Geometric Numerical Integration — TU München
Ernst Hairer — January – February 2010

Lecture 3: Backward error analysis

Table of contents

Backward error analysis is the most powerful tool for the study of the long-time behaviour of numerical integrators.

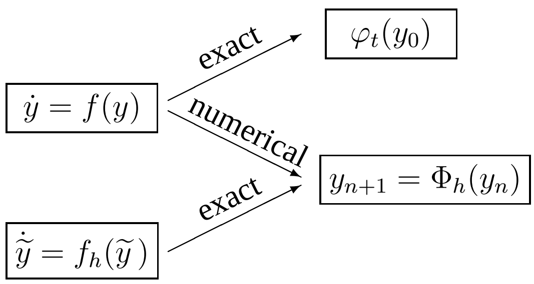
1 Modified differential equation

Consider an ordinary differential equation

$$\dot y = f(y),$$

and a numerical method $\Phi_h(y)$ which produces the approximations

$$y_0, y_1, y_2, \ldots \, .$$



A forward error analysis consists of the study of the errors $y_1 - \varphi_h(y_0)$ (local error) and $y_n - \varphi_{nh}(y_0)$ (global error) in the solution space. The idea of backward error analysis is to search for a *modified differential equation* $\dot{\widetilde y} = f_h(\widetilde y\,)$ of the form

$$\dot{\widetilde y} = f(\widetilde y\,) + h f_2(\widetilde y\,) + h^2 f_3(\widetilde y\,) + \ldots , \qquad (1)$$

such that $y_n = \widetilde y(nh)$, and to study the difference of the vector fields $f(y)$ and $f_h(y)$. We remark that the series in (1) usually diverges and that one has to truncate it suitably for a rigorous analysis. For the moment we content ourselves with a formal analysis without taking care of convergence issues.

For the computation of the modified equation (1) we put $y := \widetilde{y}(t)$ for a fixed t, and we expand the solution of (1) into a Taylor series

$$\begin{aligned} \widetilde{y}(t+h) &= y + h\big(f(y) + hf_2(y) + h^2 f_3(y) + \ldots\big) \\ &+ \frac{h^2}{2!}\big(f'(y) + hf_2'(y) + \ldots\big)\big(f(y) + hf_2(y) + \ldots\big) + \ldots . \end{aligned} \tag{2}$$

We assume that the numerical method $\Phi_h(y)$ can be expanded as

$$\Phi_h(y) = y + hf(y) + h^2 d_2(y) + h^3 d_3(y) + \ldots \tag{3}$$

(the coefficient of h is $f(y)$ for consistent methods). The functions $d_j(y)$ are known and are typically composed of $f(y)$ and its derivatives. To get $\widetilde{y}(nh) = y_n$ for all n, we must have $\widetilde{y}(t+h) = \Phi_h(y)$. Comparing like powers of h in the expressions (2) and (3) yields recurrence relations for the functions $f_j(y)$:

$$\begin{aligned} f_2(y) &= d_2(y) - \frac{1}{2!} f'f(y) \\ f_3(y) &= d_3(y) - \frac{1}{3!}\Big(f''(f,f)(y) + f'f'f(y)\Big) - \frac{1}{2!}\Big(f'f_2(y) + f_2'f(y)\Big). \end{aligned} \tag{4}$$

Example 1 Consider the scalar differential equation $\dot{y} = y^2$, $y(0) = 1$ with exact solution $y(t) = 1/(1-t)$. It has a singularity at $t = 1$. We apply the explicit Euler method $y_{n+1} = y_n + hf(y_n)$ with step size $h = 0.02$. The above procedure for the computation of the modified equation is implemented as a Maple script

```
> fcn := y -> y^2:
> nn := 6:
> fcoe[1] := fcn(y):
> for n from 2 by 1 to nn do
>    modeq := sum(h^j*fcoe[j+1], j=0..n-2):
>    diffy[0] := y:
>    for i from 1 by 1 to n do
>       diffy[i] := diff(diffy[i-1],y)*modeq:
>    od:
>    ytilde := sum(h^k*diffy[k]/k!, k=0..n):
>    res := ytilde-y-h*fcn(y):
>    tay := convert(series(res,h=0,n+1),polynom):
>    fcoe[n] := -coeff(tay,h,n):
> od:
> simplify(sum(h^j*fcoe[j+1], j=0..nn-1));
```

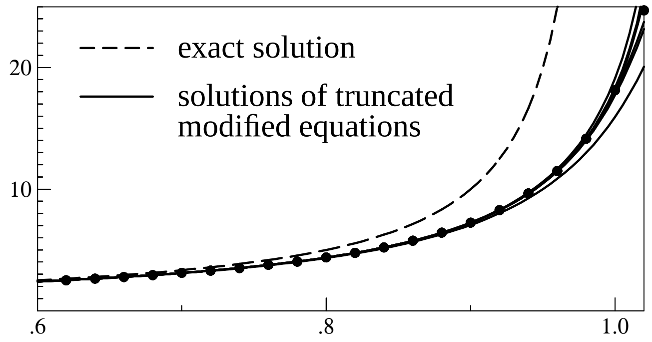


Figure 1: Solutions of the modified equation for the problem $\dot{y} = y^2$, $y(0) = 1$.

Its output is

$$\dot{\widetilde{y}} = \widetilde{y}^2 - h\widetilde{y}^3 + h^2\,\frac{3}{2}\,\widetilde{y}^4 - h^3\,\frac{8}{3}\,\widetilde{y}^5 + h^4\,\frac{31}{6}\,\widetilde{y}^6 - h^5\,\frac{157}{15}\,\widetilde{y}^7 \pm \ldots\,. \tag{5}$$

Figure 1 presents the exact solution (dashed curve), the numerical solution (thick dots), and the solution of the modified equation, when truncated after 1, 2, 3, and 4 terms. We observe an excellent agreement of the numerical solution with the exact solution of the modified equation.

A similar program for the implicit midpoint rule computes the modified equation

$$\dot{\widetilde{y}} = \widetilde{y}^2 + h^2\,\frac{1}{4}\,\widetilde{y}^4 + h^4\,\frac{1}{8}\,\widetilde{y}^6 + h^6\,\frac{11}{192}\,\widetilde{y}^8 + h^8\,\frac{3}{128}\,\widetilde{y}^{10} \pm \ldots\,, \tag{6}$$

and for the classical (explicit) Runge–Kutta method of order 4

$$\dot{\widetilde{y}} = \widetilde{y}^2 - h^4\,\frac{1}{24}\,\widetilde{y}^6 + h^6\,\frac{65}{576}\,\widetilde{y}^8 - h^7\,\frac{17}{96}\,\widetilde{y}^9 + h^8\,\frac{19}{144}\,\widetilde{y}^{10} \pm \ldots\,. \tag{7}$$

We observe that the perturbation terms in the modified equation are of size $\mathcal{O}(h^r)$, where r is the order of the method. This is true in general.

Theorem 1 *Suppose that the method $y_{n+1} = \Phi_h(y_n)$ is of order r, i.e.,*

$$\Phi_h(y) = \varphi_h(y) + h^{r+1}\delta_{r+1}(y) + \mathcal{O}(h^{r+2}),$$

where $\varphi_t(y)$ denotes the exact flow of $\dot{y} = f(y)$, and $h^{r+1}\delta_{r+1}(y)$ the leading term of the local truncation error. The modified equation then satisfies

$$\dot{\widetilde{y}} = f(\widetilde{y}\,) + h^r f_{r+1}(\widetilde{y}\,) + h^{r+1} f_{r+2}(\widetilde{y}\,) + \ldots, \qquad \widetilde{y}(0) = y_0 \tag{8}$$

with $f_{r+1}(y) = \delta_{r+1}(y)$.

