Fields Institute Communications
Volume 10, 1996

Backward Error Analysis of Symplectic Integrators

J. M. Sanz-Serna
Departamento de Matemática Aplicada y Computación
Universidad de Valladolid, Valladolid, Spain

Abstract. In this expository paper we study how numerical solutions to differential equations can be interpreted as exact solutions of nearby differential equations. The emphasis is on the interpretation of symplectic numerical solutions of Hamiltonian systems as exact solutions of a perturbed Hamiltonian problem.

1 Introduction

In numerical analysis, given a problem $\mathcal{P}$ with true solution $\mathcal{S}$ and given an approximate solution $\widetilde{\mathcal{S}}$, *forward error analysis* consists of estimating the distance between $\widetilde{\mathcal{S}}$ and $\mathcal{S}$. *Backward error analysis* consists of showing that $\mathcal{S}$ is the true solution of a problem $\widetilde{\mathcal{P}}$ which is close to $\mathcal{P}$. While backward error analysis has played a role of paramount importance in areas like numerical linear algebra, error analyses of numerical methods for differential equations have essentially been of the forward variety (see nevertheless [Beyn [1991]], [Sanz-Serna [1992]], [Eirola [1993]]).

Backward error analysis would be specially helpful for *long-time* integrations. In the long-time scenario, the outcome of any forward error analysis is that errors (in the traditional *forward* sense) are huge, regardless of the numerical method being used. On the other hand, a successful backward error analysis could show that a numerical simulation of a given system provides the true evolution of a nearby, perturbed system.

In this paper we are concerned with backward error analyses of numerical methods for the initial value problem

$$\dot{u} = f(u), \qquad t \geq 0, \tag{1.1}$$

$$u(0) = \alpha \in \mathcal{R}^D, \tag{1.2}$$

where for simplicity it is assumed that the vector field f is $\mathcal{C}^\infty$ in the whole of $\mathcal{R}^D$. Of particular significance for us is the Hamiltonian case where $D = 2d$, $u = [p, q]$,

1991 *Mathematics Subject Classification.* Primary: 65L05; Secondary: 70H05.

$p, q \in \mathcal{R}^D$ and $f = f(p,q)$ has components

$$f_i = -\frac{\partial H}{\partial q_i}, \quad f_{d+i} = \frac{\partial H}{\partial p_i}, \qquad i = 1, \ldots, d, \tag{1.3}$$

for a suitable real-valued Hamiltonian function $H = H(p,q)$. We assume that (1.1)–(1.2) is integrated by a one-step method

$$u_{n+1} = \psi_{h,f}(u_n), \qquad n = 0, 1, \ldots, \tag{1.4}$$

$$u_0 = \alpha, \tag{1.5}$$

where h is the (constant) step size and u_n is the numerical approximation at time $t_n = nh$. For instance

$$\psi_{h,f}(u_n) = u_n + hf(u_n) \tag{1.6}$$

corresponds to the well-known Euler rule. The numerical method $\psi_{h,f}$ is consistent of order $r \geq 1$ if, for all u in $\mathcal{R}^D$,

$$\psi_{h,f}(u) - \phi_{h,f}(u) = O(h^{r+1}), \qquad h \to 0, \tag{1.7}$$

where $\phi_{h,f}$ denotes the flow of (1.1), so that $\phi_{h,f}(u)$ is the value at time $t = h$ of the solution of (1.1) with initial value u at time $t = 0$. From the local error estimate (1.7), it follows ([Butcher [1987]], [Hairer et al. [1987]]), that, as $h \to 0$, the global errors $u_n - u(t_n)$ are $O(h^r)$ uniformly in bounded intervals $0 \leq t \leq t_{max}$ contained in the interval of existence of the true solution of (1.1)–(1.2).

The initial-value problem (1.1)–(1.2) has both the vector field f and the initial condition α as data. Therefore for the backward error analysis of (1.4)–(1.5) we may try (at least) two approaches. In one approach we look for a perturbed initial condition $\widetilde{\alpha}$ and compare the numerical points u_n, $n = 0, 1, \ldots$, with the values $\phi_{t_n,f}(\widetilde{\alpha})$ of the solution of (1.1) with initial condition $\widetilde{\alpha}$. This is the approach that leads to the idea of shadowing (see [Sanz-Serna and Larsson [1993]] and its references) and will not be considered further in this paper. In a second, alternative approach, we keep α as an initial condition and look for a perturbed vector field $\widetilde{f}$ so that $u_n = \phi_{t_n,\widetilde{f}}(\alpha)$. This procedure is very much related to the method of modified equations [Griffiths and Sanz-Serna [1986]], [Warming and Hyett [1974]], a tool sometimes used in the analysis of numerical methods for evolutionary partial differential equations.

