UNIVERSITY OF CAMBRIDGE

Numerical Analysis Reports

The method of iterated commutators for ordinary differential equations on Lie groups

Antonella Zanna

DAMTP 1996/NA12
August 1996

Department of Applied Mathematics and Theoretical Physics
Silver Street
Cambridge
England CB3 9EW

The method of iterated commutators for ordinary differential equations on Lie groups

Antonella Zanna *

Abstract

We construct numerical methods to integrate ordinary differential equations that evolve on Lie groups. These schemes are based on exponentials and iterated commutators, they are explicit and their order analysis is relatively simple. Thus we can construct group-invariant integrators of arbitrarily high order. Among other applications we show that this approach can be used to obtain new symplectic schemes when applied to Hamiltonian problems. Some numerical experiments are presented.

1 Introduction

The last few years have witnessed a growing interest in *geometrical* or *structural* methods for the integration of Ordinary Differential Equations (ODEs). Many numerical analysts are becoming increasingly convinced that discretization methods that share structural attributes with the underlying equation produce generally numerical approximations that are in some sense superior to those obtained with general-purpose methods. A well known example arises in the integration of Hamiltonian systems, where symplectic integrators or energy conserving integrators are nowadays considered to be a step ahead of other methods: usually geometric integrators are accompanied by other pleasant side effects, like better error propagation, as well as the reproduction of the correct qualitative behaviour of the solution (cf. for instance [23, 6, 9]). Moreover, they are also mathematically æsthetical. For a more detailed discussion on qualitative numerical methods for ODEs, we refer the reader to [14].

In this paper we are interested in developing structural methods for the integration of differential equations on Lie groups. Lie groups are a special case of differentiable manifolds but differential equations that evolve on such structures possess enough generality and ubiquity to justify the development of *ad-hoc* numerical methods. In the last years, progress in this direction has been made and a number of Lie-group invariant methods have been already developed. To our knowledge, the first such methods were introduced by Crouch and Grossman in [7]. These methods are Runge–Kutta-type schemes, they are explicit but their order analysis is particularly difficult and the authors were not able to devise methods of order $p \geq 4$. The same approach is followed also by Owren and Marthinsen who have presented systematic analysis of order conditions, thereby devising a method of order four [16]. A slightly different approach is being currently pursued by

*Newnham College, Cambridge

Munthe-Kaas. The proposed methods are also Runge–Kutta type schemes but his formalism allows for an elegant extension of the Butcher order analysis for standard RK methods [17, 18]. Thus, his methods can be obtained as modification of classical schemes with the introduction of correction functions on the Lie algebra $\mathfrak{g}$, that guarantee that the order is retained in the non-commutative case. In [18] the author presents methods up to order four.

The numerical methods we present in this paper are Lie-group invariant schemes based on the *method of iterated commutators,* which has been introduced by Iserles in [13]. Although the method in [13] was developed for linear ODEs systems with variable coefficients, it is based on very basic operations on Lie groups and, in effect, the algorithm is related to Lie's method of reduction, as explained in [19]. Therefore, if adequately amended, it can be implemented in a group-invariant manner. Our methods are explicit and their order analysis is relatively simple, hence it is possible to develop high order schemes more easily than with Runge–Kutta-type invariant schemes.

The paper is organized as follows. In the remaining part of this section we recall the definition of a Lie group and present some examples.

In §2 we demonstrate how to obtain the method of iterated commutators. We show that the method uses Lie-group techniques and therefore is suitable for integration of differential equations on such manifolds. We discuss numerical methods up to order six and sketch the general way to derive higher order schemes.

In §3 we present applications of the Lie-group techniques to Hamiltonian systems. We show that the integrators on Lie groups presented in §2 can be applied to such systems to find new symplectic integrators. In particular, the Jacobian matrix of the transformation can be decomposed into two parts. The first is a symplectic matrix while the contribution of the second part is the error that originates from violation of symplecicity. Therefore, if the second term is zero, we obtain a symplectic integrator.

Finally, in §4 we present several numerical experiments. Other issues are thereby discussed.

1.1 Lie groups

A Lie group is a differentiable manifold $\mathcal{G}$ that possesses a group structure with respect to an operation that we will assume to be a multiplication. An interested reader might refer to [3, 20, 21, 25] for notation, background, theory and further examples.

Differential equations on $\mathcal{G}$ can be written in the form

$$Y' = A(t, Y)Y, \qquad Y(0) = Y_0 \in \mathcal{G}, \tag{1}$$

where the function $A(t, Y)$ lies in $\mathfrak{g}$, the Lie algebra of the group $\mathcal{G}$. When A is independent of Y then (1) is an equation of *Lie type* [3]. In general we will assume that (1) is nonlinear. The Lie algebra $\mathfrak{g}$ is the tangent space to $\mathcal{G}$ at the identity and 'contains', in some sense, all the information associated with the group. We will refer mainly to matrix groups and the multiplication will simply be the matrix multiplication. Some recurring examples of Lie groups are:

- the *general linear group* $GL(d, \mathbb{R})$ of invertible matrices;

- the *special orthogonal group* $SO(d, \mathbb{R})$ of orthogonal matrices with the determinant equal to one;
- the *special linear group* $SL(d, \mathbb{R})$ of invertible matrices with the determinant equal to one;

with the Lie algebras

- $\mathfrak{gl}(d, \mathbb{R})$, the set of $d \times d$ real matrices;
- $\mathfrak{so}(d, \mathbb{R})$, the set of $d \times d$ skew-symmetric matrices;
- $\mathfrak{sl}(d, \mathbb{R})$, the set of $d \times d$ matrices with zero trace,

respectively. We will also consider the *symplectic group.* Having introduced the $2d \times 2d$ matrix

$$J = \begin{bmatrix} O_d & I_d \\ -I_d & O_d \end{bmatrix}, \tag{2}$$

we recall that a matrix R is *symplectic* [3] if

$$R^T J R = J. \tag{3}$$

The symplectic group $Sp(2d, \mathbb{R})$ is the set of all $(2d) \times (2d)$ real symplectic matrices. The corresponding Lie algebra is $\mathfrak{sp}(2d, \mathbb{R})$, where

$$A \in \mathfrak{sp}(2d, \mathbb{R}) \qquad \Longleftrightarrow \qquad A^T J + JA = O. \tag{4}$$

Finally we mention the main tools that we will use throughout the rest of this paper. The first is the *commutator* operator or *Lie bracket*,

$$[\,\cdot\,, \cdot\,] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}. \tag{5}$$

which, in the matrix case, is the commutator of two matrices, namely $[A, B] = AB - BA$.

The second is the *exponential* map,

$$\exp(\,\cdot\,) : \mathfrak{g} \to \mathcal{G}, \tag{6}$$

defined as

$$\exp A = \sum_{k=0}^{\infty} \frac{A^k}{k!}. \tag{7}$$

The exponential map and the commutator are the basic tools of our analysis.

2 The method of iterated commutators

The method of iterated commutators was introduced by Iserles in [13] to solve linear matrix systems with variable coefficients,

$$Y' = A(t)Y, \qquad Y(0) = Y_0, \tag{8}$$

where $A(t)$ is an analytic function. When the matrix A commutes with its derivative, the exact solution of (8) is given by

$$Y(t) = \exp\left(\int_0^t A(\tau)\,\mathrm{d}\tau\right) Y_0,$$

similarly to the scalar case. However, in the non-commutative case, in general

$$Y(t) \neq \exp\left(\int_0^t A(\tau)\,\mathrm{d}\tau\right) Y_0,$$

as one might have expected by a naive extrapolation. This is due to the fact that, letting $B(t) = \int_0^t A(\tau)\,\mathrm{d}\tau$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{B(t)} \neq A(t)\mathrm{e}^{B(t)},$$

since $A(t)$ and $B(t)$ in general do not commute. In [13] it is shown that the solution $Y(t)$ of (8) can be obtained as

$$Y(t) = \lim_{m\to\infty} \mathrm{e}^{B_0(t)} \cdots \mathrm{e}^{B_m(t)} Y_0, \tag{9}$$

where the matrix functions $B_i(t), i \in \mathbb{Z}^+$, are evaluated in the following way. First of all, we let $A_0(t) = A(t)$ and evaluate

$$B_0(t) = \int_0^t A_0(\tau)\,\mathrm{d}\tau. \tag{10}$$

Next we construct the sequence of commutators

$$C_{0,k+1}(t) = [C_{0,k}(t), B_0(t)], \quad k \geq 0, \qquad C_{0,0}(t) = A_0(t), \tag{11}$$

where

$$C_{0,k} = [C_{0,k-1}, B_0] = [[\cdots[A_0, \underbrace{B_0] \ldots, B_0], B_0]}_{k \text{ times}}, \quad k \geq 1.$$

Letting

$$A_1(t) = \sum_{k=1}^{\infty} \frac{k}{(k+1)!} C_{0,k}(t),$$

we set

$$B_1(t) = \int_0^t A_1(\tau)\,\mathrm{d}\tau.$$

The procedure can be repeated in an iterative manner. Assume that $A_i(t)$ and $B_i(t)$ are known, $i \geq 0$; then $C_{i,0}(t) = A_i(t)$ and, for $k \geq 0$, $C_{i,k+1}(t) = [C_{i,k}(t), B_i(t)]$. Thus

$$A_{i+1}(t) = \sum_{k=1}^{\infty} \frac{k}{(k+1)!} C_{i,k}(t), \tag{12}$$

and finally

$$B_{i+1}(t) = \int_0^t A_{i+1}(\tau)\,\mathrm{d}\tau. \tag{13}$$

Munthe-Kaas [19] shows that the method of iterated commutators for linear systems with variable coefficients is related to the Lie reduction method for equations of Lie type. The method of iterated commutators can be viewed as a 'numerical implementation' of the merely theoretical algorithm proposed by Sophus Lie. What is remarkable in the work of Iserles is not only that he derives his methods by a completely different approach, but also that an accurate analysis of the order of approximation is introduced for those groups, like matrix groups, that can be imbedded in $\mathbb{R}^q$ for some integer $q > 0$. This order analysis, as we shall see in §2.2, is valid also in a nonlinear case.