Proof. The construction of the functions $f_j(y)$ (see the beginning of this section) shows that $f_j(y) = 0$ for $2 \le j \le r$ if and only if $\Phi_h(y) - \varphi_h(y) = \mathcal{O}(h^{r+1})$. □

Example 2 We next consider the Lotka–Volterra equations

$$\dot{q} = q(p-1), \qquad \dot{p} = p(2-q),$$

and we apply (a) the explicit Euler method, and (b) the symplectic Euler method, both with constant step size $h = 0.1$. The first terms of their modified equations are

$$\begin{aligned} \text{(a)} \qquad \dot{q} &= q(p-1) - \frac{h}{2}\, q(p^2 - pq + 1) + \mathcal{O}(h^2), \\ \dot{p} &= -p(q-2) - \frac{h}{2}\, p(q^2 - pq - 3q + 4) + \mathcal{O}(h^2), \\ \text{(b)} \qquad \dot{q} &= q(p-1) - \frac{h}{2}\, q(p^2 + pq - 4p + 1) + \mathcal{O}(h^2), \\ \dot{p} &= -p(q-2) + \frac{h}{2}\, p(q^2 + pq - 5q + 4) + \mathcal{O}(h^2). \end{aligned}$$

Figure 2 shows the numerical solutions for initial values indicated by a thick dot. In the pictures to the left they are embedded in the exact flow of the differential equation, in those to the right they are embedded in the flow of the modified differential equation, truncated after the h^2 terms. For the symplectic Euler method, the solutions of the truncated modified equation are periodic, as is the case for the unperturbed problem.

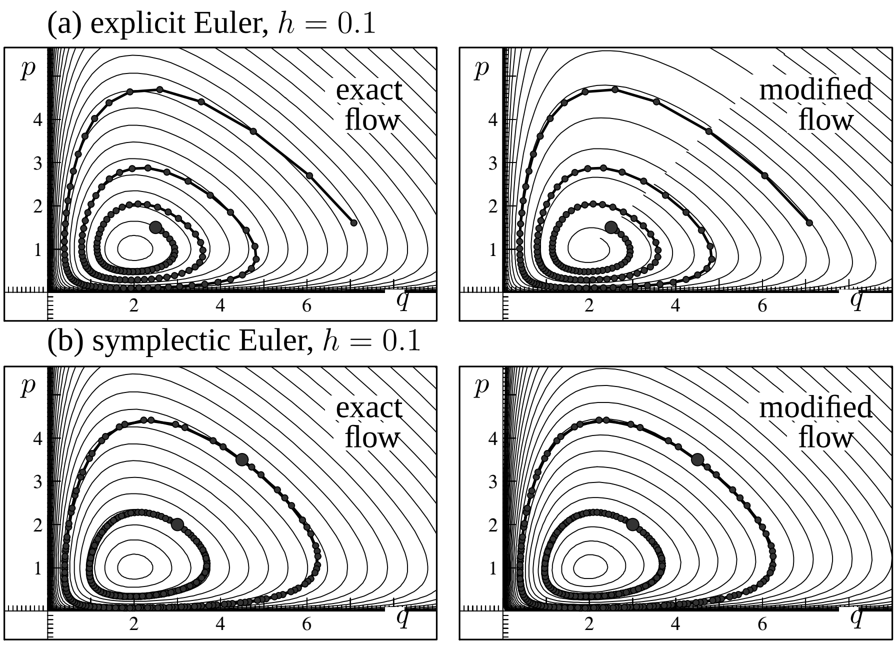


Figure 2: Numerical solution compared to the exact and modified flows

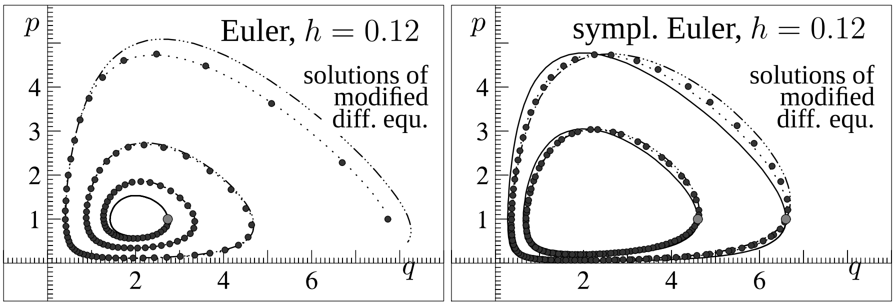


Figure 3: Study of the truncation in the modified equation

In Figure 3 we present the numerical solution and the exact solution of the modified equation, once truncated after the h terms (dashed-dotted), and once truncated after the h^2 terms (dotted). The exact solution of the problem is included as a solid curve.

Example 3 For a linear differential equation with constant coefficients

$$\dot{y} = Ay, \qquad y(0) = y_0$$

we consider numerical methods which yield $y_{n+1} = R(hA)y_n$, where $R(z)$ is the stability function of the method. In this case we get $y_n = R(hA)^n y_0$, so that $y_n = \widetilde{y}(nh)$, where $\widetilde{y}(t) = R(hA)^{t/h}y_0 = \exp\big(\frac{t}{h}\ln R(hA)\big)y_0$ is the solution of the modified differential equation

$$\dot{\widetilde{y}} = \frac{1}{h}\,\ln R(hA)\,\widetilde{y} = (A + hb_2A^2 + h^2b_3A^3 + \ldots)\,\widetilde{y} \tag{9}$$

with suitable constants $b_2, b_3, \ldots$. Since $R(z) = 1 + z + \mathcal{O}(z^2)$ and $\ln(1 + x) = x - x^2/2 + \mathcal{O}(x^3)$ both have a positive radius of convergence, the series (9) converges for $|h| < h_0$ with some $h_0 > 0$. This is an exceptional situation.

2 Modified Hamiltonian for symplectic integrators

We consider a Hamiltonian system $\dot{y} = J^{-1}\nabla H(y)$ with smooth $H(y)$, and we show that the modified equation of symplectic methods is also Hamiltonian.

Theorem 2 (Existence of a local modified Hamiltonian) *If a symplectic method $\Phi_h(y)$ is applied to a Hamiltonian system $\dot{y} = J^{-1}\nabla H(y)$ with a smooth Hamiltonian $H : U \to \mathbb{R}$, then the modified equation (1) is locally Hamiltonian.*

More precisely, for every $y_0 \in U$ and for all j there exist smooth functions $H_j(y)$, such that $f_j(y) = J^{-1}\nabla H_j(y)$ on a suitable neighborhood of y_0.