Let us illustrate the method of modified equation with the linear scalar equation

$$\dot{u} = \lambda u \tag{1.8}$$

integrated by Euler's rule (1.6). Clearly

$$\psi_{h,f}(u) - \phi_{h,f}(u) = (u + h\lambda u) - \exp(h\lambda)(u) = -\frac{h^2\lambda^2}{2}u - \frac{h^3\lambda^3}{6}u - \ldots,$$

so that, in (1.7), $r = 1$ (first order of consistency). Is it possible to find a perturbed vector field $\widetilde{f}_2$ so that Euler's rule is consistent of the *second order* with the equation $\dot{u} = \widetilde{f}_2(u)$? In symbols

$$\psi_{h,f}(u) - \phi_{h,\widetilde{f}_2}(u) = O(h^3), \qquad h \to 0. \tag{1.9}$$

On using the ansatz

$$\dot{u} = \widetilde{f}_2(u) = \widetilde{f}_2^h(u) = \lambda u + hF_2(u),$$

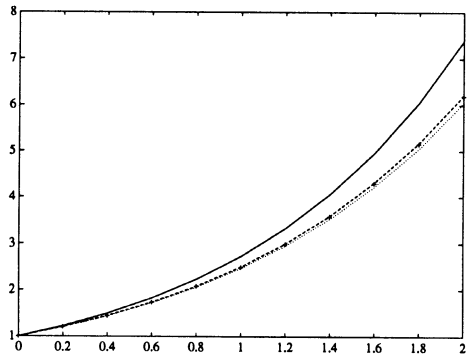

Figure 1 The initial value problem $du/dt = u$, $u(0) = 2$. Solid line: true solution.

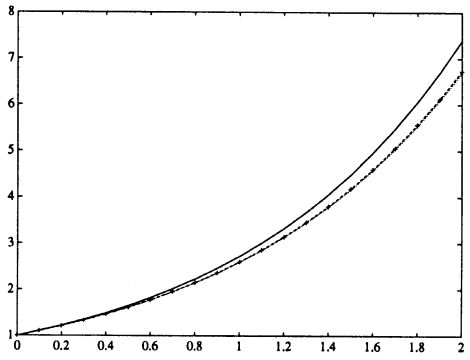

Figure 2 The case $h = 0.1$.

which implies

$$\ddot{u} = (\lambda + hF_2'(u))\dot{u} = (\lambda + hF_2'(u))\,(\lambda u + hF_2(u)),$$

we have

$$\begin{aligned}\psi_{h,f}(u) - \phi_{h,f}(u) &= (u + h\lambda u) \\ &\quad - \left(u + h[\lambda u + hF_2(u)] + \frac{h^2}{2}[\lambda^2 u + O(h)] + O(h^3)\right) \\ &= -\frac{h^2}{2}[2F_2(u) + \lambda^2 u] + O(h^3), \qquad h \to 0,\end{aligned}$$

and (1.9) leads to $F_2(u) = -\lambda^2 u/2$. By going from local errors to global errors as before, we now conclude that the Euler solution for (1.8) with $u(0) = \alpha$ is $O(h^2)$ away from the solution of the modified problem

$$\dot{u} = \left(\lambda - \frac{h\lambda^2}{2}\right) u, \quad u(0) = \alpha. \tag{1.10}$$

Figure 1 corresponds to $\lambda = 1$, $\alpha = 1$, $t_{max} = 2$ and $h = 0.2$. The solid line gives the exact solution $u(t) = \exp(t)$ and the crosses are the Euler solution. The dotted line provides the solution of (1.10). Figure 2 is identical to Figure 1, except that

now $h = 0.1$. We see that the computed points, meant to approximate (1.8) provide a better approximation to the solution of the modified problem (1.10).

Better modified equations exist for our example. We may consider

$$\dot{u} = \widetilde{f}_3^h(u) = \left(\lambda - \frac{h\lambda^2}{2}\right) u + h^2 F_2(u) \tag{1.11}$$

and determine $F_3(u)$ to ensure

$$\psi_{h,f}(u) - \phi_{h,\widetilde{f}_3^h}(u) = O(h^4), \qquad h \to 0,$$

(consistency of the third order). Straightforward algebra leads to $F_3(u) = (\lambda^3/3)u$. The solution of (1.11) with $u(0) = 1$ is shown in Figs. 1–2 by means of a dashed line. We see that, within plotting accuracy, the Euler simulation of (1.8) gives the solution of the problem

$$\dot{u} = \lambda^h u, \quad \lambda^h = \lambda - \frac{h\lambda^2}{2} + \frac{h^2\lambda^3}{3}, \quad u(0) = \alpha$$

and a backward error interpretation is possible. If we imagine a modelling situation where the value of λ in (1.8) is not known exactly but rather arises from some measurement, then the backward error interpretation is useful because it is telling us that the effect of the numerical integration is to change the value of λ into a nearby value λ^h. If $|\lambda - \lambda^h|$ is of the order of the uncertainty in the measurement of λ, then the Euler's solution is as accurate as one may wish in this setting.