2.1 Order analysis

We wish to approximate (9) by

$$Y(t) \approx \mathrm{e}^{B_0(t)} \cdots \mathrm{e}^{B_M(t)} Y_0, \tag{14}$$

truncating the infinite product after some given index $M \in \mathbb{Z}^+$. Denoting by $\|\cdot\|$ the Frobenius norm on matrices, the order of approximation of the scheme (14) is given by the following result.

Theorem 1 (Theorem 3, [13]) *Let $Y^{[M]}(t) = \mathrm{e}^{B_0(t)} \cdots \mathrm{e}^{B_M(t)} Y_0$, where the matrices $B_i(t)$, $0 \le i \le M$, are evaluated as explained in §2. Then, if $A(t)$ is analytic, it is true that*

$$\|Y^{[M]}(t) - Y(t)\| = \mathcal{O}\left(t^{2^{M+2}-1}\right). \tag{15}$$

Therefore, for every integer $M \ge 0$ the formula (14) is an approximation of order $p = 2^{M+2} - 2$. □

The order remains the same if the integrals are replaced by quadrature formulæ of sufficiently high order and the infinite series are truncated to an appropriate finite number of terms. We will assume that the matrix exponentials are evaluated exactly. Thus, with a single exponential we can approximate up to order two, with two up to order six, etc. The order increases exponentially with the number of exponentials we use.

In the case when A commutes with its derivative, we have

$$B_1(t) = \cdots = B_M(t) = O$$

and the order of approximations depends only on the quadrature formula that we employ in approximating the integral $B_0(t)$.

2.2 Application to nonlinear problems and $\mathcal{G}$-invariance

The next step consists in adapting the method of iterated commutators to find the solution of the nonlinear problem

$$Y' = A(t, Y)Y, \qquad Y(0) = Y_0. \tag{16}$$

In this case, we set $A_0(t) = A(t, Y(t))$ and for all $i \ge 0$ let

$$\begin{aligned} B_i(t) &= \int_0^t A_i(\tau)\,\mathrm{d}\tau, \\ C_{i,k+1}(t) &= [C_{i,k}(t), B_i(t)], \quad k \ge 0, \qquad C_{i,0}(t) = A_i(t), \end{aligned}$$

and finally

$$A_{i+1}(t) = \sum_{k=1}^{\infty} \frac{k}{(k+1)!} C_{i,k}(t).$$

All the above functions depend implicitly on the unknown function $Y(t)$. For the time being, however, we assume that this function is known.[1] All the proofs in [13] still hold, assuming that A_0 can be expressed as an analytic (for our purposes it is sufficient that A_0 is sufficiently smooth) function of t. Therefore it is also true that

$$Y^{[M]}(t) = \exp(B_0) \cdots \exp(B_M) Y_0$$

approximates the true solution $Y(t)$ to order $p = 2^{M+2} - 2$. In particular, we recall from [13] that the approximate solution $Y^{[M]}$ is the exact solution of the perturbed problem

$$Y^{[M]\prime} - A(t,Y) Y^{[M]} = -Y^{[M]} A_{M+1}(t),$$

where $A_{M+1}(t) = \mathcal{O}\left(t^{2^{M+2}-2}\right)$. This implies that the error in the approximation is given by the integral formula

$$\|Y^{[M]}(t) - Y(t)\| = \left\| - \int_0^t Y(t-\tau) Y^{[M]}(\tau) A_{M+1}(\tau)\, \mathrm{d}\tau \right\| = \mathcal{O}\left(t^{2^{M+2}-1}\right),$$

which is valid in a compact interval $0 \le t \le t^*$.

Our main observation is that the method of iterated commutators uses exclusively basic operations of Lie algebras and Lie groups. Therefore it allows us to construct $\mathcal{G}$-invariant numerical methods, as explained in the following result.

Theorem 2 *If Y_0 is in a Lie group $\mathcal{G}$ with Lie algebra $\mathfrak{g}$ and $A(t,Y) \in \mathfrak{g}$ for $Y \in \mathcal{G}$, then the method of iterated commutators is $\mathcal{G}$-invariant, namely also the approximate solution evolves in $\mathcal{G}$.*

Proof. Since $A(t, Y(t)) \in \mathfrak{g}$, it follows that also $B_0 \in \mathfrak{g}$. Moreover, for every $x, y \in \mathfrak{g}$, also $[x,y] \in \mathfrak{g}$ (cf. [3, 20]) and we conclude that $C_{0,k} \in \mathfrak{g}$. Thus, by induction on $i \in \mathbb{Z}^+$, we deduce that $A_i, B_i \in \mathfrak{g}$, therefore $\exp(B_i) \in \mathcal{G}$. Hence the theorem follows from the group properties of $\mathcal{G}$. $\square$

In practice, the integrals B_i will be replaced by quadrature. As long as the latter is of the form

$$B_i(t) = t \sum_{j=1}^{\nu} b_j A(c_j t, Y_j),$$

where $b_j, c_j \in \mathbb{R}$, for $j = 1, 2 \ldots, \nu$, and $Y_j \in \mathcal{G}$ is an approximation of $Y(t_j)$, then $B_i \in \mathfrak{g}$ and Theorem 2 remains valid.

[1] In the sequel we show that it is possible to obtain suitable approximations for requisite values of t.

2.3 $\mathcal{G}$-invariant numerical methods

In this section we show in detail how to generate $\mathcal{G}$-invariant methods from the method of iterated commutators when applied to nonlinear systems. We will assume for the time being, that $A(t,Y) \in \mathfrak{g}$ for $Y \in \mathcal{G}$ while $A(t,Y) \notin \mathfrak{g}$ for some $Y \notin \mathcal{G}$.

We denote by Y_n the approximation produced by the numerical method at time t_n, where $0 = t_0 < t_1 < \cdots < t_n < \cdots < t_N$ is a partition of $[0, t^*]$, the interval of integration. For the order analysis, we assume that $Y_n = Y(t_n)$. The stepsize of integration is $h_n = t_{n+1} - t_n$.

- **Orders one and two.** From the order analysis in §2.1, methods of order one and two can be obtained by means of a single exponential transformation.

  - **order one:** Consider the scheme

    $$Y_{n+1} = \exp(h_n A_n) Y_n, \qquad A_n = A(t_n, Y_n). \tag{17}$$

    Assuming that $Y_n \in \mathcal{G}$, it follows that $A_n \in \mathfrak{g}$ hence $Y_{n+1} \in \mathcal{G}$. Therefore, by induction on n, the method is $\mathcal{G}$-invariant provided that the initial condition satisfies $Y_0 \in \mathcal{G}$. The order of the method is one because the integral $\int_{t_n}^{t_{n+1}} A(t, Y(t))\, dt$ is approximated by first-order quadrature.

    In passing we mention that a version of this method is presented in [15], where an associated orthogonal flow is solved in order to approximate the solution of a *double-bracket* flow $H' = [H,[H,N]]$, an isospectral flow introduced by Brockett [1]. Isospectral flows can be obtained as orbits of group actions [3, 20, 2, 21, 24], and their numerical solution can be reduced to finding the solution of a differential equation on a Lie group.

  - **order two:** Consider the scheme

    $$Y_{n+1} = \exp\left(h_n A\left(t_n + \tfrac{1}{2}h_n, Y(t_n + \tfrac{1}{2}h_n)\right)\right) Y_n. \tag{18}$$

    Note that the integral B_0 is replaced by an order-two quadrature rule (the midpoint rule). However we need an approximation in $\mathcal{G}$ for $Y(t_n + \frac{1}{2}h_n)$. To this end, we use an order-one formula. Thus we evaluate $Y_{n+\frac{1}{2}} = \exp(\frac{1}{2}h_n A_n) Y_n = Y(t_n + \frac{1}{2}h_n) + \mathcal{O}(h_n^2)$. Instead of (18) we employ

    $$Y_{n+1} = \exp\left(h_n A_{n+\frac{1}{2}}\right) Y_n, \qquad A_{n+\frac{1}{2}} = A\left(t_n + \tfrac{1}{2}h_n, Y_{n+\frac{1}{2}}\right), \tag{19}$$

    where $A_{n+\frac{1}{2}} = A\left(t_n + \frac{1}{2}h_n, Y(t_n + \frac{1}{2}h_n)\right) + \mathcal{O}(h_n^2)$, and multiplication by h_n in the exponential yields the correct order.

- **Order $3 \le p \le 6$.** For higher-order schemes we need to consider the product of two exponentials. Thus, we approximate

  $$Y_{n+1} = \exp(B_0)\exp(B_1) Y_n.$$

– **order three:** We approximate B_0 using a two-point Gauss–Legendre quadrature rule. Thus,

$$Y_{n+1} = \exp\left(\frac{h_n}{2}(K_1 + K_2)\right)\exp(B_1)Y_n, \tag{20}$$

where

$$K_1 = A\left(t_n + c_1^{(0)}h_n, X_1\right), \qquad K_2 = A\left(t_n + c_2^{(0)}h_n, X_2\right), \tag{21}$$

$c_1^{(0)}$ and $c_2^{(0)}$ being the nodes of the Gauss–Legendre quadrature in $[0, 1]$,

$$c_1^{(0)} = \tfrac{1}{2} - \tfrac{\sqrt{3}}{6}, \qquad c_2^{(0)} = \tfrac{1}{2} + \tfrac{\sqrt{3}}{6}.$$

The weights of the quadrature formula are $b_1^{(0)} = b_2^{(0)} = \frac{1}{2}$. The matrices X_1 and X_2 in (21) are order-two approximations to the theoretical values $Y(t_n+c_1^{(0)}h_n)$ and $Y(t_n+c_2^{(0)}h_n)$ respectively and can be evaluated with the order-two formula which has been already described. Although the two-point Gauss–Legendre rule has order four, the order of the method will be three, since X_1 and X_2 are just order-two approximations.

