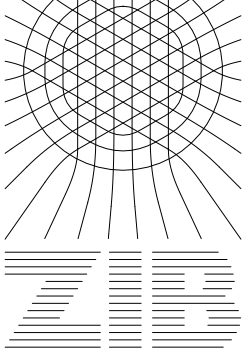
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Heilbronner Str. 10, D-10711 Berlin - Wilmersdorf



Sebastian Reich

Backward error analysis for numerical integrators

Preprint SC 96-21 (July 1996)

Backward error analysis for numerical integrators

Sebastian Reich
Konrad-Zuse-Zentrum
Heilbronner Straße 10
D–10711 Berlin, Germany
e–mail: na.reich@na-net.ornl.gov

July 31, 1996

Abstract

We consider backward error analysis of numerical approximations to ordinary differential equations, i.e., the numerical solution is formally interpreted as the exact solution of a modified differential equation. A simple recursive definition of the modified equation is stated. This recursion is used to give a new proof of the exponentially closeness of the numerical solutions and the solutions to an appropriate truncation of the modified equation. We also discuss qualitative properties of the modified equation and apply these results to the symplectic variable step-size integration of Hamiltonian systems, the conservation of adiabatic invariants, and numerical chaos associated to homoclinic orbits.

1 Introduction

In this paper, we consider the relationship between solutions to a given system of ordinary equations

$$\frac{d}{dt}\, x \;=\; Z(x)\,,$$

numerical approximations

$$x_{n+1} \;=\; G_{\Delta t}\,(x_n) \tag{1}$$

to them, and solutions to associated modified equations

$$\frac{d}{dt}\, x \;=\; X_i\,(x) \qquad (i \geq 1)\,.$$

The vector fields X_i are chosen such that the numerical solution can formally be interpreted, with increasing index i, as the more and more accurate solution of the modified equation. Previous papers on backward error analysis for differential equations include those by Warming & Hyett [31], Griffiths & Sanz-Serna [7], Feng [12], Fiedler & Scheurle [13], and Sanz-Serna [27].

More recently, formulas for the computation of the modified vector fields X_i have been derived by Hairer [14], Calvo, Murua & Sanz-Serna [4], Benettin & Giorgilli [3], and Reich [24]. In papers by Neishtadt [23], Benettin & Giorgilli [3], and Hairer & Lubich [16], the question of closeness of the numerical approximations and the solutions of the modified equations has been addressed. It has also been shown by Neishtadt [23], Hairer [14], Calvo, Murua, Sanz-Serna [4], Reich [24], and Benettin & Giorgilli [3] that for symplectic discretizations, the modified vector fields X_i are Hamiltonian.

Our approach to backward error analysis is based on a simple recursive formulation of the modified vector fields [24]. This allows us to simplify/generalize some of the proofs/results in earlier papers. In particular, in Section 2, we consider general diffeomorphisms that are close to the identity on a compact subset of phase space. We show that, restricted to this compact subset, our recursion yields a vector field with its flow-one-map exponentially close to the given diffeomorphism. Although this result is not new, our proof is different from the ones in [3],[16] and, hopefully, provides new insight. Section 3 is devoted to backward error analysis of general constant step-size one-step methods while, in Section 4, we discuss the question of conservative schemes and its backward error analysis in a general Lie algebraic setting. Finally, in Section 5, we discuss three applications for Hamiltonian equations of motion; namely: the conservation of adiabatic invariants, symplectic variable step-size integration, and numerical chaos associated to homoclinic orbits. We like to point out that, following [25], the results of this paper can be generalized to vector fields $Z : \mathcal{M} \subset R^n \to R^n$ on submanifolds $\mathcal{M}$ of R^n. As shown by Hairer & Stoffer [18], backward error analysis can also be extended to variable step-size methods.

2 Approximation of mappings near the identity

Numerical discretization of a differential equation by a one step method (1) yields a mapping $G_{\Delta t}$, $\Delta t > 0$, that is close to the identity map id for sufficiently small step-sizes Δt. Hence, in this section, we address the approximation of mappings near the identity by flows of vector fields from a rather general point of view. In the following section, we will then come back to the special case of mappings corresponding to numerical discretization of differential equations.

Let $G : U \subset R^n \to R^n$ be an analytic map on an open subset U of R^n. We assume that

$$|| \, G(x) - x \, || \;\; < \;\; \epsilon \, M \tag{2}$$

for all $x \in K$, $K \subset U$ a compact subset of U; $|| \, . \, ||$ the l^∞-norm on R^n. Here $\epsilon > 0$ is a small number and $M > 0$ is a constant of order one with respect to ϵ. In other words, G is an analytic map ϵ close to the identity map on $K \subset U$. Our aim is to find an analytic vector field $X : V \subset R^n \to R^n$ on an appropriate open subset V of R^n such that the corresponding flow map $\exp(X) : V \to R^n$ satisfies

$$\exp(X)(x) \;\; \approx \;\; G(x)$$

for all $x \in K$. For that reason, let us consider the recursion

$$\Delta X_{i+1} \;\; := \;\; G - \exp(X_i) \, , \tag{3}$$

$$X_{i+1} \;\; := \;\; X_i + \Delta X_{i+1} \tag{4}$$

with $X_0 = 0$ and $i = 0, 1, \ldots, s$.

Remark. Note that (3)-(4) can be considered as a simplified Newton method applied to the "nonlinear equation"

$$0 \;\; = \;\; G - \exp(X) \tag{5}$$

in the "unknown" X. The exact Newton method would lead to the equation

$$G(x_0) - \exp(X_i)(x_0) \;\; = \;\; \int_0^1 W(s, x_0) \, \Delta X_{i+1}(x(s)) \, ds \, , \qquad (x_0 \in U) \, . \tag{6}$$

Here $x(t)$ denotes the solution of the differential equation

$$\frac{d}{dt} x \;\; = \;\; X_i(x)$$

with initial value $x(0) = x_0$ and $W(t, x_0)$ is the Wronskian matrix of the variational equation

$$\frac{d}{dt} u \;\; = \;\; \left[\frac{d}{dx} X_i(x(t)) \right] u \, .$$

Note that (6) is, in general, not solvable for ΔX_{i+1} [19]. However, it would be certainly of interest to identify cases for which (6) is invertible and (5) has a solution. In fact, this question is closely related to Kolmogorov's method of proving KAM theory [5] (see Appendix B). In any case, one has

$$\int_o^1 W(s,x_0)\,\Delta X_{i+1}\,(x(s))\,ds \;=\; \Delta X_{i+1}\,(x_0) + \mathcal{O}(\epsilon)$$

which leads to the simplified Newton iteration (3)-(4).

In the sequel, we will implicitly use the representation of the vector field

$$Y \;:=\; G - \mathrm{id}\,,$$

$||Y(x)|| < \epsilon M$ for $x \in K$, as

$$Y \;=\; \epsilon\, Y_0\,, \qquad Y_0 \;:=\; (G-\mathrm{id})/\epsilon\,,$$

$||Y_0(x)|| < M$ on K. This allows us to formally consider the vector fields X_i and ΔX_i, $i,\ldots,s$, as functions of ϵ. Obviously, we have

$$\Delta X_1 \;=\; Y \tag{7}$$

and, using Lie series representation [8],[30] of the exponential function $\exp(X_1)$, i.e.,

$$\exp\,(X_1) \;=\; \mathrm{id} + \sum_{i=1}^{\infty} \frac{1}{i!}\,(L_{X_1})^{i-1}\,X_1 \;=\; \mathrm{id} + \sum_{i=1}^{\infty} \frac{\epsilon^i}{i!}\,(L_{Y_0})^{i-1}\,Y_0\,,$$

we obtain

$$\Delta X_2 \;=\; -\frac{\epsilon^2}{2}\,L_{Y_0}\,Y_0 + \mathcal{O}(\epsilon^3) \;=\; -\frac{1}{2}\,L_Y\,Y + \mathcal{O}(\epsilon^3)\,. \tag{8}$$

Here $L_Y\,Y$ denotes the Lie derivative of Y with respect to Y and the $(L_Y)^i\,Y$ are recursively defined through [8],[30]

$$(L_Y)^i\,Y \;=\; \left[\frac{\partial}{\partial x}\,(L_Y)^{i-1}\,Y\right] Y\,.$$

