UNIVERSITY OF CAMBRIDGE

Numerical Analysis Reports

Runge–Kutta methods for orthogonal and isospectral flows

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

DAMTP 1996/NA08
May 1996

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

Runge–Kutta methods for orthogonal and isospectral flows

M.P. Calvo,[1] A. Iserles[2] and A. Zanna[3]

Abstract Orthogonal and isospectral flows occur in many applications and they possess important invariants. However, a naive application of Runge–Kutta methods is bound to render these invariants incorrectly. In this paper we describe how to retain relevant invariants with Runge–Kutta methods or, alternatively, incur an error in the rendition of the invariants which is significantly smaller than the overall numerical error.

[1]Departamento de Matemática Aplicada y Computación, Universidad de Valladolid, Valladolid, Spain. *E-mail: Maripaz@mac.cie.uva.es*

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

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

1 Orthogonal and isospectral flows

An increasing attention has been paid in the last decade to the issue of solving specific systems of ordinary differential equations in a manner that retains their qualitative features [13, 16]. This has led to a fast-growing body of knowledge that is already providing a range of novel algorithmic tools of relevance to a wide range of applications.

This paper is concerned with the numerical solution of orthogonal and isospectral flows and it is only appropriate to commence it with a brief review of such systems, their applications and their analytic features. We will employ in the sequel standard notation of matrix analysis [11] and symplectic geometry [9]: $\mathrm{M}_d[\mathbb{R}]$ denotes the set of all $d \times d$ real matrices, while $\mathrm{O}_d[\mathbb{R}]$, $\mathrm{S}_d[\mathbb{R}]$ and $\mathfrak{o}_d[\mathbb{R}]$ stand for the sets of $d \times d$ real orthogonal, symmetric and skew-symmetric matrices respectively.

A differential equation in $\mathrm{M}_d[\mathbb{R}]$,

$$Y' = G(t, Y), \quad t \geq 0, \qquad Y(0) = Y_0 \in \mathrm{M}_d[\mathbb{R}], \tag{1.1}$$

say, is said to be an *orthogonal flow* if $Y_0 \in \mathrm{O}_d[\mathbb{R}]$ implies that $Y(t) \in \mathrm{O}_d[\mathbb{R}]$ for all $t \geq 0$. It is an easy exercise to characterize orthogonal flows [14].

Theorem 1 *The flow* (1.1) *is orthogonal if and only if* $G(t,Y) = F(t,Y)Y$, *where* $F : \mathbb{R} \times \mathrm{O}_d[\mathbb{R}] \to \mathfrak{o}_d[\mathbb{R}]$. $\square$

Applications of orthogonal flows are listed in [8], a paper which, to our knowledge, is the first to consider a discretization of such flows by numerical methods that retain their orthogonality, i.e. produce a solution sequence $\{Y_n\}$, where $Y_n \approx Y(t_n)$, such that $Y_n \in \mathrm{O}_d[\mathbb{R}]$, $n \in \mathbb{Z}^+$.

The equation (1.1) is said to be an *isospectral flow* if $\sigma(Y(t)) = \sigma(Y_0)$, $t \geq 0$, where $\sigma(\,\cdot\,)$ denotes the spectrum [15, 17]. We prove the following theorem in detail, both for completeness and because its proof turns out to be an important clue to isospectral discretization.

Theorem 2 *If*

$$G(t,Y) = [B(t,Y), Y] = B(t,Y)Y - YB(t,Y), \tag{1.2}$$

where $B \in \mathrm{Lip}[\mathbb{R} \times \mathrm{M}_d[\mathbb{R}] \to \mathrm{M}_d[\mathbb{R}]]$, *the space of Lipschitz functions mapping* $\mathbb{R} \times \mathrm{M}_d[\mathbb{R}]$ *to* $\mathrm{M}_d[\mathbb{R}]$, *then the flow* (1.1) *is isospectral. Moreover, if* (1.1) *is an isospectral flow and* $Y(t)$ *is similar to* Y_0 *for every* $t \geq 0$ *then there exists a function* $B \in \mathrm{Lip}[\mathbb{R} \times \mathrm{M}_d[\mathbb{R}] \to \mathrm{M}_d[\mathbb{R}]]$ *such that* (1.2) *holds.*