Next, we need to evaluate the matrix

$$B_1 = \int_{t_n}^{t_{n+1}} A_1(t)\,\mathrm{d}t.$$

Since by construction A_1 is of order $\mathcal{O}(h_n^2)$ and integration yields $B_1 = \mathcal{O}(h_n^3)$, we need to evaluate the integral B_1 with at least an order-two quadrature rule, and, given the particular form of A_1, we choose the weight function to be $w(t) = t^2$. Thus the error will be subsumed in the $\mathcal{O}(h_n^4)$ term, consistently with our order analysis. Also, for order three, it is sufficient to truncate the expansion of A_1 at $C_{0,2}$. Therefore, we set

$$A_1 = \tfrac{1}{2}C_{0,1} + \tfrac{1}{3}C_{0,2}.$$

The matrix functions $A_1(t), C_{0,1}(t)$ and $C_{0,2}(t)$ need all be evaluated at the nodes of the second quadrature rule. It is convenient to use the *Gauss–Radau* formula for integration in $[0, 1]$ with two nodes (one of which fixed at 1) and the weight function $w(t) = t^2$. The nodes need be linearly shifted to the interval $[t_n, t_{n+1}]$. Note that by constraining one node to the end of the interval we obtain for free the value of the function B_0 at the end of the interval $[t_n, t_{n+1}]$. Imposing order two, we obtain the following equations to be satisfied by the node $c_1^{(1)}$ and the weights $b_1^{(1)}, b_2^{(1)}$ of the second quadrature rule:

$$\begin{aligned} b_1^{(1)}\left(c_1^{(1)}\right)^2 + b_2^{(1)} &= \tfrac{1}{3}, \\ b_1^{(1)}\left(c_1^{(1)}\right)^3 + b_2^{(1)} &= \tfrac{1}{4}. \end{aligned} \tag{22}$$

The system (22) has an infinite number of solutions.[2] A particular solution is obtained by choosing $c_1^{(1)} = \frac{1}{2}$, which yields $b_1^{(1)} = \frac{2}{3}$ and finally $b_2^{(1)} = \frac{1}{6}$. Recall that $c_2^{(1)} = 1$.

– **order four:** As for order three, we approximate B_0 using the two-point Gauss–Legendre scheme. However we need to evaluate X_1 and X_2 with order three and to this end we can use the order-three method. Moreover, we require that the integral B_1 is evaluated with a third-order scheme and again we may wish to use a Gauss–Radau formula. Thus we need to add an extra order condition to (22), which is given by

$$b_1^{(1)} \left(c_1^{(1)}\right)^4 + b_2^{(1)} = \tfrac{1}{5}. \tag{23}$$

This results in

$$\begin{aligned} c_1^{(1)} &= \tfrac{3}{5}, & b_1^{(1)} &= \tfrac{125}{216}, \\ c_2^{(1)} &= 1, & b_2^{(1)} &= \tfrac{1}{8}, \end{aligned} \tag{24}$$

which is the unique solution consistent with our requirements.

The same procedure generalizes for order $p = 5, 6$ methods. The integral B_0 can be evaluated using order-six Gauss–Legendre quadrature rule and, similarly, for the integral B_1 we need to use a Gauss–Radau formula with three nodes, with the constraint $c_3^{(1)} = 1$. The quadrature nodes and weights are given respectively by

$$\begin{aligned} c_1^{(0)} &= \tfrac{1}{2} - \tfrac{\sqrt{15}}{10}, & b_1^{(0)} &= \tfrac{5}{18} \\ c_2^{(0)} &= \tfrac{1}{2}, & b_2^{(0)} &= \tfrac{4}{9} \\ c_3^{(0)} &= \tfrac{1}{2} + \tfrac{\sqrt{15}}{10}, & b_3^{(0)} &= \tfrac{5}{18} \\ c_1^{(1)} &= \tfrac{4-\sqrt{2}}{7}, & b_1^{(1)} &= \tfrac{184+11\sqrt{2}}{480}, \\ c_2^{(1)} &= \tfrac{4+\sqrt{2}}{7}, & b_2^{(1)} &= \tfrac{184-11\sqrt{2}}{480}, \\ c_3^{(1)} &= 1, & b_3^{(1)} &= \tfrac{1}{15}. \end{aligned} \tag{25}$$

For order higher than six we need to add another exponential. The procedure generalizes in a similar manner, and the quadrature rule will have weight functions $w(t) = 1$ for B_0, $w(t) = t^2$ for B_1, $w(t) = t^4$ for B_2, and so on.

A detailed implementation of our methods is explained in §4.1.

As we explain in §2.4, the product of exponentials can be replaced with a single exponential using the *Baker–Campbell–Haursdoff* formula [23, 25] and truncating it to an appropriate order. Many other variants are possible, considering various quadrature formulæ which constrain one or more nodes, for instance Gauss–Radau, as we have already

[2] A ν-point Gauss–Radau quadrature rule yields order $2\nu - 1$, therefore with two points we can achieve order three, while (22) imposes order two only. Since the system is overdetermined, we can choose $b_2^{(1)} = 0$, in which case we do not have anymore Radau quadrature. In this case, the order is exactly two and the number of nodes is reduced. The values of the remaining parameters are: $c_1^{(1)} = \frac{4}{3}, b_1^{(1)} = \frac{3}{16}$.

seen, or Gauss–Lobatto. This might reduce the computational costs of the algorithm. However, such issues are outside the scope of this paper and will be addressed in future work.

In Figure 1 we have plotted the global error against stepsize for the methods of order one, two, three and four which have been previously described. Since both axes are drawn to a logarithmic scale, the order corresponding to the method in question can be found by measuring the slope of its global error line. The circles joined by the solid line correspond to the method of order one, the stars joined by the dotted line to the method of order two, the plus signs joined by the dashed line correspond to the order-three scheme, and finally then times signs joined by the dash-dotted line to the scheme of order four. The numerical experiments (in `Matlab`) were performed evaluating the global error at $t^* = 1$ for an orthogonal flow, the integration being implemented with constant stepsizes $h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}$. The matrix function A used in the experiments (in `Matlab` formalism) was `A(t, Y) = diag(diag(Y,+1),+1) - diag(diag(Y,+1),-1)` while the initial condition was generated taking the matrix Q from `[Q,R] = qr(rand(5))`.

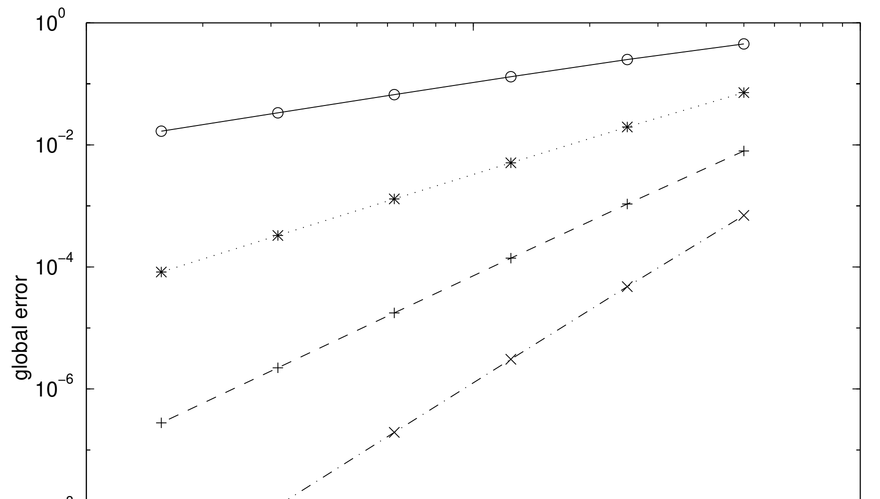

Figure 1: Global error against stepsize in logarithmic scale for methods of order one (circles joined by solid line), order two (asterisks joined by dotted line), order three (plus signs joined by dashed line) and order four (times signs joined by dash-dotted line)

Although the schemes might appear expensive, their implementation can be simplified when $A(t, Y) \in \mathfrak{g}$ for all $d \times d$ matrices Y, in which case we need not use an order-$(p-1)$ $\mathcal{G}$-invariant method to obtain correct order approximations to the X_i's in (21) and arbitrary methods can be used instead. The simplification is particularly striking when A depends only on time t.

Note that the described schemes are explicit, in the sense that Y_{n+1} can be explicitely evaluated without solving any nonlinear systems of equations.

2.4 The Baker–Campbell–Hausdorff formula

The Baker–Campbell–Hausdorff formula [23, 25] allows us to express the product of exponentials in terms of an exponential of a sum. Assume that X and Y are two given matrices, then

$$\begin{aligned}\exp(X) &= I + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots,\\ \exp(Y) &= I + Y + \frac{Y^2}{2!} + \frac{Y^3}{3!} + \cdots,\end{aligned}$$

thus

$$\exp(X)\exp(Y) = I + X + Y + \frac{X^2}{2} + XY + \frac{Y^2}{2} + \cdots.$$

We can write $\exp(X)\exp(Y) = \exp(Z)$, where Z is given by

$$Z = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}([X,[X,Y]] + [Y,[Y,X]]) + \cdots. \tag{26}$$

The equation (26) is known as the *Baker–Campbell–Hausdorff (BCH) formula.* Since all the terms in (26) are just X, Y and iterated commutators of X and Y [23], it follows that if $X, Y \in \mathfrak{g}$ then $Z \in \mathfrak{g}$ and therefore $\exp(Z) \in \mathcal{G}$.