Continuing this process, we obtain

Lemma 1. The vector fields X_i, $i = 1,2,\ldots,s$, satisfy

$$G - \exp\,(X_i) \;=\; \mathcal{O}(\epsilon^{i+1})\,.$$

Proof. We have to show that, if

$$G - \exp\,(X_i) \;=\; \mathcal{O}(\epsilon^{i+1})\,,$$

then

$$G - \exp(X_{i+1}) \;=\; \mathcal{O}(\epsilon^{i+2})\,.$$

Now, with $\Delta X_{i+1} = \mathcal{O}(\epsilon^{i+1})$,

$$\begin{aligned}\exp(X_{i+1}) &= \exp(X_i + \Delta X_{i+1}) \\ &= (\mathrm{id} + \Delta X_{i+1}) \circ \exp(X_i) + \mathcal{O}(\epsilon^{i+2})\end{aligned}$$

and

$$\begin{aligned}G - \exp(X_{i+1}) &= G - (\mathrm{id} + \Delta X_{i+1}) \circ \exp(X_i) + O(\epsilon^{i+2}) \\ &= (\Delta X_{i+1} - \Delta X_{i+1}) + \mathcal{O}(\epsilon^{i+2}) \\ &= \mathcal{O}(\epsilon^{i+2})\,.\end{aligned}$$

□

From Lemma 1 it follows that

$$\Delta X_i \;=\; -\frac{\epsilon^i}{i!}\left[\frac{\partial^i}{\partial \epsilon^i}\,\exp(X_{i-1})_{\epsilon=0}\right] + \mathcal{O}(\epsilon^{i+1})\,.$$

From now on we will drop the higher order ϵ terms in ΔX_i and simply use

$$\Delta X_i \;:=\; -\frac{\epsilon^i}{i!}\left[\frac{\partial^i}{\partial \epsilon^i}\,\exp(X_{i-1})_{\epsilon=0}\right]$$

instead of (3).

The sequence $\{\Delta X_i\}$ does not, in general, converge to zero. Thus we are looking for the integer i_* such that

$$||\, G - \exp(X_{i_*})\,||_\infty \;=\; \text{Min!}$$

where $||\,.\,||_\infty$ denotes the maximum norm on K, i.e.,

$$||\, G - \exp(X_i)\,||_\infty \;:=\; \max_{x\in K} ||\, G(x) - \exp(X_i(x))\,||$$

and

$$X_i \;=\; \sum_{j=1}^{i} \Delta X_j\,.$$

Let $B_R(x_0) \subset \mathbb{C}^n$ denote the complex ball of radius $R > 0$ around $x_0 \in R^n$ and define

$$||\, z\,|| \;:=\; \max_{i=1,\ldots,n} |z_i|\,, \qquad (z \in \mathbb{C}^n)\,.$$

Under the assumption that the real analytic vector field Y is bounded by $M > 0$ on a complex ball of radius $R > 0$ around each $x_0 \in K \subset R^n$, i.e.,

$$|| \, Y \, ||_R \;\; = \;\; \max_{x \in B_R(x_0)} || \, Y(x) \, || \;\; \leq \;\; M \, , \qquad (x_0 \in K) \, ,$$

one can prove the following result (the proof can be found in Appendix A; see also [3],[20]):

Theorem 1. Let $G : U \subset R^n \to R^n$ be an analytic map ϵ close to the identity on a compact set $K \subset U$, i.e.,

$$|| \, G(x) - x \, || \;\; < \;\; \epsilon \, M \, , \qquad (x \in K) \, .$$

Then there exists a vector field $X : V \subset R^n \to R^n$ such that

$$|| \, G(x) - \exp \, (X) \, (x) \, || \;\; \leq \;\; \epsilon \, M \, e^{-c/\epsilon} \, , \qquad (x \in K) \, , \tag{9}$$

with $c \leq R/(8Me)$ and $R > 0$ such that, for all $x_0 \in K$,

$$|| \, G \, (x) - x \, || \;\; \leq \;\; \epsilon \, M$$

on the complex ball of radius R around x_0.

3 Perturbed vector fields for numerical integration

Let us now consider a smooth vector field

$$\frac{d}{dt} \, x \;\; = \;\; Z \, (x) \, , \tag{10}$$

$Z : U \subset R^n \to R^n$ and its discretization by a one step method

$$x_{n+1} \;\; = \;\; G_{\Delta t} \, (x_n) \;\; = \;\; x_n + \Delta t \, \psi(x_n, \Delta t) \, . \tag{11}$$

We assume that $G_{\Delta t} : U \subset R^n \to R^n$ is a method of order $p \geq 1$, i.e.

$$|| \, \exp \, (\Delta t \, Z) - G_{\Delta t} \, || \;\; = \;\; \mathcal{O}(\Delta t^{p+1}) \, .$$

As in Section 2, we look for a vector field X such that

$$\exp \, (X) \;\; \approx \;\; G_{\Delta t}$$

and consider the recursion

$$\begin{aligned} \Delta X_{i+1} \;\; &:= \;\; G_{\Delta t} - \exp \, (X_i) \, , \\ X_{i+1} \;\; &= \;\; X_i + \Delta X_{i+1} \, , \end{aligned}$$

$i = 0, \ldots, s$, which, upon replacing $G_{\Delta t}$ by G, is equivalent to (3)-(4).

There are two obvious choices for the initial vector field X_0. Following the discussion of the previous section, one could take $X_0 = 0$ or, taking into account that $G_{\Delta t}$ is a discretization of (10), one could define $X_0 = \Delta t\, Z$. While $X_0 = \Delta t\, Z$ immediately yields $\Delta X_1 = \mathcal{O}(\Delta t^{p+1})$, the choice $X_0 = 0$ requires p iterations to obtain an $\mathcal{O}(\Delta t^{p+1})$ approximation to the modified vector field X. However, $X_0 = 0$ allows us to apply Theorem 1 with $G = G_{\Delta t}$ and $\epsilon = \Delta t$. Specifically:

Corollary 1. Let $G_{\Delta t} : U \subset R^n \to R^n$ be a real analytic map close to the identity on a compact set $K \subset U$, i.e.,

$$|| \, G_{\Delta t}(x) - x \, || \;\; < \;\; \Delta t \, M \qquad (x \in K) \, .$$

Then there exists a vector field $X : V \subset R^n \to R^n$ such that

$$|| \, G_{\Delta t}(x) - \exp\,(X)\,(x) \, || \;\; \leq \;\; \Delta t \, M \, e^{-c/\Delta t} \, , \qquad (x \in K) \, ,$$

with $c \leq R/(8Me)$ and $R > 0$ such that, for all $x_0 \in K$,

$$|| \, G_{\Delta t}(x) - x \, || \;\; \leq \;\; \Delta t \, M \tag{12}$$

on the complex ball of radius R around $x_0 \in K$.

Remark. If, instead of (12), only a corresponding estimate $||Z(x)|| < m$ for the vector field Z in (10) is available, then any consistent one step method $G_{\Delta t}$ certainly satisfies the condition (12) with $M = 2m$.

Remark. If, instead of (12), only a corresponding estimate $||Z(x)|| < m$ for the vector field Z in (10) is available, then any consistent one step method $G_{\Delta t}$ will certainly satisfy the condition (12) with $M = 2m$.

Remark. The discrete evolution (11) can now be considered as the discretization of the modified vector field X (as long as the numerical solution does not leave the compact set K). According to Corollary 1 and standard results in numerical analysis, the global error

$$e_n \, (x) \;\; := \;\; \exp\,(nX)\,(x) - [\, G_{\Delta t} \,]^n \; (x)$$

after n steps with step-size Δt is bounded by

$$|| \, e_n \, (x) \, || \;\; \leq \;\; \frac{M}{\tilde{L}} \, \left(e^{n \, \Delta t \, \tilde{L}} - 1 \right) \, e^{-c/\Delta t}$$

where $L = \Delta t \tilde{L} \geq 0$ is the Lipschitz constant of the modified vector field X on K. Thus the global error e_n remains exponentially small over a time interval $T = n\Delta t < c/(2\Delta t \, \tilde{L})$.

Remark. In [13], Fiedler & Scheurle showed that $G_{\Delta t}$ is equivalent to the time-Δt-flow of a non-autonomous differential equation

$$\frac{d}{dt} \, x \;\; = \;\; Z(x) + F(x, t, \Delta t)$$

with F a vector field Δt-periodic in t,

$$|| F(x,t,\Delta t) || \;=\; \mathcal{O}(\Delta t^p)\,,$$

and $p \geq 1$ the order of $G_{\Delta t}$. In view of Corollary 1, we can use the same construction to show that $G_{\Delta t}$ is equivalent to the time-Δt-flow of the non-autonomous differential equation

$$\frac{d}{dt}\,x \;=\; \frac{1}{\Delta t}\,X\,(x) + F(x,t,\Delta t)$$

where X is the modified vector field of Corollary 1 and F is a vector field Δt-periodic in t. Furthermore, because of Corollary 1,

$$|| F(x,t,\Delta t) || \;=\; \mathcal{O}(e^{-c/\Delta t})$$

for $x \in K$ and $t \in [0, \Delta t]$.