Proof Let us suppose first that (1.2) is true and let the function U be the solution of the system

$$U' = B(t, Y(t))U, \quad t \geq 0, \qquad U(0) = I. \tag{1.3}$$

We set $L(t) := U(t)Y_0U^{-1}(t)$, $t \geq 0$. Therefore $L(0) = Y_0$ and

$$L' = U'Y_0U^{-1} - UY_0U^{-1}U'U^{-1} = BL - LB, \qquad t \geq 0.$$

In other words, the function L obeys both the differential equation and the initial condition. Since it is Lipschitz, we deduce that $Y \equiv L$. However, L is similar to Y_0, therefore isospectrality.

Next we assume that $Y(t)$ is similar to Y_0 (therefore both matrices share the same eigenvalues). Hence, there exists a function $U \in \mathrm{C}^1[\mathbb{R} \to \mathrm{M}_d[\mathbb{R}]]$ such that $\det U(t) \neq 0$ and $Y(t) = U(t)Y_0U^{-1}(t)$, $t \geq 0$. The form (1.2) follows by substituting the expression for Y into the differential equation and defining $B(\,\cdot\,, Y(\,\cdot\,)) := U'(\,\cdot\,)U^{-1}(\,\cdot\,)$. □

Note that if the function B is skew-symmetric then, according to Theorem 1, (1.3) is an orthogonal flow. In that case $Y(t) = U(t)Y_0U^T(t)$. Consequently, if $Y_0 \in \mathrm{S}_d[\mathbb{R}]$ and $B(\mathbb{R} \times \mathrm{S}_d[\mathbb{R}]) \subseteq \mathfrak{o}_d[\mathbb{R}]$ then $Y(t) \in \mathrm{S}_d[\mathbb{R}]$ for all $t \geq 0$. Most of isospectral flows in applications are symmetric, hence we are devoting much attention in the sequel to this case.

Perhaps the oldest, and certainly the best-known, example of isospectral flows is represented by *Toda lattice equations,* which describe the motion of a one-dimensional lattice of material particles. Assuming nearest-neighbour interactions, the equations of motion are

$$\begin{aligned}
y_1'' &= \phi'(y_2 - y_1), \\
y_\ell'' &= \phi'(y_{\ell+1} - y_\ell) - \phi'(y_\ell - y_{\ell-1}), \qquad \ell = 2, 3, \ldots, d-1, \\
y_d'' &= -\phi'(y_d - y_{d-1}),
\end{aligned}$$

where y_ℓ is the displacement of the ℓth particle (of unit mass) and $\phi(r) = \mathrm{e}^{-r} + r$, $r \in \mathbb{R}$, is the interaction potential. These equations can be written in a Hamiltonian form as

$$\begin{aligned}
p_1' &= -\mathrm{e}^{-(q_2 - q_1)}, \\
p_\ell' &= \mathrm{e}^{-(q_\ell - q_{\ell-1})} - \mathrm{e}^{-(q_{\ell+1} - q_\ell)}, \qquad \ell = 2, 3, \ldots, d-1, \\
p_d' &= \mathrm{e}^{-(q_d - q_{d-1})}, \\
q_\ell' &= p_\ell, \qquad \ell = 1, 2, \ldots, d
\end{aligned}$$

– in other words, with Hamiltonian function

$$H(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{2}\boldsymbol{p}^T\boldsymbol{p} + \sum_{k=1}^{d-1} \mathrm{e}^{-(q_{k+1}-q_k)}.$$

Following Flaschka [17], we let

$$\alpha_\ell := \frac{1}{2}\mathrm{e}^{-(q_{\ell+1}-q_\ell)/2}, \qquad \ell = 1,2,\ldots,d-1,$$
$$\beta_\ell := \frac{1}{2}p_\ell, \qquad \ell = 1,2,\ldots,d,$$

and set

$$\left.\begin{array}{l} L_{k,j} = \begin{cases} \beta_j, & k=j, \\ \alpha_{\min\{k,j\}}, & |k-j|=1, \\ 0, & \text{otherwise}, \end{cases} \\ B_{k,j}(L) = \begin{cases} \alpha_j, & k=j+1, \\ -\alpha_j & k=j-1, \\ 0 & \text{otherwise}, \end{cases} \end{array}\right. \qquad k,j=1,2,\ldots,d,$$

whereby the Toda lattice equations can be written in the isospectral form $L' = [B(L), L]$. Note that $L(0) \in \mathrm{S}_d[\mathbb{R}]$ and $B : \mathrm{S}_d[\mathbb{R}] \to \mathfrak{o}_d[\mathbb{R}]$.