For (1.6) and (1.8), the modified equation ($N \geq 1$)

$$\dot{u} = \widetilde{f}_N^h(u) = \left(\lambda - \frac{h\lambda^2}{2} + \frac{h^2\lambda^3}{3} - \cdots \pm \frac{h^{N-1}\lambda^N}{N}\right) u,$$

is of order N, i.e.,

$$\psi_{h,f}(u) - \phi_{h,\widetilde{f}_N^h}(u) = O(h^{N+1}), \qquad h \to 0. \tag{1.12}$$

In this example, as $N \to \infty$, the vector fields $\widetilde{f}_N^h(u)$ converge to $(\log(1+h\lambda)/h)u$. Then

$$\dot{u} = \widetilde{f}_\infty^h(u) = \frac{\log(1+h\lambda)}{h} u$$

is an exact modified equation with $\psi_{h,f} \equiv \phi_{h,\widetilde{f}_\infty^h}$ [Beyn [1991]].

For general f it is still possible to find, for each $N \geq 1$, a modified equation $\dot{u} = \widetilde{f}_N^h(u)$ so that (1.12) holds. However, in general, the fields $\widetilde{f}_N^h$ do not converge as $N \to \infty$ and $\widetilde{f}_\infty^h$ cannot be defined. This is consistent with the fact that, for nonlinear f, $\psi_{h,f}$ is likely to present features (such as transversal crossings of separatrices) that cannot appear in any flow ϕ (see Section 4.10 of [Sanz-Serna [1991]]). Nevertheless, under suitable analyticity conditions, Neishtadt [Neishtadt [1984]] has proved that by letting the order N increase like $O(h^{-1})$, it is possible to have flows $\phi_{h,\widetilde{f}_{N(h)}^h}$ that approximate $\psi_{h,f}$ with exponentially small errors $O(\exp(-k/h))$, $k > 0$, as $h \to \infty$.

An outline of the remainder of the paper is as follows. In Section 2 we show how to systematically construct modified vector fields $\widetilde{f}_N^h$. The results presented are due to Hairer [Hairer [1994]], but the simple methodology used in the derivation of the formulae is taken from the Ph.D. thesis of A. Murua [Murua [1994]]. Section 3 looks

at the particular case of Hamiltonian problems integrated by symplectic (canonical) methods [Sanz-Serna and Calvo [1994]]. The main result is that a method is symplectic if and only if the modified fields $\widetilde{f}_N^h$ corresponding to Hamilonian vector fields f are also Hamiltonian. Thus symplectic simulations change the Hamiltonian function, as distinct from nonsymplectic simulations, whose effect is to perturbe the Hamiltonian equation so as to render it non-Hamiltonian. Section 4 contains an application of the modified equation method to the study of error growth in the numerical solution of nonlinear oscillations. In the final Section 5 we consider an alternative approach where the numerical method is shown to exactly solve a nonautonomous system $\dot{u} = \widehat{f}(u,t)$, where, as $t \to 0$, the function $\widehat{f}$ approaches the true autonomous f.

2 Constructing modified equations

It is well known [Butcher [1987]], [Hairer et al. [1987]] that (rooted) trees are an important tool in the analysis of one-step methods. The trees with four of fewer nodes are depicted in Fig. 3. The symbol τ_1 denotes the only tree with one node. It is common to denote by $[\tau^1, \tau^2, \ldots, \tau^m]$ the tree that consists of the root and m leaving edges to which the trees τ^1, τ^2, $\ldots$, τ^m are attached. Thus in Fig. 3, $\tau_2 = [\tau_1]$, $\tau_{31} = [\tau_1, \tau_1]$, $\tau_{32} = [\tau_2]$, etc. For each tree τ, the integers $\rho(\tau)$ and $\alpha(\tau)$ respectively denote its order (number of nodes) and number of monotonic labellings. These functions can be computed recursively by the formulae $\rho(\tau_1) = \alpha(\tau_1) = 1$ and, for $\tau = [\tau^1, \ldots, \tau^m]$,

$$\begin{aligned} \rho(\tau) &= 1 + \rho(\tau^1) + \cdots + \rho(\tau^m), \\ \alpha(\tau) &= \frac{(\rho(\tau)-1)!}{\rho(\tau^1)! \cdots \rho(\tau^m)!} \alpha(\tau^1) \ldots \alpha(\tau^m) \frac{1}{\mu_1! \mu_2! \cdots}. \end{aligned}$$

The integers μ_i count the number of equal trees among τ^1, $\ldots$, τ^m. Finally, in connection with the system (1.1), an $\mathcal{R}^D$-valued function $F(\tau)(u)$ (elementary differential) is associated with each tree τ. The recursive definition of the $F(\tau)(u)$'s is $F(\tau_1)(u) = f(u)$ and for $\tau = [\tau^1, \ldots, \tau^m]$

$$F(\tau)(u) = f^{(m)}(u)(F(\tau^1)(u), \ldots, F(\tau^m)(u)),$$

where $f^{(m)}(u)$ represents the m-th Fréchet derivative of f evaluated at u.