Let us now discuss the Taylor series expansion of the modified vector field X in terms of Δt in more detail. This will be useful in Section 4 when we consider geometric properties of X. In this context it is more appropriate to use $X_0 = \Delta t\, Z$ which immediately implies $\Delta X_1 = \mathcal{O}(\Delta t^{p+1})$. It follows from Lemma 1 with $\epsilon = \Delta t$ that, for $X_0 = \Delta t\, Z$, we have

$$\Delta X_{i+1}\,(\Delta t) \;=\; \Delta t^{i+p+1}\,\Delta \hat{X}_{i+1} + O(\Delta t^{i+p+2})\,, \qquad (i \geq 0)\,,$$

where $\Delta \hat{X}_{i+1}$ is an appropriate vector field. Thus we consider the limit

$$\lim_{t\to 0} \frac{1}{t^{i+p+1}}\,t^{i+p+1}\,\Delta \hat{X}_{i+1} \;=\; \lim_{t\to 0} \frac{G_t - \exp\,(X_i(t))}{t^{i+p+1}}$$

where

$$X_i(t) \;=\; X_i(\Delta t = t)$$

and

$$G_t\,(x) \;:=\; x + t\,\psi(x,t)\,.$$

This yields

$$\begin{aligned}
\Delta \hat{X}_{i+1} \;&:=\; \lim_{t\to 0} \frac{G_t - \exp\,(X_i(t))}{t^{i+p+1}} \\
&=\; \frac{1}{(i+p+1)!}\left[\frac{\partial^{i+p+1}}{\partial t^{i+p+1}}\,G_t - \frac{\partial^{i+p+1}}{\partial t^{i+p+1}}\,\exp\,(X_i(t))\right]_{t=0}
\end{aligned}$$

which leads us to the modified recursion

$$\Delta X_{i+1} \;:=\; \Delta t^{i+p+1} \lim_{t\to 0} \frac{G_t - \exp\,(X_i(t))}{t^{i+p+1}}\,, \tag{13}$$

$$X_{i+1} \;:=\; X_i + \Delta X_{i+1} \tag{14}$$

with $X_0 = \Delta t\, Z$.

Remark. The approach of Section 2 is recovered from the iteration (13)-(14) by using

$$G_t(x) \;:=\; x + t\,\psi(x, \Delta t)\,,$$

$p = 0$, and $X_0 = 0$, i.e., the vector field Y_0 of Section 2 is now given by $Y_0 = \psi(x, \Delta t)$. Let us denote the corresponding vector fields X_i by X_i^a and those corresponding to the iteration (13)-(14) with $G_t(x) = x + t\psi(x,t)$ and $X_0 = \Delta t\, Z$ by X_i^b. Then

$$|| \, G(x) - \exp\left(X_{i+p}^a\right)(x) \, || \;=\; \mathcal{O}(\Delta t^{i+p+1})$$

and

$$|| \, G(x) - \exp\left(X_{i}^b\right)(x) \, || \;=\; \mathcal{O}(\Delta t^{i+p+1})\,.$$

From Lemma 4 in Appendix A we have

$$|| \, G(x) - \exp\left(X_{i+p}^a\right)(x) \, || \;\leq\; \Delta t\, M \left(\frac{8\,(i+p)\,\Delta t\, M}{R} \right)^{i+p}$$

for $\epsilon \leq R/(4(i+p)M)$, $x \in K$, and, since the leading $\mathcal{O}(\Delta t^{i+p+1})$ term must be the same in both cases, we certainly also have

$$|| \, G(x) - \exp\left(X_{i}^b\right)(x) \, || \;\leq\; \Delta t\, M \left(\frac{8\,(i+p)\,\Delta t\, M}{R} \right)^{i+p}$$

for $\epsilon \leq R/(4(i+p)M$. Thus the vector field $X = X_*^a$ in Corollary 1 (with i_* appropriately chosen) can be replaced by $X = X_{i_*-p}^b$.

4 Geometric properties of backward error analysis

In this section, we consider differential equations (10) whose corresponding vector field Z belongs to a certain Lie subalgebra $\mathbf{g}$ of the infinite dimensional Lie algebra of smooth vector fields on R^n [19],[1]. Let us assume that there is a corresponding subgroup $\mathbf{G}$ of the group of diffeomorphisms on R^n such that

$$\mathbf{g} \;=\; T_{\mathrm{id}}\,\mathbf{G}\,.$$

(Note that $\mathbf{G}$ is a Frechet manifold but that $\mathbf{G}$ is, in general, not a Lie group [19].) For the Lie algebra of Hamiltonian vector fields this is, for example, the subgroup of canonical transformations. An important aspect of those differential equations is that the corresponding flow map $\exp(tZ)$ forms a one-parametric subgroup in $\mathbf{G}$ [19],[1]. Especially in the context of long term integration, it is desirable to discretize differential equations of this type in such a way that the corresponding iteration map $G_{\Delta t}$ belongs to the same subgroup $\mathbf{G}$ as $\exp(tZ)$. We will call those integrators Lie-algebraic integrators.

The following result concerning the backward error analysis of Lie-algebraic integrators can be derived [24]:

Theorem 2. Let us assume that the vector field Z in

$$\frac{d}{dt}\, x \;=\; Z(x)$$

belongs to a Lie subalgebra $\mathbf{g}$ of the Lie algebra of all smooth vector fields on R^n. Let us assume furthermore that

$$x_{n+1} \;=\; G_{\Delta t}\,(x_n) \;=\; x_n + \Delta t\, \psi(x_n, \Delta t)$$

is a Lie-algebraic integrator for this subalgebra $\mathbf{g}$, i.e., $G_{\Delta t} \in \mathbf{G}$ for all $\Delta t \geq 0$ sufficiently small. Then the perturbed vector fields X_i, $i = 1, \ldots, s$, defined through the recursion (13)-(14) satisfy

$$X_i \;\in\; \mathbf{g}$$

and the vector field X in Corollary 1 can be chosen such that $X \in \mathbf{g}$.

Proof. The statement is certainly true for $X_1 = \Delta t\, Z$. Let us assume that it also holds for X_i, i.e., $X_i(\Delta t) \in \mathbf{g}$ for all $\Delta t \geq 0$ sufficiently small. Since

$$G_t\,(x) \;=\; x + t\, \psi(x,t) \;\in\; \mathbf{G}$$

and

$$\exp\,(X_i(t)) \;\in\; \mathbf{G}\,,$$

as well as

$$G_{t=0} \;=\; \exp\,(X_i(t))_{t=0} \;=\; \mathrm{id},$$

we have

$$\Delta X_{i+1} \;=\; \Delta t^{i+p+1} \lim_{t\to 0} \frac{G_t - \exp\,(X_i(t))}{t^{i+p+1}} \;\in\; T_{\mathrm{id}}\, \mathbf{G}\,.$$

and $\Delta X_{i+1}(\Delta t) \in \mathbf{g}$ for all $\Delta t \geq 0$ sufficiently small. This implies $X_{i+1}(\Delta t) \in \mathbf{g}$ as required. □

Let us discuss two examples:

Example. Consider the Lie subalgebra of all vector fields that preserve a particular first integral $F : R^n \to R$. If $G_{\Delta t}$ satisfies

$$F(\, G_{\Delta t}(x)\,) \;=\; F(x)$$

for all $x \in R^n$, then the modified vector fields X_i possess F as a first integral as well. The same result was recently derived by Gonzales & Stuart [9] by a contradiction argument.

