

 Open access • Journal Article • DOI:10.1137/0916010

On the numerical integration of ordinary differential equations by symmetric composition methods — [Source link](#)





Robert I. McLachlan

Published on: 01 Jan 1995 - SIAM Journal on Scientific Computing (Society for Industrial and Applied Mathematics)

Topics: Exponential integrator, Numerical partial differential equations, Backward differentiation formula, Ordinary differential equation and Differential equation

Related papers:

- [Construction of higher order symplectic integrators](#)
- [Numerical Hamiltonian Problems](#)
- [Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations](#)
- [Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations](#)
- [Fourth-order symplectic integration](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/on-the-numerical-integration-of-ordinary-differential-1rew0erady>

ON THE NUMERICAL INTEGRATION OF ORDINARY DIFFERENTIAL EQUATIONS BY SYMMETRIC COMPOSITION METHODS

ROBERT I. McLACHLAN

ABSTRACT. Differential equations of the form $\dot{x} = X = A + B$ are considered, where the vector fields A and B can be integrated exactly, enabling numerical integration of X by composition of the flows of A and B . Various symmetric compositions are investigated for order, complexity, and reversibility. Free Lie algebra theory gives simple formulae for the number of determining equations for a method to have a particular order. A new, more accurate way of applying the methods thus obtained to compositions of an arbitrary first-order integrator is described and tested. The determining equations are explored, and new methods up to 100 times more accurate (at constant work) than those previously known are given.

1. Composition methods.

Composition methods are particularly useful for numerically integrating differential equations when the equations have some special structure which it is advantageous to preserve. They tend to have larger local truncation errors than standard (Runge-Kutta, multistep) methods [4,5], but this defect can be more than compensated for by their superior conservation properties.

Capital letters such as X will denote vector fields on some space with coordinates x , with flows $\exp(tX)$, *i.e.*, $\dot{x} = X(x) \Rightarrow x(t) = \exp(tX)(x(0))$. The vector field X is given and is to be integrated numerically with fixed time step t . Composition methods apply when one can write

$$X = A + B$$

in such a way that $\exp(tA)$, $\exp(tB)$ can both be calculated explicitly. Then the most elementary such method is the map (essentially the “Lie-Trotter” formula [26])

$$\varphi : x \mapsto x' = \exp(tA) \exp(tB)(x) = x(t) + \mathcal{O}(t^2). \quad (1.1)$$

The advantage of composing exact solutions in this way is that many geometric properties of the true flow $\exp(tX)$ are preserved: *group properties* in particular. If X , A , and B are Hamiltonian vector fields then both $\exp(tX)$ and the map φ

1991 *Mathematics Subject Classification.* 65L05, 70–08, 58D07, 17B66.

Key words and phrases. Initial value problems, composition methods, operator splitting, symplectic integrators, Hamiltonian systems.

are symplectic; if X , A , and B are skew-Hermitian (as in quantum mechanics, for example) then φ is automatically unitary; if X , A , and B are divergence-free then φ is volume-preserving. If $X = J(x)\nabla(H_1 + H_2)$ is a Poisson system and we split it as $X = J(x)\nabla H_1 + J(x)\nabla H_2$ then φ is a *Poisson map*, such maps being difficult to generate by other means. In addition, if X , A , and B have any common first integrals then they are shared by φ . Composition methods are also a convenient way of preserving any reversibilities of the flow, which we discuss in section 3.

More generally, methods which write $X = A + B$ and compose *approximations* of $\exp(tA)$ and $\exp(tB)$ are known as *operator-splitting* methods. These have a long history, going back to Yanenko [28] and Strang [22], who dealt with the special case known as *dimensional splitting*. Here X is the spatial discretization of a PDE with multiple space dimensions, and in the equations determining the approximation x' of $\exp(tA)$, say, x' is coupled in one spatial direction only. For example, A might contain differences in only one spatial direction. Arbitrary compositions of sets of methods have also been considered with the goal of increasing order or stability; see for example Stetter [21] on cyclic composition of linear multistep methods. An early work on compositions of one-step methods which predates the application to Hamiltonian systems is Iserles [8].

For us the primary requirement is to be able to solve $\dot{x} = A(x)$ and $\dot{x} = B(x)$ exactly. (Actually, as we shall see later, this can be relaxed—without needing to find new time-stepping coefficients—to finding first-order approximations to $\exp(tA)$ and $\exp(tB)$, or to $\exp(tX)$.) Most currently used methods split X into linear vector fields or into parts with linear flows such as shears $\dot{x}_1 = 0$, $\dot{x}_2 = f(x_1)$, but in principle one can choose pieces from any of the repertoire of known integrable systems. In this way Wisdom [27] split the differential equations for the n -planet solar system into n Sun-planet two-body problems, each an integrable Kepler problem, and $n(n-1)/2$ planet-planet interaction terms, each a shear. McLachlan [15] gave a class of Lie-Poisson systems which can be split into linear systems, and showed that a model of the 2D Euler equations for the flow of an incompressible perfect fluid falls into this class. Many PDE's arising in physics can be split after pseudospectral spatial discretization [14]. Zhang has given examples for the unitary [30] and volume-preserving [31] cases.

The method (1.1) is only first order. The order can be increased to p , say, by composing many such stages [18]:

$$\dots \exp(b_n t B) \exp(a_n t A) \dots \exp(b_1 t B) \exp(a_1 t A) \quad (1.2)$$

with the coefficients a_i, b_i chosen so that above composition approximates $\exp(tX)$ with error $\mathcal{O}(t^{p+1})$. (This will require some smoothness: A and B must be C^r for some $r \geq p$.) Like Iserles [8] and Suzuki [24] we shall also consider compositions of an arbitrary first-order method $\varphi(t)$ and its inverse:

$$\dots \varphi^{\pm 1}(w_2 t) \varphi^{\pm 1}(w_1 t). \quad (1.3)$$

If nothing more than $\varphi(t) = 1 + tX + \mathcal{O}(t^2)$ is known about φ , this appears to be significantly more general than (1.2). We shall see in Theorems 1 and 2 that this is not so: in fact, any method of the form (1.2) directly generates a method of the form (1.3) of the same order, and vice versa. Thus the work of finding high-order methods, which for historical reasons has been concentrated on type (1.2), need not be repeated.

A fundamental result, useful in analyzing compositions such as (1.2) and (1.3), is the Baker-Campbell-Hausdorff (BCH) formula [3, p.160]: formally at least,

$$\exp A \exp B = \exp \left(A + B + \frac{1}{2}[A, B] + \frac{1}{12}([A, [A, B]] + [B, [B, A]]) + \dots \right), \quad (1.4)$$

where $[A, B] = A \cdot \nabla B - B \cdot \nabla A$ is the commutator bracket of vector fields. The composition (1.2) may then be expanded as a asymptotic series

$$\exp(tX_1 + t^2X_2 + \dots) \quad (1.5)$$

where $X_n \in L^n(A, B)$, the elements of degree n of the free Lie algebra generated by A and B , that is, the vector space spanned by all commutators of degree n of A and B . Let the dimension of $L^n(A, B)$ be $c(n)$ (see Eq. (2.1)). On choosing a basis for $L^n(A, B)$, the coordinates of X_n are polynomials of degree n in the a_i 's and b_i 's. If these are zero then $X_n = 0$ for all A and B . Therefore for a method to have order p (i.e., $X_1 = X$, $X_2 = \dots X_p = 0$) there are $\sum_{n=1}^p c(n)$ *determining equations*.

The simplest example of (1.2) is *leapfrog*:

$$\exp(\tfrac{1}{2}tA) \exp(tB) \exp(\tfrac{1}{2}tA) \quad (1.6)$$

which is second order. Note that if many steps are performed without output, only one evaluation each of A and B is required per time step. An important property of leapfrog is its (time) symmetry; we say a map S depending on a time step t is *symmetric* if

$$S(-t)S(t) = 1. \quad (1.7)$$

If a method has this property, only odd powers of t appear in the expansion (1.5) so only the odd-order determining equations need to be solved [29].

There are several ways of deriving the determining equations, which we shall not go into in detail here. Yoshida [29] does a direct expansion of the X_n in (1.5) using the BCH formula (1.4), simplified using symmetry arguments, and Suzuki [23,24] has built a general theory along these lines. Sanz-Serna, Abia, and Calvo (see [19] for a review) have extended the graph-theoretic approach, standard in the numerical solution of initial value problems, to the symplectic and symmetric cases.

