Section 3.2 a numerical experiment illustrating the second approach, which avoids the computation with quadruple precision and is therefore more efficient.

3. Numerical experiments

The aim of this section is to illustrate the effect of the algorithms described in the previous section. This can best be done with simple examples, where further sources of round-off errors are minimized.

3.1. Effect of compensated summation

To illustrate the effect of the algorithm of Section 2.3 we consider a Hamiltonian system without constraints. In this case the method is explicit, no nonlinear equation has to be solved. We consider the mathematical pendulum

$$H(p, q) = \frac{1}{2}\, p^2 - \cos q$$

and apply the 8th order symmetric linear multistep method SY8 of Quinlan and Tremaine [9] (see also [4, page 603]) with step size $h = 0.01$. Accurate starting approximations are computed with a high order implicit Runge–Kutta method. The step size is small enough, so that the discretization error of the linear multistep method is below round-off.

Figure 1 shows the error in the Hamiltonian as a function of time in double logarithmic scale. The use of compensated summation reduces this error by more than a factor 10. The broken grey lines in the figure are of slope $1/2$ and indicate a growth that is proportional to $t^{1/2}$. For the standard implementation as well as for that with compensated summation the error increases not worse than $ct^{1/2}$, which indicates a random walk behavior.

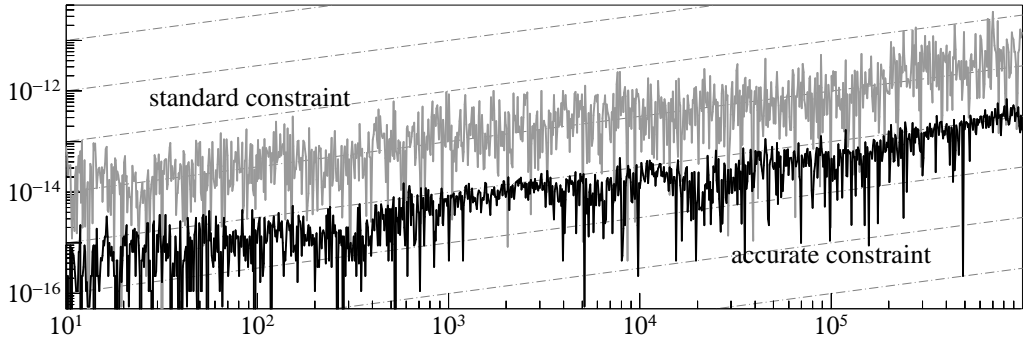

Figure 2: (Two-body problem on the sphere). Error in the Hamiltonian of an 8th order symmetric multistep method as function of time without (grey) and with (black) the technique of Section 2.5; step size $h = 0.001$; iteration until convergence (Section 2.4) is applied.

3.2. Improving round-off by solving accurately the constraints

The algorithms of Sections 2.4 and 2.5 concern the treatment of the algebraic constraint in the problem (1). We were surprised to see that without the ideas of Section 2.5 the stopping criterion of the simplified Newton iterations for the computation of λ_{n+k-1} has only little effect on the roundoff error. The algorithm of Section 2.5, however, has a big influence.

As example we consider the two-body problem on the sphere. It describes the motion of two particles on the unit sphere respecting the force given by the potential (see [8])

$$U(q) = -\frac{\cos\theta}{\sin\theta}, \qquad \cos\theta = \langle Q_1, Q_2 \rangle,$$

where $q \in \mathbb{R}^6$ collects the position coordinates of the two particles $Q_1 = (q_1, q_2, q_3)^\mathsf{T}$ and $Q_2 = (q_4, q_5, q_6)^\mathsf{T}$. The constraints are quadratic

$$g_1(q) = Q_1^\mathsf{T} Q_1 - 1, \qquad g_2(q) = Q_2^\mathsf{T} Q_2 - 1$$

and ideally suited for the algorithm of Section 2.5. For our numerical experiment we use initial values as in [2], but the qualitative behavior is independent of them. We apply the symmetric linear multistep method with generating polynomials

$$\rho(\zeta) = (\zeta - 1)^2 \prod_{j=1}^{3} (\zeta^2 + 2a_j\zeta + 1)$$

($a_1 = -0.8$, $a_2 = -0.4$, $a_3 = 0.7$), and $\sigma(\zeta)$ such that the method is explicit and of order 8. Starting approximations are computed using a high order implicit Runge–Kutta method.