Example. Let $\{.,.\}$ denote the Poisson bracket of a (linear) Poisson manifold $V = R^n$. Then the Lie algebra of Hamiltonian vector fields on V is given by

$$\frac{d}{dt}\, x \;\; = \;\; \{\, \mathrm{id},\, H \,\}\,(x)$$

where $H : V \to R$ is an arbitrary smooth function. The corresponding group $\mathbf{G}$ is given by the set of smooth diffeomorphisms on V that preserve the Poisson bracket $\{.,.\}$. If the discrete evolution (11) satisfies $G_{\Delta t} \in \mathbf{G}$ for all $\Delta t > 0$, then the modified vector fields X_i are Hamiltonian vector fields on V. If we assume furthermore, that the Hamiltonian H is analytic and the discrete evolution $G_{\Delta t}$ satisfies the conditions of Corollary 1, then one has

$$|\, \tilde{H}([G_{\Delta t}]^n(x)) - \tilde{H}(x)\,| \;\; = \;\; \mathcal{O}(\Delta t^{p+1})\,, \tag{15}$$

$\tilde{H}$ the Hamiltonian of the modified vector field, i.e., $X = \{\mathrm{id}, \tilde{H}\}$, as well as

$$|\, H([G_{\Delta t}]^n(x)) - H(x)\,| \;\; = \;\; \mathcal{O}(\Delta t^{p})$$

over time intervals

$$T \;\; = \;\; \Delta t\, n \;\; = \;\; \mathcal{O}(\Delta t^{p+1}\, e^{c/\Delta t})\,.$$

Note that the Hamiltonian $\tilde{H}$ of the modified vector field X satisfies

$$\tilde{H}(x) - \Delta t\, H(x) \;\; = \;\; \mathcal{O}(\Delta t^{p+1})\,, \qquad (x \in K)\,,$$

with $p \geq 1$ the order of the discretization (11). The estimate (15) follows from the fact that the global error in $\tilde{H}(x_n)$, $x_n = [G_{\Delta t}]^n(x_0)$, grows only linearly with $n \geq 1$ [3],[16] and that after one step

$$\tilde{H}(G_{\Delta t}(x)) - \tilde{H}(x) \;\; = \;\; \mathcal{O}(\Delta t\, e^{-c/\Delta t})\,.$$

Remark. Note that time-reversible differential equations, i.e., differential equations (10) with

$$Z\,(R\,(x)) \;\; = \;\; -R\,Z\,(x)$$

where R is an invertible linear transformation with $R\,R = I$, do not form a Lie algebra. Also time-reversible maps G, i.e., maps G satisfying

$$R\,G\,(x) \;\; = \;\; [G]^{-1}\,(R\,(x))\,,$$

do not form a group. Hence Theorem 2 cannot be applied to this class of problems. However, time-reversible vector fields form a linear subspace in the Lie algebra of vector fields and the derivative of any one-parametric family G_t of time-reversible maps with $G_{t=0} = \mathrm{id}$ with respect to t at $t = 0$ is an element of this subspace, i.e., is a time-reversible vector field. Thus the proof of Theorem 2 can be generalized to time-reversible integration. For details see Hairer & Stoffer [18].

5 Applications

5.1 Adiabatic invariants

Let us consider a time-dependent Hamiltonian system on R^2 with Hamiltonian $H(q,p,t)$, $q,p \in R$. Using the extended Hamiltonian

$$E(q,p,s,e) \;:=\; H(q,p,s) - e\,,$$

the corresponding equations of motion

$$\begin{aligned}
\frac{d}{dt}q &= +\nabla_p E(q,p,s,e) \;=\; +\nabla_q H(q,p,s)\,,\\
\frac{d}{dt}p &= -\nabla_q E(q,p,s,e) \;=\; -\nabla_p H(q,p,s)\,,\\
\frac{d}{dt}e &= +\nabla_s E(q,p,s,e) \;=\; +\nabla_s H(,q,p,s)\,,\\
\frac{d}{dt}s &= -\nabla_e E(q,p,s,e) \;=\; 1
\end{aligned}$$

are Hamiltonian in the extended phase space R^4. We assume that the Hamiltonian H is of the form

$$H(q,p,s) \;=\; \frac{1}{2}p^2 + V(q,s)\,.$$

For example,

$$V(q,s) \;=\; \frac{1}{2}\omega(s)^2 q^2\,. \tag{16}$$

Then the equations of motion can be discretized by the well-known Verlet method, i.e.,

$$\begin{aligned}
q_{n+1} &= q_n + \Delta t\, p_{n+1/2}\,,\\
p_{n+1/2} &= p_n - \frac{\Delta t}{2}\nabla_q V(q_n,s_n)\,,\\
p_{n+1} &= p_{n+1/2} - \frac{\Delta t}{2}\nabla_q V(q_{n+1},s_{n+1})\,,\\
e_{n+1} &= e_n + \frac{\Delta t}{2}\left[\nabla_s V(q_n,s_n) + \nabla_s V(q_{n+1},s_{n+1})\right],\\
s_{n+1} &= s_n + \Delta t\,.
\end{aligned}$$

This discretization is symplectic and, therefore, according to Theorem 2, there exists a modified Hamiltonian vector field X with modified Hamiltonian $\tilde{E}$ such that its time one flow is exponentially close to the discrete evolution $G_{\Delta t}$ given by the Verlet discretization. Furthermore, because the equation of motion in the variable s is integrated exactly, the modified Hamiltonian $\tilde{E}$ is again of the form

$$\tilde{E}(q,p,s,e) \;=\; \tilde{H}(q,p,s) + e\,,$$

$\tilde{H}(q,p,t)$ an appropriate function. Let us assume now that, for fixed t, the Hamiltonian $H(q,p,t)$ has periodic solutions and that $H(t)$ varies very slowly in time compared to the periodic motion in (q,p), i.e.,

$$|\,\frac{\partial}{\partial t}\,H(q,p,t)\,| \;\leq\; \delta$$

for all t and $\delta > 0$ sufficiently small. Then the corresponding equations of motion possess an adiabatic invariant

$$J \;=\; \frac{1}{2\pi}\oint p\,dq$$

which remains almost constant over an exponentially long period of time [23], i.e.,

$$|\,J(t) - J(0)\,| \;\leq\; \delta\,, \qquad t = \mathcal{O}(e^{c/\delta})\,.$$

For the time-dependent potential energy function (16), the adiabatic invariant is $J(t) = E(t)/\omega(t)$. Now, for fixed t, the perturbed Hamiltonian $\tilde{H}(q,p,t)$ will also possess periodic solutions and the derivative of $\tilde{H}$ with respect to t will be small, i.e.,

$$|\,\frac{\partial}{\partial t}\,\tilde{H}(q,p,t)\,| \;\leq\; \tilde{\delta}$$

with $\tilde{\delta} = \delta + \mathcal{O}(\Delta t^2)$ since the Verlet method is second order. Thus the perturbed Hamiltonian equations of motion also have

$$J \;=\; \frac{1}{2\pi}\oint p\,dq$$

as an adiabatic invariant. Let us write $x = (q,p)$ and $x_j = [G_{\Delta t}]^j(x_0)$. Then

$$\begin{aligned}
|\,J(x_n) - J(x_0)\,| \;&=\; |\sum_{j=0}^{n-1} J(x_{j+1}) - J(x_j)\,| \\
&=\; |\sum_{j=0}^{n-1} J(G_{\Delta t}(x_j)) - J(\exp(X)(x_j)) + J(\exp(X)(x_j)) - J(x_j)\,| \\
&\leq\; \sum_{j=0}^{n-1} |\,J(G_{\Delta t}(x_j)) - J(\exp(X)(x_j))\,| + \sum_{j=0}^{n-1} |\,J(\exp(X)(x_j)) - J(x_j)\,| \\
&\leq\; n\left[\lambda\, c_1\, e^{-c_2/\Delta t} + c_3\, e^{-c_4/\tilde{\delta}}\right]
\end{aligned}$$

with $c_i > 0$, $i = 1,\ldots,4$, appropriate constants and $\lambda > 0$ the Lipschitz constant of J. Thus one can conclude that symplectic integrators do not only approximately conserve total energy over exponentially long periods of time but adiabatic invariants are approximately conserved over exponentially long periods of time as well. This has been confirmed by numerical experiments conducted by Shimada & Yoshida [28]. An interesting application of this result is related to Hamiltonian systems containing a strong convex potential [26]. Here the highly oscillatory motion about the minima of the strong potential gives rise to an adiabatic invariant

that should also be preserved under numerical discretization. We will discuss this in more detail in a forthcoming publication.