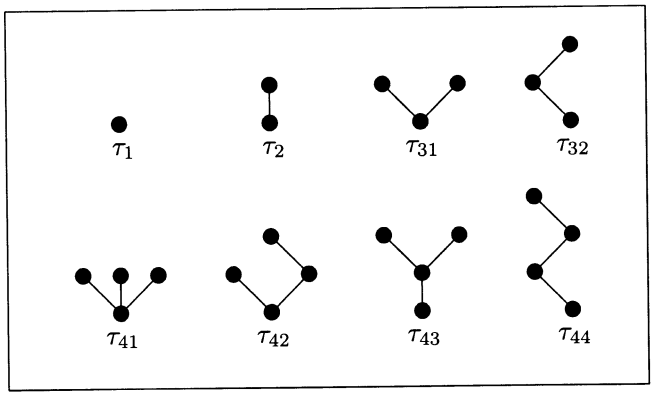

Figure 3 Trees of order ≤ 4.

With these notations, the formal Taylor expansion of the flow $\phi_{h,f}$ in powers of h is given by

$$\phi_{h,f}(u) = u + \sum_{\tau \in T} \frac{h^{\rho(\tau)}}{\rho(\tau)!} \alpha(\tau) F(\tau)(u),$$

where T denotes the set of all trees.

Numerical methods can be Taylor expanded in a similar way. For instance, for a Runge-Kutta method

$$\begin{aligned} U_i &= u_n + h \sum_{j=1}^{s} a_{ij} f(U_j), \\ u_{n+1} &= u_n + h \sum_{i=1}^{s} b_i f(U_i), \end{aligned}$$

the Taylor expansion is

$$\psi_{h,f}(u) = u + \sum_{\tau \in T} \frac{h^{\rho(\tau)}}{\rho(\tau)!} \alpha(\tau) \left(\gamma(\tau) \sum_{i=1}^{s} b_i \Phi_i(\tau) \right) F(\tau)(u),$$

where the recursive definitions of γ and Φ_i are $\gamma(\tau_1) = 1$, $\Phi_i(\tau_1) = 1$ and

$$\begin{aligned} \gamma(t) &= \rho(\tau)\gamma(\tau^1)\dots\gamma(\tau^m), \\ \Phi_i(\tau) &= \sum_{j_1,\dots,j_m} a_{ij_1}\Phi_{j_1}(\tau^1)\dots a_{ij_m}\Phi_{j_m}(\tau^m). \end{aligned}$$

In view of the Taylor expansions above, [Hairer and Wanner [1974]] introduced the notion of a B-series. Given a real-valued mapping a defined in the union of T and the set $\{\emptyset\}$, a B-series $B(a,u)$ is a formal power series

$$a(\emptyset)u + \sum_{\tau \in T} \frac{h^{\rho(\tau)}}{\rho(\tau)!} \alpha(\tau) a(\tau) F(\tau)(u).$$

(The B-series depends on h and f, but this dependence is not made explicit in the notation.) Thus, the results just quoted imply that the true flow $\phi_{h,f}$ corresponds to $a \equiv 1$, while for a Runge-Kutta method $a(\emptyset) = 1$ and, for τ in T,

$$a(\tau) = \gamma(\tau) \sum_{i=1}^{s} b_i \Phi_i(\tau).$$

The Taylor expansion of most one-step methods used in practice is also a B-series. In the remainder of the paper we assume that we are dealing with a method $\psi_{h,f}(u)$ corresponding to a suitable B-series $B(a,u)$,

$$\psi_{h,f}(u) = B(a,u), \tag{2.1}$$

without specifying the exact nature of the method. We suppose throughout that the method is consistent, i.e., at least of order 1:

$$a(\emptyset) = 1, \qquad a(\tau_1) = 1. \tag{2.2}$$

Our aim is to construct a formal power series $\widetilde{f}$

$$\sum_{\tau\in T}\frac{h^{\rho(\tau)-1}}{\rho(\tau)!}\alpha(\tau)b(\tau)F(\tau)(u) \tag{2.3}$$

so that for each integer $N \geq 1$

$$\frac{du}{dt} = \widetilde{f}_N^h(u) = \sum_{1\leq\rho(\tau)\leq N}\frac{h^{\rho(\tau)-1}}{\rho(\tau)!}\alpha(\tau)b(\tau)F(\tau)(u) \tag{2.4}$$

provides a modified equation of order N.

An essential tool for our purposes is the formula for composition of B-series, see Theorem 11.6 in [Hairer et al. [1987]]. If a and b are B-series coefficients with $a(\emptyset) = 1$ then the composition $B(b, B(a, y))$ is a again a B-series $B(ab, y)$ whose coefficients $ab(\tau)$ can be found in a systematic way from the a's and b's. The formulae for the first $ab(\tau)$ are

$$\begin{aligned}
ab(\emptyset) &= b(\emptyset),\\
ab(\tau_1) &= b(\emptyset)a(\tau_1) + b(\tau_1),\\
ab(\tau_2) &= b(\emptyset)a(\tau_2) + 2b(\tau_1)a(\tau_1) + b(\tau_2),\\
ab(\tau_{31}) &= b(\emptyset)a(\tau_{31}) + 3b(\tau_1)a(\tau_1)^2 + 3b(\tau_2)a(\tau) + b(\tau_{31}),\\
ab(\tau_{32}) &= b(\emptyset)a(\tau_{32}) + 3b(\tau_1)a(\tau_2) + 3b(\tau_2)a(\tau) + b(\tau_{32}).
\end{aligned} \tag{2.5}$$

