UNIVERSITY OF CAMBRIDGE

Numerical Analysis Reports

Conservative methods for the Toda lattice equations

M.P. Calvo, A. Iserles and A. Zanna

DAMTP 1996/NA18
October 1996

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

Conservative methods for the Toda lattice equations

M. P. Calvo[*], A. Iserles[†] and A. Zanna[‡]

Abstract

We are concerned with the numerical integration of the Toda lattice equations by using different *conservative* methods. Numerical experiments suggest that the global error for isospectral schemes decreases exponentially with time but it is almost constant for either symplectic or more general integrators. We provide a theoretical explanation for these experimental findings.

1 Introduction

In the last years there has been a growing interest in studying methods for the numerical integration of systems of ordinary differential equations that preserve some characteristic features of the exact flow of the system. An instance of such methods are the well-known *symplectic* integrators [12] which preserve the symplectic structure of phase space for Hamiltonian systems. A second example are *isospectral* methods [2] which preserve the eigenvalues of the solution of the differential system

$$\frac{d}{dt}L(t) = [B(L(t)), L(t)], \quad t \geq 0, \tag{1.1}$$

[*]Departamento de Matemática Aplicada y Computación, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain, *Email: maripaz@mac.cie.uva.es*

[†]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom. *Email: A.Iserles@amtp.cam.ac.uk*

[‡]Newnham College, University of Cambridge, Cambridge, United Kingdom. *Email: A.Zanna@amtp.cam.ac.uk*

where $L(0)$ is a given $d \times d$ real matrix, $B(L)$ is a matrix function of L and $[\cdot, \cdot]$ denotes the commutator of two matrices.

Both special integrators are relevant to the numerical integration of the Toda lattice equations. The Hamiltonian system models the motion of a finite number of particles on the line with exponential interaction between nearest neighbours. The associated Hamiltonian function is

$$H(\mathbf{p}, \mathbf{q}) = \frac{1}{2}\sum_{i=1}^{d} p_i^2 + \sum_{i=1}^{d-1} \exp\left(2(q_i - q_{i+1})\right),$$

where q_i is the position of the ith particle and p_i is its momentum, $1 \leq i \leq d$. The equations of motion are

$$\frac{d}{dt}q_i = \frac{\partial H}{\partial p_i}, \qquad \frac{d}{dt}p_i = -\frac{\partial H}{\partial q_i}, \qquad 1 \leq i \leq d. \tag{1.2}$$

After introducing the new variables [8]

$$\begin{aligned} a_i &= -p_i, \qquad 1 \leq i \leq d, \\ b_i &= \exp\left(q_i - q_{i+1}\right), \qquad 1 \leq i \leq d-1, \end{aligned} \tag{1.3}$$

the differential system may be rewritten in the form

$$\begin{aligned} \frac{d}{dt}a_i &= 2(b_i^2 - b_{i-1}^2), \qquad 1 \leq i \leq d, \qquad b_0 = b_d = 0, \\ \frac{d}{dt}b_i &= b_i(a_{i+1} - a_i), \qquad 1 \leq i \leq d-1, \end{aligned} \tag{1.4}$$

or, equivalently, as (1.1) where $L(t)$ is the symmetric tridiagonal matrix with elements $L_{i,i}(t) = a_i(t)$, $1 \leq i \leq d$ and $L_{i,i+1}(t) = L_{i+1,i}(t) = b_i(t)$, $1 \leq i \leq d-1$ and $B(L) = L_+ - L_-$ is the difference between the upper and lower triangular parts of the matrix L. This corresponds to the Lax representation of the Toda lattice equations. Flaschka has shown [8] that the problem (1.1) is equivalent to solving

$$\begin{aligned} \frac{d}{dt}U(t) &= B(L(t))U(t), \quad t \geq 0, \\ U(0) &= Id, \end{aligned} \tag{1.5}$$

where Id denotes the identity matrix, in tandem with

$$L(t) = U(t)L(0)U^{-1}(t), \quad t \geq 0. \tag{1.6}$$

Note that, if L is a symmetric matrix (as for the Toda lattice equations), $B(L)$ is skew-symmetric and the solution of (1.5) is orthogonal. Thus, in (1.6), $U^{-1}(t)$ can be replaced by $U^T(t)$. The eigenvalues of the symmetric matrix $L(0)$ are a set of d integrals of the motion which are in involution [10], making (1.2) a completely integrable Hamiltonian system.

Symplectic integrators have proved themselves very efficient for the numerical integration of Hamiltonian systems [4], [12]. Furthermore, as (1.2) is a completely integrable Hamiltonian system, the error in the solution will grow at most linearly with time [1]. On the other hand, using an isospectral method for the numerical integration of (1.1), d independent first integrals of motion are preserved. Then, a good behaviour of the errors should also be expected in this case.

In Section 2 two families of isospectral methods are introduced. Section 3 is devoted to present a numerical comparison among different isospectral methods and standard ODE integrators when a lattice with three particles is considered. Finally, in Section 4 we analyze the experimental findings and present a theoretical study of the performance of general one-step methods in the integration of (1.4).

This is perhaps the place to mention that the computation of isospectral flows is an important special case of the numerical analysis of *homogeneous spaces*. The theme of the present paper is numerical techniques that address themselves specifically to isospectral flows. A more general computational approach has been recently advocated, based on a solution of differential equations in a Lie group [9, 11, 14, 15]. However, it is perhaps premature to analyse convergence of these new methods within the framework of this paper.