Linear algebra is a rich source of isospectral flows. Perhaps the best-known example are the *QR flows,* $B(Y) = f(Y)_- - f(Y)_+$, where $f : \mathbb{C} \to \mathbb{C}$ is analytic on $\sigma(Y_0)$ and

$$\left.\begin{array}{l} (V_-)_{k,j} = \begin{cases} v_{k,j}, & k \ge j+1, \\ 0, & \text{otherwise}, \end{cases} \\ (V_+)_{k,j} = \begin{cases} v_{k,j}, & k \le j-1, \\ 0, & \text{otherwise}, \end{cases} \end{array}\right. \qquad k,j = 1,2,\ldots,d \quad \text{for} \quad V \in \mathrm{M}_d[\mathbb{R}].$$

While $f(z) = z$ corresponds to the Toda lattice, $f(z) = \log z$ is more interesting, since the set $\{Y(n)\}_{n\in\mathbb{Z}^+}$, namely the exact solution sampled at integer time intervals, is precisely the sequence of QR iterates, as applied to a positive definite $Y_0 \in \mathrm{S}_d[\mathbb{R}]$ [18].

Among many other examples of isospectral flows in linear algebra and optimization, we just mention the *inverse eigenvalue problem* for *Toeplitz* matrices: given $\lambda_1, \lambda_2, \ldots, \lambda_d \in \mathbb{R}$, we seek a Toeplitz matrix $T \in \mathrm{S}_d[\mathbb{R}]$ such that $\sigma(T) = \{\lambda_\ell\}_{\ell=1,2,\ldots,d}$. Let $Y_0 = \mathrm{diag}\,\{\lambda_1, \lambda_2, \ldots, \lambda_d\}$ and set

$$B_{j,k}(Y) := \begin{cases} y_{j,k-1} - y_{j+1,k}, & j < k, \\ 0, & j = k, \\ y_{j+1,k} - y_{j,k-1}, & j > k, \end{cases} \qquad j, k = 1, 2, \ldots, d.$$

Hence $Y_0 \in \mathrm{S}_d[\mathbb{R}]$ and $B : \mathrm{S}_d[\mathbb{R}] \to \mathfrak{o}_d[\mathbb{R}]$. Chu and Driessel proved that the equilibria of this flow are Toeplitz matrices. Therefore, as long as $\lim_{t\to\infty} Y(t)$ exists, it shares the eigenvalues of Y_0 and is a solution of the inverse eigenvalue problem [6].

Numerical discretization of isospectral flows presents an appealing alternative to more conventional methods of linear algebra. However, there is little point in discretizing an isospectral flow for the inverse eigenvalue problem, say, unless the underlying ODE retains the eigenvalues.

Runge–Kutta methods present the most promising vehicle for the retention of orthogonality or isospectrality. We proved in [5] that no multistep or multiderivative methods may respect quadratic invariants and, for reasons that will become clearer in the next section, this excludes such methods.

In Section 2 we review a number of results on Runge–Kutta methods and their retention of orthogonality or isospectrality. Section 3 is devoted to a new material on the application of general Runge–Kutta methods to either orthogonal or isospectral flows. In the first case orthogonality is lost but, by alternating between the original and an adjoint equation, the orthogonality error can be substantially reduced below the error in the solution of the underlying ODE. In the case of isospectral flows our method correctly retains eigenvalues but it may destroy symmetry. The breakdown in symmetry can be again substantially reduced by using an adjoint equation.

2 Invariant Runge–Kutta schemes

Invariant Runge–Kutta methods on the orthogonal group $\mathrm{O}_d[\mathbb{R}]$ have been characterized by Dieci, Russell and Van Vleck [8] and their result has been further sharpened in [2] by removing an unnecessary condition. An alternative approach identifies invariance on $\mathrm{O}_d[\mathbb{R}]$ with a special case of satisfaction of *quadratic conservation laws,* thereby employing a theorem due to Cooper [7].