Numerical example. Let us consider a one-dimensional harmonic oscillator with a slowly varying frequency. The Hamiltonian is

$$H(q,p,\epsilon t) \;=\; \frac{1}{2}\,p^2 + \frac{1}{2}\,\omega(\epsilon t)^2\,q^2$$

where

$$\omega\,(\epsilon t) \;=\; 2\pi\,(1+\delta\,\sin\,(2\pi\epsilon t))$$

with $\delta = 0.1$ and $\epsilon = 0.001$ [28]. The adiabatic invariant is

$$J \;=\; \frac{1}{2\pi}\,\oint p\,dq \;=\; \frac{H\,(\epsilon t)}{\omega\,(\epsilon t)}\,.$$

We integrated the corresponding equations of motion by the symplectic implicit midpoint rule with step-sizes $\Delta t = 1.0$ and $\Delta t = 10.0$. Note that the period of the "fast" oscillations in (q,p) is $T = 1$. Thus, for step-sizes $\Delta t > 0.2$, those oscillations are non longer correctly reproduced. However, the implicit midpoint rule is stable for arbitrary step-sizes when applied to an unperturbed harmonic oscillator and one could expect that one can also use larger step-sizes for the harmonic oscillator with slowly varying frequency. However, as our numerical results indicate, one has to be very cautious with this statement (see Fig. 1). This can be explained as follows: For larger step-sizes, the implicit midpoint rule is equivalent to the exact solution of a harmonic oscillator with lower frequency $\tilde{\omega} < \omega$. Thus, for very large step-sizes, the adiabatic invariance condition [28]

$$2\pi\,|\,\frac{d}{dt}\,\tilde{\omega}\,|\,\frac{1}{\tilde{\omega}^2} \;\ll\; 1$$

is not necessarily satisfied anymore and the quantity $J(t)$ can start to drift arbitrarily.

5.2 Symplectic variable step-size integration

According to a result by Stoffer & Nipp [29], classical variable step-size methods asymptotically reduce to a sequence of mappings

$$x_{n+1} \;=\; G_{\Delta t(x_n)}\,(x_n)\,, \tag{17}$$

$$t_{n+1} \;=\; t_n + \Delta t\,(x_n) \tag{18}$$

with $\Delta t(x)$ an appropriate function determined by the step-size selection criteria. Typically, we have

$$\Delta t\,(x) \;=\; \delta\,s(x,\delta)$$

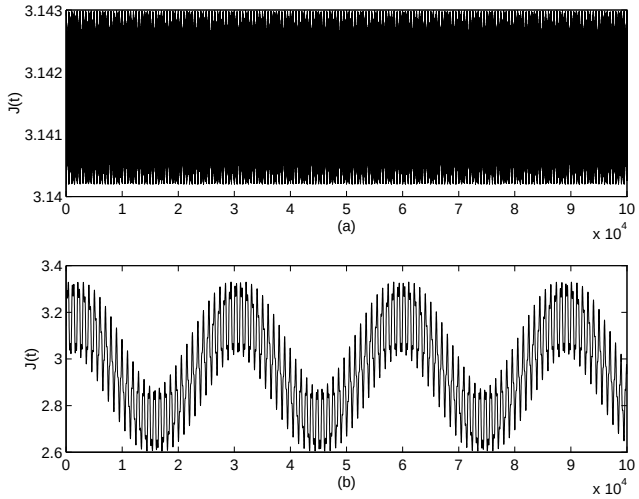


Figure 1: Numerical values of the adiabatic invariant $J(t)$ for step-sizes $\Delta t = 1.0$ (a) and $\Delta t = 10.0$ (b).

with $\delta = \mathrm{TOL}^{1/p}$, $\mathrm{TOL} \ll 1$ a given parameter and p the order of the method $G_{\Delta t}$. Let us now take a different point of view: The variable step-size method (17)-(18) can be viewed as a constant step-size discretization with step-size δ applied to the scaled differential equation

$$\frac{d}{d\tau} x \;=\; \rho(x)\, Z(x)\,, \tag{19}$$

$\rho(x) \approx s(x,\delta)$. As advocated by Huang & Leimkuhler [20] in the context of time-reversible integration, one could, in fact, take (19) as the starting point, i.e., define an appropriate scaling function $\rho(x)$ and discretize the scaled differential equation by, for example, a time-reversible method. That this can lead to highly efficient methods has been demonstrated in [20] for time-reversible Hamiltonian systems of the form

$$\begin{aligned} \frac{d}{dt} q &= p\,, \\ \frac{d}{dt} p &= -\nabla_q V\,(q)\,, \end{aligned}$$

$q, p \in R^n$, with Hamiltonian

$$H(q,p) \;=\; \frac{p^T p}{2} + V(q)\,.$$

As suggested in [20], the scaling function is defined by

$$\rho\,(q,p) \;=\; \frac{1}{\sqrt{p^T p + (\nabla_q V(q))^T \nabla_q V(q)}}\,. \tag{20}$$

Note that this choice makes a lot of sense in the context of Corollary 1: The constant M there is proportional to the maximum over $||Z(x)||$, $x \in K$, $||\,.\,||$ the Euclidian norm in R^n. Using (19) with the scaling function (20), which corresponds to $\rho(x) = 1/||Z(x)||$, yields that the

constant M for the scaled differential equation (19) is close to one uniformly on the compact set K.

It has not been shown yet that reversible (non-symplectic) methods show the same excellent long-term behavior as symplectic methods do; namely: conservation of energy over exponentially long periods of time. Thus it seems reasonable to look for a symplectic discretization of th e scaled Hamiltonian equations of motion: First we introduce the modified Hamiltonian function

$$E\,(q,p,t,e) \;:=\; \rho(q,p)\,(H(q,p)-e)$$

with corresponding equations of motion

$$\begin{aligned}
\frac{d}{d\tau}\,q &= \rho(q,p)\,p + (H(q,p)-e)\,\nabla_p\,\rho(q,p)\,,\\
\frac{d}{d\tau}\,p &= -\rho(q,p)\,\nabla_q\,V(q) - (H(q,p)-e)\,\nabla_q\,\rho(q,p)\,,\\
\frac{d}{d\tau}\,t &= \rho(q,p)\,,\\
\frac{d}{d\tau}\,e &= 0
\end{aligned}$$

in extended phase space $R^{2n}\times R^2$. In particular, let us consider the case $e=H(q(0),p(0))$ and ρ only a function of q. Then

$$\begin{aligned}
\frac{d}{d\tau}\,q &= \rho(q)\,p\,,\\
\frac{d}{d\tau}\,p &= -\rho(q)\,\nabla_q\,V(q) - (H(q,p)-e)\,\nabla_q\,\rho(q) \;=\; -\rho(q)\,\nabla_q\,V(q)\,,\\
\frac{d}{d\tau}\,t &= \rho\,(q)\,,\\
\frac{d}{d\tau}\,e &= 0
\end{aligned}$$

which is just our scaled Hamiltonian vector field and can be discretized by the symplectic Euler method, i.e.

$$\begin{aligned}
q_{n+1} &= q_n + \Delta\tau\,\rho(q_n)\,p_{n+1}\,,\\
p_{n+1} &= p_n - \Delta\tau\,\rho(q_n)\,\nabla_q\,V(q_n) - \Delta\tau\,(H(q_n,p_{n+1})-e)\,\nabla_q\,\rho(q_n)\,,\\
t_{n+1} &= t_n + \Delta\tau\,\rho\,(q_n)\,.
\end{aligned}$$

Note that, for symplecticity, one has to keep the term $(H(q_n,p_{n+1})-e)\,\nabla_q\,\rho(q_n)$. Let us now define our scaling function ρ. According to (20), we obtain

$$\begin{aligned}
\rho\,(q) &= \frac{1}{\sqrt{p^Tp + (\nabla_qV(q))^T\nabla_qV(q)}}\\
&= \frac{1}{\sqrt{2(e-V(q)) + (\nabla_qV(q))^T\nabla_qV(q)}}\,.
\end{aligned} \tag{21}$$

The method is explicit in the variable q. Unfortunately this implies that the method is only first order in Δt. However, the method is symplectic and, therefore, the Hamiltonian $E = (H(q,p) - e)\rho(q)$ is conserved to $\mathcal{O}(\Delta t)$ over exponentially long periods of time. This implies

$$H(q_n, p_n) - e \;\; = \;\; \mathcal{O}(\Delta\tau)$$

over exponentially long periods of time. Thus the proposed method seem suitable for long term, relatively low precision, variable step-size simulations as they occur, for example, in molecular dynamics. A second-order symplectic discretization could be obtained by using the second-order Lobatto IIIa-b partitioned Runge-Kutta formula [6]. The resulting scheme is implicit in $\rho(q)$. However, in many cases the scaling function $\rho(x)$ can be greatly simplified and its evaluation is cheap compared to the evaluation of the force field $F(q) = -\nabla_q V(q)$. Independently of us, the same approach to symplectic variable step-size integration has been formulated by Hairer [15].