We introduce a real parameter λ and write the flow of the vector field in (2.3) as a B-series

$$\phi_{\lambda h,\widetilde{f}}(u) = u + \sum_{\tau\in T}\frac{h^{\rho(\tau)}}{\rho(\tau)!}\alpha(\tau)a_\lambda(\tau)F(\tau)(u).$$

Next we substitute this series into the equation

$$\frac{d}{dt}\phi_{t,\widetilde{f}} = \widetilde{f}(\phi_{t,\widetilde{f}});$$

in doing so the B-series of the right hand side is computed by the formula for composing B-series. In this way we find that the $a_\lambda(\tau)$ satisfy, for each tree τ,

$$\frac{d}{d\lambda}a_\lambda(\tau) = (a_\lambda b)(\tau). \tag{2.6}$$

Furthermore at $\lambda = 0$, $\phi_{0,\widetilde{f}}(u) = u$ and hence, for each τ,

$$a_0(\tau) = 0. \tag{2.7}$$

The relations (2.6)–(2.7) allow us the computation of the $a_\lambda(\tau)$'s in terms of the $b(\tau)$'s when the latter are known. In our setting, the b coefficients are determined to ensure that, for each τ, at $\lambda = 1$,

$$a_1(\tau) = a(\tau), \tag{2.8}$$

to impose that, as formal power series, $\phi_{h,\widetilde{f}}$ and $\psi_{h,f}$ coincide.

The relations (2.6)–(2.8) make it possible to recursively compute the b coefficients. Let us illustrate this. For τ_1 we obtain from (2.6) and (2.5), $(d/d\lambda)a_\lambda(\tau_1) = b(\tau_1)$, so that, according to (2.7), $a_\lambda(\tau_1) = \lambda b(\tau_1)$. If we now impose (2.8), we obtain the relation $b(\tau_1) = a(\tau_1)$. We conclude, from the consistency assumption (2.2), that $b(\tau_1) = 1$, and therefore, as expected, $\widetilde{f}$ differs from f in $O(h)$ terms.

If we now go through the same steps for the next tree τ_2, we obtain

$$a(\tau_2) = b(\tau_1)^2 + b(\tau_2).$$

Note that for a method of order ≥ 2, $a(\tau_2) = 1$, which, in tandem with $b(\tau_1) = 1$, leads to $b(\tau_2) = 0$ and $\widetilde{f}$ and f differ in $O(h^2)$ terms.

The equations for finding $b(\tau_{32})$ and $b(\tau_{31})$ turn out to be

$$\begin{aligned} a(\tau_{31}) &= b(\tau_1)^3 + \frac{3}{2} b(\tau_2) b(\tau_1) + b(\tau_{31}), \\ a(\tau_{32}) &= b(\tau_1)^3 + 3 b(\tau_2) b(\tau_1) + b(\tau_{32}). \end{aligned}$$

From here $b(\tau_{31}) = b(\tau_{32}) = 0$ for methods of order ≥ 3.

We summarize our findings in the following theorem, due to Hairer [Hairer [1994]].

Theorem 2.1 *Assume that an order r, $r \geq 1$, one-step method (2.1) can be formally Taylor expanded into a B-series $B(a, u)$. There is a unique B-series (2.3), differing from $f(u)$ in $O(h^r)$ terms, such that, for each integer $N \geq 1$, (2.4) provides a modified system of order N. The coefficients b can be recursively found as functions of the coefficients a.*

3 The Hamiltonian case

We now consider the Hamiltonian case (1.3). There has been much recent interest [Sanz-Serna and Calvo [1994]] in simulating (1.1)–(1.3) by so-called symplectic or canonical integrators, i.e., by methods for which $\psi_{h,f}$ preserves the differential form $dp \wedge dq$ and thus reproduce the main feature of Hamiltonian flows.

Calvo and the present author [Calvo and Sanz-Serna [1994]] have shown that, for methods $\psi_{h,f}$ given by a B-series as in (2.1), it is possible to check canonicity by looking at the corresponding coefficients $a(\tau)$. We use Butcher's notation and set

$$\tau \circ \tau^* = [\tau^1, \tau^2, \ldots, \tau^m, \tau^*]$$

if τ^* and $\tau = [\tau^1, \tau^2, \ldots, \tau^m]$ are trees. Then a B-series is canonical [Calvo and Sanz-Serna [1994]], if, for any pair of trees τ^1, τ^2,

$$\frac{a(\tau^1 \circ \tau^2)}{\gamma(\tau^1 \circ \tau^2)} + \frac{a(\tau^2 \circ \tau^1)}{\gamma(\tau^2 \circ \tau^1)} = \frac{a(\tau^1)}{\gamma(\tau^1)} \frac{a(\tau^2)}{\gamma(\tau^2)}. \tag{3.1}$$

On the other hand, Hairer [Hairer [1994]] has proved that the B-series vector field (2.3) (with f given by (1.3)) is Hamiltonian if and only if, for any τ^1, τ^2,

$$\frac{b(\tau^1 \circ \tau^2)}{\gamma(\tau^1 \circ \tau^2)} + \frac{b(\tau^2 \circ \tau^1)}{\gamma(\tau^2 \circ \tau^1)} = 0. \tag{3.2}$$

When (3.2) holds, it is possible to explicitly find the formal Hamiltonian $\widetilde{H}$ whose vector field is (2.3).