Theorem 3 *Let $G(t, Y) = F(t, Y)Y$, where $F : \mathbb{R} \times \mathrm{M}_d[\mathbb{R}] \to \mathfrak{o}_d[\mathbb{R}]$. The*

Runge–Kutta method

$$\begin{aligned}\Phi_\ell &= Y_n + h\sum_{j=1}^{\nu} a_{\ell,j}K_j, \\ K_\ell &= G(t_n + c_\ell h, \Phi_\ell),\end{aligned} \qquad \ell = 1,2,\ldots,\nu$$

$$Y_{n+1} = Y_n + h\sum_{\ell=1}^{\nu} b_\ell K_\ell \tag{2.1}$$

is orthogonal if and only if $M = O$*, where* $m_{i,j} = b_i a_{i,j} + b_j a_{j,i} - b_i b_j$, $i,j = 1,2,\ldots,\nu$. □

Matters are substantially more complicated in the case of isospectral flows. Retention of eigenvalues is equivalent to the conservation of the quantities

$$\mathrm{tr}\,(Y_n)^k, \qquad k = 1,2,\ldots,d.$$

Provided that $d \geq 3$, it has been proved in [2] that no nonconfluent Runge–Kutta method can be assured to recover all these d integrals of motion. This is a special case of a failure of Runge–Kutta methods to recover all but the simplest (e.g. quadratic and symplectic) conservation laws [5].

Theorem 4 *Let* $d \geq 3$*. Given an arbitrary nonconfluent Runge–Kutta method* (2.1) *there exists an isospectral flow for which the method produces a non-isospectral solution sequence.* □

Isospectral flows can, however, be solved isospectrally with Runge–Kutta methods as long as, instead of applying such schemes to the original equation, we apply them to the system (1.3). In other words, we solve

$$U' = B(t_n + t, UY_nU^{-1})U, \quad t \geq 0, \qquad U(0) = I, \tag{2.2}$$

and let

$$Y_{n+1} = U_1 Y_n U_1^{-1}. \tag{2.3}$$

Note that, (2.3) being a similarity transformation, the numerical map is indeed isospectral. We refer to [4] for an asymptotic analysis for Runge–Kutta methods applied to (2.2),(2.3) for the special case of isospectral flows originating in the Toda lattice equations.

In the important special case $Y_0 \in \mathrm{S}_d[\mathbb{R}]$, $B(\mathbb{R} \times \mathrm{S}_d[\mathbb{R}]) \subseteq \mathfrak{o}_d[\mathbb{R}]$, solving (2.2) with an orthogonal Runge–Kutta scheme results in a symmetric solution sequence $\{Y_n\}_{n\in\mathbb{Z}^+}$. Unfortunately, it follows at once from Theorem 3 that all orthogonal Runge–Kutta methods are implicit. Therefore, as long as the cost of solution is at issue and we wish to employ an explicit Runge–Kutta scheme, we must reconcile ourselves to a breakdown in symmetry. In the next section we describe an approach that, although not preventing such a breakdown, decreases it a great deal, while not adding to the cost of computation.

3 The technique of the adjoint equation

Let us suppose that an orthogonal flow

$$Y' = F(t, Y)Y, \quad t \geq 0, \qquad Y(0) = Y_0 \in \mathrm{O}_d[\mathbb{R}], \tag{3.1}$$

is solved by a time-stepping method. We designate the matrix

$$\mathfrak{O}_n := Y_n Y_n^T - I, \qquad n \in \mathbb{Z}^+,$$

as the *orthogonality error* of the method. Given that the method is of order p, we deduce at once that $\mathfrak{O}_n = \mathcal{O}(h^{\tilde{p}})$, where $\tilde{p} \geq p$. Needless to say, $\mathfrak{O}_n \equiv O$ for orthogonal methods. The quantity

$$\mathfrak{O}_n^{\mathrm{loc}} := \mathfrak{O}_n - \mathfrak{O}_{n-1}, \qquad n \in \mathbb{N},$$

denotes the *local orthogonality error.* If we assume that $Y_{n-1} = Y(t_{n-1})$, then $\mathfrak{O}_{n-1} = O$ and $\mathfrak{O}_n^{\mathrm{loc}}$ is the orthogonality error caused by the local truncation error of the numerical method.