2 Constructing isospectral methods

It is possible to construct one-step methods for the numerical integration of (1.4) following Flaschka's ideas applied to the numerical integration of (1.1). Given an approximation L_n to the exact solution of (1.1) at time level t_n, we can advance by a step of length h in the numerical integration of the associated linear system

$$\frac{d}{dt}U = B(\hat{L}(t))U, \qquad U(t_n) = I, \qquad t \geq t_n, \tag{2.1}$$

where $\hat{L}(t)$ is the exact solution of (1.1) with the initial condition $\hat{L}(t_n) = L_n$. If U_{n+1} is the numerical approximation to the solution of (2.1) at time level $t_n + h$, we define an approximation to the solution of (1.1) at time level $t_n + h$ by

$$L_{n+1} = U_{n+1} L_n U_{n+1}^{-1}. \tag{2.2}$$

In this manner matrices L_n and L_{n+1} are similar for all n and the spectrum of the initial matrix $L(0)$ is preserved by the numerical scheme. We briefly describe two different ways of constructing isospectral methods.

2.1 Modified Gauss–Legendre methods

In [2] we have introduced a family of isospectral methods. The main idea is as follows: If $B(\hat{L}(t))$ is a skew-symmetric matrix, then the exact solution of (2.1) is a unitary matrix. As long as (2.1) is solved with a *unitary* method [7], U_{n+1} is itself a unitary matrix and (2.2) can be rewritten in the form

$$L_{n+1} = U_{n+1} L_n U_{n+1}^T. \tag{2.3}$$

The approximation L_{n+1} is orthogonally similar to L_n and both isospectrality and symmetry are preserved.

As unitary integrators for the numerical solution of (2.1) we choose the Gauss–Legendre Runge–Kutta methods (GLRK). However, as in (2.1) the function $B(\hat{L})$ depends on $\hat{L}(t)$, we need to know not just L_n and L_{n+1} but also the values of $\hat{L}$ at the Gauss–Legendre nodes $t_n + c_l h$, $l = 1, \ldots, s$.

To overcome this difficulty, the easiest strategy is to integrate from $t = 0$ the differential system

$$\begin{aligned} \frac{d}{dt}V(t) &= B(V(t)L(0)V^T(t))V(t), \quad t \geq 0 \\ V(0) &= Id, \end{aligned} \tag{2.4}$$

with the Gauss–Legendre method defining $L_n = V_n L(0) V_n^T$, $n = 1, 2, \ldots$. This is equivalent to solving (2.1) for $t \geq t_n$ replacing $\hat{L}(t)$ by $U(t) L_n U^T(t)$. Note that the intermediate stages of the method do not, in general, produce orthogonal matrices. Therefore the function B need be amended so that it maps all $d \times d$ matrices to skew-symmetric ones. This presents no problems whatsoever in the case of Toda lattice equations.

A second option, which has been considered in [2], consists of replacing the theoretical problem (2.1) by an approximate one by introducing a polynomial $\tilde{L}(t)$ which interpolates the exact flow $\hat{L}(t)$ at the points t_n and t_{n+1}. This is to say, we choose $\tilde{L}(t)$ as the Hermite interpolant of $\hat{L}(t)$ of degree $2s-1$ at $\{t_n, t_{n+1}\}$. In [2] it is proved that this procedure does not affect the order of the GLRK method.

2.2 Improved Taylor methods

In the previous subsection we have solved (1.1) by means of unitary integrators for the equation (2.1), but this is costly because these methods are implicit. An alternative, which we have addressed in greater detail in [3], is to advance the integration as follows. In each step we solve (2.1) by an arbitrary method and update the solution of (1.1) using (2.2). Since (2.2) is a similarity transformation, it follows at once that this procedure conserves the eigenvalues.

The equation (2.1) is linear but the function $B(\hat{L}(t))$ is available only at $t = t_n$. A viable alternative is to use a *Taylor* method of order r,

$$U_{n+1} = \sum_{l=0}^{r} \frac{1}{l!} U^{(l)}(t_n) h^l, \tag{2.5}$$

where $h = t_{n+1} - t_n$. Since (2.1) is linear, there exist functions $F_l(t)$, $0 \le l \le r+1$, such that $U^{(l)}(t) = F_l(t)U(t)$, where the sequence $\{F_l(t)\}$ satisfies the recursion

$$F_{l+1}(t) = F_l'(t) + F_l(t)B(\hat{L}(t)), \qquad F_0(t) = I, \qquad l \ge 0. \tag{2.6}$$

It is easy to convince ourselves that the main virtue of (2.5) is its low cost. However, the price that we are paying for this is a breakdown in symmetry. Provided that $L_n \in S(d)$, it is trivial to verify that $L_{n+1} \in S(d)$ if and only if $U_{n+1}^T U_{n+1}$ and L_n commute. Although this is, clearly, the case if U_{n+1} is orthogonal, this can be hardly expected in the case of the Taylor method (2.5), which produces perturbations of $\mathcal{O}(h^p)$ in the *symmetry error matrix*

$$E_n = \frac{1}{2}(L_n - L_n^T), \qquad n \in Z^+$$

The first observation in our quest to decrease the order of magnitude of E_n is that, alternatively to (2.1), we may solve the *adjoint equation*

$$\frac{d}{dt}V = -VB(\hat{L}(t)), \qquad V(t_n) = I, \qquad t \geq t_n. \tag{2.7}$$

This yields the adjoint Taylor method

$$L_{n+1} = V_{n+1}^{-1} L_n V_{n+1}, \tag{2.8}$$

where

$$V_{n+1} = \sum_{l=0}^{r} \frac{1}{l!} V^{(l)}(t_n) h^l \tag{2.9}$$

and $V^{(l)}(t) = V(t)G_l(t)$, $0 \leq l \leq r+1$. It is not difficult to see that as long as B is skew-symmetric, the sequence $\{G_l^T(t)\}$ satisfies the same recurrence (2.6) as the sequence $\{F_l(t)\}$.