Suppose that we use the method of iterated commutators, as explained beforehand, to advance from t_n to t_{n+1} with stepsize $h_n = t_{n+1} - t_n$. By replacing the product of exponentials with a single exponential, using the BCH formula, we can obtain cheaper methods, since evaluating the exponential is an expensive task. To retain the order of the method of iterated commutators, we need to truncate the new symbol to an appropriate order. Note that

$$[h_n^r C_r, h_n^s C_s] = h_n^{r+s}[C_r, C_s],$$

and that $B_i = \mathcal{O}\left(h_n^{2^{i+2}-1}\right)$, for $i = 0, 1, \ldots, M$. Therefore, we assume that $Y_n^{[M]} = Y(t_n)$ and for order $3 \leq p \leq 6$ we have

$$Y_{n+1}^{[M]} = \exp(B_0)\exp(B_1)Y_n^{[M]}.$$

We can replace the latter formula by

$$Y_{n+1}^{[M]} = \exp(\tilde{B})Y_n^{[M]}$$

where the new argument $\tilde{B}$ is given by

$$\begin{aligned}\tilde{B} &= B_0 + B_1, && \text{order } p = 3,\\ \tilde{B} &= B_0 + B_1 + \tfrac{1}{2}[B_0, B_1], && \text{order } p = 4,\\ \tilde{B} &= B_0 + B_1 + \tfrac{1}{2}[B_0, B_1] + \tfrac{1}{12}[B_0,[B_0,B_1]], && \text{order } p = 5, 6.\end{aligned}$$

To obtain formulæ of order $p \geq 7$, we need to find the symbol for the product of three exponentials. We use again the BCH formula to find

$$\exp X \exp Y \exp W = \exp Z \exp W = \exp \Omega,$$

where

$$\begin{aligned} \Omega &= Z + W + \frac{1}{2}[Z, W] + \frac{1}{12}([Z,[Z,W]] + [W,[W,Z]]) + \cdots \\ &= X + Y + W + \frac{1}{2}([X,Y] + [X,W] + [Y,W]) + \frac{1}{12}([X,[X,Y]] + \cdots) + \cdots . \end{aligned}$$

Furthermore, we conclude by virtue of the BCH formula that there exists a matrix $S \in \mathfrak{g}$, such that the numerical approximation after n steps can be written compactly in the form

$$Y_n = \exp(S) Y_0. \tag{27}$$

3 An application to Hamiltonian systems

We wish to solve the Hamiltonian system

$$\begin{aligned} \boldsymbol{p}' &= -\frac{\partial H}{\partial \boldsymbol{q}}, \qquad & \boldsymbol{p}(0) = \boldsymbol{p}_0, \\ \boldsymbol{q}' &= \frac{\partial H}{\partial \boldsymbol{p}}, \qquad & \boldsymbol{q}(0) = \boldsymbol{q}_0, \end{aligned} \tag{28}$$

where $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^d$. The solution $(\boldsymbol{p}, \boldsymbol{q})$ of (28) can be written in terms of the *solution operator* $\Phi_H(t, 0)$,

$$(\boldsymbol{p}(t), \boldsymbol{q}(t)) = \Phi_H(t,0)(\boldsymbol{p}_0, \boldsymbol{q}_0).$$

The operator Φ_H satisfies

$$\Phi_H(t_2, t_0) = \Phi_H(t_2, t_1)\Phi_H(t_1, t_0), \qquad \forall t_0, t_1, t_2 \in I,$$

where I is a real interval. In particular, if H is autonomous, then $\Phi_H(t, t_0)$ depends only on the elapsed time $t - t_0$ and, following [23], will be denoted by $\Phi_{H,t-t_0}$. In other words, it obeys the semigroup properties

$$\Phi_{H,t+s} = \Phi_{H,t}\Phi_{H,s}.$$

Definition 1 [23] Suppose that ψ maps $(\boldsymbol{p}, \boldsymbol{q})$ to $(\boldsymbol{p}^*, \boldsymbol{q}^*)$. Then the map ψ is symplectic if the Jacobian matrix

$$\psi' = \frac{\partial(\boldsymbol{p}^*, \boldsymbol{q}^*)}{\partial(\boldsymbol{p}, \boldsymbol{q})} = \frac{\partial \psi(\boldsymbol{p}, \boldsymbol{q})}{\partial(\boldsymbol{p}, \boldsymbol{q})}$$

is symplectic. $\diamond$

Recalling the definition of the symplectic group introduced in §1.1, according to (3), the above definition is equivalent to the requirement that

$$(\psi')^T J \psi' = J, \tag{29}$$

where J is the matrix in (2).

Theorem 3 (Theorem 2.1, [23]) *For each t, t_0, the solution operator $\Phi_H(t, t_0)$ of the Hamiltonian system* (28) *is symplectic.* □

If $A \in Sp(2d, \mathbb{R})$, it follows from (3) that A is invertible and $\det A = \pm 1$. Thus the symplectic group splits into two connected components. If $\gamma(t)$ is a curve in $Sp(2d, \mathbb{R})$ and is continuous as a function of time, then it must evolve in one of the two connected components, which is uniquely determined by the initial condition. Thus, for the Jacobian matrix of the Hamiltonian flow, we have $\Phi'_H|_{t=0} = I$ and we deduce that $\det \Phi'_H = 1$ for all $t \geq 0$. This feature is also known as *conservation of oriented volume* in $\mathbb{R}^{2d}$ (cf. [23]).

For the numerical solution produced by a group-invariant method, the following result holds.

Theorem 4 *Suppose that the Hamiltonian system* (28) *is written in the form* (1), *where $A \in \mathfrak{sp}(2d, \mathbb{R})$. Then the Jacobian matrix of the the numerical solution produced by a Lie group invariant method obeys the equality*

$$\frac{\partial \psi(\boldsymbol{p}_n, \boldsymbol{q}_n)}{\partial(\boldsymbol{p}_n, \boldsymbol{q}_n)} = \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right) + \frac{\partial \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right)}{\partial(\boldsymbol{p}_n, \boldsymbol{q}_n)} \begin{bmatrix} \boldsymbol{p}_n \\ \boldsymbol{q}_n \end{bmatrix}, \tag{30}$$

where $\exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right) \in Sp(2d, \mathbb{R})$. *Therefore if*

$$\frac{\partial \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right)}{\partial(\boldsymbol{p}_n, \boldsymbol{q}_n)} \begin{bmatrix} \boldsymbol{p}_n \\ \boldsymbol{q}_n \end{bmatrix} = O, \tag{31}$$

then the method is symplectic.

Proof. By using the BCH formula, it follows that the numerical approximation can be written in the form

$$\begin{bmatrix} \boldsymbol{p}_{n+1} \\ \boldsymbol{q}_{n+1} \end{bmatrix} = \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right) \begin{bmatrix} \boldsymbol{p}_n \\ \boldsymbol{q}_n \end{bmatrix}, \tag{32}$$

for some $\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n) \in \mathfrak{sp}(2d, \mathbb{R})$. Therefore the numerical map $\psi(\boldsymbol{p}_n, \boldsymbol{q}_n)$ is simply

$$\psi(\boldsymbol{p}_n, \boldsymbol{q}_n) = \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right) [\boldsymbol{p}_n, \boldsymbol{q}_n]^T,$$

and its Jacobian matrix is

$$\psi' = \frac{\partial \psi(\boldsymbol{p}_n, \boldsymbol{q}_n)}{\partial(\boldsymbol{p}_n, \boldsymbol{q}_n)} = \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right) + \frac{\partial \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right)}{\partial(\boldsymbol{p}_n, \boldsymbol{q}_n)} \begin{bmatrix} \boldsymbol{p}_n \\ \boldsymbol{q}_n \end{bmatrix}.$$

(Note that $\exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right)$ is a $2d \times 2d$ matrix, therefore

$$\frac{\partial \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right)}{\partial(\boldsymbol{p}_n, \boldsymbol{q}_n)}$$

is a tensor and multiplication by $[\boldsymbol{p}_n, \boldsymbol{q}_n]^T$ on the right yields a $2d \times 2d$ matrix.)

Since $\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n) \in \mathfrak{sp}(2d, \mathbb{R})$, we deduce $\exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right) \in Sp(2d, \mathbb{R})$ and it obeys the identity

$$\exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right)^T J \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right) = J,$$

therefore if (31) is satisfied then the numerical method is symplectic. □

Corollary 4.1 *Lie-group invariant methods and the method of iterated commutators applied to an Hamiltonian system (28) written in the form (1), are symplectic when the matrix function $A \in \mathfrak{sp}(2d, \mathbb{R})$ is constant or when it depends only on t.* □

In the more general case, when A depends on $\boldsymbol{p}$ and $\boldsymbol{q}$ (nonlinear systems), it is necessary to check symplecticity for each single method.

In other words, Theorem 4 says that the Jacobian matrix of the map (32) can be decomposed into two terms. The first is automatically a symplectic matrix, while the contribution of the second term can be interpreted as the departure from symplecticity.

Differentiation of the exponential is a difficult task. From the formulæ in [13, 17] we deduce

$$\frac{\partial \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right)}{\partial(\boldsymbol{p}_n, \boldsymbol{q}_n)} = \exp\left(\tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)\right) \sum_{j=0}^{\infty} \frac{1}{(j+1)!} \sum_{i=0}^{j} (-1)^i \binom{j}{i} \tilde{B}^i \tilde{B}' \tilde{B}^{j-i},$$

whereby $\tilde{B} = \tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)$ and $\tilde{B}' = \partial \tilde{B}(\boldsymbol{p}_n, \boldsymbol{q}_n)/\partial(\boldsymbol{p}_n, \boldsymbol{q}_n)$. Hence the condition (31) is quite complicated to analyse in order to verify symplecticity. However we hope that this approach will allow us to introduce explicit symplectic maps.

It is an easy exercise to verify that also the proposed methods, when symplectic, preserve oriented volume [23]. In such cases, there exists a $\tilde{S} \in \mathfrak{sp}(2d, \mathbb{R})$ so that the Jacobian matrix of the numerical methods is $\psi' = \exp(\tilde{S})$. From $\tilde{S} \in \mathfrak{sp}(2d, \mathbb{R})$ it follows that $\mathrm{tr}(\tilde{S}) = 0$. Thus

$$\begin{aligned} \det \psi' &= \det(\exp(\tilde{S})) = \prod \lambda(\exp(\tilde{S})) = \prod \exp(\lambda(\tilde{S})) \\ &= \exp(\mathrm{tr}(\tilde{S})) = 1, \end{aligned} \tag{33}$$

where the product is carried over the spectrum. Also it is clear that conservation of volume is a weaker condition than symplecticity. As a matter of fact, (33) is true for all $\tilde{S} \in \mathfrak{sl}(2d, \mathbb{R})$. The symplectic algebra is a subalgebra of $\mathfrak{sl}(2d, \mathbb{R})$.