Likewise, given that a symmetric isospectral flow

$$Y' = [B(t, Y), Y], \quad t \geq 0, \qquad Y(0) = Y_0 \in \mathrm{S}_d[\mathbb{R}], \tag{3.2}$$

where $B(\mathbb{R} \times \mathrm{S}_d[\mathbb{R}]) \subseteq \mathfrak{o}_d[\mathbb{R}]$, is solved by the method (2.2),(2.3), we call

$$\mathfrak{S}_n := Y_n - Y_n^T, \qquad n \in \mathbb{Z}^+,$$

the *symmetry error* and reserve the notation $\mathfrak{S}_n^{\mathrm{loc}} := \mathfrak{S}_n - \mathfrak{S}_{n-1}$, $n \in \mathbb{N}$, for the *local symmetry error.* Note that such a method is *always* isospectral but that, unless an orthogonal method is used to discretize (2.2), in general $\mathfrak{S}_n \neq O$.

Given an orthogonal flow (3.1), it is a straightforward consequence of Theorem 1 that the inverse $X := Y^{-1}$ obeys the *adjoint* differential equation

$$X' = -XF(t, X^{-1}), \quad t \geq 0, \qquad X(0) = Y_0^{-1}. \tag{3.3}$$

Instead of discretizing (3.1), we might apply a numerical method to (3.2) and let $Y_n = X_n^{-1}$, $n \in \mathbb{N}$. As long as we use an orthogonal method, this practice makes little sense, since it has no perceptible benefits over the original equation, while adding to the computational cost. However, insofar as general numerical methods are concerned, we have advocated in [3] to alternate between (3.1) and (3.3), using the former to time-step for even values of n, say, and the latter for odd values. Likewise, we have suggested that the isospectral flow be solved by alternating between (2.2),(2.3) and

$$\begin{aligned} V' &= -VB(t_n+t, V^{-1}Y_nV), \quad t \geq 0, \qquad V(0)=I, \\ Y_{n+1} &= V_1^{-1}Y_nV_1. \end{aligned} \tag{3.4}$$

We proved in [3] that this approach leads to $\mathfrak{S}_{2n} = \mathcal{O}(h^{p+1})$ when the underlying scheme is a Taylor method and these results can be immediately extended to orthogonal flows. However, our method of proof relied heavily on error expansions in the derivatives of the underlying flow, hence it is invalid for Runge–Kutta methods.

Herewith we present an orthogonality and symmetry error analysis which is consistent with error expansions in elementary differentials, characteristic of Runge–Kutta methods [1, 10].

We consider the solution of (3.1) by means of the Runge–Kutta method (2.1). Recalling that $\det Y_0 \neq 0$, we may assume without loss of generality that $Y_0 = I$, since we may let $\tilde{Y}(t) := Y(t)Y_0^{-1}$, $\tilde{F}(t,\tilde{Y}) = F(t,\tilde{Y}Y_0)$ and solve

$$\tilde{Y}' = \tilde{F}(t,\tilde{Y})\tilde{Y}, \quad t \geq 0, \qquad \tilde{Y}(0) = I,$$

in place of (3.1). We assume in the sequel that (3.1) is autonomous, i.e. that the function F is independent of t. This minor loss of generality is amply compensated by a significant reduction in the complexity of algebraic manipulation.

Given $X, Y_0 \in \mathrm{M}_d[\mathbb{R}]$ and $\varepsilon \in \mathbb{R}$, it is true that

$$F(Y_0+\varepsilon X) = F(Y_0) + \varepsilon F_Y(X) + \frac{1}{2}\varepsilon^2 F_{YY}(X) + \cdots, \tag{3.5}$$

where

$$F_Y(X) = \sum_{i,j=1}^{d} \frac{\partial F(Y_0)}{\partial y_{i,j}} x_{i,j}, \quad F_{YY}(X) = \sum_{i,j=1}^{d}\sum_{k,\ell=1}^{d} \frac{\partial^2 F(Y_0)}{\partial y_{i,j}\partial y_{k,\ell}} x_{i,j}x_{k,\ell}, \tag{3.6}$$

etc. In effect, this is the same as treating each component of F as real-valued vector function and evaluating its differential along the direction defined by the vector corresponding to X (*Frechet derivative* or directional derivative).