Numerical example. As a numerical example, we look at the following modified Kepler problem:

$$\begin{aligned} \frac{d}{dt} q &= p\,, \\ \frac{d}{dt} p &= -\nabla_q V(q)\,, \end{aligned}$$

$q, p \in R^2$, and

$$V(q_x, q_y) = -\frac{1}{\sqrt{(q_x/10)^2 + (q_y)^2}}\,.$$

The problem is non-integrable and, in fact, the dynamics is chaotic, i.e., can be reduced to the Bernoulli shift [21]. We chose initial values $q = (0,1)$ and $p = (1,0)$. The equations of motion are integrated using the "variable" step-size symplectic Euler method with scaling function (21) and $\Delta\tau = 0.05$. The error in energy

$$\Delta H \;\; = \;\; |\, H(q,p) - e \,|$$

and the variation in the actual step-size $\Delta t = \rho(q)\,\Delta\tau$ can be found in Fig. 2.

5.3 Homoclinic orbits and numerical chaos

In [11], Herbst and Ablowitz considered symplectic discretization of Duffing's equation

$$\begin{aligned} \frac{d}{dt} q &= p\,, \\ \frac{d}{dt} p &= -q + 2q^3. \end{aligned}$$

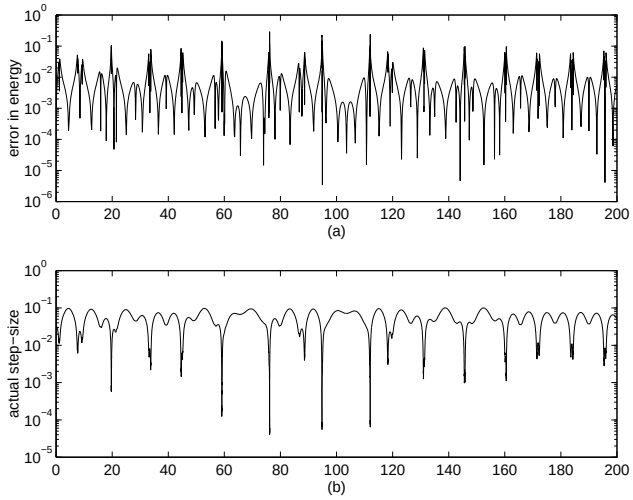


Figure 2: Error in energy (a) and actual step-size Δt (b) as a function of time.

It was noted that the homoclinic orbit associated with the hyperbolic fixed point $(q,p) = (0,0)$ can lead to numerical chaos whenever the discretization is not integrable [11],[13]. It was also shown in [11] that the associated Mel'nikov function is exponentially small which implies that the "chaotic layer" around the homoclinic orbit at the origin decreases at an exponential rate as $\Delta t \to 0$ for any symplectic discretization of Duffing's equation. Here we give a different proof of the exponentially smallness of the associated Mel'nikov function: We assume that Duffing's equation is discretized by a symplectic method. Let $\tilde{H}(q,p)$ denote the Hamiltonian corresponding to the modified vector field X of Corollary 1. For convenience, we scale the Hamiltonian $\tilde{H}$ by $1/\Delta t$, i.e., $\tilde{H} := \tilde{H}/\Delta t$. In general, the fixed point $(q,p) = (0,0)$ will also be a fixed point of $G_{\Delta t}$. Since the phase space of Duffing's equation is two-dimensional, i.e., $q, p \in R$, and $(q,p) = (0,0)$ is a hyperbolic fixed point, the modified vector field X will also possess a homoclinic orbit at the origin (for Δt sufficiently small). Furthermore, it follows from [13] (see also the corresponding remark in Section 3) that there exists a time-dependent Hamiltonian $\hat{H}(q,p,t)$, Δt-periodic in t, such that the time-Δt-flow of the Hamiltonian system

$$\begin{aligned} \frac{d}{dt}\, q &= +\nabla_p \tilde{H}(q,p,\Delta t) + \Delta t\, \nabla_p \hat{H}(q,p,t,\Delta t)\,, \\ \frac{d}{dt}\, p &= -\nabla_q \tilde{H}(q,p,\Delta t) - \Delta t\, \nabla_q \hat{H}(q,p,t,\Delta t) \end{aligned}$$

is equivalent to $G_{\Delta t}$. Furthermore, it follows from Corollary 1 that $||\nabla \hat{H}(q,p,t)||$ is exponentially small, i.e., there exists a constant $c > 0$ such that

$$|| \nabla \hat{H}(q(t),p(t),t,\Delta t) || \;=\; \mathcal{O}(e^{-c/\Delta t}) \tag{22}$$

along solution curves $(q(t),p(t))$ of

$$\frac{d}{dt}\, q \;=\; +\nabla_p \tilde{H}(q,p,\Delta t)\,, \tag{23}$$

$$\frac{d}{dt}\, p \;=\; -\nabla_q \tilde{H}(q,p,\Delta t)\,. \tag{24}$$

The Mel'nikov function $M(\tau)$ [5] corresponding to the homoclinic orbit $z(t) = (q(t), p(t))$ of (23)-(24) is given by

$$M(\tau, \Delta t) \;=\; \int_{-\infty}^{+\infty} \{\tilde{H}, \hat{H}\}(z_\tau(t))\, dt$$

with $z_\tau(t) = z(t+\tau)$, $\tau \in [0, \Delta t]$. Thus, because of (22),

$$|\, M(\tau, \Delta t)\,| \;=\; \mathcal{O}(e^{-c/\Delta t})$$

as claimed. (Note that, according to (22), higher order terms in the expansion of the splitting distance between the stable and unstable manifolds of $(q,p) = (0,0)$ will be bounded by $\mathcal{O}(\Delta t\, e^{-c/\Delta t})$ [10].) For further results on the discretization of homoclinic orbits see Fiedler & Scheurle [13].

Acknowledgement. We like to thank Ernst Hairer for the encouragement to write this paper and Christian Lubich as well as Claudia Wulff for comments on an earlier draft.

Appendix A

Proof of Theorem 1. Let us introduce some notations: (i) Let f be a real analytic function on a complex ball of radius $r > 0$ around $0 \in R$. Cauchy's inequality yields then, under the assumption that

$$|\, f(y)\,| \;\leq\; m$$

for all $|y| \leq r$, the estimate

$$|\, f^{(j)}(0)\,| \;\leq\; j!\, m\, r^{-j}\,.$$

for the jth derivative of f at $y = 0$. (ii) Let X be a real analytic mapping on a complex ball of radius $r > 0$ around a point $x_0 \in R^n$. Then one denotes

$$||\, X\,||_r \;=\; \max_{x \in B_r(x_0)} ||\, X(x)\,||$$

where $B_r(x_0) \subset \mathbb{C}^n$ is the complex ball of radius r around $x_o \in R^n$ and

$$||\, z\,|| \;=\; \max_{i=1,\ldots,n} |\, z_i\,|\,, \qquad (z \in \mathbb{C}^n)\,.$$

Let the real analytic vector field

$$Y \;:=\; G - \mathrm{id}$$

satisfy

$$||\, Y\,||_R \;\leq\; \epsilon\, M \tag{25}$$

with $R > 0$ appropriately chosen and $x_0 \in K$. We also write $Y = \epsilon Y_0$ which allows us to formally consider the vector fields X_i and ΔX_i, $i = 1, \ldots, s$, as functions of ϵ.