The key point is that if $\{a(\tau)\}$ are the coefficients associated with a numerical method as in (2.1) and $\{b(\tau)\}$ are the coefficients of the corresponding modified vector field as in Theorem 1, then (3.1) and (3.2) are equivalent [Hairer [1994]]. Therefore canonical methods can be characterized as those methods that when applied to Hamiltonian problems possess modified equations $\dot{u} = \widetilde{f}_N^h(u)$ that are Hamiltonian for all $N = 1, 2, \ldots$. The use of a symplectic integrator changes the

Hamiltonian function of the system being integrated; the use of a nonsymplectic integrator perturbes the differential equation so as to turn it non-Hamiltonian.

4 An application

We now illustrate the use of modified equations. We consider the well-known pendulum system with Hamiltonian function (energy) $H = (1/2)p^2 + 1 - \cos q$. Let (p_0, q_0) be an initial condition with energy H_0, $0 < H_0 < 2$, leading to a periodic solution. In phase plane the trajectory corresponds to the level set $H = H_0$; the period T_0 of the solution is an increasing function of H_0. Furthermore, we respectively denote by f_0 and g_0 the vector field f evaluated at (p_0, q_0) and the energy gradient at (p_0, q_0). The vectors f_0 and g_0 are mutually orthogonal by conservation of energy.

This initial value problem is integrated by a one-step method of order r with step length h, that for simplicity we assume to be of the form $h = T_0/\nu$, with ν a positive integer. Let $e_M(h)$ be the global error $u_n - u(t_n)$ after $n = M\nu$ steps, i.e. after simulating M periods of the solution. Then it is not too difficult to show, see [Calvo and Sanz-Serna [1993]], that

$$e_M(h) = Me_1(h) + \frac{1}{2}(M^2 - M)\Big(g_0, e_1(h)\Big)\,\delta_0 f_0 + O(h^{2r}), \qquad h \to 0,$$

where $(\cdot, \cdot)$ means inner product and δ_0 denotes the derivative of the period T with respect to the energy H evaluated at the initial condition. Therefore, ignoring the $O(h^{2r})$ remainder, the error $e_M(h)$ grows quadratically with M. The leading M^2 growth is in the direction of f_0, i.e., tangent to the solution at the initial point, thus corresponding to a *phase error*. However linear error growth with M is possible: if

$$(g_0, e_1(h)) = O(h^{2r}) \tag{4.1}$$

(i.e., the error after one period is almost orthogonal to the energy gradient), then

$$e_M(h) = Me_1(h) + O(h^{2r}), \qquad h \to 0.$$

In a nutshell, the way global errors build up is determined by the *direction* of the error $e_1(h)$. This is not suprising: if after one period the error $e_1(h)$ has a significant component in the direction of g_0, then the numerical solution has jumped in phase plane to a neighbouring trajectory corresponding to a different (say larger) value of the energy. Thereafter, the method, when evaluating the vector field f, picks up wrong information as to the solution period and is lead to believe that the motion is slower than it really is. As the integration proceeds the numerical solution keeps jumping to higher and higher energy levels and getting unduly slowing down. This is the mechanism leading to quadratic growth in the *phase* error. On the other hand, if $e_1(h)$ is essentially in the direction of f_0, then there is no energy error: the method is basically describing the right trajectory with a slightly distorted average velocity and errors grow linearly. These considerations apply to all nonlinear oscillators with one degree of freedom [Calvo and Sanz-Serna [1993]], to Kepler's problem [Calvo and Sanz-Serna [1993]], and even to some partial differential equations [Frutos and Sanz-Serna [1994]]. See [Cano and Sanz-Serna [1995]] for a comprehensive treatment.

We now use the method of modified equations to show that for canonical methods (4.1) holds. Let us construct a modified problem of order $2r$. By the results

in the preceding section, this is a Hamiltonian problem with Hamiltonian $\widetilde{H}$. The modified solution $\widetilde{u}$ conserves $\widetilde{H}$ exactly and hence, Taylor expanding,

$$\begin{aligned} 0 = \widetilde{H}(\widetilde{u}(T_0)) - \widetilde{H}(u_0) &= (\widetilde{g}_0, \widetilde{u}(T_0) - u_0) + O(|\widetilde{u}(T_0) - u_0|^2) \\ &= (\widetilde{g}_0, \widetilde{u}(T_0) - u_0) + O(h^{2r}). \end{aligned} \tag{4.2}$$

Here $\widetilde{g}_0$ is the gradient of $\widetilde{H}$ at the initial point u_0 and we have used that

$$\widetilde{u}(T_0) - u_0 = \widetilde{u}(T_0) - u(T_0) = O(h^r), \tag{4.3}$$

due to the periodicity of the true solution and to the fact that the true and modified vector fields differ in $O(h^r)$ terms. From (4.2)–(4.3), along with $\widetilde{g}_0 - g_0 = O(h^r)$, we obtain

$$(g_0, \widetilde{u}(T_0) - u_0) = O(h^{2r}),$$

and finally, since $\widetilde{u}$ and the numerical solution differ in $O(h^{2r})$ terms, (4.1) holds.