Adding special symmetries to the method reduces the number of determining equations to be solved, but it also reduces the number of parameters a_i , b_i available to satisfy them. Let m be the number of evaluations of B per time step. We shall distinguish the following cases:

Type NS, non-symmetric.

Historically [18], the first methods derived were of the form

$$\exp(a_{m+1}tA) \exp(b_mtB) \exp(a_mtA) \dots \exp(b_1tB) \exp(a_1tA) \quad (1.8)$$

which has $2m + 1$ parameters.

Type S, symmetric.

Imposing symmetry (1.7) gives methods of the type

$$\exp(a_1tA) \exp(b_1tB) \dots \exp(b_{(m+1)/2}tB) \dots \exp(b_1tB) \exp(a_1tA) \quad (1.9a)$$

for an odd number m of evaluations of B , or

$$\exp(a_1 t A) \exp(b_1 t B) \dots \exp(a_{(m/2)+1} t A) \dots \exp(b_1 t B) \exp(a_1 t A) \quad (1.9b)$$

for even m . In both cases there are $m + 1$ parameters.

Type SS, symmetric composed of symmetric steps.

Yoshida [29] used the composition

$$S(w_1 t) \dots S(w_{(m+1)/2} t) \dots S(w_1 t) \quad (1.10)$$

where S is any symmetric integrator (usually second order). It is advantageous to use only odd values of m . Some possible choices for the basic component S are leapfrog (1.6) (notice that consecutive steps with A may be amalgamated), a generalized leapfrog $\prod_{i=1}^r \exp(\frac{1}{2} t A_i) \prod_{i=r}^1 \exp(\frac{1}{2} t A_i)$ when $X = \sum_{i=1}^r A_i$ [7], the midpoint rule $x' = x + tX((x + x')/2)$, or a symmetrized integrator $\varphi(t/2)\varphi^{-1}(-t/2)$ where φ is any first-order method [2]. There are only $(m + 1)/2$ parameters in (1.10), but the number of determining equations is drastically reduced.

Type SB³A, Symmetric with $[B, [B, [B, A]]] = 0$.

One often finds splittings which satisfy $[B, [B, [B, A]]] = 0$, so that the coefficient of such a term in (1.5) does not need to be set to zero. This reduces the number of determining equations, which is further reduced by considering symmetric methods of the form (1.9). The most important example is the class of Hamiltonian systems of the form $\dot{q} = p$, $\dot{p} = -\nabla V(q)$ with the splitting

$$A = \begin{pmatrix} p \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ -\nabla V(q) \end{pmatrix} \quad (1.11)$$

so that in this case SB³A methods will be of Runge-Kutta-Nyström (RKN) type [19]. However, there are other applications, such as Poisson systems with constant Poisson tensor $J = \begin{pmatrix} 0 & K \\ -K^t & 0 \end{pmatrix}$, Hamiltonian $H(q, p) = A(q, p) + B(q)$ and A quadratic in p [14]. We shall see later that the distinction between cases (1.11) and $[B, [B, [B, A]]] = 0$ is not important in practice.

Section 2 applies some results from the theory of free Lie algebras to the problem of counting the number of determining equations. This knowledge is crucial, as it determines both the number of stages needed to achieve a given order and the number of free parameters then available to “tune” the method. It also suggests suitable bases in which to express the determining equations, although we do not explore that issue here. Note that the free Lie algebra approach only counts the dimensions of certain spaces, giving upper bounds for the number of stages required. There could be further simplifications in the determining equations for particular methods. For example, although at high order type NS has more determining equations than type S, type S methods also solve the NS equations. The same is true in turn for types S and SS. The only *known* simplifications are those arising from the symmetries presented here, but proofs that there are no more can only be carried out by algebraically reducing the determining equations for particular cases, as in Koseleff [9,10].

We also show in section 2 that type S methods can be adapted to the case when one only has $\varphi(t)$, an arbitrary first-order approximation to $\exp(tX)$. In section 3

Table 1. Number of determining equations and parameters

Data for types NS and S are from (2.1), for type SS from (2.9b), and for type SB³A from (2.10).

Order	Type NS	Type S	Type SS	Type SB ³ A
<i>Determining equations:</i>				
2	3	2	1	2
4	8	4	2	4
6	23	10	4	8
8	71	28	8	18
10	226	84	16	44
<i>Parameters:</i>				
	$2m + 1$	$m + 1$	$(m + 1)/2$	$m + 1$

we discuss the reversibility properties of composition methods. The solution space of the determining equations is searched in section 4 to find methods with minimum error constants, these optimized methods being illustrated with some brief examples in section 5.

2. Counting the determining equations.

As discussed above, the number of determining equations at order n for methods of the form (1.8) or (1.9) is the number of independent commutators of A and B of order n . These commutators span the subspace $L^n(A, B)$ of $L(A, B)$, the *free Lie algebra generated by A and B* , which may be thought of as the vector space spanned by A , B , and all their commutators $[A, B]$, $[A, [A, B]]$, \dots . The dimension of $L^n(A, B)$ is given by Witt's formula [3, p.140]

$$c(n) = \frac{1}{n} \sum_{d|n} \mu(d) 2^{n/d} \quad (2.1)$$

where $\mu(d)$ is the Möbius function $\mu(1) = 1$, $\mu(d) = (-1)^j$ if d is the product of j distinct primes, and $\mu(d) = 0$ otherwise.

The first 10 values of $c(n)$ are 2, 1, 2, 3, 6, 9, 18, 30, 56, and 99, giving the total numbers of determining equations shown in Table 1: $\sum_{n=1}^p c(n)$ for a type NS method of order p , and $\sum_{n=1}^{p/2} c(2n - 1)$ for a type S method of (even) order p .

(NOTE: Sanz-Serna [19] considers *partitioned Runge-Kutta* (PRK) methods, a method of integrating equations in the form $\dot{x}_1 = A(x_1, x_2)$, $\dot{x}_2 = B(x_1, x_2)$. When explicit, as they can be when $A = A(x_2)$ and $B = B(x_1)$ as arises in Hamiltonian systems, PRK methods reduce to a special case of (1.2). Note (1.2) applies even when the dependent variables are not so partitioned. The numbers of determining equations at order $1 \leq n \leq 10$ for general (*i.e.*, possibly implicit) PRK methods are $c_{\text{prk}}(n) = 2, 1, 2, 3, 6, 10, 22, 42, 94$, and 203 (see [19], Table 1, column 5). For orders $p > 5$, $c_{\text{prk}}(n) > c(n)$, that is, explicitness must create some redundancy in the PRK determining equations. It does not appear than partitioning creates any redundancy in the determining equations of (1.2).)

One may also start with an arbitrary map $\varphi(t)$ (assumed smooth in space and time) approximating $\exp(tX)$ to first order, and compose it together with its inverse

as in (1.3): $\dots \varphi^{\pm 1}(w_2 t) \varphi^{\pm 1}(w_1 t)$. Taking the logarithm of φ gives the asymptotic series

$$\varphi(t) = \exp(\alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \dots) \quad (2.2)$$

where the α_i are vector fields. Since $\varphi(t)$ is first order, $\alpha_1 = X$. Such a formulation is useful because often a “two-map” ($X = A + B$) splitting is not available. If $X = \sum_{i=1}^r A_i$ then $\varphi(t) = \prod_{i=1}^r \exp(tA_i)$ may be available, giving an “ r -map” integrator as in [7]. Witt’s formula for the number of independent commutators of order n formed from r indeterminates (here, the vector fields A_i) is $\frac{1}{n} \sum_{d|n} \mu(d) r^{n/d}$, showing that it is hopeless to generalize (1.2) to general compositions of r flows—there are just too many independent commutators. If $\exp(tA_i)$ is not available, one can use $\varphi(t) = \prod_{i=1}^r \varphi_i(t)$ where $\varphi_i(t)$ approximates $\exp(tA_i)$ to first order. In the symplectic case, a suitable φ is generated by the generating function of the third kind $q^t p - tH(q, p)$; for the Lie-Poisson case, a suitable (Poisson) φ is constructed in [6]. If one is not worried about staying in the right group, $\varphi(t)$ could be Euler’s method $1 + tX$.