Our suggestion is to employ (2.5) for even values of n and (2.9) otherwise, i.e.

$$\begin{aligned} L_{2n+1} &= U_{2n+1} L_{2n} U_{2n+1}^{-1}, \\ L_{2n+2} &= V_{2n+2}^{-1} L_{2n+1} V_{2n+2}, \end{aligned} \tag{2.10}$$

since, according to [3], this procedure results in a smaller error in symmetry error matrices E_n. In what follows this method will be referred to as *improved* Taylor method of order r. For a further analysis of these methods see [3].

3 Numerical results

In order to illustrate some of the results in this paper, let us consider a lattice with just three particles in the interval $0 \leq t \leq 640$, with the initial conditions

$$\mathbf{q}(0) = [0, 0, 0]^T, \qquad \mathbf{p}(0) = [1, -0.5, -0.5]^T.$$

Note that three is the minimun number of particles that need be considered in order to observe the different behaviour of isospectral methods and conventional ODE integrators which preserve quadratic invariants [2].

In the numerical experiments the integration has been performed using four different methods of order four. The first method is the Gauss–Legendre method of order four (GL4) applied to the differential system (1.4). This method preserves quadratic invariants of the problem but is not isospectral (see [2]). The second integrator is the modified Gauss–Legendre method of the same order (MGL4), obtained when the Gauss–Legendre method of order four is applied to the differential system (2.4), updating the L matrix at the end point as explained in Subsection 2.1. The third scheme is the fourth-order improved Taylor method (IT4) introduced in Subsection 2.2. These two methods are both isospectral. While MGL4 is implicit, IT4 is explicit but requires the computation of the inverse of a matrix at each step. This task has been done using the NAG subroutine F04AEF. The last method is the Gauss–Legendre method of order four applied to the Hamiltonian system (1.2), changing variables at the end point using (1.3). We will refer to this symplectic scheme as S4.

All the methods have been implemented with stepsizes $h = 1/4$, $1/8$, $\ldots$, $1/512$ and, for the implicit integrators, the tolerance used to terminate the fixed point iteration was 10^{-15}.

In order to compare the above methods, we have measured the errors in the variables a_i's and b_i's at $t = 5, 10, 20, 40, \ldots, 640$, using the Euclidean norm of $\mathbb{R}^5$. This is equivalent to measuring the Frobenius norm of the error in the numerical approximation to the matrix $L(t)$ at the same time levels. The exact solution for the comparison has been computed with quadruple precision by integrating the Hamiltonian system (1.2) with a symplectic explicit Runge–Kutta–Nyström method of order 8 [5] with stepsize 1/1024 and changing variables at the end point.

Figure 1 gives error against CPU time, measured in seconds, for the methods being considered and final times 5 (a) and 40 (b). The symbols used in all the figures for representing the data obtained running the different methods are the following:

(i) circles joined by a dotted line for GL4,

(ii) stars joined by a solid line for MGL4,

(iii) a dashed-dotted line with $\times$ signs for IT4, and

(iv) plus signs joined by a dashed line for S4.

We observe that at final time $t = 5$ both isospectral schemes are less efficient than the Gauss–Legendre method applied to (1.2) and (1.4) respectively. However, when the final time is 40, the isospectral integrators become