Proof. [1] [2] Assume that $f_j(y) = J^{-1}\nabla H_j(y)$ for $j = 1, 2, \ldots, r$ (this is satisfied for $r = 1$, because $f_1(y) = f(y) = J^{-1}\nabla H(y)$). We have to prove the existence of a Hamiltonian $H_{r+1}(y)$. The idea is to consider the truncated modified equation

$$\dot{\widetilde{y}} = f(\widetilde{y}) + hf_2(\widetilde{y}) + \ldots + h^{r-1}f_r(\widetilde{y}), \tag{10}$$

which is a Hamiltonian system with Hamiltonian $H(y)+hH_2(y)+\ldots+h^{r-1}H_r(y)$. Its flow $\varphi_{r,t}(y_0)$, compared to that of (1), satisfies

$$\Phi_h(y_0) = \varphi_{r,h}(y_0) + h^{r+1}f_{r+1}(y_0) + \mathcal{O}(h^{r+2}),$$

and also

$$\Phi'_h(y_0) = \varphi'_{r,h}(y_0) + h^{r+1}f'_{r+1}(y_0) + \mathcal{O}(h^{r+2}).$$

By our assumption on the method and by the induction hypothesis, Φ_h and $\varphi_{r,h}$ are symplectic transformations. Together with $\varphi'_{r,h}(y_0) = I + \mathcal{O}(h)$, this implies

$$J = \Phi'_h(y_0)^T J \Phi'_h(y_0) = J + h^{r+1}\Big(f'_{r+1}(y_0)^T J + J f'_{r+1}(y_0)\Big) + \mathcal{O}(h^{r+2}).$$

The matrix $Jf'_{r+1}(y)$ is therefore symmetric, and the (local) existence of $H_{r+1}(y)$ satisfying $f_{r+1}(y) = J^{-1}\nabla H_{r+1}(y)$ follows from the Integrability Lemma. □

The application of the Integrability Lemma shows that the modified Hamiltonian is globally defined on U, if $U = \mathbb{R}^{2d}$, or if U is star-shaped, or if U is simply connected.

3 Near conservation of the total energy

As a first application of backward error analysis we study the long-time energy conservation of a symplectic numerical scheme (of order r) applied to Hamiltonian systems $\dot{y} = J^{-1}\nabla H(y)$. It follows from Theorem 2 that the corresponding modified differential equation is also Hamiltonian. After truncation we get

$$H^{[N]}(y) \;=\; H(y) + h^r H_{r+1}(y) + \ldots + h^{N-1}H_N(y), \tag{11}$$

which we assume to be *globally* defined on the same open set as the original Hamiltonian $H(y)$.

---

[1] G. Benettin & A. Giorgilli, *On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms*, J. Statist. Phys. 74 (1994) 1117–1143.

[2] Y.-F. Tang, *Formal energy of a symplectic scheme for Hamiltonian systems and its applications (I)*, Computers Math. Applic. 27 (1994) 31–39.

Theorem 3 *Consider a Hamiltonian system with smooth $H : U \to \mathbb{R}$ (where $U \subset \mathbb{R}^{2d}$), apply a symplectic numerical method $\Phi_h(y)$ with step size h, and assume that its modified Hamiltonian is globally defined on U.*

If the numerical solution stays in the compact set $K \subset U$, then we have asymptotically for $h \to 0$

$$\begin{aligned} H^{[N]}(y_n) &= H^{[N]}(y_0) + \mathcal{O}(t\,h^N) \\ H(y_n) &= H(y_0) + \mathcal{O}(h^r) \end{aligned}$$

over time intervals of size $t = nh \le C\,h^{-N+r}$.

Proof. We let $\varphi_{N,t}(y_0)$ be the flow of the truncated modified equation. The effect of the truncation is that $\|y_{j+1} - \varphi_{N,h}(y_j)\| \le C\,h^{N+1}$. Since the truncated modified equation is Hamiltonian with $H^{[N]}(y)$ of (11), we have $H^{[N]}\big(\varphi_{N,t}(y_j)\big) = H^{[N]}(y_j)$ for all times t. Using a global h-independent Lipschitz constant for $H^{[N]}$ on the compact set K, the statement on the long-time conservation of $H^{[N]}$ is a consequence of

$$\begin{aligned} H^{[N]}(y_n) - H^{[N]}(y_0) &= \sum_{j=0}^{n-1} \Big(H^{[N]}(y_{j+1}) - H^{[N]}(y_j) \Big) \\ &= \sum_{j=0}^{n-1} \Big(H^{[N]}(y_{j+1}) - H^{[N]}\big(\varphi_{N,h}(y_j)\big) \Big) \;=\; \mathcal{O}(nh^{N+1}). \end{aligned}$$

The statement for the Hamiltonian H follows from (11), because $H_{r+1}(y) + \ldots + h^{N-r-1}H_N(y)$ is uniformly bounded on K independently of h and N. □

If the Hamiltonian and the integrator are analytic, it is possible to prove exponentially small error bounds on exponentially large time intervals. More precisely, for sufficiently small h there exists $N = N(h) \sim h^{-1}$ such that in the estimates of the theorem above h^N can be replaced by $\mathrm{e}^{-\gamma/h}$.

Example 4 The mathematical pendulum is a system with one degree of freedom having the Hamiltonian $H(p,q) = \frac{1}{2}\,p^2 - \cos q$. Since the Hamiltonian and the vector field are 2π-periodic in q, it is natural to consider q as a variable on the circle S^1. Hence, the phase space of points (p,q) becomes the cylinder $\mathbb{R} \times S^1$. Figure 4 shows some level curves of $H(p,q)$ together with numerical solutions.

Theorem 3 explains the near conservation of the Hamiltonian with the symplectic Euler method (existence of a global modified Hamiltonian will be discussed later). This is illustrated in Figure 5. The linear drift of the numerical Hamiltonian for non-symplectic methods can be explained by a computation

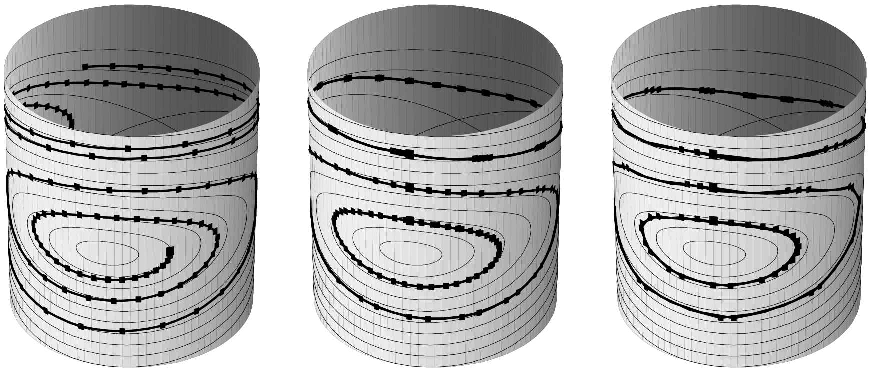


Figure 4: Solutions of the pendulum problem; explicit Euler with step size $h = 0.2$, initial value $(p_0, q_0) = (0, 0.5)$; symplectic Euler with $h = 0.3$ and initial values $q_0 = 0$, $p_0 = 0.7, 1.4, 2.1$; Störmer–Verlet with $h = 0.6$.

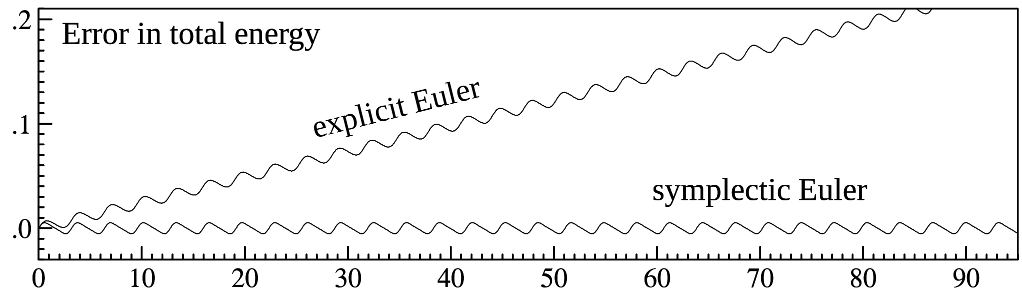


Figure 5: Error in the total energy for the explicit and symplectic Euler methods applied with step size $h = 0.005$ and initial value $(p_0, q_0) = (2.5, 0)$.