Formula (1.3) is at first sight more general than (1.2), because it contains an infinite number of indeterminates α_i instead of only two, A and B ; but we show now that (1.2) and (1.3) are really equivalent. To count the determining equations arising from (1.3), the following extension of Witt’s formula [3, p.141] was used by Suzuki [24]. It gives the dimension b of the space spanned by commutators of the indeterminates A_1, A_2, \dots , with each A_i occurring n_i times:

$$b(n_1, n_2, \dots) = \frac{1}{\sum n_i} \sum_{d|n_i \forall i} \mu(d) \frac{(\sum n_i/d)!}{\prod_i (n_i/d)!}. \quad (2.3)$$

Then the number of determining equations $d(n)$ at order n in (1.3) is the sum of the dimensions of the spaces spanned by commutators of the α_i whose total order is n : 1 at order 2 (spanned by α_2); 2 at order 3 (spanned by α_3 and $[\alpha_1, \alpha_2]$); 3 at order 4 (spanned by α_4 , $[\alpha_1, \alpha_3]$, and $[\alpha_1, [\alpha_1, \alpha_2]]$), etc.

$$d(n) = \sum_{n_1 + 2n_2 + \dots + nn_n = n} b(n_1, n_2, \dots, n_n) \quad (2.4)$$

This requires an elaborate search for partitions of n ; the formula is greatly simplified by the following

Theorem 1. $d(n) = c(n)$ for all $n > 1$.

The proof will use the *Lyndon basis* for free Lie algebras, which we describe briefly. See Lothaire [10] for more details. A *word* is a sequence of letters chosen from an alphabet \mathcal{A} ; words are multiplied by concatenation. A *Lyndon word* is a word which is not the power of another word (*i.e.*, it is *primitive*) and is lexicographically minimal (*i.e.*, would be first in a dictionary) amongst its cyclic permutations. The Lyndon words on $\mathcal{A} = \{A, B\}$ are $\{A, B, AB, AAB, ABB, AAAB, AABBB, ABBBB, AAAAB, AAABB, AABAB, \dots\}$. There is a bijection from the set of Lyndon words to a basis for $L(\mathcal{A})$ [10, p.67]. Thus the number of Lyndon words of a given length is given by Witt’s formula. We also need the *Lazard elimination method* ([10], p.85 and [3], p.132), which decomposes $L(A, B)$ as $B \oplus L(A, [A, B], [[A, B], B], \dots)$.¹

¹We have found it convenient to evaluate sums such as (2.9a), tables of Lyndon bases, determining equations, etc. in Mathematica. Contact the author for more details.

Proof of Theorem 1. Choosing $\alpha_i = [[\dots[A, B], B], \dots, B]$ ($i - 1$ B 's) immediately shows that $d(n) \geq c(n)$ for $n > 1$. In fact this mapping provides more, a *bijection* between the Lyndon bases of $L(\alpha_1, \alpha_2, \dots)$ and $L(A, B) \setminus B$, because the Lazard elimination method shows that the set of Lyndon words over the alphabet $\{A, AB, ABB, AB BB, \dots\}$ is equal to the set of Lyndon words over $\{A, B\}$, excluding B . The above bijection preserves the total order of a word, so $\sum_{i=1}^n c(n) = 1 + \sum_{i=1}^n d(n)$. Now $c(1) = 2$ (because $L^1(A, B) = \{A, B\}$) and $d(1) = 1$ (because $L^1(\alpha_1, \dots) = \{\alpha_1\}$), which proves the result. \blacklozenge

A standard way to construct higher-order methods out of φ is to let $S(t) = \varphi(t/2)\varphi^{-1}(-t/2)$ and then use an SS method [2]. However, because of Theorem 1, one can do better. Suppose some coefficients a_i, b_i have been determined which give (1.2) order p . Let $\varphi^+(t) = \varphi(t)$ and $\varphi^-(t) = \varphi^{-1}(-t)$. (In most of the examples mentioned above, φ^- is exactly as easy to compute as φ^+ .) Then one can use a (more accurate) type S method with φ^+, φ^- stages, instead of being restricted to type SS:

Theorem 2. *The determining equations for the map (1.8) and for the map*

$$\varphi^+(c_m t) \varphi^-(d_m t) \dots \varphi^+(c_1 t) \varphi^-(d_1 t) \quad (2.5)$$

to have order $p > 1$ are equivalent, where

$$\begin{aligned} d_1 &= a_1, & c_m &= a_{m+1}, & d_i + c_i &= b_i & \text{for } i = 1, \dots, m, \text{ and} \\ & & & & d_i + c_{i-1} &= a_i & \text{for } i = 2, \dots, m. \end{aligned} \quad (2.6)$$

Proof. Let

$$\varphi(t) = \exp(\alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \dots) \quad (2.7a)$$

Then

$$\varphi^{-1}(t) = \exp(-\alpha_1 t - \alpha_2 t^2 - \alpha_3 t^3 - \dots)$$

so

$$\varphi^{-1}(-t) = \exp(\alpha_1 t - \alpha_2 t^2 + \alpha_3 t^3 - \dots) \quad (2.7b)$$

At order n there is one determining equation for (2.5) for each basis element of $L^n(\alpha_1, \alpha_2, \dots)$.

Now start with a method of the form (1.8), and break it up schematically as follows:

$$\begin{array}{ccccccc} e^{a_{m+1}A} & & e^{b_m B} & & e^{a_m A} & & \dots \\ & | & / & \backslash & / & \backslash & \\ e^{c_m A} e^{c_m B} & & e^{d_m B} e^{d_m A} & & & & \dots \\ & | & & | & & & \\ \varphi^+(c_m t) & & & \varphi^-(d_m t) & & & \dots \end{array}$$

using the equations (2.6). Notice that Eqs. (2.6) are $2m + 1$ linear equations in $2m$ unknowns; the compatibility condition for these equations is $\sum_{i=1}^{m+1} a_i = \sum_{i=1}^m b_i$. This is satisfied with both sides equal to 1 when (1.8) is a consistent method, *i.e.*, when $p \geq 1$.

Taking $\varphi^+(t) = \exp(tA)\exp(tB)$ and $\varphi^-(t) = \exp(tB)\exp(tA)$ shows that (1.8) has order p when (2.5) does.

To show the converse we need to show that there are no simplifications in the determining equations of (2.5) under the particular choice $\varphi(t) = \exp(tA)\exp(tB)$, *i.e.*, when $\alpha_1 = A + B$, $\alpha_2 = \frac{1}{2}[A, B]$, \dots . This is true because, for $n > 1$, Theorem 1 states that there are the same number of independent commutators of order n of A and B as there are of the α_i . \blacklozenge

Similar counting arguments apply to compositions of symmetric methods. Now

$$S(t) = \exp(\alpha_1 t + \alpha_3 t^3 + \alpha_5 t^5 + \dots) \quad (2.8)$$

so, using (2.3), there are

$$c_s(n) = \dim(L^n(\alpha_1, \alpha_3, \dots)) = \sum_{n_1+3n_3+\dots=n} b(n_1, n_3, \dots) \quad (2.9a)$$

determining equations at order n , as given in [24]. This can be simplified greatly using a similar bijection to that in the proof of Theorem 1. Let A be an indeterminate of order 1 and B an indeterminate of order 2, and let $\beta_i = [[\dots[A, B], B], \dots, B]$ ($(i-1)/2$ B 's). Then the sets of Lyndon words over the alphabets $\mathcal{A} = \{\alpha_1, \alpha_3, \alpha_5, \dots\}$ and $\{\beta_1, \beta_3, \beta_5, \dots\} = \{A, AB, ABB, \dots\}$ are equal, and the latter is equal to the set of Lyndon words over the alphabet $\mathcal{B} = \{A, B\}$, excluding B . If a word over \mathcal{A} has order n and its image (as a word over \mathcal{B}) under the bijection has j B 's, then the image must have $n - 2j$ A 's. So

$$c_s(n) = \sum_{j=1}^{\lfloor n/2 \rfloor} b(j, n - 2j) \quad (2.9b)$$

If $n = p$ is prime, only 1 divides both j and $p - 2j$, so then

$$c_s(p) = \sum_{j=1}^{\lfloor p/2 \rfloor} \frac{1}{p-j} \binom{p-j}{j}. \quad (2.9c)$$

For type SS methods, only the odd-order determining equations need be solved; their numbers are $c_s(3) = 1$, $c_s(5) = 2$, $c_s(7) = 4$, $c_s(9) = 8$, $c_s(11) = 18$, $c_s(13) = 40$, and $c_s(15) = 90$ (the last two are given incorrectly in [24]).