Figure 2 shows the error in the Hamiltonian along the numerical solution for two implementations, both with step size $h = 0.001$, so that the truncation error is below round-off. The error indicated by "standard constraint" corresponds to an implementation, where all the improvements of Sections 2.1 to 2.4 are taken into account. The error indicated by "accurate constraint" uses in addition the technique of Section 2.5. We observe an enormous improvement with little additional work. The more accurate computation of $g(q)$ needs for a few steps an additional iteration for the solution of the nonlinear equation. For example, on an interval of length 10^3 with step size $h = 0.001$, we have 10^6 force evaluations and 4 608 497 evaluations of $g(q)$ for the

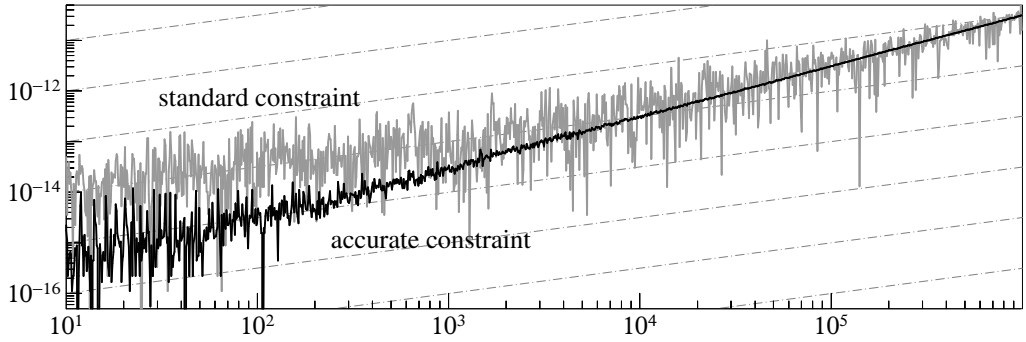

Figure 3: (Two-body problem on the sphere). Error in the Hamiltonian of an 8th order symmetric multistep method as function of time without (grey) and with (black) the technique of Section 2.5; step size $h = 0.001$; stopping criterion (Section 2.4) is $\|\Delta\lambda^{(i)}\| \leq tol$ with $tol = 10^{-15}$.

standard implementation, and 4 818 860 evaluations of $g(q)$ for the "accurate constraint" implementation. The cpu time with the Intel Fortran complier compiler IFORT on a MacBook Pro is 2.27 sec for the standard and 2.34 sec for the accurate implementation. This experiment shows that an accurate computation of the Lagrange multiplier from the algebraic constraint is essential for a reduction of round-off effects.

Our second experiment (Figure 3) is very similar to the one for Figure 2. The only difference is that instead of using the stopping criterion "iteration until convergence" (Section 2.4) we stop the iteration as soon as $\|\Delta\lambda^{(i)}\| \leq tol$ with tol as small as possible. For $tol = 10^{-16}$ we had difficulties with convergence, so that we have taken $tol = 10^{-15}$. For the "standard constraint" implementation we cannot observe a significant difference between the results of Figures 2 and 3. For the "accurate constraint" implementation, however, there is an enormous difference. Whereas the error behaves like a random walk in Figure 2, it grows linearly with time in Figure 3.

Let us mention that a direct implementation of the multistep method, based on the formulation (6) and exploiting the symmetry of the coefficients (without compensated summation and without the techniques of Sections 2.4 and 2.5) yields an error in the Hamiltonian that is nearly identical to the "standard constraint" error in Figure 3. For this reason we did not include it in our figures. Our conclusion is that all improvements discussed in this note (compensated summation, careful stopping criterion, accurate evaluation of the constraint) should be considered in the an implementation. If one of them is omitted, a significant loss in accuracy will be the consequence.

We have made further experiments with Hamiltonian problems that are subject to quadratic constraints. For example, we have considered a triple coupled pendulum in dimension $d = 8$ with $m = 5$ constraints. The results with our code are qualitatively the same as those presented here. For non-quadratic constraints the technique of Section 2.5 has to modified suitably.

Acknowledgement

This work was partially supported by the Fonds National Suisse, projects No. 200020-144313 and No. 200021-129485.

References

[1] H.C. Andersen, Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations, J. Comput. Phys. 52 (1983) 24–34.

[2] P. Console, E. Hairer, C. Lubich, Symmetric multistep methods for constrained Hamiltonian systems, Numerische Mathematik (2013) ?–?
[3] E. Hairer, C. Lubich, Symmetric multistep methods over long times, Numer. Math. 97 (2004) 699–723.
[4] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations, Springer Series in Computational Mathematics 31, Springer-Verlag, Berlin, 2nd edition, 2006.
[5] E. Hairer, R.I. McLachlan, A. Razakarivony, Achieving Brouwer's law with implicit Runge-Kutta methods, BIT 48 (2008) 231–243.
[6] E. Hairer, S.P. Nørsett, G. Wanner, Solving Ordinary Differential Equations I. Nonstiff Problems, Springer Series in Computational Mathematics 8, Springer, Berlin, 2nd edition, 1993.
[7] P. Henrici, Discrete Variable Methods in Ordinary Differential Equations, John Wiley & Sons Inc., New York, 1962.
[8] V.V. Kozlov, A.O. Harin, Kepler's problem in constant curvature spaces, Celestial Mech. Dynam. Astronom. 54 (1992) 393–399.
[9] G.D. Quinlan, S. Tremaine, Symmetric multistep methods for the numerical integration of planetary orbits, Astron. J. 100 (1990) 1694–1700.