Next we discuss how to write the Hamiltonian system (28) in the form (1). First of all, we stipulate that the Hamiltonian system has a fixed point at the origin. Suppose that A is a $2d \times 2d$ real matrix. We partition it into four $d \times d$ blocks $A_{i,j}$, $i, j = 1, 2$, so that

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}.$$

Therefore $A \in \mathfrak{sp}(2d, \mathbb{R})$ if and only if

$$A_{1,1}^T + A_{2,2} = O_d, \quad \text{and} \quad A_{1,2} = A_{1,2}^T,\ A_{2,1} = A_{2,1}^T. \tag{34}$$

Hence we must seek $A \equiv A(\boldsymbol{p}, \boldsymbol{q})$ such that

$$\begin{aligned} A_{1,1}\boldsymbol{p} + A_{1,2}\boldsymbol{q} &= -\frac{\partial H}{\partial \boldsymbol{q}}, \\ A_{2,1}\boldsymbol{p} - A_{1,1}^T\boldsymbol{q} &= \frac{\partial H}{\partial \boldsymbol{p}}. \end{aligned}$$

Since the dimension of $\mathfrak{sp}(2d, \mathbb{R})$ is $2d^2 + d$ and we impose only $2d$ conditions, in the general case we have many choices for the representation A and this freedom can be used to meet extra requirements. For instance, we could ask that the eigenvalues of A reside in the complex left half plane, since numerical experiments seem to indicate that eigenvalues with positive real part might cause the numerical method to be unstable.

If H is a *separable* Hamiltonian,

$$H(\boldsymbol{p}, \boldsymbol{q}) = T(\boldsymbol{p}) + V(\boldsymbol{q}),$$

then it follows that $A_{1,1} = O_d$. In particular, when

$$T(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T\boldsymbol{p}, \tag{35}$$

we can choose an arbitrary symmetric matrix $A_{2,1}$ that has an eigenvalue equal to one and $\boldsymbol{p}$ as a corresponding eigenvector. One particular choice is $A_{2,1} = I_d$.

The quadratic Hamiltonian is particularly simple. If also

$$V(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T\boldsymbol{q},$$

then we can choose $A_{1,2} = -I_d$. Thus, the system is equivalent to

$$\begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{bmatrix}' = -J \begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{bmatrix}, \tag{36}$$

which is the same as $[\boldsymbol{p}', \boldsymbol{q}']^T = -J\nabla H$, given the particular form of H. In this case, since J is constant, the method of iterated commutators yields the exact solution

$$\begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{bmatrix} = \mathrm{e}^{-tJ} \begin{bmatrix} \boldsymbol{p}_0 \\ \boldsymbol{q}_0 \end{bmatrix},$$

where

$$\mathrm{e}^{-tJ} = I\cos t - J\sin t.$$

Another example is the nonlinear pendulum with Hamiltonian function

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

(cf. [22]). A simple calculation allows us to write the system in the form

$$\begin{bmatrix} p \\ q \end{bmatrix}' = \begin{bmatrix} 0 & -2(1-\cos q) - q\sin q \\ 1 & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}, \tag{37}$$

where

$$A(p,q) = \left[\begin{array}{cc} 0 & -2(1-\cos q) - q\sin q \\ 1 & 0 \end{array} \right] \in \mathfrak{sp}(2,\mathbb{R}), \qquad p,q \in \mathbb{R}.$$

Although it is not guaranteed that the proposed methods are symplectic, numerical calculations seem to indicate their performance is superior with comparison to standard ODEs methods of the same order. For some numerical experiments, we refer the reader to §4.

Careful analysis should establish whether these methods share the same attribute, e.g. linear error propagation in case of periodic orbits, of standard symplectic integrators [6], or energy conserving numerical schemes [9]. In fact, these attributes of classical ODE schemes rely on the form of the error expansion. This and more is being currently investigated.

4 Numerical examples

4.1 Implementation of the method of iterated commutators

In this section we show in detail how to implement the method of iterated commutators for order $3 \le p \le 6$ (we proceed as in [13]). In a similar way the procedure generalizes to higher order methods.

For ease of notation, we identify the quantities A_i, B_i etc. with their approximations.

Let ν_0 and ν_1 be the number of quadrature nodes for the integration of B_0 and B_1 respectively. Thus $c_i^{(0)}, b_i^{(0)}$, $1 \le i \le \nu_0$, and $c_i^{(1)}, b_i^{(1)}$, $1 \le i \le \nu_1$, will denote the nodes (in $[0,1]$) and the weights of the quadrature rules. Furthermore we will denote by $\tilde{Y}(t)$ an approximation of appropriate order of the exact value $Y(t)$, as explained in §2.3. Assume that $Y_n \in \mathcal{G}$ is known. We set

$$c_0^{(1)} = 0, \qquad B_0(t_n) = O.$$

`for` $1 \le i \le \nu_1$

- Evaluate $\tilde{Y}(t_n + [c_{i-1}^{(1)} + (c_i^{(1)} - c_{i-1}^{(1)})c_j^{(0)}]h_n)$ with a $\mathcal{G}$-invariant method of order $p-1$.
- Calculate
$$\begin{aligned} B_0(t_n + c_i^{(1)}h_n) &= B_0(t_n + c_{i-1}^{(1)}h_n) + (c_i^{(1)} - c_{i-1}^{(1)})h_n \sum_{j=1}^{\nu_0} b_j^{(0)} A_0 \Big(t_n + [c_{i-1}^{(1)} \\ &\quad + (c_i^{(1)} - c_{i-1}^{(1)})c_j^{(0)}]h_n, \tilde{Y}(t_n + [c_{i-1}^{(1)} + (c_i^{(1)} - c_{i-1}^{(1)})c_j^{(0)}]h_n)\Big). \end{aligned}$$
- Evaluate $\tilde{Y}(t_n + c_i^{(1)}h_n) \in \mathcal{G}$ with order $p-1$.
- Calculate
$$C_{0,0}(t_n + c_i^{(1)}h_n) = A_0\left(t_n + c_i^{(1)}h_n, \tilde{Y}(t_n + c_i^{(1)}h_n)\right).$$
- Set $B_0(t_{n+1}) = B_0(t_n + c_{\nu_1}^{(1)}h_n)$.

- For $1 \le k \le n_c$, where n_c is the number commutators required by the formula, evaluate
$$C_{0,k}(t_n + c_i^{(1)}h_n) = \left[C_{0,k-1}(t_n + c_i^{(1)}h_n), B_0(t_n + c_i^{(1)}h_n)\right].$$
- Set
$$A_1(t_n + c_i^{(1)}h_n) = \sum_{k=1}^{n_c} \frac{k}{(k+1)!} C_{0,k}(t_n + c_i^{(1)}h_n).$$

`end`

We have all the necessary ingredients to find $B_1(t_n + c_i^{(1)}h_n)$, which is given by

$$B_1(t_n + c_i^{(1)}h_n) = h_n \sum_{i=1}^{\nu_1} b_i^{(1)} A_1(t_n + c_i^{(1)}h_n).$$

Finally, the numerical approximation to Y_{n+1} is

$$Y_{n+1} = \exp\left(B_0(t_n + h_n)\right) \exp\left(B_1(t_n + h_n)\right) Y_n,$$

whereby the product of exponentials is replaced by a single exponential by means of the BCH formula, as explained in §2.4.

4.2 Integration on $SU(d, \mathbb{C})$

One of the most illustrative examples to test Lie-group integrators consists in discretizing ordinary differential equations on the Lie group $SU(d, \mathbb{C})$ which consists of $d \times d$ unitary matrices with determinant equal to one. The corresponding Lie algebra $\mathfrak{su}(d, \mathbb{C})$ is the set of all skew-hermitian $d \times d$ complex matrices with zero trace. $SU(d, \mathbb{C})$ is a subgroup of $U(d, \mathbb{C})$, the group of unitary matrices, and $\mathfrak{su}(d, \mathbb{C})$ is a subalgebra of $\mathfrak{u}(d, \mathbb{C})$, the set of complex skew-hermitian $d \times d$ matrices.

Integration of unitary flows has been already studied in the last few years [8, 12, 14] and unitary methods are nowadays widely available. For instance, when the function $A(t, Y)$ is always skew-hermitian, it is well known that in the class of Runge–Kutta (RK) schemes, methods obeying the condition $M = O$ are unitary [12, 14]. The matrix $M = (m_{i,j})$ is obtained combining the coefficients $a_{i,j}$ and the weights b_i of an s-stage RK method as

$$m_{i,j} = b_i a_{i,j} + b_j a_{j,i} - b_i b_j, \qquad i, j = 1, 2, \ldots s.$$

In particular, Gauss–Legendre RK methods are unitary [8, 14].

Although in the real case, orthogonality of the method guarantees $SO(d)$-invariance, this is no longer the case when we wish to integrate on $SU(d, \mathbb{C})$. The orthogonality of the Gauss–Legendre RK schemes is a consequence of their retention of quadratic conservation laws. The conservation of the determinant, instead, is an order-d conservation law in the entries of the underlying matrix, and, if the eigenvalues change smoothly, also the determinant is a smooth function of the eigenvalues. Since $O(d)$ consists of two connected components, one corresponding to matrices having determinant equal to $+1$, the other to those having determinant -1, each continuous flow must evolve in one of the two connected components. Without loss of generality we can assume that this component is $SO(d)$. For the numerical scheme, provided that the stepsize used for integration is sufficiently small, the determinant will be recovered correctly, because also the numerical solution evolves in the correct connected component. Therefore orthogonality implies $SO(d)$-invariance.