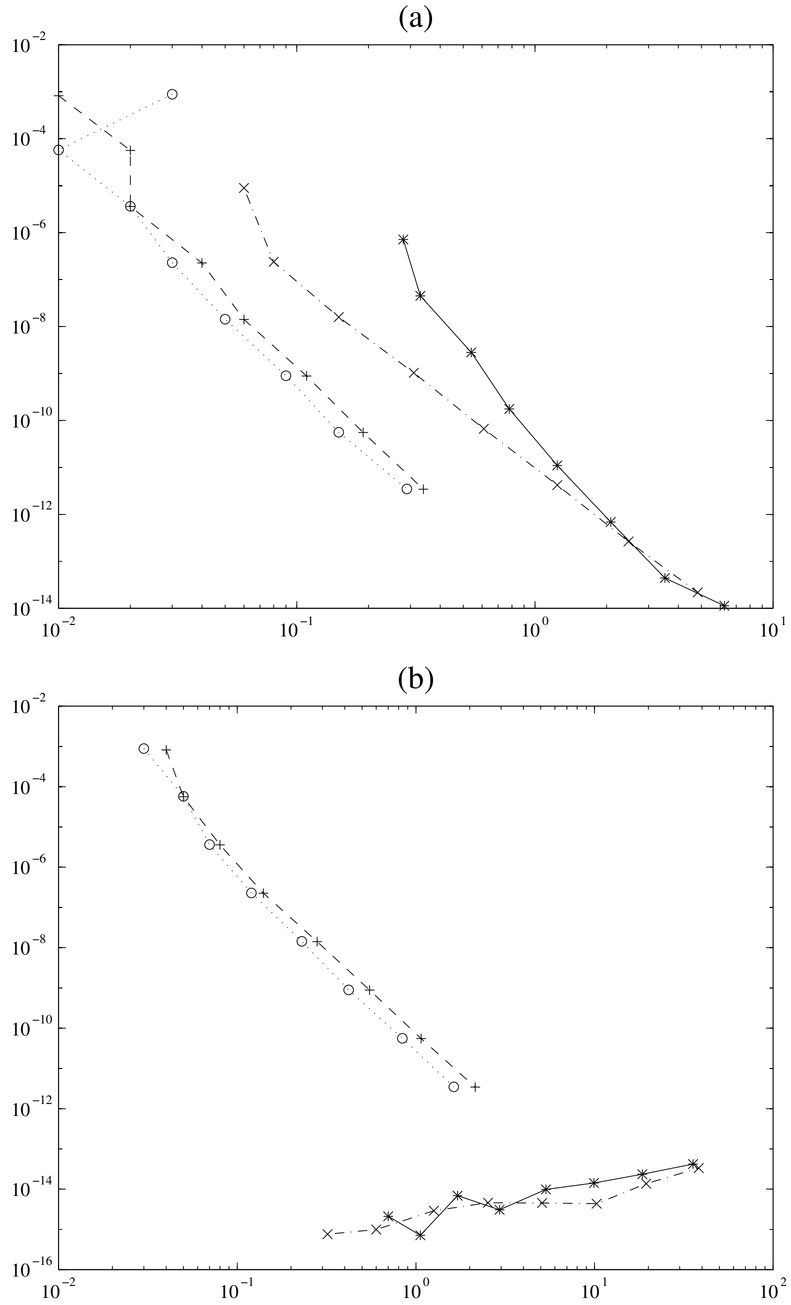

Figure 1: Error after 5 and 40 units of time

more efficient than the non-isospectral schemes. On the other hand, a comparison between both isospectral methods leads to a better performance of the explicit one.

Comparing both figures, we observe that the errors for the isospectral integrators are smaller at $t = 40$ than at $t = 5$. The errors for $t = 5$ range from 10^{-5} to 10^{-14} while the errors for $t = 40$ stay between 10^{-13} and 10^{-15}. Errors smaller than 10^{-15} should not be expected because the tolerance used for stopping the fixed point iteration is 10^{-15}. On the contrary, the errors obtained running the non-isospectral schemes are almost the same for both values of the final time, ranging from 10^{-3} for $h = 1/4$ to 10^{-12} for $h = 1/512$.

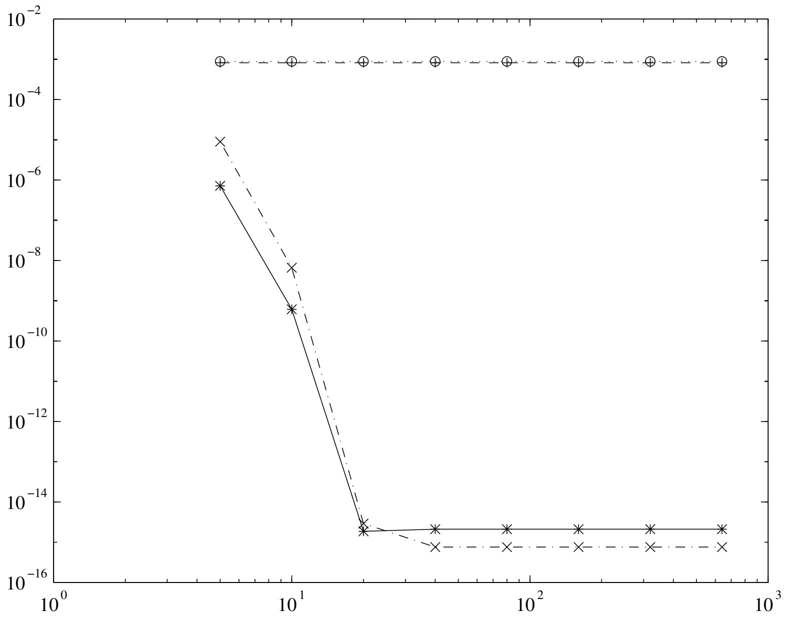

Figure 2: Error against time

In order to see this fact more clearly, we have plotted in Figure 2 the error against time for the four methods being considered when the step size is 1/4. For GL4 and S4 the errors are almost constant with respect to time as stated earlier. For the isospectral integrators, the error decreases with time at the begining, being almost constant for longer times, but they are of the same size as the tolerance used when solving the implicit equations.

In Section 4 these experimental findings are investigated theoretically.

4 Integration of the nonperiodic Toda lattice equations by one-step methods

4.1 The continuous problem

Let us consider the differential system (1.4) with initial condition $(\mathbf{a}^0, \mathbf{b}^0)$ satisfying $b_i^0 > 0, \quad 1 \le i \le d-1$.

Given $\mathbf{a} \in \mathbb{R}^d, \mathbf{b} \in \mathbb{R}^{d-1}$, let us denote by $L(\mathbf{a}, \mathbf{b})$ the $d \times d$ symmetric tridiagonal matrix

$$L(\mathbf{a}, \mathbf{b}) = \begin{bmatrix} a_1 & b_1 & & \\ b_1 & \ddots & \ddots & \\ & \ddots & \ddots & b_{d-1} \\ & & b_{d-1} & a_d \end{bmatrix},$$

and let $L_0 = L(\mathbf{a}^0, \mathbf{b}^0)$. We assume for the sake of simplicity that all the eigenvalues of L_0 are distinct.

It is well known [6] that the exact solution of (1.4) satisfies for $t \ge 0$ the bounds

$$\begin{aligned} |a_i(t) - \lambda_i| &\le c\,\mathrm{e}^{-2\mu t}, \quad 1 \le i \le d, \\ b_i^2(t) &\le c\,\mathrm{e}^{-2\mu t}, \quad 1 \le i \le d-1, \end{aligned} \tag{4.1}$$

where c is a real constant depending only on the initial condition $(\mathbf{a}^0, \mathbf{b}^0)$, $\lambda_1 > \lambda_2 > \cdots > \lambda_d$ are the eigenvalues of the symmetric matrix L_0 and $\mu = \min\limits_{1 \le i \le d-1} (\lambda_i - \lambda_{i+1}) > 0$.