The B^3A case is also clarified by the Lazard elimination method. Now the number $c_{b^3a}(n)$ of determining equations at order n is the dimension of the space spanned by commutators of order n of A and B , when $C \equiv [B, [B, [B, A]]] = 0$. Under the bijection of Theorem 1 ($C = \alpha_4$, $[B, C] = \alpha_5$, *etc.*), $\alpha_i = 0$ for $i \geq 4$. Thus, for $n > 1$,

$$c_{b^3a}(n) = \sum_{n_1+2n_2+3n_3=n} b(n_1, n_2, n_3) \quad (2.10)$$

and the first 10 values of $c_{b^3a}(n)$ are 2, 1, 2, 2, 4, 5, 10, 15, 26, and 42, giving the total numbers of determining equations shown in Table 1. An alternative construction proceeds as follows: C is independent of commutators of A and B of order less than 4. So the subspace of $L^n(A, B)$ ($n < 8$) on which an arbitrary commutator of A

and B is zero when $C = 0$ is spanned by the commutators of 1 C and $n - 4$ A 's and B 's; so

$$\begin{aligned} c(n) - c_{b^3a}(n) &= \sum_{i=0}^{n-4} b(1, i, n-4-i) \\ &= \frac{1}{n-3} \sum_{i=0}^{n-4} \frac{(n-3)!}{i!(n-4-i)!} \\ &= 2^{n-4} \end{aligned} \tag{2.11}$$

However, for $n \geq 8$, this is an overestimate because the independence assumption fails; *e.g.*, for $n = 8$, $[C, [B, [B, [B, A]]]]$ is erroneously included in (2.11). We do not have a simplification of (2.10) for $n \geq 8$.

The first savings occur at order 4, when 1 term, namely C itself, is zero. But for a symmetric method, the order 4 terms are zero anyway. Thus, *types S and SB^3A are equivalent for orders ≤ 4 .*

We now consider the question of whether the particular ‘‘RKN’’ choice

$$A = \begin{pmatrix} p \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ -\nabla V(q) \end{pmatrix} \tag{1.11}$$

leads to any further reduction in the number of determining equations. Under (1.11), we can replace commutator brackets of vector fields by Poisson brackets of the Hamiltonians $H_A = p^2/2$ and $H_B = V(q)$. One should now describe the Lie algebra generated by these two functions. We have not solved this problem, but the following bound—obtained by counting the Lyndon words over $\{A, B\}$ which are nonzero in this instance—is lower than that in [5, Tab. 1, col. 4] for $n > 8$.

First notice that $\alpha_1 = H_A$ is quadratic in p , $\alpha_2 = \{H_B, \alpha_1\} = \nabla V(q)^t p$ is linear in p , and $\alpha_3 = \{H_B, \alpha_2\}$ is independent of p . Thus $\{\alpha_3, \alpha_2\}$ is independent of p , and the order 8 term ($n_1 = 0$, $n_2 = 1$, and $n_3 = 2$ in (2.10)) $\{\alpha_3, \{\alpha_3, \alpha_2\}\}$ is identically zero. In general, the elements being bracketed in a term in (2.10) have total degree $2n_1 + n_2$ in p , which is reduced by one for each bracket; the final bracket will be zero if $2n_1 + n_2 - (n_1 + n_2 + n_3 - 1) = n_1 - n_3 + 1 < 0$. Such a term should then be dropped from the sum (2.10).

$$c_{\text{rkn}}(n) \leq \sum_{\substack{n_1+2n_2+3n_3=n \\ n_1 \geq n_3-1}} b(n_1, n_2, n_3) \tag{2.12}$$

At orders 8, 9, and 10 there is a reduction of 1, 0, and 2 in c . These seem to be the only such simplifications, so we conjecture that (2.12) is an equality, implying the

Conjecture. *The solution sets of the determining equations of the method (1.2) under the assumptions $[B, [B, [B, A]]] = 0$ and (1.11) are identical for nonsymmetric methods of order ≤ 7 , and for symmetric methods of order ≤ 10 .*

Note that the bases used in deriving (2.10) and (2.12) would be convenient ones in which to derive the determining equations themselves.

One could also consider a type ‘‘SSB³A’’; but this leads to no further simplification, as the following theorem shows.

Theorem 3. *Consider a B^3A method of order p (not necessarily symmetric) given by the composition of leapfrog steps $\exp(\frac{1}{2}w_itA)\exp(w_itB)\exp(\frac{1}{2}w_itA)$, where $[B, [B, [B, A]]] = 0$. Then this method also has order p when the leapfrog steps are replaced by any symmetric method, such as leapfrog with arbitrary A, B .*

Proof. We must show that the determining equations for a method given by the composition of symmetric steps S do not change when S is replaced by leapfrog with $[B, [B, [B, A]]] = 0$. Consider the method $S(w_mt)\dots S(w_1t)$, with S as in (2.8). Now replace S by leapfrog with arbitrary A, B . We have

$$\alpha_1 = A + B, \quad \alpha_3 = \frac{1}{12}[B, [B, A]] - \frac{1}{24}[A, [A, B]], \dots,$$

each α_i being a weighted sum of all commutators of order i of A and B . One could collect commutators of A and B in the expansion of $S(w_mt)\dots S(w_1t)$, duplicating many of the determining equations. Some of these may be dropped when $[B, [B, [B, A]]] = 0$, but because A and B appear symmetrically in the α_i (up to constants), there will always be an identical equation with A and B swapped which cannot be dropped. Thus there is no simplification in the determining equations.

◆

To conclude this section we cover three compositions which do *not* prove useful. First, the form $\prod \varphi_A(a_it)\varphi_B(b_it)$, where φ_A and φ_B are first-order integrators for the differential equations $\dot{x} = A$ and $\dot{x} = B$, respectively. Writing $\varphi_A = \exp(tA + t^2\alpha_2 + \dots)$ and $\varphi_B = \exp(tB + t^2\beta_2 + \dots)$ shows that there are far too many independent terms in such a composition—eight at third order, for example. Even if φ_A and φ_B are symmetric then one will not be able to do better than working with $\varphi(t) = \varphi_A(t)\varphi_B(t)$.

Second, the composition $\prod \varphi(w_it)$, suggested in [24], which would be useful because it does not involve φ^{-1} . The determining equations at order two are $\sum w_i = 1$ and $\sum w_i^2 = 0$, which have no real solutions. We do not know of any applications of complex solutions of the determining equations: even for complex equations such as the nonlinear Schrödinger equation, stepping in the imaginary time direction would bring severe stability problems. For real equations, one could add new determining equations to make the method real overall; this is likely to remove any advantage of the complex solutions.

Third, the nonsymmetric method $\prod S(w_it)$ where $S(t)$ is symmetric. Counting the free parameters is encouraging (two for fourth order at $m = 5$, for example, when type SS has only one parameter free) but a limited search of the solution set indicated that the truncation error was minimized at the type SS solutions. A cautionary tale, given the emphasis on counting free parameters in this paper.

3. Reversibility.

The overall symmetry (1.7) of both (1.9) and (1.10) is not just useful for simplifying the determining equations; it can also imply that the maps inherit reversibility properties of the differential equations. A vector field X is *reversible* under an involution R (a map with the property $R^2 \equiv 1$) if $XR = -RX$; its flow φ then has the property $R\varphi = \varphi^{-1}R$. That is, changing variables to $y = Rx$ is the same as reversing the direction of time. A system can be reversible with respect to more than

one involution. Define the symmetry set $\Sigma = \{x : R(x) = x\}$. When the dimension of Σ is half the dimension of the phase space, near Σ reversible systems “look like” Hamiltonian ones: they have a KAM theorem [12] and their eigenvalues have the same restrictions as those of Hamiltonian systems. Therefore one should definitely use a reversible integrator on such systems. In the Hamiltonian case, the further restriction of reversibility when R is an anti-symplectic map further restricts the dynamics. The generic codimension of fixed points with multiple eigenvalues depends on reversibility [12]. Symmetric orbits (those mapped into their time-reversal by R) intersect Σ twice, which makes them easier to find numerically; orbits bifurcating from them at eigenvalues different from one are also symmetric. All of these properties will be inherited by a reversible symplectic integrator.