similar to that of the proof of Theorem 3. From a Lipschitz condition of the Hamiltonian and from the standard local error estimate, we obtain $H(y_{n+1}) - H(\varphi_h(y_n)) = \mathcal{O}(h^{r+1})$. Since $H(\varphi_h(y_n)) = H(y_n)$, a summation of these terms leads to

$$H(y_n) - H(y_0) = \mathcal{O}(th^r) \qquad \text{for } t = nh. \tag{12}$$

Example 5 In the numerical experiment of Figure 6 we study the effect of "large" step sizes for the pendulum problem. We have drawn $200\,000$ steps of the numerical solution of the implicit midpoint rule for various step sizes h and for initial values $(p_0, q_0) = (0, -1.5)$, $(p_0, q_0) = (0, -2.5)$, $(p_0, q_0) = (1.5, -\pi)$, and $(p_0, q_0) = (2.5, -\pi)$. They are compared to the contour lines of the truncated

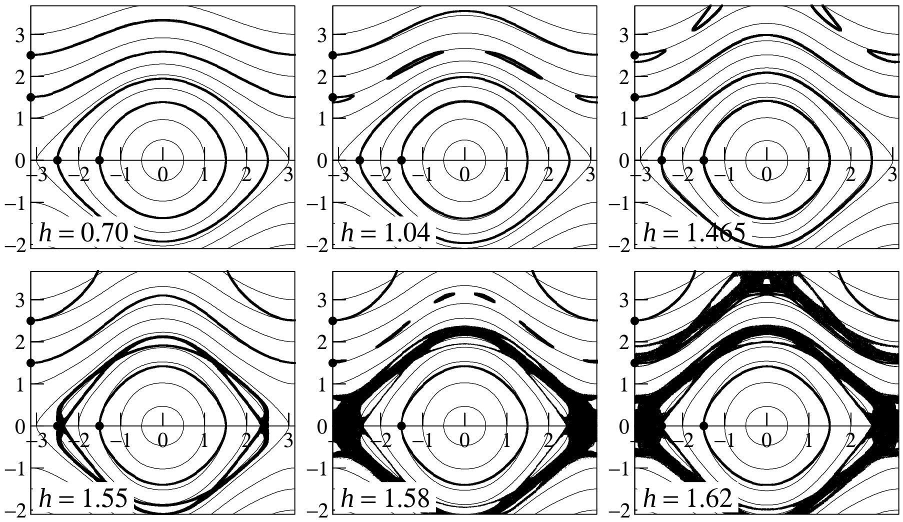


Figure 6: Numerical solutions of the implicit midpoint rule with large step sizes

modified Hamiltonian

$$H^{[4]}(p,q) = \frac{p^2}{2} - \cos q + \frac{h^2}{48}\Big(\cos(2q) - 2p^2\cos q\Big).$$

This shows that for step sizes as large as $h \le 0.7$ the Hamiltonian $H^{[4]}$ is extremely well conserved. Beyond this value, the dynamics of the numerical method soon turns into chaotic behaviour.

Example 6 (Modified Hamiltonian of the Störmer–Verlet method) We present the modified Hamiltonian for a separable Hamiltonian $H(p,q) = T(p) + U(q)$ with $T(p) = \frac{1}{2}\,p^{\mathsf T}p$. Using the notation $\{A,B\} = \nabla_q A^{\mathsf T}\nabla_p B - \nabla_p A^{\mathsf T}\nabla_q B$ for the Poisson bracket of two functions depending on p and q, it is given by

$$\begin{aligned}\widetilde H \;=\; & T + U + h^2\Big(\frac{1}{12}\big\{T,\{T,U\}\big\} - \frac{1}{24}\big\{U,\{U,T\}\big\}\Big)\\ & + h^4\Big(-\frac{1}{720}\big\{T,\big\{T,\big\{T,\{T,U\}\big\}\big\}\big\} + \frac{1}{360}\big\{U,\big\{T,\big\{T,\{T,U\}\big\}\big\}\big\}\\ & \quad -\frac{1}{480}\big\{U,\big\{U,\big\{T,\{T,U\}\big\}\big\}\big\} + \frac{1}{120}\big\{T,\big\{T,\big\{U,\{U,T\}\big\}\big\}\big\}\Big) + \mathcal{O}(h^6).\end{aligned}$$

This modified Hamiltonian can be computed with the recursive procedure explained in the beginning of this lecture. A more elegant derivation is by the use of the symmetric Baker–Campbell–Hausdorff formula (cf. Sect. III.4.2 of the monograph on “Geometric Numerical Integration”). Since $\widetilde H(p,q)$ it is composed of derivatives of T and U, it is globally defined.

4 Counter-example: Takahashi–Imada integrator

We consider a second order differential equation $\ddot{q} = f(q)$, and we assume that it is Hamiltonian, i.e., $f(q) = -\nabla U(q)$. Takahashi & Imada[3] noticed that the accuracy of the Störmer–Verlet discretization is greatly improved by adding the expression $\alpha h^2 f'(q) f(q)$ with $\alpha = \frac{1}{12}$ to every force evaluation. This yields

$$\begin{aligned} p_{n+1/2} &= p_n + \frac{h}{2}\big(I + \alpha h^2 f'(q_n)\big) f(q_n) \\ q_{n+1} &= q_n + h\, p_{n+1/2} \\ p_{n+1} &= p_{n+1/2} + \frac{h}{2}\big(I + \alpha h^2 f'(q_{n+1})\big) f(q_{n+1}), \end{aligned} \tag{13}$$

which is still symplectic, because for $f(q) = -\nabla U(q)$ it can be interpreted as applying the Störmer–Verlet method to the Hamiltonian system with potential

$$V(q) = U(q) - \frac{\alpha}{2}\, h^2\, \|\nabla U(q)\|^2.$$

Simplified Takahashi–Imada method. To avoid the derivative evaluation of the vector field $f(q)$, which corresponds to a Hessian evaluation for $f(q) = -\nabla U(q)$, we replace $\big(I + \alpha h^2 f'(q)\big) f(q)$ with $f\big(q + \alpha h^2 f(q)\big)$ and thus consider[4]

$$\begin{aligned} p_{n+1/2} &= p_n + \frac{h}{2}\, f\big(q_n + \alpha h^2 f(q_n)\big) \\ q_{n+1} &= q_n + h\, p_{n+1/2} \\ p_{n+1} &= p_{n+1/2} + \frac{h}{2}\, f\big(q_{n+1} + \alpha h^2 f(q_{n+1})\big) \end{aligned} \tag{14}$$

which is a $\mathcal{O}(h^5)$ perturbation of the method (13). The method is volume preserving and thus symplectic for problems with one degree of freedom. To see this, note that (14) is a composition of shears (mappings of the form $(p,q) \mapsto (p + a(q), q)$ and $(p,q) \mapsto (p, q + hp)$), and shears always preserve the volume.

Modified differential equation. Substituting in the modified Hamiltonian of Example 6 the potential U with $U - \frac{\alpha}{2} h^2 \|\nabla U\|^2 = U - \frac{\alpha}{2} h^2 \big\{U, \{U, T\}\big\}$ yields the modified Hamiltonian of the symplectic Takahashi–Imada integrator (13):

$$\widehat{H}(p,q) = H(p,q) + h^2 H_3(p,q) + h^4 H_5(p,q)$$

[3] M. Takahashi and M. Imada. *Monte Carlo calculation of quantum systems. II. Higher order correction.* J. Phys. Soc. Jpn., 53:3765–3769, 1984.