Let $e_\ell \in \mathcal{I} \times \mathcal{I}$, where $\mathcal{I} = \{1, 2, \ldots, d\}$ and $\ell \in \{1, 2, \ldots, q\}$ for some $q \in \mathbb{N}$. Setting $\boldsymbol{e} = (e_\ell)_{\ell=1}^q$, we denote

$$F_e^{(q)} := \frac{\partial^q F(Y)}{\partial y_{e_1} \partial y_{e_2} \cdots \partial y_{e_q}}.$$

It is trivial to observe that every elementary differential of $F(Y)Y$ can be written in the form

$$\sum \mu_r F_{e_{r,1}}^{(j_{r,1})} F_{e_{r,2}}^{(j_{r,2})} \cdots F_{e_{r,s(r)}}^{(j_{r,s(r)})},$$

where the nonnegative integers $j_{r,1}, \ldots, j_{r,s(r)}$, the positive integers $s(r)$ and the real numbers μ_r are independent of the function F. Therefore the local error of an order-p Runge–Kutta method (2.1) can be written in the form

$$Y_n = \tilde{Y}(t_n) + h^{p+1} E_n + \mathcal{O}(h^{p+2}),$$

where $t_n = nh$, $n \in \mathbb{Z}^+$, $\tilde{Y}$ is the exact solution of $Y' = F(Y)Y$, $t \geq t_{n-1}$, with the initial condition $\tilde{Y}(t_{n-1}) = Y_{n-1}$ and

$$E_n = \sum \sigma_r F_{e_{r,1}}^{(i_{r,1})} F_{e_{r,2}}^{(i_{r,2})} \cdots F_{e_{r,q(r)}}^{(i_{r,q(r)})}. \tag{3.7}$$

Proposition 5 *Given that the orthogonal flow* (3.1) *is solved by a numerical method of order p, the orthogonality error obeys the identity*

$$\mathfrak{O}_n^{\mathrm{loc}} = h^{p+1}[E_n Y^T(t_n) + Y(t_n) E_n^T] + \mathcal{O}(h^{p+2}). \tag{3.8}$$

Proof Since the exact solution is orthogonal, it follows from the definition of orthogonality error and $\tilde{Y}(t) = Y(t) + \mathcal{O}(h^p)$ that $\tilde{Y}(t_n)^{-1} = Y^T(t_n) + \mathcal{O}(h^p)$, therefore

$$\begin{aligned}
\mathfrak{O}_n &= [\tilde{Y}(t_n) + h^{p+1}E_n][\tilde{Y}(t_n) + h^{p+1}E_n]^T - I + \mathcal{O}(h^{p+2}) \\
&= [I + h^{p+1}E_n Y^T(t_n)]\tilde{Y}(t_n)\tilde{Y}^T(t_n)[I + h^{p+1}Y(t_n)E_n^T] - I + \mathcal{O}(h^{p+2}) \\
&= [I + h^{p+1}E_n Y^T(t_n)](I + \mathfrak{O}_{n-1})[I + h^{p+1}Y(t_n)E_n^T] - I + \mathcal{O}(h^{p+2}) \\
&= \mathfrak{O}_{n-1} + h^{p+1}[E_n Y^T(t_n) + Y(t_n)E_n^T] + \mathcal{O}(h^{p+2}).
\end{aligned}$$

This, in tandem with $\mathfrak{O}_n^{\mathrm{loc}} = \mathfrak{O}_n - \mathfrak{O}_{n-1}$, prove the proposition. □

Since $Y(t) = I + \mathcal{O}(h^p)$ and F is skew-symmetric, the formulae (3.7),(3.8) imply that

$$\begin{aligned}\mathfrak{O}_n^{\rm loc} \;=\; h^{p+1} \sum \sigma_r \Big[& F_{e_{r,1}}^{(i_{r,1})} F_{e_{r,2}}^{(i_{r,2})} \cdots F_{e_{r,q(r)}}^{(i_{r,q(r)})} \\ & + (-1)^{q(r)} F_{e_{r,q(r)}}^{(i_{r,q(r)})} \cdots F_{e_{r,2}}^{(i_{r,2})} F_{e_{r,1}}^{(i_{r,1})} \Big] + \mathcal{O}(h^{p+2}).\end{aligned} \tag{3.9}$$