Theorem 4. *Let the vector fields X , A and B be reversible under the involution R . Then a symmetric method φ of type S (Eq. (1.9)) is also reversible.*

$$\begin{aligned}
\text{Proof.} \quad R\varphi &= R \exp(a_1 t A) \exp(b_1 t B) \dots \exp(b_1 t B) \exp(a_1 t A) \\
&= \exp(-a_1 t A) R \exp(b_1 t B) \dots \exp(b_1 t B) \exp(a_1 t A) \\
&\quad \vdots \\
&= \exp(-a_1 t A) \exp(-b_1 t B) \dots \exp(-b_1 t B) \exp(-a_1 t A) R \\
&= \varphi^{-1} R \quad \blacklozenge
\end{aligned}$$

Example. Hamiltonian systems with Hamiltonian $\frac{1}{2}p^2 + V(q)$ ($p, q \in \mathbb{R}^n$) are reversible under $R : (q, p) \mapsto (q, -p)$, and, if V is even, also under $R' : (q, p) \mapsto (-q, p)$. Under the splitting (1.11), both A and B inherit these reversibilities, hence so do symmetric integrators of the form (1.9).

Similarly, for methods of the form (1.10) one needs S to be reversible. It is for the midpoint rule φ when R is linear:

$$\begin{aligned}
x \xrightarrow{\varphi} x' &= x + tX((x + x')/2) \xrightarrow{R} x'' = R(x + tX((x + Rx'')/2)) \\
&= Rx - tX((Rx + x'')/2)
\end{aligned}$$

and

$$x \xrightarrow{R} x' = Rx \xrightarrow{\varphi^{-1}} x'' = Rx - tX((Rx + x'')/2)$$

showing that $R\varphi = \varphi^{-1}R$.

Otherwise one should start with a first order map ψ , project onto reversible maps with $\varphi(t) = R\psi^{-1}(t/2)R\psi(t/2)$ [20], and apply Theorem 2 to φ .

4. Minimum-error methods.

Ideally one would like to know the fastest method for a given problem with a specified accuracy; in practice we can only determine good all-round methods of each particular order. We classify methods by their type, order, and number m of evaluations of B per time step. (The number of evaluations of A is one more when output is desired.)

There are many possible ways in which error constants can be defined. For standard integrators one uses some norm of the coefficients of the elementary differentials appearing in the first term of the local truncation error; an alternative is to measure the errors in the defining equations at the next highest $(p + 1)$ order. When working with Hamiltonian systems with Hamiltonian H , McLachlan and Atela [13] defined the *Hamiltonian truncation error* as $H - \hat{H}(t)$, where the map $x'(t)$ satisfies $d(x')/dt = J\nabla\hat{H}(t)$, and then measured the Euclidean norm of the elementary differentials multiplying the coefficient of t^p in $H - \hat{H}(t)$. One can also work with the energy error $H(x') - H(x)$ or the *autonomous Hamiltonian truncation error*, defined as the Hamiltonian of the vector field X_{p+1} where

$$\dots \exp(b_1 t B) \exp(a_1 t A) = \exp(X + t^{p+1} X_{p+1} + \dots). \quad (4.1)$$

The asymptotic series on the right hand side of (4.1) does not usually converge, but keeping only the first two terms is a good approximation when t is small enough. An advantage of this approach is that X_{p+1} is easy to calculate using the BCH formula and Poisson brackets [14,29].

There is a certain arbitrariness, not only in the choice of criterion, but also in the weighting of each term in the error or in the defining equations. In fact, which criterion is used does not matter much, because its only application is to compare similar methods to choose the “best overall” independently of any particular test vector field. The “optimal” methods are very similar under any of the criteria. Here we use the Hamiltonian truncation error of [13], because we are primarily interested in the symplectic case.

It is also important to compensate for the differing number m of evaluations of B in different methods. For example, if the amount of work to integrate to a fixed time is given, the time step for leapfrog ($m = 1$) will be half that of an $m = 2$ method. This will reduce the error in leapfrog by a factor of 4. Thus, for a method of order p with error constant E requiring m evaluations of B per time step, we shall use the *effective error constant* $(m/p)^p E$. (The normalizing factor p^{-p} is only present so that the error constants do not get confusingly large). All errors stated below are effective error constants. We have carried out searches for the best methods of each of types S, SS, and SB³A, for orders 2, 4, 6, and 8, and various values of m . For comparison, we also report the error constants when the methods are applied to the RKN case (1.11). Note that even if the effective error constant decreases as m increases, it may still not be advantageous to use the method with larger m for finite time steps. This will depend on the system being integrated and on the error required. The calculations reported below are analytic for orders 2 and 4, and numerical for orders 6 and 8. In most cases only the results are stated.

TYPE S, SYMMETRIC

Order 2. It may come as a surprise that the popular leapfrog (1.6) can be beaten, just. It has an error constant of 0.070. Taking $m = 2$, we get the family of second-order methods

$$\exp(z t A) \exp(\tfrac{1}{2} t B) \exp((1 - 2z) t A) \exp(\tfrac{1}{2} t B) \exp(z t A). \quad (4.2)$$

The error constant reaches a minimum of 0.026 at $z = (y^2 + 6y - 2)/12y \approx 0.1932$, where $y = (2\sqrt{326} - 36)^{1/3}$. Notice that all stages are in the $+t$ direction, so

that this method is also suitable for equations unstable in the $-t$ direction, such as discretizations of parabolic PDE's. Substeps in the $-t$ direction do not destroy stability in such cases, but can degrade it.

We illustrate the above discussion on error measurement for this case. It turns out that the errors in the order 3 defining equations, the local truncation error, the autonomous Hamiltonian truncation error, and the Hamiltonian truncation error are all minimized at the same value of z , 0.1932. For example, X_3 in (4.1) is $((6z - 1)[B, [B, A]] + (2 - 12z + 12z^2)[A, [A, B]])/24$ and we minimize $(6z - 1)^2 + (2 - 12z + 12z^2)^2$. For a Hamiltonian system split as $H = A(p) + B(q)$, the energy error is minimized at $z \approx 0.1912$.

Order 4. The most well-known method, $\varphi(z t) \varphi((1 - 2z)t) \varphi(z t)$ where $z = 1/(2 - \sqrt[3]{2})$ and φ is leapfrog (1.6), has an error constant of 0.098. We know it's worth looking at $m > 3$ because Suzuki's method [23] $\varphi(y t)^2 \varphi((1 - 4y)t) \varphi(y t)^2$, $y = 1/(4 - \sqrt[3]{4})$, has a smaller error constant, 0.055. We therefore explore the cases $m = 4$ and $m = 5$.

For $m = 4$ we have 5 unknowns and 4 determining equations, which can be reduced to a quadratic. Let the free parameter be b_1 . There are two solutions for each b_1 not in $[0, \frac{1}{2}]$. The minimum error is 0.014 near $b_1 = \frac{6}{11}$ (see Table 2).

$m = 5$ gives two free parameters; let them be b_1 and b_2 . There are several local minima of the error, all roughly equal. The absolute minimum is 0.0037, but to get simple coefficients we take the nearby values $b_1 = \frac{2}{5}$, $b_2 = -\frac{1}{10}$, which gives an error of 0.0046. This is 21 times smaller than the $m = 3$ method so we recommend it for all uses.

Orders ≥ 6 . From Table 1, note that order 6 requires $m = 9$. But if we take $m = 9$ then solutions of type SS will have one free parameter. This makes it extremely difficult to locate the isolated solutions of type S, and it seems unlikely that they would be more accurate than the best of type SS. Therefore we call a halt at order 4.

TYPE SS, SYMMETRIC COMPOSED OF SYMMETRIC STEPS

Order 2. These methods are composed of m steps, each already of order 2. $m = 2$ reduces to two identical steps, equivalent to halving the step size; $m = 3$ allows fourth order.