Let R be a positive real number such that $R < \min\{\mu/5, 3\mu/(10d)\}$ and let us consider the space $\mathrm{M}_d[\mathbb{R}]$ of $d \times d$ real matrices, endowed with the norm $\|M\| = \max_{1 \le i,j \le d} |m_{i,j}|$, and the ball B_R of radius R, centred at the diagonal matrix $\Lambda = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_d)$. If $M = (m_{i,j})_{1 \le i,j \le d} \in B_R$, then it follows from the bound on R that $m_{1,1} > m_{2,2} > \cdots > m_{d,d}$ and $m_{i,i} - m_{i+1,i+1} \ge (3\mu)/5, 1 \le i \le d-1$.

Next we prove the following lemmas that will be used in the sequel.

Lemma 1 *Let us consider a solution $L(t) = (l_{i,j}(t))_{1 \le i,j \le d}$ of the differential system (1.1) associated with (1.4), without necessarily stipulating that the initial condition $L(0)$ is symmetric, and let us assume that $L(t) \in B_R$, $t \ge 0$.*

Then there exists $\hat{\mu} > 0$ such that the function $g(t) = \max_{1 \le i \ne j \le d} l_{i,j}^2(t)$ satisfies the inequality

$$g(t) \le g(0)\,\mathrm{e}^{-\hat{\mu}t}, \qquad t \ge 0. \tag{4.2}$$

Proof From (1.1) we see that, letting $B(L) = (b_{i,j})_{1 \le i,j \le d}$, it is true that

$$\frac{d}{dt}(l_{i,j}^2) = 2l_{i,j} \sum_{k=1}^{d} (b_{i,k} l_{k,j} - l_{i,k} b_{k,j}).$$

Exploiting $b_{i,i} = 0, 1 \le i \le d$ and $b_{i,j} = l_{i,j}$ for $i < j$, we deduce that if $i < j$ then

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(l_{i,j}^2) &= 2l_{i,j}\left[\left(-2\sum_{k=1}^{i-1} l_{i,k} l_{k,j}\right) - l_{i,i} l_{i,j} + l_{i,j} l_{j,j} + \left(2\sum_{k=j+1}^{d} l_{i,k} l_{k,j}\right)\right] \\
&= 2l_{i,j}^2 (l_{j,j} - l_{i,i}) + 4l_{i,j}\left(\sum_{k=j+1}^{d} l_{i,k} l_{k,j} - \sum_{k=1}^{i-1} l_{i,k} l_{k,j}\right) \qquad (4.3) \\
&\le \left(\frac{-6\mu}{5}\right) l_{i,j}^2 + 4l_{i,j}\left(\sum_{k=j+1}^{d} l_{i,k} l_{k,j} - \sum_{k=1}^{i-1} l_{i,k} l_{k,j}\right).
\end{aligned}$$

The same relation is also valid for $i > j$, using the fact that $b_{i,j} = -l_{i,j}$ for $i > j$.

The function g is continuous and has right derivative at every $t \ge 0$. Furthermore, given $t \ge 0$, there exist an interval $[t, t+\delta]$ and indices $I \ne J$ in $\{1, 2, \ldots, d\}$ such that $g(s) = l_{I,J}^2(s), t \le s \le t + \delta$. Using (4.3) we derive the following bound for the right derivative of g,

$$\frac{\mathrm{d}^+}{\mathrm{d}t} g(t) = \frac{\mathrm{d}}{\mathrm{d}t} l_{I,J}^2(t) \le -\frac{6\mu}{5} l_{I,J}^2(t) + 4dRg(t) = \left(-\frac{6\mu}{5} + 4dR\right) g(t).$$

The inequality (4.2) follows with $\hat{\mu} = 6\mu/5 - 4dR > 0$. □

In the particular case when the initial condition for (1.1) is a symmetric tridiagonal matrix, the solution $L(t)$ is also symmetric and tridiagonal for $t \ge 0$. If $L(t) \in B_R$ for $t \ge 0$, we deduce from (1.4) the bound

$$b_i(t) \le b_i(0)\,\mathrm{e}^{-3\mu t/5}, \qquad 1 \le i \le d-1, \qquad t \ge 0. \tag{4.4}$$

Lemma 2 *Let R' be a positive constant satisfying $R'(1 + dR'/\hat{\mu}) < R$ and let us consider a solution $L(t) = (l_{i,j}(t))_{1\leq i,j\leq d}$ of the differential system (1.1) associated with (1.4). Again, we do not stipulate that the initial condition $L(0)$ is symmetric. If $L(0) \in B_{R'}$, then $L(t) \in B_R$ for $t \geq 0$.*

Proof Let us assume that there exists $\tilde{t} > 0$ such that $L(t) \in B_R$ for $0 \leq t < \tilde{t}$ but $L(\tilde{t}) \notin B_R$.

If there exists an index $i \in \{1, 2, \dots, d\}$ such that $|l_{i,i}(\tilde{t}) - \lambda_i| > R$ then

$$\begin{aligned} |l_{i,i}(\tilde{t}) - \lambda_i| &\leq |l_{i,i}(\tilde{t}) - l_{i,i}(0)| + |l_{i,i}(0) - \lambda_i| \leq \int_0^{\tilde{t}} |l'_{i,i}(t)|\,\mathrm{d}t + R' \\ &\leq \int_0^{\tilde{t}} \left(\sum_{k\neq i} |l_{i,k}(t)||l_{k,i}(t)| \right) \mathrm{d}t + R'. \end{aligned}$$

Using now (4.2), which is valid in the range $0 \leq t < \tilde{t}$, we obtain

$$|l_{i,i}(\tilde{t}) - \lambda_i| \leq R' + d(R')^2 \int_0^{\tilde{t}} \mathrm{e}^{-\hat{\mu}t}\,\mathrm{d}t \leq R' \left(1 + \frac{dR'}{\hat{\mu}}\right) < R,$$

a contradiction.