[4] J. Wisdom, M. Holman, and J. Touma. *Symplectic correctors.* In J. E. Marsden, G. W. Patrick, and W. F. Shadwick, editors, *Integration Algorithms and Classical Mechanics*, pages 217–244. Amer. Math. Soc., Providence R. I., 1996.

where (for one degree of freedom)

$$H_3(p,q) = \frac{1}{12}\,p^2\,U_{qq}(q) - \Big(\frac{1}{24}+\frac{\alpha}{2}\Big)\,U_q(q)U_q(q) \tag{15}$$

$$H_5(p,q) = -\frac{1}{720}\,p^4\,U_{qqqq}(q) - \frac{1}{120}\,p^2\,U_q(q)U_{qqq}(q) + \ldots \tag{16}$$

Since the method (14) differs from (13) only by a $\mathcal{O}(h^5)$ perturbation, we find that the modified differential equation of (14) satisfies

$$\begin{aligned} \dot p &= -\nabla_q \widehat{H}(p,q) + \frac{1}{2}\,h^4\alpha^2\,\big(f''(f,f)\big)(q) + \mathcal{O}(h^6) \\ \dot q &= \nabla_p \widehat{H}(p,q) + \mathcal{O}(h^6) \end{aligned} \tag{17}$$

Notice that for potentials $U(q)$ that are 2π-periodic, the function $\widehat{H}(p,q)$ is also 2π-periodic in q, but the integral of the perturbation need not be 2π-periodic.

To study energy conservation of the simplified Takahashi–Imada integrator (14) we notice that along the (formal) exact solution of its modified differential equation (17) we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\widehat{H}(p,q) = \frac{1}{2}\,h^4\alpha^2\,p^{\mathsf{T}} f''(f,f) + \mathcal{O}(h^6) = -\frac{1}{2}\,h^4\alpha^2\,U_{qqq}(p,U_q,U_q) + \mathcal{O}(h^6), \tag{18}$$

where we have used $f(q) = -\nabla U(q) = -U_q(q)$. In general, the expression on the right-hand side of (18) cannot be written as a total differential. However, this formula yields much insight into the long-time energy conservation of (14).[5]

- *Bounded error in the energy without any drift.* For problems with one degree of freedom the right-hand side of (18) can be written as $\frac{d}{dt}F\big(q(t)\big)$ with $F'(q) = U_{qqq}(q)U_q(q)^2$ (neglecting terms of size $\mathcal{O}(h^6)$). If $U(q)$ is T-periodic in q, and $\int_0^T U_{qqq}(q)U_q(q)^2 dq = 0$, then the antiderivative (or indefinite integral) $F(q)$ of $U_{qqq}(q)U_q(q)^2$ is globally defined on the circle S^1. This is the situation for the pendulum, where $U(q) = -\cos q$. We expect that also higher order terms can be treated in this way, so that no energy drift can be observed in this situation (see Figure 7, where $h = 0.2$, and initial values $q(0) = 0$, $p(0) = 2.5$ are used).
- *Linear energy drift.* In general, without any particular assumption on the potential, the numerical energy will have a drift of size $\mathcal{O}(th^4)$. E.g., for a non-symmetrically perturbed pendulum with $U(q) = -\cos q + 0.2\,\sin(2q)$, for which $\int_0^{2\pi} U_{qqq}(q)U_q(q)^2 dq \approx 3.77$ (see Figure 7).

[5] E. Hairer, R.I. McLachlan, and R.D. Skeel. *On energy conservation of the simplified Takahashi–Imada method.* ESAIM: M2AN 43:631644, 2009.

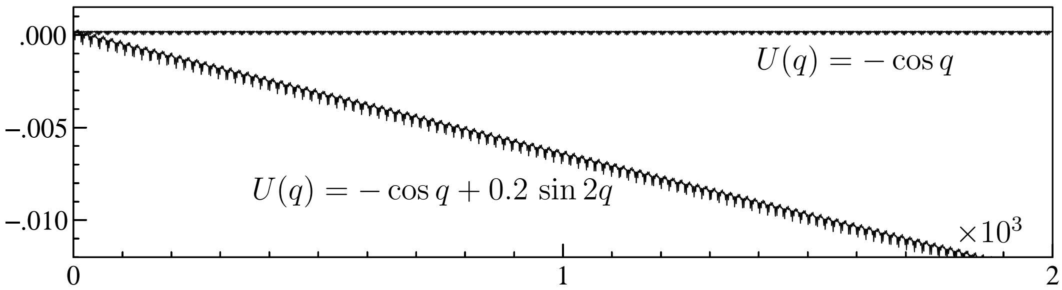


Figure 7: Error in the shifted modified energy $H(p_n, q_n) + h^2\, H_3(p, q)$ along the numerical solution of the simplified Takahashi–Imada method.

- *Random walk behavior – drift like square root of time.* Under the assumption that the solution of the modified differential equation is ergodic on an invariant set A with respect to an invariant measure μ, we have

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t F\big(p(s), q(s)\big)\,\mathrm{d}s = \int_A F(x)\,\mu(\mathrm{d}x), \tag{19}$$

where $x = (p, q)$ and the function $F(x)$ is the right-hand side of (18). Again there will be a linear drift of size $\mathcal{O}(th^4)$ in general. However, in the presence of symmetries, the integral of the right-hand side in (19) is likely to vanish and the numerical Hamiltonian will look like a random walk, so that there will be a drift of size $\mathcal{O}(\sqrt{t}\, h^4)$.

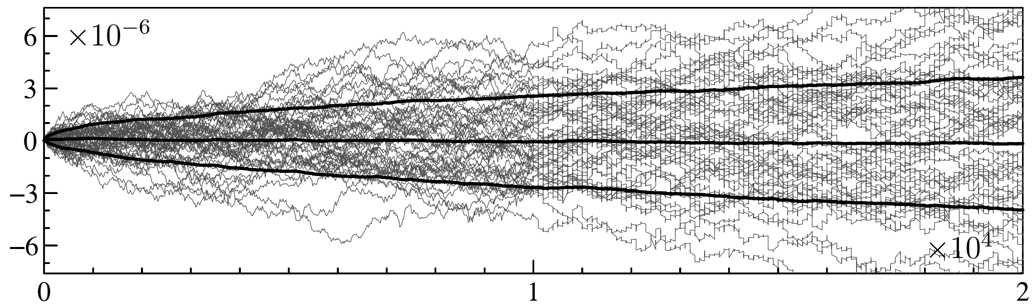


Figure 8: Error in the modified energy $H(p_n, q_n) + h^2 H_3(p_n, q_n)$ of the N-body problem ($N = 9$. Lennard–Jones potential) along the numerical solution of the simplified Takahashi–Imada method (50 trajectories). Included is the average over 400 trajectories as a function of time ($\mu = 0.091 \times 10^{-6}$ at $t = 2 \times 10^4$) and the standard deviation ($\sigma = 3.08 \times 10^{-6}$ at $t = 2 \times 10^4$).

5 Existence of global modified Hamiltonian

The theory of generating functions shows that every symplectic one-step method $\Phi_h : (p, q) \mapsto (P, Q)$ can be locally expressed in terms of a function $S(P, q, h)$ as

$$p = P + \nabla_q S(P, q, h), \qquad Q = q + \nabla_p S(P, q, h). \tag{20}$$

This property allows us to prove the existence of a globally defined modified Hamiltonian, whenever the generating function is globally defined.