Lemma 2. The Lie derivatives $(L_Y)\, Y$, $i \geq 0$, satisfy the estimate

$$\frac{1}{(i+1)!} \,||\, (L_Y)^i\, Y \,||_{R/2} \;\; \leq \;\; \epsilon\, M \, \left(\frac{2\,\epsilon\, M}{R} \right)^i .$$

Proof. Since

$$||\, \exp{(t\, Y)}(x) - x \,|| \;\; \leq \;\; \int_0^{|t|} ||\, Y(\exp{(\tau\, Y)}(x)) \,||\, |d\tau| \,,$$

the map $\exp(tY)$ certainly satisfies

$$\exp{(t\, Y)}(x) - x \;\; \in \;\; B_R\,(x_0)$$

for all $|t| \leq R/(2\epsilon M)$ and all $x \in B_{R/2}(x_0)$. For $x \in B_{R/2}(x_0)$, define

$$f\,(t) \;\; := \;\; \exp{(t\, Y)}\,(x) - x \,.$$

Since

$$||\, f\,(t) \,|| \;\; = \;\; ||\, \exp{(t\, Y)}\,(x) - x \,|| \;\; \leq \;\; \frac{R}{2}$$

for $|t| \leq R/(2\epsilon M)$ and $x \in B_{R/2}(x_0)$ as well as

$$(L_Y)^i\, Y\,(x) \;\; = \;\; \frac{\partial^{i+1}}{\partial t^{i+1}} \, \exp{(t\, Y)}_{t=0}\,(x) \;\; = \;\; f^{(i+1)}\,(t=0)\,,$$

it follows from Cauchy's estimate that

$$||\, (L_Y)^i\, Y\,(x) \,|| \;\; \leq \;\; (i+1)!\, \epsilon\, M \, \left(\frac{2\,\epsilon\, M}{R} \right)^i$$

for all $x \in B_{R/2}(x_0)$. □

Next we have to derive an estimate for $||\Delta X_i||$, $i = 1, \ldots, s$. According to (7), (8), and Lemma 2, we have

$$||\, \Delta X_1 \,||_{R/2} \;\; \leq \;\; \epsilon\, M$$

and

$$||\, \Delta X_2 \,||_{R/2} \;\; \leq \;\; \epsilon\, M \, \left(\frac{2\,\epsilon\, M}{R} \right) .$$

Lemma 3. The vector fields ΔX_i satisfy

$$|| \Delta X_i ||_{R/2} \;\leq\; (i-1)\,\epsilon\,M \left(\frac{2\,(i-1)\,\epsilon\,M}{R}\right)^{i-1} \tag{26}$$

for $i > 1$.

Proof. The statement is true for $i = 2$. We know that ΔX_i and X_i, $i = 1, \ldots, s$, are analytic functions of ϵ. To not confuse the argument ϵ with the constant ϵ in (2), we write $\Delta X_i(\xi)$, $X_i(\xi)$ instead of $\Delta X_i(\epsilon)$, $X_i(\epsilon)$. Let us assume that (26) holds for $i = 1, \ldots, j$. Then

$$\begin{aligned} || X_j(\xi) ||_{R/2} &\leq \sum_{i=1}^{j} || \Delta X_i(\xi) ||_{R/2} \\ &\leq \xi\,M \left[1 + \sum_{i=2}^{j} (i-1) \left(\frac{2\,(i-1)\,\xi\,M}{R}\right)^{i-1}\right] \end{aligned}$$

which implies

$$|| X_j(\xi) ||_{R/2} \;\leq\; j\,\xi\,M \tag{27}$$

for

$$\xi \;\leq\; \frac{R}{2\,j\,M}$$

where we have used that

$$1 + \sum_{i=1}^{j} (i-1) \left(\frac{i-1}{j}\right)^{i-1} \;\leq\; j$$

for $j \geq 2$. Let us now consider the vector-valued function

$$f(\xi) \;:=\; \exp\left(X_j(\xi)\right)(x_0) - x_0\,.$$

Since

$$|| \exp(X_j)(x_0) - x_0 || \;\leq\; \int_0^1 || X_j(\exp(\tau\,X_j)(x_0)) ||\,|d\tau|$$

and (27), we have (similar to the proof of Lemma 2)

$$|| f(\xi) || \;\leq\; \frac{R}{2}$$

for $|\xi| \leq R/(2jM)$ which, by Cauchy's estimate implies

$$\begin{aligned} || \Delta X_{j+1}(\xi)(x_0) || &= \frac{\xi^{j+1}}{(j+1)!} || f^{(j+1)}(\xi = 0) || \\ &\leq j\,\xi\,M \left(\frac{2\,j\,\xi\,M}{R}\right)^{j}. \end{aligned} \tag{28}$$

To derive this estimate, we have used that $||Y(\bar{x})|| \leq \xi M$ on $B_{R/2}(x)$ for each $x \in B_{R/2}(x_0)$ (see also the proof of Lemma 2). Now, with this assumption, we get identical Cauchy estimates for the coefficients of the Taylor series expansion of Y around each $x \in B_{R/2}(x_0)$ which also implies identical estimates for $|| \Delta X_{j+1}(\xi)(x) ||$ if we would explicitly compute $\Delta X_{j+1}(\xi)(x)$ from the Taylor expansion of Y around $x \in B_{R/2}(x_0)$. Since, for $x = x_0$, this estimate has to be bounded by the (Cauchy type) estimate (28), the estimate (28) is, in fact, valid for all $x \in B_{R/2}(x_0)$, i.e.,

$$|| \Delta X_{j+1}(\xi) ||_{R/2} \;\leq\; j\,\xi\,M \left(\frac{2\,j\,\xi\,M}{R}\right)^j$$

as claimed (see also the remark below). □

Remark. Successive Taylor series expansion of the exponential functions $\exp(X_i)$, $i = 1, \ldots, j$, reveals that each exponential function $\exp(X_i)$ can be written as an appropriate linear combination of the elementary differentials $f_{i,j}$ of the vector field Y [14] (G corresponds to the forward Euler discretization of the vector field Y_0 with "step-size" ϵ and, for fixed $i \geq 1$, the $f_{i,j}$'s denote the elementary differentials of "order" $\mathcal{O}(\epsilon^i)$). Since

$$\Delta X_i(x) \;=\; -\frac{\epsilon^i}{i!}\left[\frac{\partial^i}{\partial \epsilon^i}\,\exp\left(X_{i-1}\right)_{\epsilon=0}\right],$$

this implies that ΔX_i is a linear combination of the elementary differentials of order $\mathcal{O}(\epsilon^i)$, i.e.,

$$\Delta X_i(x) \;=\; \frac{1}{i!}\sum_j d_{i,j}\,f_{i,j}(x)\,. \tag{29}$$

Now, each Lie derivative $(L_Y)^{i-1}Y$ is a linear combination (with weights equal or greater than one) of the elementary differentials $f_{i,j}$ of order $\mathcal{O}(\epsilon^i)$ as well [17], i.e.,

$$\frac{1}{i!}\,(L_Y)^{i-1}\,Y(x) \;=\; \frac{1}{i!}\sum_j a_{i,j}\,f_{i,j}(x)\,,$$

$a_{i,j} \geq 1$. By Lemma 2, we know that

$$\frac{1}{i!}\,||\,(L_Y)^{i-1}\,Y\,||_{R/2} \;\leq\; \epsilon\,M\left(\frac{2\,\epsilon\,M}{R}\right)^{i-1}$$

and, therefore,

$$||\,f_{i,j}\,||_{R/2} \;\leq\; i!\,\epsilon\,M\left(\frac{2\,\epsilon\,M}{R}\right)^{i-1}$$

for all elementary differentials $f_{i,j}$ of order $\mathcal{O}(\epsilon^i)$. (For a different derivation of this fact see [20]). Thus there exist appropriate constants $b_i > 0$ such that

$$||\,\Delta X_i\,||_{R/2} \;\leq\; b_i\,\epsilon\,M\left(\frac{2\,\epsilon\,M}{R}\right)^{i-1}$$

and, according to Lemma 3, $b_i = (i-1)^i$. Using the recursive formulae for the coefficients $d_{i,j}$ in (29) [14], a similar estimate has been derived in [20].

Next we need an estimate for the difference between $G(x_0)$ and the exponential map $\exp(X_i)(x_0)$, $x_0 \in K$. This is the subject of the following

Lemma 4. Whenever the constant ϵ in (2) satisfies

$$\epsilon \;\leq\; \frac{R}{4\,i\,M}\,,$$

then

$$||\, G(x_0) - \exp(X_i)(x_0)\,|| \;\leq\; \epsilon\, M \left(\frac{8\,i\,\epsilon\,M}{R}\right)^i .$$

Proof. We know that

$$\begin{aligned} ||\, G_\epsilon(x_0) - \exp(X_i(\epsilon))(x_0)\,|| &= ||\, \Delta X_{i+1}(\epsilon)(x_0)\,|| + \mathcal{O}(\epsilon^{i+2}) \\ &= \frac{\epsilon^{i+1}}{(i+1)!}\,||\, f^{(i+1)}(\xi=0)\,|| + \mathcal{O}(\epsilon^{i+2}) \end{aligned}$$

with $G = G_\epsilon = \mathrm{id} + \epsilon Y_0$ and, as in the proof of Lemma 3,

$$f(\xi) \;:=\; \exp(X_i(\xi))(x_0) - x_0\,.$$

Following standard Taylor series expansion, we have

$$\frac{\epsilon^{i+1}}{(i+1)!}\,||\, f^{(i+1)}(\xi=0)\,|| + \mathcal{O}(\epsilon^{i+2}) \;\leq\; \frac{\epsilon^{i+1}}{(i+1)!}\sup_{0\leq\xi_0\leq\epsilon}||\, f^{(i+1)}(\xi=\xi_0)\,||\,.$$