Turning our attention to off-diagonal terms, for every indices $i \neq j$ between 1 and d such that $|l_{i,j}(\tilde{t})| > R$, we deduce in a similar manner that

$$\begin{aligned} |l_{i,j}(\tilde{t})| &\leq |l_{i,j}(\tilde{t}) - l_{i,j}(0)| + |l_{i,j}(0)| \leq \int_0^{\tilde{t}} |l'_{i,j}(t)|\,\mathrm{d}t + R' \\ &\leq \int_0^{\tilde{t}} \left(\sum_{k\neq i,j} |l_{i,k}(t)||l_{k,j}(t)| \right) \mathrm{d}t + R' \leq R' \left(1 + \frac{dR'}{\hat{\mu}}\right) < R, \end{aligned}$$

which is also a contradiction. $\square$

4.2 The numerical method

Suppose we have a consistent one-step method for the numerical integration of the differential system (1.4), with the initial condition $(\mathbf{a}^0, \mathbf{b}^0)$ or, equivalently, a one-step method for (1.1) with initial condition given by the symmetric tridiagonal matrix $L_0 = L(\mathbf{a}^0, \mathbf{b}^0)$. We write $L(t) = D(t) + O(t)$, where $D(t)$ and $O(t)$ are the diagonal and off-diagonal parts of the solution $L(t)$, respectively.

Given an approximation L_n to the solution of (1.1) at time level t_n, we similarly split $L_n = D_n + O_n$, $n = 0, 1, \ldots$. We can therefore write the one-step method as

$$D_{n+1} = \psi_h^1(D_n, O_n), \qquad O_{n+1} = \psi_h^2(D_n, O_n), \qquad n = 0, 1, \ldots.$$

Let us assume that there exist positive constants $M, \tilde{\mu}$ such that for sufficiently small h, if $L_n = D_n + O_n \in B_{R'}$, then

$$\|D_{n+1} - D_n\| \leq Mh\|O_n\|, \qquad \|O_{n+1}\| \leq \mathrm{e}^{-\tilde{\mu}h}\|O_n\|. \tag{4.5}$$

The first condition ensures that diagonal matrices are equilibria of the numerical method, as is the case with the continuous problem, while the second one implies exponential decay of the off-diagonal elements to zero. Both conditions are not too restrictive and they are satisfied by very general one-step methods, as we see in Subsections 4.3 and 4.4.

Under assumptions (4.5), let $R'' = \min\{R', \tilde{\mu}R'/(M + 2\tilde{\mu})\}$ and consider $B_{R''}$, the ball in $\mathrm{M}_d[\mathbb{R}]$ of radius R'' centered at the diagonal matrix Λ. As the symmetric matrix $L(\mathbf{a}(t), \mathbf{b}(t))$ associated with the solution of (1.4) converges to the diagonal matrix Λ as $t \to +\infty$ (see [6]), there exists $T_0 > 0$ such that $L(t) \in B_{R''/2}$ for all $t \geq T_0$. Standard error estimates show that, reducing the step-size if necessary, if $T_0 = n_0 h$, then the numerical approximation L_{n_0} to $L(T_0)$ satisfies

$$L_{n_0} \in B\left(L(T_0), R''/2\right),$$

therefore

$$L_{n_0} \in B_{R''} \subset B_{R'} \subset B_R.$$

Furthermore, it is also possible to choose h in such a way that

$$\frac{Mh}{1 - \mathrm{e}^{-\tilde{\mu}h}} < \frac{M}{\tilde{\mu}} + 1.$$

We will now prove the following lemma.

Lemma 3 *Subject to the previous assumptions, it is true that $L_n \in B_{R'}$ for all $n \geq n_0$.*

Proof Let us suppose that there exists $n' > n_0$ such that $L_n \in B_{R'}$ for all $n_0 \leq n < n'$ but $L_{n'} \notin B_{R'}$.

If $D_{n'} \notin B(\Lambda, R')$, then

$$\begin{aligned}
\|D_{n'} - \Lambda\| &\leq \sum_{j=n_0}^{n'-1} \|D_{j+1} - D_j\| + \|D_{n_0} - \Lambda\| \leq Mh \sum_{j=n_0}^{n'-1} \|O_j\| + R'' \\
&\leq Mh\|O_{n_0}\| \sum_{j=0}^{n'-n_0-1} \mathrm{e}^{-\tilde{\mu} jh} + R'' \leq R''\left(1 + Mh\frac{1}{1-\mathrm{e}^{-\tilde{\mu}h}}\right) \\
&< R''(2 + \frac{M}{\tilde{\mu}}) \leq R',
\end{aligned}$$

which is a contradiction.

If $O_{n'} \notin B(0, R')$, using (4.5) we obtain

$$\|O_{n'}\| \leq \mathrm{e}^{-\tilde{\mu}(n'-n_0)h}\|O_{n_0}\| < \|O_{n_0}\| \leq R'' \leq R',$$

which is again a contradiction.

Therefore, the sequence $\{L_n\}_{n=n_0}^{\infty}$ lies in $B_{R'}$. □

As a consequence of the above lemma, the estimate (4.5) is valid for all $n \geq n_0$. It follows that the sequence $\{O_n\}_{n=n_0}^{\infty}$ converges to the zero matrix as n tends to $+\infty$.

On the other hand, let be $k > j \geq n_0$. In this case

$$\begin{aligned}
\|D_k - D_j\| &\leq \sum_{l=j}^{k-1} \|D_{l+1} - D_l\| \leq Mh \sum_{l=j}^{k-1} \|O_l\| \\
&\leq Mh\|O_j\| \sum_{l=0}^{k-j-1} \mathrm{e}^{-\tilde{\mu} lh} \leq MhR' \frac{\mathrm{e}^{-\tilde{\mu}(j-n_0)h}}{1-\mathrm{e}^{-\tilde{\mu}h}},
\end{aligned}$$

which tends to the zero matrix as j tends to $+\infty$. Hence, the sequence $\{D_n\}_{n=n_0}^{\infty}$ has a limit D_∞. We have proved the main result of this paper.