To illustrate our concepts, and also to derive an interesting interim result, we next focus our attention on the special case of a second-order explicit two-stage Runge–Kutta scheme. Its application to (3.1), performing one single time stepping, results in

$$\begin{aligned} \Phi_1 &= I, & K_1 &= F(\Phi_1)\Phi_1, \\ \Phi_2 &= I + c_2 h K_1, & K_2 &= F(\Phi_2)\Phi_2, \\ Y_1 &= I + h(b_1 K_1 + b_2 K_2), \end{aligned}$$

where $b_1 + b_2 = 1$, $b_2 c_2 = \frac{1}{2}$. Expansion in Frechet derivatives (cf. (3.5),(3.6)) yields

$$\begin{aligned} K_2 \;=\; & F + c_2 h \left(\sum_{i,j} \frac{\partial F}{\partial y_{i,j}} F_{i,j} + F^2 \right) \\ & + c_2^2 h^2 \left[\frac{1}{2} \sum_{i,j} \sum_{k,\ell} \frac{\partial^2 F}{\partial y_{i,j} \partial y_{k,\ell}} F_{i,j} F_{k,\ell} + \left(\sum_{i,j} \frac{\partial F}{\partial y_{i,j}} F_{i,j} \right) F \right] + \cdots, \end{aligned} \tag{3.10}$$

where all the functions (otherwise explicitely mentioned) are evaluated at $Y_0 = I$. Using the standard Runge–Kutta formalism [10], we can write K_2 as

$$K_2 = F + c_2 h \left[F_Y(F) + F^2 \right] + c_2^2 h^2 \left[\frac{1}{2} F_Y(F_Y(F)) + F_Y(F) F \right] + \cdots,$$

where the F_Ys should be understood as the Frechet derivatives (3.6). However, since matrix products do not commutate, we prefer to use (3.10). Therefore

$$\begin{aligned} Y_1 \;=\; & I + h(b_1 + b_2) F + h^2 b_2 c_2 \left(\sum_{i,j} \frac{\partial F}{\partial y_{i,j}} F_{i,j} + F^2 \right) \\ & + h^3 b_2 c_2^2 \left[\frac{1}{2} \sum_{i,j} \sum_{k,\ell} \frac{\partial^2 F}{\partial y_{i,j} \partial y_{k,\ell}} F_{i,j} F_{k,\ell} + \left(\sum_{i,j} \frac{\partial F}{\partial y_{i,j}} F_{i,j} \right) F \right] + \cdots. \end{aligned}$$

Substituting the values of b_1 and b_2, we thus deduce that

$$E_1 = \left(\frac{1}{2}c_2 - \frac{1}{3}\right) D_1 - \frac{1}{6}D_2,$$

where

$$\begin{aligned} D_1 &:= \frac{1}{2}\sum_{i,j}\sum_{k,\ell} \frac{\partial^2 F}{\partial y_{i,j}\partial y_{k,\ell}} F_{i,j}F_{k,\ell} + \left(\sum_{i,j}\frac{\partial F}{\partial y_{i,j}}F_{i,j}\right) F, \\ D_2 &:= F\sum_{i,j}\frac{\partial F}{\partial y_{i,j}}F_{i,j} + F^3, \end{aligned}$$

and, by virtue of skew-symmetry,

$$\mathfrak{O}_1^{\rm loc} = \frac{1}{2}h^3(c_2-1)\left[\left(\sum_{i,j}\frac{\partial F}{\partial y_{i,j}}F_{i,j}\right)F + F\left(\sum_{i,j}\frac{\partial F}{\partial y_{i,j}}F_{i,j}\right)\right] + \mathcal{O}(h^4).$$

In other words, as long as $c_2 = 1$, by induction on n we obtain a method with $\mathfrak{O}_n^{\rm loc} = \mathcal{O}(h^4)$, therefore $\mathfrak{O}_n = \mathcal{O}(h^3)$. This results in the scheme

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

which, among all two-stage explicit Runge–Kutta methods of order two, is the only one that produces an orthogonality error of $\mathcal{O}(h^3)$ in an interval $[0, t^*]$, whereby all other explicit two stage methods of order two produce an error of $\mathcal{O}(h^2)$. This analysis, however, cannot be easily generalized to higher order explicit Runge–Kutta. For example, among all three stage explicit methods of order three, none has $\mathfrak{O}_n^{\rm loc} = \mathcal{O}(h^5)$.