Similar to the proof of Lemma 3, we obtain that

$$||\, f(\xi)\,|| \;\leq\; \frac{R}{2}$$

for $|\xi_0 - \xi| \leq R/(4iM)$ and $\xi_0 \leq R/(4iM)$. Thus, Cauchy's estimate implies

$$\begin{aligned} ||\, f^{(i+1)}(\xi=\xi_0)\,|| &\leq (i+1)!\,\frac{R}{2}\left(\frac{4\,i\,M}{R}\right)^{i+1} \\ &\leq (i+1)!\,M\left(\frac{8\,i\,M}{R}\right)^i \end{aligned}$$

and, for $\epsilon \leq R/(4iM)$, the desired estimate

$$||\, G_\epsilon(x_0) - \exp(X_i(\epsilon))(x_0)\,|| \;\leq\; \epsilon\, M\left(\frac{8\,i\,\epsilon\,M}{R}\right)^i$$

follows. □

Starting with $i = 1$, Lemma 4 yields now

$$|| G(x_0) - \exp(X_i)(x_0) || \;\leq\; \epsilon M \left(\frac{8\, i\, \epsilon\, M}{R}\right)^i, \qquad (x_0 \in K),$$

provided $\epsilon \leq R/(4iM)$ for all $i = 1, \ldots, s$. The expression $(8iM\epsilon)^i/R$ is a convex function in $i > 0$ with its global minimum at

$$i_o \;=\; \frac{R}{8\,\epsilon\, M\, e}.$$

Let i_* be the integer closest to i_o and $i_* \leq i_o$. Note that this choice of i_* certainly implies

$$\epsilon \;\leq\; \frac{R}{4\, i_*\, M}.$$

Then, for all $x_0 \in K$,

$$\begin{aligned} ||G(x_0) - \exp(X_i)(x_0)|| &\leq \epsilon M e^{-i_*}, \\ &\leq \epsilon M e^{-c/\epsilon} \end{aligned}$$

where $c = i_*\epsilon \leq R/(8Me)$. Thus we have proved Theorem 1.

Appendix B

Let us consider a one degree of freedom (analytic) Hamiltonian system

$$\begin{aligned} \frac{d}{dt} q &= p, \\ \frac{d}{dt} p &= -\nabla_q V(q), \end{aligned}$$

$q, p \in R$, for which the level sets of constant energy are compact submanifolds of R^2 (at least on an appropriate subdomain of R^2). Let us discretize this system by a time-reversible and exactly energy-conserving method $G_{\Delta t}$. We consider restriction of $G_{\Delta t}$ to a level set of constant energy and ask ourselves if, on this level set, $G_{\Delta t}$ is equivalent to the time-Δt-flow of a autonomous vector field X. For simplicity, we assume the level set to be a one dimensional manifold diffeomorph to the unit circle. Thus the restricted $G_{\Delta t}$ is diffeomorph to a diffeomorphism on the unit circle. Let $\mu \geq 0$ denote the corresponding rotation number [5]. As can be shown by Kolmogorov's method [5], the circle map is diffeomorph to a rotation whenever μ is "sufficiently" irrational [5]. This implies in turn that $G_{\Delta t}$ restricted to this particular level set of constant energy is equivalent to the time-Δt-flow of a vector field X. In other words, (5) has a solution on this level set. In terms of our Newton iteration (6), Kolmogorov's method can be interpreted as follows: If the rotation number μ is irrational, then the system (6) (or some nearby system) can be solved for ΔX_{i+1}. However, this system is ill-conditioned due to

small denominators. The condition of "sufficient irrationality" of μ insures that Newton's (or Kolmogorov's) method converges nevertheless [5]. For the symplectic and/or time-reversible discretization of the above one degree of freedom Hamiltonian system a similar statement can be made: For a generic (analytic) Hamiltonian with periodic solutions, most level curves of constant energy will "survive" under symplectic/time-reversible discretization (although slightly deformed). These are the level curves for which the period of the motion of the analytic problem is again "sufficiently" irrational. This is a consequence of KAM theory for symplectic/time-reversible maps in the plane [2],[22]. On these invariant curves the map $G_{\Delta t}$ is diffeomorph to a rotation and, therefore, can be represented as the time-Δt-flow of a vector field X on the corresponding curve. Furthermore, even if the integration is started away from an invariant curve, the energy will be (approximately) conserved for all $t \geq 0$ and the numerical orbit is stable. This explains, for example, the excellent numerical results for symplectic/time-reversible integration of Kepler's equation (in the reduced one degree of freedom formulation).

References

[1] Adams, M., Ratiu, T., and Schmid, R., The Lie group structure of diffeomorphism groups and invertible Fourier integrals operators with applications, in: *Infinite dimensional groups with applications*, Kac, V. (editor), Springer Verlag, 1985.

[2] Arnold, V.I., *Mathematische Methoden der klassischen Mechanik*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1988.

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

[4] Calvo, M.P., Murua, A., and Sanz-Serna, J.M., Modified equations for ODEs. *Contemporary Mathematics* **172**, 63–74, 1994.

[5] de Almeida, A.M.O., *Hamiltonian systems: Chaos and quantization*. Cambridge University Press, Cambridge, 1988.

[6] Geng, S., Symplectic partitioned Runge-Kutta methods, preprint, 1992.

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

[8] Gröbner, W., *Lie Reihen und ihre Anwendungen*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1990.

[9] Gonzales, O. and Stuart, A., Remarks on the qualitative properties of modified equations, preprint, 1996.

[10] Guckenheimer, J. and Holmes, P., *Nonlinear oscillations, dynamical systems and bifurcations of vector fields*. Springer Verlag, New York, 1983.

[11] Herbst, B.M., and Ablowitz, M.J., Numerical chaos, symplectic integrators, and exponentially small splitting distances. *J. Comput. Phys.* **105**, 122–133, 1993.

[12] Feng Kang, Formal power series and numerical algorithms for dynamical systems. Proceedings of international conference on scientific computation, Hangzhou, China, Eds. Tony Chan & Zhong-Ci Shi, *Series on Appl. Math* **1**, 28–35, 1991.

[13] Fiedler, B. and Scheurle, J., Discretization of homoclinic orbits, rapid forcing and "invisible" chaos. Preprint SC 91–5, to appear in AMS Memoirs.

[14] Hairer, E., Backward analysis of numerical integrators and symplectic methods. *Annals of Numerical Mathematics* **1**, 107–132, 1994.

[15] Hairer, E., Variable time step integration with symplectic methods. manuscript, Geneva, 1996.

[16] Hairer, E. and Lubich, Ch., The life-span of backward error analysis for numerical integrators. *Num. Math.*, to appear.

[17] Hairer, E., Norsett, S.P., and Wanner, G., *Solving ordinary differential equations I. Nonstiff problems.* second revised edition, Springer Verlag, 1993.

[18] Hairer, E. and Stoffer, D., Reversible tong-term integration with variable step sizes. *SIAM J. Sci. Comput.*, to appear.

[19] Hamilton, R., The inverse function theorem of Nash and Moser. *Bull. Am. Math. Soc.* **7**, 65-222, 1982.

[20] Huang, W. and Leimkuhler, B., The adaptive Verlet method. *SIAM J. Sci. Comput.*, to appear.

[21] Kirchgraber, U., and Stoffer, D., On the definition of chaos. *ZAMM* **69**, 175–185, 1989.

[22] Moser, J., *Stable and random motion in dynamical systems.* Princeton University Press, Princeton, 1973.

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

[24] Reich, S., Numerical integration of generalized Euler equations. preprint, 1993.

[25] Reich, S., On higher-order semi-explicit symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems. *Num. Math.*, to appear.

[26] Rubin, H. and Ungar, P., Motion under a strong constraining force. *Comm. Pure Appl. Math.* **10**, 65–87, 1957.

[27] Sanz-Serna, J.M., Symplectic integrators for Hamiltonian problems: an overview. *Acta Numerica* **1**, 243–286, 1992.

[28] Shimada, M. and Yoshida, H., Long-term conservation of adiabatic invariants by using symplectic integrators. *Publ. Astron. Soc. Japan* **48**, 147–155, 1996.

[29] Stoffer, D. and Nipp, K., Invariant curves for variable step size integrators. *BIT* **31**, 169–180, 1991.

[30] Varadarajan, V.S., *Lie groups, Lie algebras, and their representation*, Prentice-Hall, Englewood Cliffs, 1974.

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