5 An alternative approach

In an alternative approach to modified equations, we may compare $\psi_{h,f}(\alpha)$ with the value $\Phi_{\widehat{f}}(h,0)\alpha$, at time $t = h$, of the solution of a nonautonomous system $\dot{u} = \widehat{f}(u,t)$ with initial condition $u = \alpha$ at $t = 0$. It is not difficult to see that $\widehat{f}$ can be chosen to ensure that for all h and α,

$$\psi_{h,f}(u) = \Phi_{\widehat{f}}(h,0)\alpha. \tag{5.1}$$

In fact, consider the family $\mathcal{F}$ of curves $t \mapsto \psi_{t,f}\alpha$ (α is the parameter in the family). Differentiation with respect to t and elimination of α lead to a differential equation $\dot{u} = \widehat{f}(u,t)$ satisfied by all curves in the family. Then (5.1) holds.

For Euler's rule (1.6) applied to (1.8), the family $\mathcal{F}$ is

$$u = \alpha + t\lambda\alpha;$$

differentiation leads to

$$\dot{u} = \lambda\alpha$$

and, eliminating α, we find

$$\dot{u} = \widehat{f}(u,t) = \frac{\lambda}{1+t\lambda}u. \tag{5.2}$$

For a general system (1.1) integrated by a B-series method (2.1) we look for an $\widehat{f}$ of the form (cf. (2.3))

$$\widehat{f}(u,t) = \sum_{\tau\in T} \frac{t^{\rho(\tau)-1}}{\rho(\tau)!}\alpha(\tau)\widehat{b}(\tau)F(\tau)(u) \tag{5.3}$$

Differentiation with respect to t of the B-series for $\psi_{t,f}$ and substitution in $\dot{u} = \widehat{f}(u,t)$ show that for each tree τ

$$\rho(\tau)a(\tau) = (a\widehat{b})(\tau). \tag{5.4}$$

The $\widehat{b}$'s can be recursively found from (5.4). The formulae for the first $\widehat{b}$'s are

$$\begin{aligned}
\widehat{b}(\tau_1) &= a(\tau_1),\\
2\widehat{b}(\tau_1)a(\tau_1)+\widehat{b}(\tau_2) &= 2a(\tau_2),\\
3\widehat{b}(\tau_1)a(\tau_1)^2+3\widehat{b}(\tau_2)a(\tau_1)+\widehat{b}(\tau_{31}) &= 3a(\tau_{31}),\\
3\widehat{b}(\tau_1)a(\tau_2)+3\widehat{b}(\tau_2)a(\tau_1)+\widehat{b}(\tau_{32}) &= 3a(\tau_{32}).
\end{aligned}$$

Note that $\widehat{b}(\tau_1)=1$ because we assume $a(\tau_1)=1$ (consistency). For a method of order ≥ 2, $a(\tau_2)=1$ and then $b(\tau_2)=0$. If the order is ≥ 3, then $\widehat{b}(\tau_{31})=\widehat{b}(\tau_{32})=0$. In general for a method of order r, $\widehat{f}$ differs from f in terms $O(t^r)$ as $t\to 0$. Furthermore, it is easy to check that, for a method of order r, the $\widehat{b}$'s corresponding to trees of order $r+1$ are related to the b's of Theorem 1 through the relation

$$\widehat{b}(\tau)=(r+1)b(\tau).$$

Hence, at the leading $O(h^r)$ order, $\widehat{f}(u,h)-f(u)$ and $\widetilde{f}^h_N(u)-f(u)$, $N>r$, only differ in a factor $r+1$. We have proved the following result.

Theorem 5.1 *Assume that an order r, $r\geq 1$, one-step method $\psi_{t,f}$ can be formally Taylor expanded into a B-series $B(a,u)$ (2.1). There is a unique $\widehat{f}(u,t)$ (5.3), differing from $f(u)$ in $O(t^r)$ terms ($t\to 0$), such that, for all stepsizes h and all points α, $\psi_{h,f}\alpha$ is the value at time $t=h$ of the solution of $\dot{u}=\widehat{f}(u,t)$ with initial condition $u(0)=\alpha$. The coefficients $\widehat{b}$ can be recursively found as functions of the coefficients a.*

Either by using the ideas in Hairer [Hairer [1994]] or by a general argument presented in [Sanz-Serna and Calvo [1994]], it is possible to prove that, in the Hamiltonian case, a method is canonical if and only if $\widehat{f}(u,t)$ turns out to be Hamiltonian. McLachlan and Atela [Mclachlan and Atela [1991]] then use the discrepancy between the true Hamiltonian H and the Hamiltonian $\widehat{H}$ of $\widehat{f}(u,t)$ as a measure of the accuracy of the numerical method.