Order 4. For most applications, the (more accurate) type S methods will be preferred. If, however, the proposed symmetric stage is the midpoint rule, then SS may be required. It is not possible to do much better than Suzuki's method given above: the error constant can be reduced from 0.055 to 0.033, at $w_1 = 0.28$.

We have found no cases in which it is advantageous to take m even in an SS method, because doing so only provides the same number of unknowns as one obtains with $m - 1$ evaluations of B . For example, consider $S(z t) S((\frac{1}{2} - z)t)^2 S(z t)$. For this to be fourth order ($\sum w_i^3 = 0$) requires $12z^2 - 6z + 1 = 0$, which has no real solutions. At $m = 6$ the best method has error 0.31.

Order 6. $m = 7$ is required and leaves no free parameters. Yoshida [29] gave 3 solutions of the determining equations. The best is his method A, with error 0.063. At $m = 9$ (one free parameter) the optimal method has error 0.0115, and at $m = 11$ (two free parameters), 0.0087. This is a marginal reduction so we recommend the $m = 9$ method, given in Table 2.

Order 8. $m = 15$ is required. Yoshida gave 5 solutions, of which the best ("method

Table 2. Coefficients of symmetric composition methods.

Missing coefficients $w_{(m+1)/2}$ *etc.* are defined by the first-order conditions $\sum a_i = \sum b_i = \sum w_i = 1$. All numbers are correct to 20 digits. Type SB³A (see (1.9), (1.11)) may be used when $X = A + B$ and $[B, [B, [B, A]]] = 0$; type S (see (1.9)) may be used with any splitting $X = A + B$ or with an arbitrary first-order map, see Theorem 2; type SS may be used with these, or with any symmetric map $S(t)$, see (1.7), (1.10).

Order 2.

SS, $m = 1$: error 0.070, $w_1 = 1$ (leapfrog)

S, $m = 2$: error 0.026,

$$a_1 = \frac{y^2 + 6y - 2}{12y} \approx 0.1932, b_1 = \frac{1}{2}, y = (2\sqrt{326} - 36)^{1/3}$$

Order 4.

SS, $m = 3$: error 0.098 (as RKN, 0.087) $w_1 = (2 - \sqrt[3]{2})^{-1}$

SS, $m = 5$: error 0.055 (as RKN, 0.044) $w_1 = w_2 = (4 - \sqrt[3]{4})^{-1}$

SS, $m = 5$: error 0.033 (as RKN, 0.032) $w_1 = 0.28$,
 $w_2 = 0.62546642846767004501$

S, $m = 4$: error 0.014 (as RKN, 0.0081)

$$b_1 = \frac{6}{11}, a_1 = \frac{642 + \sqrt{471}}{3924}, a_2 = \frac{121}{3924}(12 - \sqrt{471})$$

S, $m = 5$: error 0.0046 (as RKN, 0.0045)

$$b_1 = \frac{2}{5}, b_2 = -\frac{1}{10}, a_1 = \frac{14 - \sqrt{19}}{108}, a_2 = \frac{20 - 7\sqrt{19}}{108},$$

SB³A, $m = 4$: error 0.0084,

$$b_1 = 1, b_2 = -\frac{1}{2}, a_1 = \frac{1}{2} - z, a_2 = -\frac{1}{3} + z, a_3 = \frac{2}{3}, z = \frac{\sqrt{7/8}}{3}$$

SB³A, $m = 5$: error 0.0011,

$$b_1 = -3/73, b_2 = 17/59, \\ a_1 = 0.40518861839525227722, a_2 = -0.28714404081652408900$$

Order 6.

SS, $m = 7$: error 0.063 (as RKN, 0.054)

$$w_1 = 0.78451361047755726382, w_2 = 0.23557321335935813368, \\ w_3 = -1.17767998417887100695$$

SS, $m = 9$: error 0.025 (as RKN, 0.023)

$$w_1 = 0.1867, w_2 = 0.55549702371247839916, \\ w_3 = 0.12946694891347535806, w_4 = -0.84326562338773460855$$

SB³A, $m = 7$: error 0.0013,

$$\begin{aligned} b_1 &= 0.00016600692650009894, b_2 = -0.37962421426377360608, \\ b_3 &= 0.68913741185181063674, b_4 = 0.38064159097092574080, \\ a_1 &= -1.01308797891717472981, a_2 = 1.18742957373254270702, \\ a_3 &= -0.01833585209646059034, a_4 = 0.34399425728109261313 \end{aligned}$$

Order 8.

SS, $m = 15$: error 0.14 (as RKN, 0.057)

$$\begin{aligned} w_1 &= 0.74167036435061295345, w_2 = -0.40910082580003159400, \\ w_3 &= 0.19075471029623837995, w_4 = -0.57386247111608226666, \\ w_5 &= 0.29906418130365592384, w_6 = 0.33462491824529818378, \\ w_7 &= 0.31529309239676659663 \end{aligned}$$

SS, $m = 17$: error 0.050 (as RKN, 0.023)

$$\begin{aligned} w_1 &= 25/194, w_2 = 0.58151408710525096243, \\ w_3 &= -0.41017537146985013753, w_4 = 0.18514693571658773265, \\ w_5 &= -0.40955234342085141934, w_6 = 0.14440594108001204106, \\ w_7 &= 0.27833550039367965131, w_8 = 0.31495668391629485789 \end{aligned}$$

D”) has error 5.00. But there are many more solutions to the determining equations. A computer search (over “reasonable” parameter ranges) found 100, of which the best has error 0.14 (this method was also found recently by Suzuki [25]). These solutions are discussed more later. With $m = 17$, optimizing over the free parameter gave error 0.05 (see Table 2), and with $m = 19$, error 0.06. Here we stopped.

TYPE SB³A, SYMMETRIC WITH $[B, [B, [B, A]]] = 0$

Order 2. This is identical to type S, above.

Order 4. As discussed previously, the determining equations are the same as for type S in this case, so one may use the methods derived above. However, the error terms are *not* the same, and the optimal methods are found at different parameter values.

For $m = 4$ (one free parameter, b_1) we found the bizarre situation that the error hardly depends on b_1 at all: in fact, it tends to 0.0096 as $b_1 \rightarrow \pm\infty$, compared to the minimum of 0.0078 at $b_1 = \frac{3}{5}$. Numerical methods do not usually allow their step sizes to tend to infinity!

Let $b_1 \rightarrow \infty$. Then the solution considered has $a_1 = (3 - \sqrt{3})/6 + \mathcal{O}(1/b_1)$, $a_2 = (\sqrt{3} - 2)/(24b_1^2) + \mathcal{O}(1/b_1^4)$, and $b_2 = \frac{1}{2} - b_1$. Consider the suspect steps $\exp(b_1 tB) \exp(a_2 tA) \exp(-b_1 tB) : (q, p) \mapsto (q', p'')$, where

$$\begin{aligned} p' &= p + b_1 t \nabla V(q) \\ q' &= q + \frac{\varepsilon}{b_1^2} t p', \quad \text{where } \varepsilon = \frac{\sqrt{3} - 2}{24} \approx -0.011 \\ p'' &= p' - b_1 t \nabla V(q') \end{aligned}$$

and we see that the two large steps in p almost cancel one another out because $b_1(\nabla V(q') - \nabla V(q)) = \mathcal{O}(1/b_1)$.

This seems to be just a curiosity, though, and for practical use we recommend either $b_1 = \frac{6}{11}$ as for type S, or $b_1 = 1$, which has the simple coefficients $b_2 = -\frac{1}{2}$, $a_1 = \frac{1}{2} - z$, $a_2 = -\frac{1}{3} + z$ where $z = \sqrt{7/8}/3$, $a_3 = \frac{2}{3}$, and error 0.0084.

For $m = 5$ one can do about four times better than the optimal type S method, with $b_1 \approx -0.04$, $b_2 \approx 0.29$, error 0.0011 (see Table 2).

There have been two previous searches for optimal order 4 RKN (equivalently, B³A) methods which did not impose symmetry (and thus were not reversible), *i.e.*, which used type NS. For example, $m = 4$ gives one free parameter because one of the eight determining equations (see (2.10)) is identically zero in the RKN case. Calvo and Sanz-Serna [4] optimized this case and found a method with error 0.0019; McLachlan and Atela [13] set $a_1 = 0$ and found a method with error 0.0024. These are both better than the symmetric $m = 4$ methods, but worse than the symmetric $m = 5$ method above. One could consider beating it by going to $m = 5$, but then [17] order 5 is possible.