Theorem 4 *Assume that the symplectic method Φ_h has a generating function*

$$S(P, q, h) \;=\; h\, S_1(P, q) + h^2 S_2(P, q) + h^3 S_3(P, q) + \ldots \tag{21}$$

with smooth $S_j(P, q)$ defined on an open set U. Then, the modified differential equation is a Hamiltonian system with

$$\widetilde{H}(p, q) = H(p, q) + h\, H_2(p, q) + h^2 H_3(p, q) + \ldots\ , \tag{22}$$

where the functions $H_j(p, q)$ are defined and smooth on the whole of U.

Proof. The exact solution $(P, Q) = \big(\widetilde{p}(t), \widetilde{q}(t)\big)$ of the Hamiltonian system corresponding to $\widetilde{H}(p, q)$ is given by

$$p = P + \nabla_q \widetilde{S}(P, q, t), \qquad Q = q + \nabla_P \widetilde{S}(P, q, t),$$

where $\widetilde{S}$ is the solution of the Hamilton–Jacobi differential equation

$$\partial_t \widetilde{S}(P, q, t) \;=\; \widetilde{H}\big(P, q + \nabla_P \widetilde{S}(P, q, t)\big), \qquad \widetilde{S}(P, q, 0) = 0. \tag{23}$$

Since $\widetilde{H}$ depends on the parameter h, this is also the case for $\widetilde{S}$. Our aim is to determine the functions $H_j(p, q)$ such that the solution $\widetilde{S}(P, q, t)$ of (23) coincides for $t = h$ with (21).

We first express $\widetilde{S}(P, q, t)$ as a series

$$\widetilde{S}(P, q, t) \;=\; t\, \widetilde{S}_1(P, q, h) + t^2 \widetilde{S}_2(P, q, h) + t^3 \widetilde{S}_3(P, q, h) + \ldots\ ,$$

insert it into (23) and compare powers of t. This allows us to obtain the functions $\widetilde{S}_j(p, q, h)$ recursively in terms of derivatives of $\widetilde{H}$:

$$\begin{aligned}
\widetilde{S}_1(p, q, h) \;&=\; \widetilde{H}(p, q) \\
2\, \widetilde{S}_2(p, q, h) \;&=\; \Big(\frac{\partial \widetilde{H}}{\partial q} \cdot \frac{\partial \widetilde{S}_1}{\partial P}\Big)(p, q, h) \\
3\, \widetilde{S}_3(p, q, h) \;&=\; \Big(\frac{\partial \widetilde{H}}{\partial q} \cdot \frac{\partial \widetilde{S}_2}{\partial P}\Big)(p, q, h) + \frac{1}{2}\bigg(\frac{\partial^2 \widetilde{H}}{\partial q^2}\Big(\frac{\partial \widetilde{S}_1}{\partial P}, \frac{\partial \widetilde{S}_1}{\partial P}\Big)\bigg)(p, q, h).
\end{aligned} \tag{24}$$

We then write $\widetilde{S}_j$ as a series

$$\widetilde{S}_j(p,q,h) = \widetilde{S}_{j1}(p,q) + h\,\widetilde{S}_{j2}(p,q) + h^2\widetilde{S}_{j3}(p,q) + \dots\ ,$$

insert it and the expansion (22) for $\widetilde{H}$ into (24), and compare powers of h. This yields $\widetilde{S}_{1k}(p,q) = H_k(p,q)$ and for $j > 1$ we see that $\widetilde{S}_{jk}(p,q)$ is a function of derivatives of H_l with $l < k$.

The requirement $S(p,q,h) = \widetilde{S}(p,q,h)$ finally shows $S_1(p,q) = \widetilde{S}_{11}(p,q)$, $S_2(p,q) = \widetilde{S}_{12}(p,q) + \widetilde{S}_{21}(p,q)$, etc., so that

$$S_j(p,q) = H_j(p,q) + \text{“function of derivatives of } H_k(p,q) \text{ with } k < j\text{”}.$$

For a given generating function $S(P,q,h)$, this recurrence relation allows us to determine successively the $H_j(p,q)$. We see from these explicit formulas that the functions H_j are defined on the same domain as the S_j. □

Example 7 (Symplectic Euler Method) The symplectic Euler method is nothing other than (20) with $S(P,q,h) = h\,H(P,q)$. Following the constructive proof of Theorem 4 we obtain

$$\widetilde{H} \;=\; H - \frac{h}{2}H_pH_q + \frac{h^2}{12}\Big(H_{pp}H_q^2 + H_{qq}H_p^2 + 4H_{pq}H_qH_p\Big) + \dots\ . \tag{25}$$

as the modified Hamiltonian of the symplectic Euler method.

Theorem 5 *A symplectic Runge–Kutta method (i.e., $b_ia_{ij} + b_ja_{ji} = b_ib_j$ for all i,j) applied to a system with smooth Hamiltonian $H : U \to \mathbb{R}$ (with $U \subset \mathbb{R}^{2d}$ an arbitrary open set) has a modified Hamiltonian (22) with smooth functions $H_j(y)$, defined globally on U.*

Proof. Let (P_i, Q_i) be the internal stages of an implicit Runge–Kutta method $(p,q) \mapsto (P,Q)$. It follows from implicit differentiation of the Runge–Kutta equations that, under the condition $b_ia_{ij} + b_ja_{ji} = b_ib_j$ for all i,j, the Runge–Kutta formulas can be written as (20) with (an idea of Lasagni 1988)

$$S(P,q,h) = h\sum_{i=1}^{s} b_iH(P_i,Q_i) - h^2\sum_{i,j=1}^{s} b_ia_{ij}H_q(P_i,Q_i)^TH_p(P_j,Q_j).$$

This shows that the coefficient functions $S_j(P,q)$ can be expressed in terms of derivatives of $H(P,q)$. □

This theorem extends to partitioned Runge–Kutta methods including the Störmer–Verlet integrators. Theorem 4 implies that all methods based on generating functions (e.g., variational integrators) have a globally defined modified Hamiltonian.

6 Completely integrable Hamiltonian systems

There is an interesting class of Hamiltonian problems, for which symplectic integrators have an improved long-time behavior of the global error. Here and in the following, we denote the standard d-dimensional torus by

$$\mathbb{T}^d = \mathbb{R}^d/2\pi\mathbb{Z}^d = \{(\theta_1 \bmod 2\pi, \dots, \theta_d \bmod 2\pi)\ ;\ \theta_i \in \mathbb{R}\}.$$

Definition 1 *We call a Hamiltonian system* completely integrable*, if, for every (p_0, q_0) in the domain of $H(p,q)$, there exists a symplectic diffeomorphism*

$$(p,q) = \psi(a,\theta), \quad 2\pi\text{-periodic in } \theta$$

between $V \times \mathbb{T}^d$ and $U \subset \mathbb{R}^{2d}$ (where U is a neighborhood of (p_0, q_0), and V is open), such that the Hamiltonian in the new variables becomes

$$H(p,q) \ = \ H(\psi(a,\theta)) = K(a).$$

The variables $(a,\theta) = (a_1, \dots, a_d, \theta_1 \bmod 2\pi, \dots, \theta_d \bmod 2\pi)$ are called *action-angle variables*. In these variables, the system becomes

$$\dot a_i = 0, \quad \dot\theta_i = \omega_i(a), \quad i = 1, \dots, d$$

with $\omega_i(a) = \frac{\partial}{\partial a_i} K(a)$, and can be solved directly $a_i(t) = a_{i0}$, $\theta_i(t) = \theta_{i0} + \omega_i(a_0)\, t$, so that we get periodic (or quasi-periodic) flow

$$\big(p(t), q(t)\big) = \psi\big(a_0, \theta_0 + \omega(a_0)\, t\big).$$

The following theorem gives a practical characterization of integrability[6].