A possible drawback of the approach in this section is that advancing n steps with the numerical method is not the same as going from $t=0$ to $t=nh$ with the solution of $\dot{u}=\widehat{f}(u,t)$, because, since $\widehat{f}$ is nonautonomous, for the solution operator Φ

$$\Phi_{\widehat{f}}(h,0)\circ\cdots\circ\Phi_{\widehat{f}}(h,0)\neq\Phi_{\widehat{f}}(nh,0).$$

There is a way around this problem: for $0\leq t<h$ we keep the function $\widehat{f}(u,t)$ found before and for $h\leq t<2h$, $2h\leq t<3h$, dots we repeat it periodically to get an h-periodic *discontinous* function $\widehat{f}^h(u,t)$. Now n steps with the method $\psi_{h,f}$ are equivalent to advancing from $t=0$ to $t=nh$ with the differential system $\dot{u}=\widehat{f}^h(u,t)$. For a symplectic method applied to a Hamiltonian problem, the corresponding nonautonomous Hamiltonian $\widehat{H}^h$ has delta functions at the points $t=nh$, n integer. This is similar to the situation in [Wisdom and Homan [1991]].

Let us finally point out a connection with the approach in Section 2: the autonomous vector field $\widetilde{f}^h_N$ considered there is the result of eliminating *by averaging* the terms t^k, $k\geq N$ of the periodic function $\widehat{f}^h$ (cf. [Neishtadt [1984]]). For instance, from (5.2),

$$\dot{u}=[\lambda-t\lambda^2+O(t^2)]u,$$

and averaging the term $-t\lambda^2$ over the period $0 \leq t \leq h$ we obtain $-h\lambda^2/2$ as in (1.10).

Acknowledgements

This research has been supported by project DGICYT PB92-254.

References

Beyn, W.-J. [1991], *Numerical methods for dynamical systems*, In Advances in Numerical Analysis (Light, W., ed.), **I**, 175–236. Clarendon, Oxford.

Butcher, J. C. [1987], *The Numerical Analysis of Ordinary Differential Equations.* John Wiley, Chichester.

Calvo, M. P. and Sanz-Serna, J. M. [1993], *The development of variable-step symplectic integrators, with application to the two-body problem*, SIAM J. Sci. Comput., **14**, 936–952.

Calvo, M. P. and Sanz-Serna, J. M. [1994], *Canonical B-series*, Numer. Math, **67**, 161–175.

Cano, B. and Sanz-Serna, J. M. [1995], *Error growth in the numerical integration of periodic orbits, with application to Hamiltonian and reversible systems*, SIAM J. Numer. Anal., to appear.

Eirola, T. [1993], *Aspects of backward error analysis of numerical ODEs*, J. Comput. Appl. Math., **45**, 65–73.

de Frutos, J. and J. M. Sanz-Serna, J. M. [1994], *Erring and being conservative*, In Numerical Analysis (Griffiths, D. F. and Watson, G. A. eds.), Longman, London, 75–88.

Griffiths, D. F. and Sanz-Serna, J. M. [1986], *On the scope of the method of modified equations*, SIAM J. Sci. Comput., **7**, 994–1008.

Hairer, E. [1994], *Backward analysis of numerical integrators and symplectic schemes*, Annals Numer. Math., **1**, 107–132.

Hairer, E., Nørsett, S. P., and Wanner, G. [1987], *Solving Ordinary Differential Equations I, Nonstiff Problems.* Springer, Berlin.

Hairer, E. and G. Wanner, G. [1974], *On the Butcher group and general multivalue methods*, Computing, **13**, 1–15.

McLachlan, R. I. and Atela, P. [1991], *The accuracy of symplectic integrators*, Nonlinearity, **5**, 541–562.

Murua, A. [1994], *Métodos simplécticos desarrollables en P-series*, Tesis, Universidad de Valladolid.

Neishtadt, A. I. [1984], *The separation of motions in systems with rapidly rotating phase*, J. Appl. Math. Mech., **48**, 133-139.

Sanz-Serna, J. M. [1991], *Two topics on nonlinear stability*, In Advances in Numerical Analysis (Light, W. ed.), Clarendon, Oxford. **I**, 147–174.

Sanz-Serna, J. M. [1992], *Numerical ordinary differential equations vs. dynamical systems*, In The Dynamics of Numerics and the Numerics of Dynamics (Broomhead, D.S. and Iserles, A., eds.), Clarendon, Oxford, 81–106.

Sanz-Serna, J. M. and Calvo, M. P. [1994], *Numerical Hamiltonian Problems.* Chapman and Hall, London.

Sanz-Serna, J. M. and Larsson, S. [1993], *Shadows, chaos and saddles.* Appl. Numer. Math., **13**, 181–190.

Warming, R.F. and Hyett, B.J. [1974], *The modified equation approach to the stability and accuracy of finite difference methods*, J. Comput. Phys., **14**, 159–179.

Wisdom, J. and Homan, M. [1991], *Symplectic maps for the N-body problem*, Astron. J., **102**, 1528–1538.