Theorem 1 *Let us consider a one-step method for the numerical integration of the differential system (1.1) associated with (1.4) and let us assume that (4.5) holds. Then the sequence $\{L_n\}_{n=0}^{\infty}$ converges to a diagonal matrix D_∞ as n tends to $+\infty$.*

4.3 The case of Runge–Kutta methods

Let us suppose that the numerical integration of (1.4) is performed using a Runge–Kutta method of order r. Therefore, if $(\mathbf{a}^n, \mathbf{b}^n)$ is the numerical solution of (1.4) at time level t_n and we factorize $L_n = D_n + O_n = L(\mathbf{a}^n, \mathbf{b}^n)$, it is clear that $\|D_n\| = \|\mathbf{a}^n\|_\infty$ and $\|O_n\| = \|\mathbf{b}^n\|_\infty$.

Given $L_n = L(\mathbf{a}^n, \mathbf{b}^n)$ in $B_{R'}$,

$$\|O_{n+1}\| = \|\mathbf{b}^{n+1}\|_\infty \leq \|\mathbf{b}^{n+1} - \hat{\mathbf{b}}(t_n + h)\|_\infty + \|\hat{\mathbf{b}}(t_n + h)\|_\infty,$$

where $(\hat{\mathbf{a}}(t), \hat{\mathbf{b}}(t))$ is the exact solution of (1.4) with the initial condition $\hat{\mathbf{a}}(t_n) = \mathbf{a}^n$, $\hat{\mathbf{b}}(t_n) = \mathbf{b}^n$.

Since $(\mathbf{a}^n, \mathbf{b}^n) \in B_{R'}$, Lemma 2 ensures that $(\hat{\mathbf{a}}(t), \hat{\mathbf{b}}(t)) \in B_R$ for $t \geq t_n$, and $\hat{\mathbf{b}}$ satisfies (4.4). Then,

$$\|\hat{\mathbf{b}}(t_n + h)\|_\infty \leq \mathrm{e}^{-3\mu h/5}\|\mathbf{b}^n\|_\infty = \mathrm{e}^{-3\mu h/5}\|O_n\|.$$

The first term $\mathbf{b}^{n+1} - \hat{\mathbf{b}}(t_n+h)$ is the local truncation error that can be written as a linear combination of elementary differentials of the vector field for (1.4) evaluated at $(\mathbf{a}^n, \mathbf{b}^n)$. The components of these elementary differentials are sums of products of components of the vector field and components of partial derivatives of the vector field. In the problem we are dealing with, partial derivatives of the vector field are polynomials in the components of $(\mathbf{a}^n, \mathbf{b}^n)$, which are uniformly bounded in the ball $B_{R'}$. As the vector field itself can be bounded in the ball by the product of some positive constant and $\|\mathbf{b}^n\|_\infty$, we may write

$$\|\mathbf{b}^{n+1} - \hat{\mathbf{b}}(t_n + h)\|_\infty \leq Ch^{r+1}\|\mathbf{b}^n\|_\infty = Ch^{r+1}\|O_n\|,$$

and therefore the second inequality of (4.5) holds for some positive constant $\tilde{\mu}$.

On the other hand,

$$\|D_{n+1} - D_n\| = \|\mathbf{a}^{n+1} - \mathbf{a}^n\|_\infty \leq \|\mathbf{a}^{n+1} - \hat{\mathbf{a}}(t_n + h)\|_\infty + \|\hat{\mathbf{a}}(t_n + h) - \mathbf{a}^n\|_\infty.$$

In this case, we deduce from (1.4) that, for $1 \leq i \leq d$,

$$\hat{a}_i(t_n + h) - a_i^n = 2\int_0^h \left\{[\hat{b}_i(t_n + s)]^2 - [\hat{b}_{i-1}(t_n + s)]^2\right\} \mathrm{d}s.$$

As $0 \leq \hat{b}_i(t_n + s) \leq b_i^n$ for $0 \leq s \leq h, \quad 1 \leq i \leq d-1$ (cf. (4.4)), it is true that

$$\|\hat{\mathbf{a}}(t_n + h) - \mathbf{a}^n\|_\infty \leq 4hR'\|\mathbf{b}^n\|_\infty = 4hR'\|O_n\|.$$

The term $\mathbf{a}^{n+1} - \hat{\mathbf{a}}(t_n + h)$ corresponds to the local truncation error, that can be bounded in the ball by $Ch^{r+1}\|\mathbf{b}^n\|_\infty = Ch^{r+1}\|O_n\|$ as before. Thus, the first inequality of (4.5) also holds when the integration of (1.4) is performed using a Runge–Kutta method.

4.4 One-step methods via Flaschka's formalism

The only assumption we impose on the one-step method for solving (2.1) is that U_{n+1} can be written as a linear combination of elementary differentials. Note that both Runge–Kutta and Taylor series methods satisfy this condition.

Due to the special structure of systems (1.1), (2.1), if L_n lies in $B_{R'}$, then these elementary differentials are uniformly bounded by some positive constant, scaled by $\|B(L_n)\|$. If $L_n = D_n + O_n$, it is clear that $\|B(L_n)\| = \|O_n\|$. In what follows we will write $B_n = B(L_n)$.

We will see that, for methods constructed in this way, conditions (4.5) hold.