Theorem 6 (Arnold-Liouville) *Suppose that for the Hamiltonian $H(p,q)$ there exist smooth functions $F_1 = H, F_2, \dots, F_d : U \to \mathbb{R}$, $U \subset \mathbb{R}^{2d}$ satisfying*

- (I1) *$F_1, \dots, F_d$ are in involution, i.e., $\{F_i, F_j\} = 0$, where the Poisson bracket is given by $\{F, G\} = \nabla_q F^{\mathsf{T}} \nabla_p G - \nabla_p F^{\mathsf{T}} \nabla_q G$,*
- (I2) *the gradients of $F_1, \dots, F_d$ are everywhere linearly independent,*
- (I3) *the solution trajectories of the Hamiltonian systems with Hamiltonian F_i (for $i = 1, \dots, d$) exist for all times and remain in U.*

If, in addition, the level sets $\{(p,q) \in U\ ;\ F_i(p,q) = c_i, i = 1, \dots, d\}$ are compact, then the Hamiltonian system is completely integrable.

[6] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, 1978, second edition 1989.

Example 8 (Motion in central field) Consider the Hamiltonian

$$H = \frac{1}{2}\,(p_1^2 + p_2^2) + V(r), \qquad r = \sqrt{q_1^2 + q_2^2},$$

with a potential $V(r)$ that is defined and smooth for $r > 0$. The Kepler problem corresponds to $V(r) = -1/r$, and the perturbed Kepler problem to $V(r) = -1/r - \mu/(3r^3)$. Changing to polar coordinates (see Exercises 3 and 4)

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}, \qquad \begin{pmatrix} p_r \\ p_\varphi \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\cos\varphi \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \tag{26}$$

this becomes

$$H(p_r, p_\varphi, r, \varphi) = \frac{1}{2}\Big(p_r^2 + \frac{p_\varphi^2}{r^2}\Big) + V(r).$$

The system has the angular momentum $L = p_\varphi$ as a first integral, since H does not depend on φ. Clearly, $\{H, L\} = 0$ everywhere. The gradients of H and L are linearly independent unless both $p_r = 0$ and $p_\varphi^2 = r^3 V'(r)$. By inserting $p_\varphi^2 = 2r^2(H - V(r))$ and eliminating r this becomes a condition of the form $\alpha(H, L) = 0$, which for the Kepler problem reads $L^2(1 + 2HL^2) = 0$. The conditions of Theorem 6 are thus satisfied on the domain

$$U = \{(p_r, p_\varphi, r, \varphi)\ ;\ r > 0,\ \alpha(H, L) \neq 0\}.$$

Example 9 (Toda lattice) This is a system of particles on a line interacting pairwise with exponential forces. The motion is determined by the Hamiltonian

$$H(p, q) \;=\; \sum_{k=1}^{n} \Big(\frac{1}{2}\, p_k^2 \;+\; \exp(q_k - q_{k+1}) \Big)$$

with periodic boundary conditions: $q_{n+1} = q_1$. With the notation $a_k = -\frac{1}{2} p_k$, $b_k = \frac{1}{2}\exp\big(\frac{1}{2}(q_k - q_{k+1})\big)$, all n eigenvalues of the matrix

$$L \;=\; \begin{pmatrix} a_1 & b_1 & & & b_n \\ b_1 & a_2 & b_2 & 0 & \\ & b_2 & \ddots & \ddots & \\ & 0 & \ddots & a_{n-1} & b_{n-1} \\ b_n & & & b_{n-1} & a_n \end{pmatrix}$$

are first integrals of the system. It can be shown (here without proof) that this system is completely integrable

Many important problems in celestial mechanics are small perturbations of integrable systems, e.g., planetary motion.

7 Linear error growth for integrable systems

We consider a completely integrable Hamiltonian system

$$\dot{p} = -\nabla_q H(p,q), \qquad \dot{q} = \nabla_p H(p,q). \tag{27}$$

with real analytic Hamiltonian. We let $(p,q) = \psi(a,\theta)$ be the symplectic diffeomorphism that transforms (27) to action-angle variables, and we denote the inverse transformation by $(a,\theta) = \big(I(p,q), \Theta(p,q)\big)$. Consequently, the components $I_1, \ldots, I_d$ of I are first integrals of the system, i.e., $I(p(t),q(t)) = I(p_0,q_0)$ for all t. In the action-angle variables, the Hamiltonian is $K(a) = H(p,q)$, and we denote the vector of frequencies by $\omega(a) = \nabla K(a)$. We consider this in a neighbourhood of some $a^* \in \mathbb{R}^d$.

The aim of this section is to prove that for symplectic methods applied to completely integrable systems we have simultaneously

- linear growth of the global error,
- near conservation of all first integrals depending only on the action variables.

These properties are illustrated in Figure 9. Non-symplectic methods, like the classical Runge–Kutta method of order 4, show a quadratic growth of the global error and a linear growth of the error in the action variables.

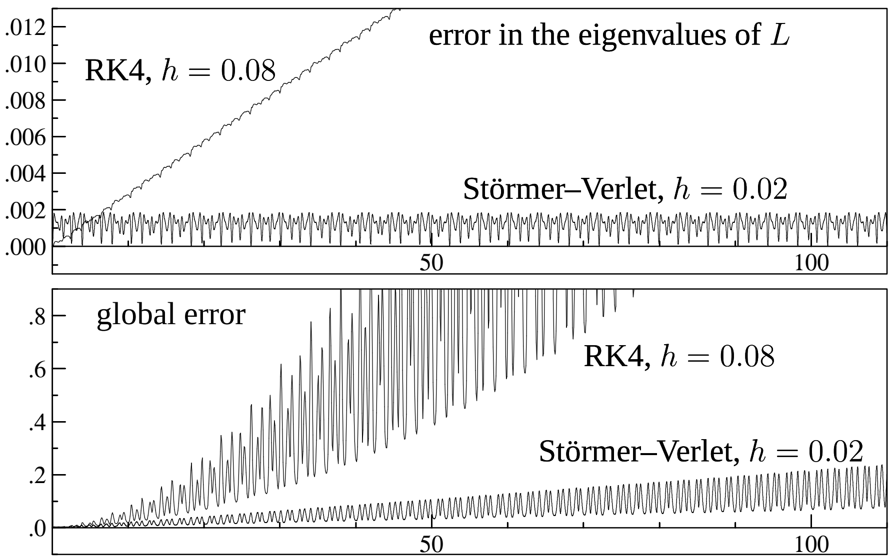


Figure 9: Euclidean norm of numerical errors for the Toda lattice with $n = 3$; initial values are $p_1 = -1.5$, $p_2 = 1$, $p_3 = 0.5$, and $q_1 = 1$, $q_2 = 2$, $q_3 = -1$.

Theorem 7 *Consider*

- *completely integrable Hamiltonian system with real-analytic Hamiltonian*
- *symplectic integrator of order r with globally defined modified Hamiltonian*
- *strong non-resonance condition for $\omega(a^*)$*

$$|k \cdot \omega(a^*)| \geq \gamma |k|^{-\nu}, \quad k \in \mathbb{Z}^d,\ k \neq 0 \tag{28}$$

- $\|I(p_0, q_0) - a^*\| \leq Const\, |\log h|^{-\nu-1}$

Then, there exist constants C, h_0 such that for $h \leq h_0$ and for $t = nh \leq h^{-r}$ the numerical solution satisfies

$$\|(p_n, q_n) - (p(t), q(t))\| \leq C\, t\, h^r$$
$$\|I(p_n, q_n) - I(p_0, q_0)\| \leq C\, h^r.$$

Proof. Let us mention (without proof) that for $\nu > d - 1$ the set of frequencies in a fixed ball that do not satisfy (28) has Lebesgue measure bounded by $c\gamma$. Therefore, almost all frequencies satisfy (28) for some $\gamma > 0$.