This is not any more the case when $\mathcal{G} = SU(d,\mathbb{C})$. Recall that a unitary matrix has a determinant of modulus equal to one but which can have an arbitrary argument. Therefore $U(d,\mathbb{C})$ consists of a unique connected component.

Although with a unitary RK method unitarity can be preserved, as a consequence of Theorem 1 in [5], the determinant will not be preserved, being free to move on the unit circle in the complex plane as long as we consider a system with $d \geq 3$. In other words: RK methods can be $U(d,\mathbb{C})$-invariant but not $SU(d,\mathbb{C})$-invariant.

In figure 2 we have plotted the evolution of the determinant for the implicit midpoint rule (IMR) applied to the ODE (1) with the matrix function A given by

$$A(Y) = (\mathrm{Re}\, Y - \mathrm{Re}\, Y^T) + \mathrm{i}(\mathrm{Im}\, Y + \mathrm{Im}\, Y^T - 2\mathrm{diag}(Y). \tag{38}$$

Since IMR obeys $M = O$ and $A(Y) \in \mathfrak{su}(d,\mathbb{C}) \subset \mathfrak{u}(d,\mathbb{C})$ for all Y, the scheme is unitary. The numerical approximations (corresponding to various initial conditions in $SU(d,\mathbb{C})$ for various d) should be thought as starting at $(1,0,0)$ and evolving on a cylinder of radius equal to one. The axis of the cylinder corresponds to time.

Note that, since the error for unitary methods is tangent to the cylinder, small errors in the real part of the determinant cause larger errors in its imaginary part: writing $\det Y_n = x_n + \mathrm{i}z_n$ and letting $x_n = 1 + \varepsilon_n$, the identity $x_n^2 + z_n^2 = 1$ implies that

$$|z_n| = \sqrt{1 - (1 - \varepsilon_n)^2} \approx \sqrt{2}\varepsilon_n^{1/2}.$$

In this example all matrices tend to a limit. Such limits are diagonal matrices. Thus the trajectories in figure 2, produced by the IMR, converge to spurious limits.

Instead, for the Lie group-invariant methods, numerical experiments confirm that the departure from $SU(d,\mathbb{C})$ is of the order of machine accuracy.

We remark that RK schemes with $M = O$ are already superior to other general-purpose numerical methods. For example, integrating with Forward Euler, using $h = 1/100$, ten times smaller than the stepsize used for the IMR, we obtain a picture of the solution which is completely wrong. In figure 3 we display the projection of the determinant on the complex plane for Forward Euler and the IMR. The error of the IMR is clearly tangent to the unit circle.

4.2.1 Comparison with Runge–Kutta type unitary methods.

We compare the Lie group-invariant methods proposed in Section 2 with the Runge–Kutta type orthogonal integrators proposed by Higham in [12].

Assume that (1) is solved by the numerical methods

$$Y_{n+1} = \left(I - \tfrac{1}{2}h_n A_n\right)^{-1} \left(I + \tfrac{1}{2}h_n A_n\right) Y_n, \qquad A_n = A(t_n, Y_n), \tag{39}$$

$$Y_{n+1} = \left(I - \tfrac{1}{2}h_n A_{n+\frac{1}{2}}\right)^{-1} \left(I + \tfrac{1}{2}h_n A_{n+\frac{1}{2}}\right) Y_n, \tag{40}$$

$$A_{n+\frac{1}{2}} = A(t_n + \tfrac{1}{2}h_n, Y_{n+\frac{1}{2}}), \quad Y_{n+\frac{1}{2}} = \left(I - \tfrac{1}{4}h_n A_n\right)^{-1} \left(I + \tfrac{1}{4}h_n A_n\right) Y_n,$$

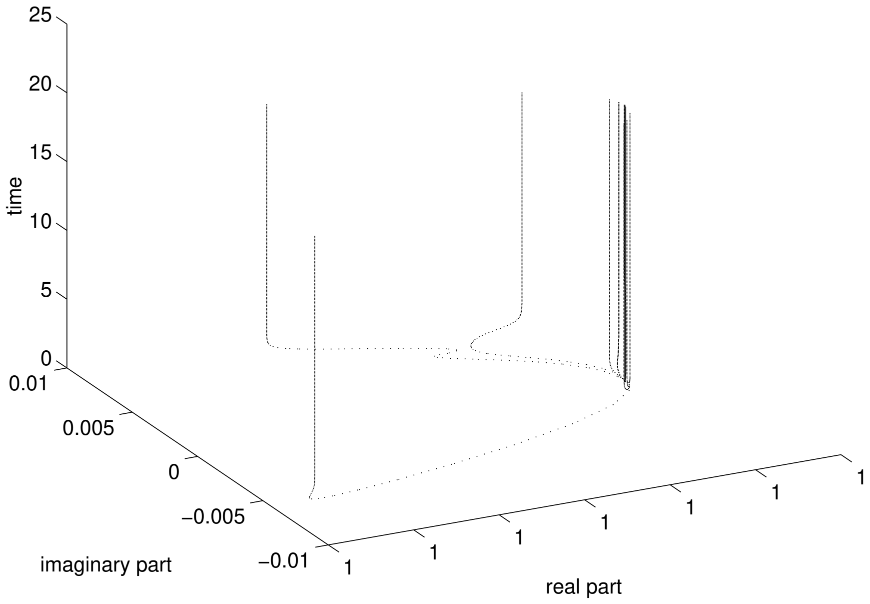

Figure 2: The determinant of the numerical solutions for the IMR method applied to the problem (1) with matrix A as in (38). The different trajectories correspond to various stepsizes, initial conditions, and dimension d.

of order one and two respectively. If $\mathcal{G} = SO(d)$, then the methods (39)–(40) produce orthogonal flows even if $A(t, Y)$ is skew-symmetric only for orthogonal argument Y, as proved in [12]. We intend to show that (39) and (40) can be obtained by the method of iterated commutators (17)–(19) respectively, by approximating the exponential with a [1/1] (diagonal) Padé approximant. Orthogonality (in fact unitarity) is retained because diagonal Padé approximants map skew-symmetric (resp. skew-hermitian) matrices into orthogonal (unitary) matrices, as explained in the following result.

Theorem 5 *Let $R(z)$ be the rational function $R(z) = P(z)/P(-z)$, $z \in \mathbb{C}$, where P is a real polynomial. Then*

$$R(\mathfrak{u}(d, \mathbb{C})) \subseteq U(d, \mathbb{C}). \tag{41}$$

Proof. Let $A \in \mathfrak{u}(d, \mathbb{C})$ and let A^* denote its hermitian ($A^* = \bar{A}^T$). Recall that A is skew-hermitian if $A + A^* = O$. We have

$$\begin{aligned} R^*(A)R(A) &= P^*(A)\left(P^{-1}(-A)\right)^* P^{-1}(-A)P(A) \\ &= P(-A)P^{-1}(A)P^{-1}(-A)P(A), \end{aligned}$$

where we have used the relation $P^*(A) = P(A^*)$, since P has real coefficients. Taking into account that $P(-A), P^{-1}(A), P^{-1}(-A), P(A)$ commute, we deduce that $R^*(A)R(A) = I$ and the theorem follows. $\square$

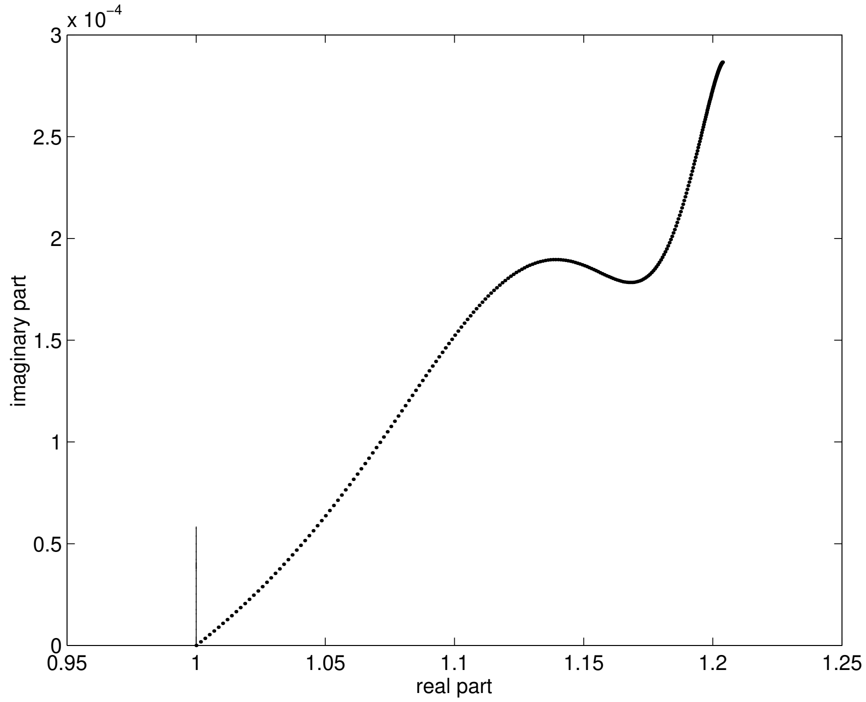

Figure 3: The determinant of the numerical solutions of the problem (1) with matrix A as in (38). The dotted line corresponds to Forward Euler, with stepsize $h = 1/100$ and the solid line to the IMR with stepsize $h = 1/10$. The dimension of the problem is $d = 10$ and the initial condition was generated randomly in $SU(10, \mathbb{C})$.

As a corollary, we conclude that such approximants are orthogonal, by restricting to the real case. This constitutes an alternative proof to the orthogonality of (39) and (40).

Next, assume that $R(z)$ is an analytic approximant to the exponential function.We require that, given $d > 2$ arbitrary complex numbers $z_1, z_2, \ldots, z_n$ whose sum $\sum_{i=1}^{d} z_i = 0$, $R(z)$ obeys the identity $\prod_{i=1}^{d} R(z_i) = 1$. Unfortunately, Feng and Shang have proven in [10] that the only analytic approximation consistent with the exponential function that maps $\mathfrak{sl}(d, \mathbb{C})$ into $SL(d, \mathbb{C})$ is the exponential function itself. Therefore, since $\mathfrak{su}(d, \mathbb{C}) = \mathfrak{sl}(d, \mathbb{C}) \cap \mathfrak{u}(d, \mathbb{C})$, we conclude that the methods proposed by Higham are unitary but do not retain flows in the special unitary group $SU(d, \mathbb{C})$, since they are not $SL(d, \mathbb{C})$-invariant.