It is not difficult to deduce from (2.2) that

$$\begin{aligned} L_{n+1} - \hat{L}(t_n + h) &= [U_{n+1} - \hat{U}(t_n + h)]L_n U_{n+1}^{-1} \\ &- \hat{L}(t_n + h)[U_{n+1} - \hat{U}(t_n + h)]U_{n+1}^{-1}, \end{aligned}$$

where $\hat{U}(t)$ is the exact solution of (2.1). In order to bound $\|U_{n+1}^{-1}\|$, it is enough to observe that U_{n+1} has a Taylor expansion that can be written in terms of elementary differentials, whose behaviour is very well known. Then, $U_{n+1} = I + M_n h$, where $\|M_n\| \leq C\|B_n\|$. After reducing h if necessary, so that $hCR < 1$, we conclude that

$$\|U_{n+1}^{-1}\| = \|(I + M_n h)^{-1}\| = \|\sum_{k=0}^{\infty}(-M_n h)^k\| \leq \sum_{k=0}^{\infty}(hC\|B_n\|)^k \leq \frac{1}{1 - hCR}.$$

Therefore, the local truncation error $L_{n+1} - \hat{L}(t_n + h)$ behaves as the local truncation error $U_{n+1} - \hat{U}(t_n + h)$, that is to say, can be uniformly bounded in the ball $B_{R'}$ by $Kh^{r+1}\|O_n\|$ for a positive constant K.

Next, in order to get (4.5), we apply the same argument as used in Section 3. Let us denote by $\hat{D}(t)$ and $\hat{O}(t)$ the diagonal and off-diagonal elements of $\hat{L}(t)$. As $L_n \in B_{R'}$, $\hat{L}(t) \in B_R$ for $t \geq t_n$ and we can use (4.2),

$$\|D_{n+1} - D_n\| \leq \|D_{n+1} - \hat{D}(t_n + h)\| + \|\hat{D}(t_n + h) - D_n\|.$$

The first term on the right is the local truncation error and it can be bounded by $Kh^{r+1}\|O_n\|$ provided that $L_n \in B_{R'}$. Moreover, the ith diagonal component of the last term obeys the identity

$$\begin{aligned} \hat{D}_{i,i}(t_n + h) - D_{n,i,i} &= \int_0^h \hat{D}'_{i,i}(t_n + s)\,\mathrm{d}s \\ &= \int_0^h \left(-\sum_{k=1}^{i-1} \hat{l}_{i,k}\hat{l}_{k,i} + \sum_{k=i+1}^{d} \hat{l}_{i,k}\hat{l}_{k,i} \right)(t_n + s)\,\mathrm{d}s. \end{aligned}$$

Hence, using (4.2), we deduce that

$$\|\hat{D}(t_n + h) - D_n\| \leq dhR'\|O_n\|.$$

On the other hand, for the off-diagonal elements we write

$$\|O_{n+1}\| \leq \|O_{n+1} - \hat{O}(t_n + h)\| + \|\hat{O}(t_n + h)\|.$$

The first term is again the local truncation error which is uniformly bounded in $B_{R'}$ by $Kh^{r+1}\|O_n\|$, and the second one, as indicated by (4.2), is bounded by $\mathrm{e}^{-\hat{\mu}h/2}\|O_n\|$.

Thus, estimates (4.5) also hold for methods constructed via the Flaschka formalism. As a consequence, the matrix sequence $\{L_n\}_{n=0}^{\infty}$ has a limit as n tends to $+\infty$, which is a diagonal matrix. As numerical methods constructed in this way are isospectral (see [2]), the diagonal elements of the limit matrix are the eigenvalues of the initial condition L_0.

Acknowledgements. The authors are grateful to Prof. J. M. Sanz-Serna for his suggestions in connection with the theoretical analysis included in the paper. This research has been supported by project DGICYT PB92-254.

References

[1] M. P. Calvo & E. Hairer, "Accurate long-term integration of dynamical systems", *Appl. Numer. Maths* **18** (1995), 95–105.

[2] M. P. Calvo, A. Iserles & A. Zanna, "Numerical solution of isospectral flows", DAMTP Tech. Rep. 1995/NA03 (1995), University of Cambridge.

[3] M. P. Calvo, A. Iserles & A. Zanna, "Semi-explicit methods for isospectral flows", in preparation.

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

[5] M. P. Calvo & J. M. Sanz-Serna, "High order symplectic Runge–Kutta–Nyström methods", *SIAM J. Sci. Comput.* **14** (1993), 1237–1252.

[6] P. Deift, T. Nanda & C. Tomei, "Ordinary differential equations and the symmetric eigenvalue problem", *SIAM J. Numer. Anal.* **20** (1983), 1–22.

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

[8] H. Flaschka, "The Toda lattice", *Phys. Rev. B* **9** (1974), 1924–1925.

[9] A. Iserles & S.P. Nørsett, "On the computation of differential equations on matrix Lie groups", to appear.

[10] J. Moser, "Finitely many mass points on the line under the influence of an exponential potential – An integrable system", in *Dynamical Systems Theory and Applications*, (J. Moser, ed.) Springer-Verlag, Berlin (1975), 467–497.

[11] H. Munthe-Kaas, "Runge–Kutta methods on Lie groups", University of Bergen Tech. Rep. (1996).

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

[13] M. Toda, *Theory of Nonlinear Lattices*, Springer-Verlag, Berlin, (1981).

[14] A. Zanna, "The method of iterated commutators for ordinary differential equations on Lie groups", DAMTP Tech. Rep. 1996/NA12 (1996), University of Cambridge.

[15] A. Zanna & H. Munthe-Kaas, "Iterated commutators, Lie's reduction method and ordinary differential equations on matrix Lie groups", to appear in *Foundations of Computational Mathematics* (F. Cucker, ed.), Springer-Verlag, New York.