The main steps of the proof are illustrated in Figure 10. The missing part is

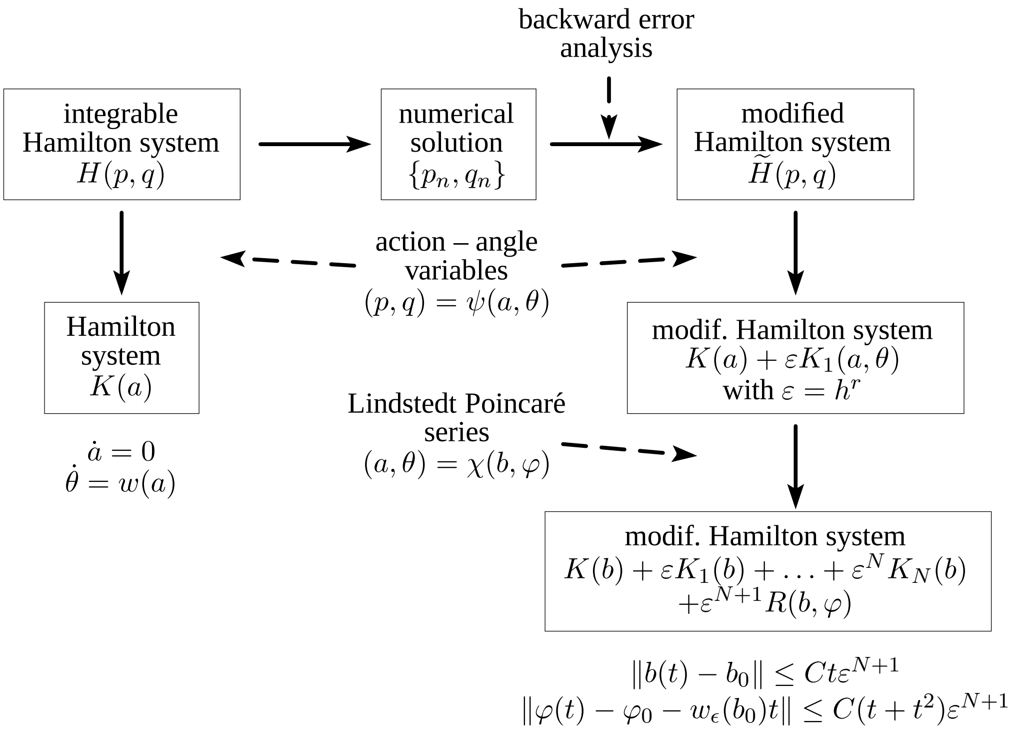


Figure 10: Idea of the proof for the linear error growth of symplectic integrators.

the recursive elimination of the angle-variables by the use of a Lindstedt–Poincaré series. This will be discussed in the following (omitting technical details).

We consider a perturbed Hamiltonian

$$\widetilde{K}(a, \theta) = K(a) + \varepsilon K_1(a, \theta).$$

The problem is to find a symplectic transformation $(a, \theta) = \chi(b, \varphi)$ which eliminates the angle variable θ as far as possible. We look for a transformation of the form

$$b = a - \nabla_\theta S(b, \theta), \quad \varphi = \theta + \nabla_b S(b, \theta),$$

where the generating function is given by a truncated series

$$S(b, \theta) = \varepsilon S_1(b, \theta) + \varepsilon^2 S_2(b, \theta) + \ldots + \varepsilon^N S_N(b, \theta)$$

with coefficient functions that are 2π-periodic in θ. Such a transformation is $\mathcal{O}(\varepsilon)$ close to the identity. In the new variables the Hamiltonian is

$$\begin{aligned} \widetilde{K}(\chi(b, \varphi)) \;&=\; \widetilde{K}(a, \theta) = \widetilde{K}(b + \nabla_\theta S(b, \theta), \theta) \\ &=\; K(b) + \varepsilon\big(\omega(b) \cdot \nabla_\theta S_1(b, \theta) + K_1(b, \theta)\big) + \mathcal{O}(\varepsilon^2) \end{aligned}$$

where $\omega(b) = \nabla K(b)$. We aim in finding S_1 such that

$$\omega(b) \cdot \nabla_\theta S_1(b, \theta) + K_1(b, \theta)$$

does not depend on θ. Expanding the periodic functions into Fourier series

$$S_1(b, \theta) = \sum_{k \in \mathbb{Z}^d} s_k(b) e^{\mathrm{i}\, k \cdot \theta}, \quad K_1(b, \theta) = \sum_{k \in \mathbb{Z}^d} h_k(b) e^{\mathrm{i}\, k \cdot \theta},$$

we obtain a formal solution from

$$s_k(b) \;=\; -\frac{h_k(b)}{\mathrm{i}\, k \cdot \omega(b)}, \quad k \neq 0.$$

At this point we are struck by the *problem of small denominators*. For any values of the frequencies $\omega_j(b)$, the denominator $k \cdot \omega(b) = k_1\omega_1(b) + \cdots + k_d\omega_d(b)$ becomes arbitrarily small for some $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$, and even vanishes if the frequencies are rationally dependent. Using the non-resonace condition (28) and the fast decay of Fourier coefficients for analytic functions permits to overcome this difficulty.

Further coefficients of the generating function $S(b, \theta)$ can be constructed in a similar way. □

8 Exercises

1. Change the Maple script of Example 1 in such a way that the modified equations for the implicit Euler method, the implicit midpoint rule, or the trapezoidal rule are obtained. Observe that for symmetric methods one gets expansions in even powers of h.

2. Compute a first integral of the Lotka–Volterra equations (Example 2) and of the truncated modified equation for the symplectic Euler method.

3. Let $Q = \chi(q)$ be a change of position coordinates. Prove that the relation $p = \chi'(q)^{\mathsf{T}} P$ extend this to a symplectic mapping $(p, q) \mapsto (P, Q)$.
   Hint. Consider the generating function $S(P, q) = P^{\mathsf{T}} \chi(q)$.

4. Let $y = \psi(z)$ be a symplectic change of coordinates. Prove that it transforms $\dot{y} = J^{-1} \nabla H(y)$ into $\dot{z} = J^{-1} \nabla K(z)$ with $K(z) = H(y) = H\big(\psi(z)\big)$.

5. (Field & Nijhoff 2003)[7] Apply the symplectic Euler method to the system with Hamiltonian $H(p, q) = \ln(\alpha + p) + \ln(\beta + q)$. Compute the modified Hamiltonian and prove that the series converges for sufficiently small step sizes.
   Hint. The method conserves exactly $I(p, q) = (\alpha + p)(\beta + q)$. Find linear two-term recursions for $\{p_n\}$ and $\{q_n\}$, and use the ideas of Example 3.
   Result.
   $$\widetilde{H}(p, q) = H(p, q) - \sum_{k \geq 1} \frac{h^k \, I(p, q)^{-k}}{k(k+1)}.$$

6. Consider a differential equation $\dot{y} = f(y)$ with a divergence-free vector field, and apply a volume-preserving integrator. Show that every truncation of the modified equation has again a divergence-free vector field.
   Hint. Adapt the proof by induction of Theorem 2.

[7] C.M. Field & F.W. Nijhoff, *A note on modified Hamiltonians for numerical integrations admitting an exact invariant*, Nonlinearity 16 (2003) 1673–1683.