Order 6. There is a coincidence that $m = 7$ gives isolated solutions for both types SS and SB³A, although the determining equations are different in the two cases, the solution set of SB³A containing that of SS. Okunbor and Skeel [17] found 16 methods and we do not find any more. Their best, given in Table 2, has an error constant of 0.0013; this cannot be substantially decreased by increasing m .

Order 8. We have not explored this case in detail. There is the problem that taking $m = 17$, to get isolated SB³A solutions, means that type SS solutions will have one free parameter. Still, Okunbor [16] does give three sets of coefficients which are SB³A and not SS. One can consider optimizing the SS methods for RKN, by minimizing their errors when $[B, [B, [B, A]]] = 0$; this did not lead to substantial improvements.

Calvo and Sanz-Serna [5] develop an optimized symmetric eighth-order RKN method with $m = 24$ which they found to be superior to Yoshida's method D in tests. We calculate its error constant to be 0.43 (as an RKN method, 0.19), larger than the error-0.05 (0.02 as RKN) method found above. However, they also imposed the additional constraint that the method be a composition of leapfrog steps (1.6). From Theorem 4, this means that their method is of type SS—that is, it works for all splittings $X = A + B$, not just for those of the form (1.11). The $m = 17$ SS method of Table 2 will be superior in the general and in the RKN case.

It would be useful to have some simple function of the stage lengths which characterized the accuracy of these methods. Figure 1 illustrates two possibilities. We have taken the 100 type SS, order 8, $m = 15$ methods and compared their errors to (a) $M_{\text{length}} = \sum_{i=1}^m |w_i|$, the “total distance traveled”, and (b) $M_{\text{neg}} = \min_{i=1}^m w_i$, the most negative stage. Clearly M_{length} and M_{neg} are strongly correlated. Although there is a unique method which minimizes M_{length} and maximizes M_{neg} , and has error very close to the minimum, there is substantial scatter even at the “good” end of these figures. Other methods with errors $2.5\times$ larger have very similar M_{length} and M_{neg} ; there are also very accurate methods with M_{length} throughout the range 6.5–10. Methods with $M_{\text{length}} \approx 10$ have errors varying by a factor of 1000. Here these heuristics can only be used to select a set of potentially accurate methods.

In the case of free parameters, consider type SS, order 6, $m = 9$ methods (which have one free parameter, say b_1). Here the heuristics are more promising: although they do not identify *globally* best methods, they do quite well *locally*, at least within

the arbitrariness of the error measurement. Figure 1(c) shows two solution paths for this case. The minimum error, 0.025, is at $b_1 = 0.19$, $M_{\text{length}} = 4.37$, $M_{\text{neg}} = -0.843$; but the latter are best at $b_1 = 0.39$, $M_{\text{length}} = 3.82$, $M_{\text{neg}} = -0.706$, where the error is 0.036. This suggests a search procedure in which one locally minimizes M_{length} from successive starting points, each local minimum being tested for its error constant. This strategy was used interactively to locate the above methods.

5. Numerical examples.

We shall illustrate the above methods and error calculations with some brief examples, considering only the symplectic case. As usual we use the energy error as an indicator of the degree to which phase space structures are preserved by the integrator.

An entirely separate issue is to consider the growth of the *pointwise* error in the solution, which we comment on briefly using some ideas from the theory of Hamiltonian systems [1], and test in the last example. Consider an n -degree-of-freedom ($2n$ -dimensional) Hamiltonian system to be integrated over a time interval T with time step t . If the system is integrable, the numerical integrator is near-integrable. Then chaotic numerical orbits occupy an exponentially small region of phase-space volume and can be ignored. ‘Most’ orbits are constrained to n -dimensional invariant tori which are $\mathcal{O}(t^p)$ away from the tori of the original system. The angular velocities on these tori are $\mathcal{O}(t^p)$ away from the correct ones, leading to an $\mathcal{O}(T)$ error in the solution. This has been observed in symplectic integrations of the Kepler problem [19]. Because the errors in nonsymplectic integrators are $\mathcal{O}(T^2)$ [19] (the actions and hence the frequencies drift linearly in time, leading to quadratic growth in the angle errors), a symplectic integrator will always beat a nonsymplectic integrator over sufficiently long time intervals.

At the other extreme there are fully chaotic orbits. Now nearby orbits diverge like $\mathcal{O}(\exp(\lambda T))$, where $\lambda > 0$ is a Lyapunov exponent; thus errors in the numerical solution will grow at the same rate. The numerical value of the error at a fixed time T only depends on the truncation error, which can be smaller for nonsymplectic methods.

In between there is a range of mixed behavior. Consider an elliptic fixed point (or periodic orbit) to which KAM theory applies. The region around this point has a positive density of invariant tori in both the original and the numerical systems. On these the error is still $\mathcal{O}(T)$. In the chaotic bands between these tori, the Lyapunov exponents are $\mathcal{O}(d^{j/2})$ where d is the distance from the fixed point and j is the order of the resonance driving the chaos. As $d \rightarrow 0$ and $j \rightarrow \infty$ there are increasingly longer time intervals in which the $\mathcal{O}(\exp(d^{j/2}T))$ component of the error is numerically smaller than the additional $\mathcal{O}(T^2)$ error term in a nonsymplectic integrator. Symplectic integrators can be competitive here too.

A separable Hamiltonian. Let

$$H = \frac{1}{2}(p_1^2 + p_2^2 + q_1^2 + q_2^2) + q_1^2 q_2 - \frac{1}{3} q_1^3,$$

the Hamiltonian for the Hénon-Heiles system. We have integrated this to $T = 500$ with initial conditions $(q_1, q_2, p_1, p_2) = (0.1, 0.1, 0, 0)$, using the splitting $H =$

$H_A(p) + H_B(q)$ (more energetic initial conditions gave similar results). The energy errors, which do not grow in time, are shown in Figure 2 and confirm the analyses of the preceding section. The non-RKN methods are included as they would be needed on systems not of this type.

Note the large advantage of the $m = 2$ method over leapfrog, and of type S over type SS in general. For the $m = 5$ SB³A method, sixth-order errors dominate the fourth-order errors for fewer than 3500 function evaluations (corresponding to the large step size $t = 0.7$ and errors larger than 10^{-5}). This highlights a shortcoming of only considering the leading term in the errors. One could consider decreasing the sixth-order error at some expense in the fourth-order error, to obtain a method more accurate at very large step sizes, perhaps even going to $m = 7$. However, there is no unique way to do this, as the breakeven point will depend on the system being integrated. The fact that the effective sixth-order error of our $m = 5$ method is already smaller than that of the $m = 7$ sixth-order SB³A method suggests that there is not much scope for improvement.

In Figure 2(c) we compare the most accurate methods of each type. Notice that the breakeven errors (the error at which one should switch to a higher-order method) are smaller for S than for SB³A methods. In the symplectic case, high-order composition methods will always be beaten by Gaussian Runge-Kutta (GRK, see [19]) for small enough step sizes. This is because, for order p , the latter only require $\frac{1}{2}p(1 + \mathcal{O}(t))$ evaluations of X , and have smaller truncation errors as well. For example, consider eighth-order methods. GRK8 needs four evaluations of X per iteration and has an error constant about 1000 times smaller than our $m = 17$ composition method, so will be superior if it converges to the required accuracy (say 10^{-16}) in fewer than $1000^{1/8}17/4 \approx 10$ iterations. In this problem, this occurs when $t < 0.25$. For GRK4, the breakeven is 7.3 iterations against type S, and 4.6 iterations against type SB³A. Very large problems will favor composition methods more, but problems in which $\prod_{i=1}^n \exp(tA_i)$ is more complicated than X (e.g., if the former involves non-elementary functions) will favor Gaussian Runge-Kutta.