From the result of Feng and Shang we also deduce that, for invariant integration on spaces like $SL(d, \mathbb{C})$ and $SU(d, \mathbb{C})$ it is necessary to evaluate the exponential exactly. As soon as the exact exponential is replaced with an analytic (in a neighbourhood of the origin) approximant, we generally depart from the group and invariance is lost.

4.3 Hamiltonian systems

In this section we present some numerical experiments concerning the analysis of the Hamiltonian equations which has been presented in §3. We have already mentioned that when the Hamiltonian function is quadratic, then the system is equivalent to a constant coefficient differential equation and therefore all consistent Lie-group invariant methods sample the exact solution, irrespective of their order. This is for instance the case for the system (36).

In the periodic linear (variable coefficients) case, subject to commutativity, e.g. when $A(t) \in \mathfrak{sp}(2d, \mathbb{R})$ for all t, it commutes for different values of t and the solution of the system is periodic, it is possible to show that the method of iterated commutators samples exactly a *nearby* periodic solution, that can be found in a closed form.

Theorem 6 *Consider the system*

$$Y' = A(t)Y, \qquad Y(0) = Y_0, \tag{42}$$

where $A(t)$ is a periodic function with period T such that $\int_0^T A(t+\tau)\,\mathrm{d}\tau = O$, for all $t \geq 0$, where $\left(\int_0^T A(t+\tau)\,\mathrm{d}\tau\right)_{i,j} = \int_0^T A_{i,j}(t+\tau)\,\mathrm{d}\tau$. Moreover, assume that $A(t_1)$ commutes with $A(t_2)$ for all $t_1, t_2 \geq 0$. Then the solution of (42) is periodic and the numerical solution produced by the method of iterated commutators implemented with constant stepsize h samples exactly a nearby periodic solution with period T.

Proof. The exact solution of (42) is given explicitly by

$$Y(t) = \exp\left(\int_0^t A(\tau)\,\mathrm{d}\tau\right) Y_0,$$

and, because of commutativity, also the numerical solution with the method of iterated commutators will produce a single exponential. The order of the numerical approximation depends solely on the order of the quadrature formula that we choose to approximate the integral of a function f,

$$\int_0^t f(\tau)\,\mathrm{d}\tau \approx \sum_{j=1}^r b_j f(c_j t), \qquad \sum_{j=1}^r b_j = 1,$$

where the b_js and c_js are the weights and the nodes of the quadrature rule respectively. Since $A(t)$ is periodic, it posses a Fourier series (for simplicity, we assume $T = 2\pi$),

$$A(t) = \sum_{k=-\infty}^{\infty} A_k \mathrm{e}^{\mathrm{i}kt}, \qquad A_0 = O,$$

where all the matrices A_k commute. It suffices to consider the system

$$Y' = A(t)Y, \qquad Y(0) = I.$$

Therefore

$$\int_0^t A(\tau)\,\mathrm{d}\tau = \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \frac{A_k}{\mathrm{i}k}\mathrm{e}^{\mathrm{i}kt} - \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \frac{A_k}{\mathrm{i}k},$$

and the exact solution is given by

$$Y(t) = \exp\left(\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \frac{A_k}{\mathrm{i}k}\mathrm{e}^{\mathrm{i}kt} - \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \frac{A_k}{\mathrm{i}k}\right). \tag{43}$$

Next we evaluate the numerical solution. Since in each interval $[t_m, t_{m+1}]$ we approximate

$$\int_{t_m}^{t_{m+1}} A(\tau)\,\mathrm{d}\tau = h\sum_{j=1}^{r} b_j A(h(m+c_j)) + \mathcal{O}\left(h^{p+1}\right) \approx h\sum_{j=1}^{r} b_j \sum_{k=-\infty}^{\infty} A_k \mathrm{e}^{\mathrm{i}k(m+c_j)h},$$

the numerical solution at the nth step is given by

$$Y_n = \prod_{m=0}^{n-1} \exp\left[h\sum_{j=1}^{r} b_j \sum_{k=-\infty}^{\infty} A_k \mathrm{e}^{\mathrm{i}k(m+c_j)h}\right] = \exp\left[h\sum_{m=0}^{n-1}\sum_{j=1}^{r} b_j \sum_{k=-\infty}^{\infty} A_k \mathrm{e}^{\mathrm{i}k(m+c_j)h}\right].$$

Exchanging the order of summation,

$$\sum_{m=0}^{n-1}\sum_{j=1}^{r} b_j \sum_{k=-\infty}^{\infty} A_k \mathrm{e}^{\mathrm{i}k(m+c_j)h} = \sum_{k=-\infty}^{\infty} A_k \sum_{j=1}^{r} b_j \mathrm{e}^{\mathrm{i}kc_jh} \sum_{m=0}^{n-1} \mathrm{e}^{\mathrm{i}kmh}.$$

Summing up the finite geometric series

$$\sum_{m=0}^{n-1} \mathrm{e}^{\mathrm{i}kmh} = \frac{\mathrm{e}^{\mathrm{i}nkh}-1}{\mathrm{e}^{\mathrm{i}kh}-1}, \qquad k \neq 0,$$

we obtain

$$Y_n = \exp\left[h\sum_{\substack{k=-\infty\\k\neq 0}}^{\infty} \frac{A_k}{\mathrm{e}^{\mathrm{i}kh}-1}\left(\sum_{j=1}^{r} b_j \mathrm{e}^{\mathrm{i}c_jkh}\right)\mathrm{e}^{\mathrm{i}knh} - h\sum_{\substack{k=-\infty\\k\neq 0}}^{\infty} \frac{A_k}{\mathrm{e}^{\mathrm{i}kh}-1}\left(\sum_{j=1}^{r} b_j \mathrm{e}^{\mathrm{i}c_jkh}\right)\right]. \quad (44)$$

Comparing (44) with (43), in which we have substituted $t_n = nh$, we observe that the amplitude of each kth harmonic is modified by

$$\left(\frac{1}{\mathrm{i}k} - \frac{h}{\mathrm{e}^{\mathrm{i}kh}-1}\sum_{j=1}^{r} b_j \mathrm{e}^{\mathrm{i}c_jkh}\right) A_k, \qquad k \neq 0.$$

Such deformation of the wave depends solely on the weights and nodes of the quadrature rule. Note that (44) has the same period as (43) and, when $h \to 0$, for fixed $t_n = nh$ the numerical solution (44) tends to the exact solution (43) for every consistent quadrature formula. □

Theorem 6 does not exploit at all the fact that $A \in \mathfrak{sp}(2d, \mathbb{R})$ and it is true for all periodic systems of the form (42). Also, the error is a periodic function with the same period T.

In the linear non-commutative case, i.e. subject to $A \in \mathfrak{sp}(2d, \mathbb{R})$ being function of time only, by Corollary 4.1 the proposed methods are symplectic when implemented with constant stepsize. An analysis of the error terms should however establish whether these methods fit in the framework of Cano and Sanz-Serna [6] with regard to error growth for periodic systems.

The analysis is far more complicated in the nonlinear non-commutative case. It is however illustrative to consider the nonlinear pendulum (37). Although we have not undertaken any analysis of symplecticity for the group-invariant method of order four, numerical integration (performed with constant stepsize $h = 0.1$) indicates that the group-invariant schemes are superior to general-purpose explicit schemes of the same order. A phase portrait for this system is plotted in figure 4, where the fourth-order group-invariant numerical solution is compared with that obtained with the explicit fourth-order Runge–Kutta method whose coefficients are given by the tableau

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} . \qquad (45)$$

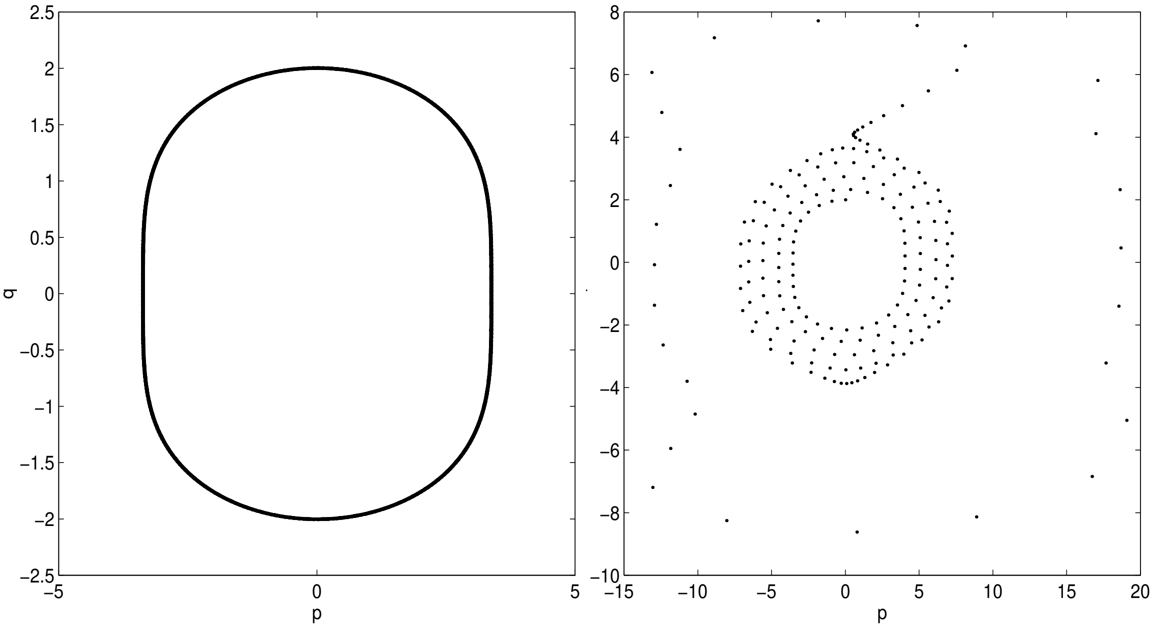

Figure 4: Phase portrait of the nonlinear pendulum (37). The first plot corresponds to the group-invariant method of order four, the second to the explicit order four RK method whose tableau is given by (45). Although both integrations are performed with the same stepsize $h = 0.1$ on the interval $[0, 1000]$, for the second plot we display only integration over $[0, 20]$ since for higher values of t the numerical approximation blows up. Note that different scales are used for the axes of the two phase portraits.