A non-separable Hamiltonian. We next illustrate Theorem 2 for a more complicated splitting. Add a non-separable term to the Hénon-Heiles Hamiltonian:

$$\begin{aligned} H &= \frac{1}{2}(p_1^2 + p_2^2 + q_1^2 + q_2^2) + q_1^2 q_2 - \frac{1}{3}q_1^3 + (q_1 p_1)^2 \\ &= H_1(p) + H_2(q) + H_3(q, p) \end{aligned}$$

and let $\varphi(t) = \exp(tJ\nabla H_1)\exp(tJ\nabla H_2)\exp(tJ\nabla H_3)$ where $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$. Such a “3-map” splitting was first proposed by Forest and Ruth [7] and is also used in [25]. Clearly some consecutive terms in (2.5) can be amalgamated. We obtain the following results for the energy errors at constant work, for the initial condition $(q_1, q_2, p_1, p_2) = (0.1, 0.5, 0, 0)$, leading to a quasiperiodic orbit:

- S, $m = 2$: 4.6 \times better (more accurate) than leapfrog;
- S, $m = 5$: 6 \times better than the best SS $m = 5$; 19 \times better than SS $m = 3$;
- SB³A, $m = 5$: 21 \times better than the best SS $m = 5$.

The last item suggests that optimization we performed for the SB³A case is also relevant here, although we have not explored this issue. Other initial conditions lead to similar results.

Finally, we also checked the error in the solution itself. This depended more strongly on the initial condition. At $T = 500$, the results are

- S, $m = 2$: $1.4\times$ worse than leapfrog;
- S, $m = 5$: $34\times$ better than SS $m = 5$; $337\times$ better than SS $m = 3$.

Changing to a chaotic orbit starting at $(q_1, q_2, p_1, p_2) = (0.2, 0.5, 0, 0)$ increased the errors by a factor of about 4×10^6 :

- S, $m = 5$: $19\times$ better than SS $m = 5$; $57\times$ better than SS $m = 3$;

but the second-order methods could not integrate so far accurately. At $T = 100$,

- S, $m = 2$: $1.4\times$ better than leapfrog.

Although these numbers do not exactly mirror the error constants in Table 2, we still conclude that the composition (2.5) is advantageous. One could also use nonsymmetric methods in (2.5).

Acknowledgements. I am greatly indebted to Clint Scovel for our many useful and momentum-building discussions. I also thank H. Hermes, A. Iserles, M. Kowski, J. Meiss, J. M. Sanz-Serna, R. Skeel, and M. Suzuki for their advice and comments. The use of the Lazard elimination method in the proof of Theorem 1 was suggested by G. Melancon. Since preparing this manuscript, I have had useful talks with P.-V. Koseleff, who has also been studying free Lie algebras in numerical methods [9,10]. Theorem 1 is also found in [9, p.10]. The comments and references provided by two anonymous referees are gratefully acknowledged.

REFERENCES

1. V. I. Arnol'd, *Mathematical Methods of Classical Mechanics*, 2nd ed., Springer-Verlag, New York, 1989.
2. S. Benzel, Ge Zhong and C. Scovel, *Elementary construction of higher order Lie-Poisson integrators*, Phys. Lett. A **174** (1993), 229-232.
3. N. Bourbaki, *Lie Groups and Lie Algebras, Chapters 1-3*, Springer-Verlag, New York, 1989.
4. M. P. Calvo and J. M. Sanz-Serna, *The development of variable-step symplectic integrators, with applications to the two-body problem*, SIAM J. Sci. Comput. **14** (1993), 936-952.
5. M. P. Calvo and J. M. Sanz-Serna, *High order symplectic Runge-Kutta-Nyström methods*, SIAM J. Sci. Comput. **14** (1993), 1237-1252.
6. P. J. Channell and J. C. Scovel, *Integrators for Lie-Poisson dynamical systems*, Physica D **50** (1991), 80-88.
7. E. Forest and R. Ruth, *Fourth-order symplectic integration*, Physica D **43** (1990), 105-117.
8. A. Iserles, *Composite methods for numerical solution of stiff systems of ODE's*, SIAM J. Numer. Anal. **21** (1984), 340-351.
9. P.-V. Koseleff, *Calcul formel pour les méthodes de Lie en mécanique hamiltonienne*, Thesis, École Polytechnique, 1993.
10. P.-V. Koseleff, *Relations among formal Lie series and construction of symplectic integrators*, Applied Algebra, Algebraic Algorithms and Error-correcting Codes, AAECC-10 (Puerto Rico, 1993) (G. Cohen, T. Mora, and O. Moreno, eds.), Springer, Berlin New York, 1993.
11. M. Lothaire, *Combinatorics on Words*, Addison-Wesley, Massachusetts, 1983.
12. R. MacKay, *Some aspects of the dynamics and numerics of Hamiltonian systems*, The dynamics of numerics and the numerics of dynamics (D. S. Broomhead and A. Iserles, eds.), Oxford, 1992.
13. R. I. McLachlan and P. Atela, *The accuracy of symplectic integrators*, Nonlinearity **5** (1992), 541-562.
14. R. I. McLachlan, *Symplectic integration of Hamiltonian wave equations*, Numer. Math. (to appear).
15. R. I. McLachlan, *Explicit Lie-Poisson integration and the Euler equations*, Phys. Rev. Lett. **71** (1993), 3043-3046.
16. D. I. Okunbor, *Canonical integration methods for Hamiltonian dynamical systems*, Thesis, Comp. Sci., U. Illinois Urbana-Champaign (1992).
17. D. I. Okunbor and R. D. Skeel, *Canonical Runge-Kutta-Nyström methods of orders 5 and 6*, preprint.

18. R. D. Ruth, *A canonical integration technique*, IEEE Trans. Nucl. Sci. **NS-30** (1983), 2669–2671.
19. J. M. Sanz-Serna, *Symplectic integrators for Hamiltonian problems: An overview*, Acta Numerica 1992, Cambridge University Press, 1992, pp. 243–286.
20. J. C. Scovel, *Symplectic numerical integration of Hamiltonian systems*, The Geometry of Hamiltonian Systems (T. Ratiu, ed.), Springer-Verlag, New York, 1991.
21. H. J. Stetter, *Analysis of Discretization Methods for Ordinary Differential Equations*, Springer-Verlag, Berlin Heidelberg, 1973.
22. G. Strang, *Accurate partial difference methods. I: Linear Cauchy problems*, Arch. Rat. Mech. **12** (1963), 392–402.
23. M. Suzuki, *Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations*, Phys. Lett. A **146** (1990), 319–323.
24. M. Suzuki, *General theory of higher-order decomposition of exponential operators and symplectic integrators*, Phys. Lett. A **165** (1992), 387–395.
25. M. Suzuki, *Symplectic decomposition theory of time-evolution operators in nonseparable Hamiltonian systems*, preprint.
26. H. F. Trotter, *On the product of semi-groups of operators*, Proc. Am. Math. Soc. **10** (1959), 545–551.
27. J. Wisdom and M. Holman, *Symplectic maps for the N-body problem*, Astron. J. **102**(4) (1991), 1528–1538.
28. N. N. Yanenko, *A difference method of solution in the case of the multidimensional equation of heat conduction*, Dokl. Akad. Nauk USSR **125** (1959), 1207–1210. (Russian)
29. H. Yoshida, *Construction of higher order symplectic integrators*, Phys. Lett. A **150** (1990), 262–269.
30. M.-Q. Zhang, *Explicit unitary schemes to solve quantum operator equations of motion*, J. Stat. Phys. **65** (1991), 793–799.
31. M.-Q. Zhang, *Algorithms that preserve the volume amplification factor for linear systems*, Appl. Math. Lett. **6** (1993), 59–61.

Figure 1. Effect of stage lengths on the error constant of type SS methods. 100 eighth-order, $m = 15$ methods are compared for the effect of (a) M_{length} , the total distance traveled, and (b) M_{neg} , the most negative stage. (c) shows the connection between these variables for the most accurate families of type SS, sixth order, $m = 9$ methods.

Figure 2. Efficiency of various composition methods applied to the Hénon-Heiles system. (a), methods of order 2 and 4; (b), methods of order 6 and 8; (c), best methods of each type compared. The abscissae measure the number of evaluations of the force B over an integration time of 500.

PROGRAM IN APPLIED MATHEMATICS, UNIVERSITY OF COLORADO AT BOULDER, BOULDER, CO 80309-0526

Current address: Forschungsinstitut für Mathematik, ETH-Zentrum, 8092 Zürich, Switzerland

E-mail address: rxm@boulder.colorado.edu