As for orthogonal and unitary matrices, we consider the issue of replacing the exponential by a rational approximant and whether such a rational approximant retains symplecticity. Recall that, if $A \in Sp(2d, \mathbb{R})$ then

i) $-A \in Sp(2d, \mathbb{R})$.

ii) A is invertible and $A^{-1} \in Sp(2d, \mathbb{R})$.

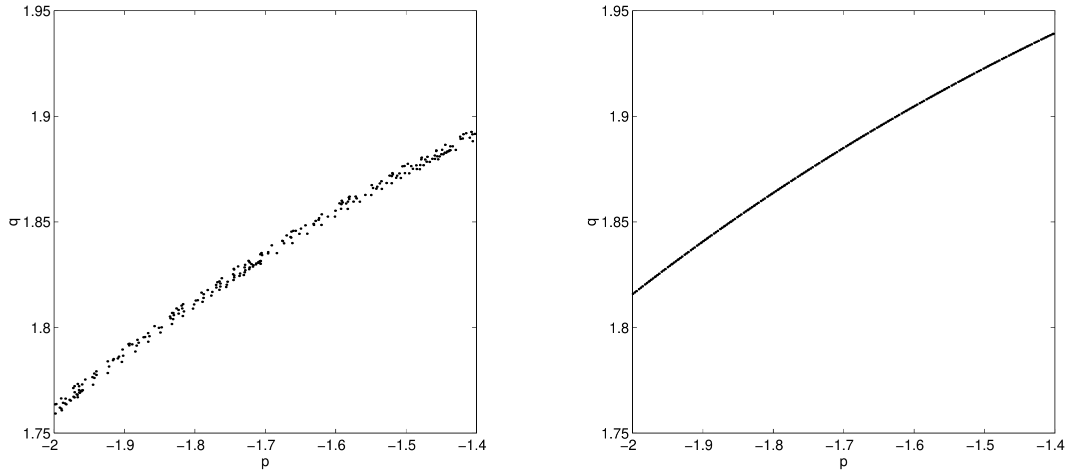

Figure 5: Zoom of a section of the nonlinear pendulum (37). The first plot corresponds to the group-invariant method of order four, the second to the (symplectic) IMR. Both integrations are performed with the same stepsize $h = 0.1$ on the interval $[0, 1000]$.

Therefore we can prove the following result:

Theorem 7 *For the symmetric rational function* $R(z) = P(z)/P(-z)$ *it is true that*

$$R(\mathfrak{sp}(2d, \mathbb{R})) \subseteq Sp(2d, \mathbb{R}). \tag{46}$$

Moreover, if $P(z)$ *has real coefficients, then* $\det(R(A)) = \pm 1$*, for all* $A \in \mathfrak{sp}(2d, \mathbb{R})$*.*

Proof. Let $A \in \mathfrak{sp}(2d, \mathbb{R})$. We wish to prove that

$$R^T(A) J R(A) = J, \tag{47}$$

or, equivalently, that

$$P^T(A) J P(A) = P^T(-A) J P(-A). \tag{48}$$

We write the polynomial $P(z)$ in terms of its roots,

$$P(z) = \prod_{i=1}^{n} (z - \xi_i).$$

Therefore,

$$\begin{aligned} P^T(A) J &= \left[\prod_{i=1}^{n} (A^T - \xi_i I)\right] J = \left[\prod_{i=1}^{n-1} (A^T - \xi_i I)\right] (A^T J - \xi_n J) \\ &= -\left[\prod_{i=1}^{n-1} (A^T - \xi_i I)\right] J (A + \xi_n I), \end{aligned} \tag{49}$$

where we have used the identity $A^T J = -JA$. Thus, by induction on s,

$$P^T(A) J = (-1)^s \left[\prod_{i=1}^{n-s} (A^T - \xi_i I)\right] J \left[\prod_{j=n-s+1}^{n} (A + \xi_j I)\right].$$

Letting $s = n$ we obtain
$$P^T(A)J = (-1)^n JP(-A),$$
which yields
$$P^T(A)JP(A) = (-1)^n JP(-A)P(A).$$
Replacing A by $-A$, a similar procedure leads to

$$P^T(-A)JP(-A) = (-1)^n JP(-A)P(A),$$

and (48) follows. Thus the first part of the theorem is proved. The second part, is an immediate consequence of (47).[3] $\square$

From Theorem 7 we deduce that if the group-invariant method is symplectic, then replacing the exponential by a rational approximant of the form $R(z) = P(z)/P(-z)$ we generate other symplectic schemes.

Numerical examples seem to indicate that the group-invariant equivalent of the IMR is symplectic when applied to an Hamiltonian system written in the form (1). We do not know whether this implies or follows from the symplecticity of the IMR for (28) since the two implementations produce different numerical approximations.

Other issues, for example the best way to reformulate the Hamiltonian problem in the form (1) and many other related topics are being currently investigated.

Acknowledgement. I would like to thank Arieh Iserles and Hans Munthe-Kaas for many valuable comments and discussion on the subject matter of this paper. Thanks to Robert MacKay for a discussion of Hamiltonian systems and symplecticity. This work has been in part supported by the SYNODE project, `http://www.imf.unit.no/num/synode/`.

References

[1] R.W. Brockett, "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems", *Lin. Alg. Appl.*, **146**, 79–91, 1991.

[2] R.W. Brockett, "The double brackett equation as the solution of a variational problem", *Fields Institute Communications,* **3**, 69–76, 1994.

[3] R.L. Bryant, "An introduction to Lie groups and symplectic geometry", in *Geometry and Quantum Field Theory,* (D.S. Freed and K.K. Uhlenbeck eds) AMS IAS/Park City Math. Series, vol. 1, 1995.

[4] M.P. Calvo, A. Iserles and A. Zanna, "Numerical solution of isospectral flows", DAMTP Techn. Rep. NA03/95, University of Cambridge, 1995.

[5] M.P. Calvo, A. Iserles and A. Zanna, "Runge–Kutta methods on manifolds", in *A.R. Mitchell* 75^{th} *Birthday Volume,* (D.F. Griffiths & G.A. Watson eds), World Scientific, Singapore, 57–70, 1996.

[3]if $P(z)$ has complex coefficients, (47) implies only that $[\det(R(A)]^2 = 1$

[6] B. Cano and J.M. Sanz-Serna, "Error growth in the numerical integration of periodic orbits, with application to Hamiltonians and reversible systems", Tech. Rep. 1, University of Valladolid, 1995.

[7] P.E. Crouch and R. Grossman, "Numerical integration of ordinary differential equations on manifolds", *J. Nonlinear Sci.,* **3**, 1-33, 1993.

[8] L. Dieci, D.R. Russell and E.S. van Vleck, "Unitary integrators and applications to continuous orthonormalization techniques", *SIAM J. Numer. Anal.*, **31**, 261–281, 1994.

[9] D.J. Estep and A.M. Stuart, "The rate of error growth in Hamiltonian-conserving integrators", *Z. angew. Math. Phys.,* **46**, Birkhäuser Verlag, Basel, 407–418, 1995.

[10] K. Feng and Z.-J. Shang, "Volume-preserving algorithms for source-free dynamical sysyems", *Numer. Math.,* **71**, 451–463, 1995.

[11] E. Hairer, S.P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I. Nonstiff Problems,* 2nd revised edition, Springer, Berlin, 1993.

[12] D.J. Higham, "Runge–Kutta type methods for orthogonal integration", Tech. Rep. NA/168, Department of Mathematics and Computer Science, University of Dundee, 1996.

[13] A. Iserles, "Solving ordinary differential equations by exponentials of iterated commutators", *Numer. Math.,* **45**, 183–199, 1984.

[14] A. Iserles and A. Zanna, " Qualitative numerical analysis of ordinary differential equations", DAMTP Tech. Rep. 1995/NA05, to appear in *Lectures in Applied Mathematics* (J. Renegar, M. Shub & S. Smale eds), AMS, Providence RI, 1995.

[15] J.B. Moore, R.E. Mahony and U. Helmke, "Numerical gradient algorithms for eigenvalues and singular value calculations", *SIAM J. Matrix Anal. Appl.*, **17**, 3, 881–902, 1994.

[16] A. Marthinsen and B. Owren, "Order conditions for integration methods based on rigid frames", to appear.

[17] H. Munthe-Kaas, "Lie–Butcher theory for Runge–Kutta methods", *BIT,* **35**, 572–587, 1995.

[18] H. Munthe-Kaas, "Runge–Kutta methods on Lie groups", to appear.

[19] H. Munthe-Kaas, "Numerical solution of differential equations on manifolds", to appear.

[20] P.J. Olver, *Applications of Lie Groups to Differential Equations,* 2nd edition, Springer-Verlag, New York, 1991.

[21] A.M. Peremolov, *Integrable Systems of Classical Mechanics and Lie Algebras,* vol. 1, Birkhäuser Verlag, Basel, 1990.

[22] G.P. Ramaswami, *Numerical Solution of Separable Hamiltonians,* Ph.D. Thesis, University of Cambridge, Cambridge, 1996.

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

[24] S.T. Smith, "Dynamical systems that perform the singular value decomposition", *Systems & Control Letters,* **16**, 319–327, 1991

[25] V.S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations,* Springer-Verlag, New York, 1984.