Returning to the general case we observe next that if, instead of (3.1), we apply a Runge–Kutta method to the equation

$$X' = XH(X), \quad t \ge t_n, \qquad X(t_n) = Y_n^{-1},$$

and let $X_{n+1} = X(t_{n+1}) + h^{p+1}\tilde{E}_{n+1} + \mathcal{O}(h^{p+2})$, then

$$\tilde{E}_{n+1} = \sum \sigma_r H_{e_{r,q(r)}}^{(i_{r,q(r)})} \cdots H_{e_{r,2}}^{(i_{r,2})} H_{e_{r,1}}^{(i_{r,1})}.$$

Note that the above estimate of $\tilde{E}_{n+1}$ remains true when H or $X(t_n)$ are perturbed by $\mathcal{O}(h)$, since the effects of such perturbations are relegated to higher-order terms.

Let us thus suppose that $H(X) = -F(X^{-1}) + \mathcal{O}(h)$. Because of skew-symmetry, we deduce that the local orthogonality error $\tilde{\mathfrak{O}}_n^{\rm loc}$ for the equation (3.3) is

$$\begin{aligned}\tilde{\mathfrak{O}}_n^{\rm loc} &= h^{p+1} \sum \sigma_r \Big[F_{e_{r,1}}^{(i_{r,1})} F_{e_{r,2}}^{(i_{r,2})} \cdots F_{e_{r,q(r)}}^{(i_{r,q(r)})} \\ &\qquad + (-1)^{q(r)} F_{e_{r,q(r)}}^{(i_{r,q(r)})} \cdots F_{e_{r,2}}^{(i_{r,2})} F_{e_{r,1}}^{(i_{r,1})} \Big] + \mathcal{O}(h^{p+2}).\end{aligned} \tag{3.11}$$

Theorem 6 *Suppose that the equation* (3.1) *is solved by an order-p Runge–Kutta method to advance from* t_{2n} *to* t_{2n+1}*, while progress from* t_{2n+1} *to* t_{2n+2} *is achieved by means of the adjoint equation* (3.3) *with* $H(X) = -F(X^{-1})$*. Given that* $F(\mathrm{O}_d[\mathbb{R}]) \subseteq \mathfrak{o}_d[\mathbb{R}]$*,* $Y_0 \in \mathrm{O}_d[\mathbb{R}]$*, it is true that*

$$\mathfrak{O}_{2n} = \mathcal{O}(h^{p+1}), \qquad n \in \mathbb{Z}^+.$$

Proof The orthogonality of X and $X(t_{2n+2}) = I + \mathcal{O}(h)$ imply that

$$\begin{aligned} X_{2n+2}^{-1} &= [X(t_{2n+2}) + \tilde{E}_{2n+2} h^{p+1} + \mathcal{O}(h^{p+2})]^{-1} \\ &= X^T(t_{2n+2}) - \tilde{E}_{2n+2} h^{p+1} + \mathcal{O}(h^{p+2}). \end{aligned}$$

Since the initial condition at t_{n+1} need be perturbed by the error in Y_{2n+1}, we deduce that the error across both steps is

$$Y_{2n+2} - Y(t_{2n+2}) = X_{2n+2}^{-1} - X^T(t_{2n+2}) = (E_{2n+1} - \tilde{E}_{2n+2}) h^{p+1} + \mathcal{O}(h^{p+2}).$$

Consequently,

$$\mathfrak{O}_{2n+2} - \mathfrak{O}_{2n} = [(E_{2n+1} + E_{2n+1}^T) - (\tilde{E}_{2n+2} + \tilde{E}_{2n+2}^T)] h^{p+1} + \mathcal{O}(h^{p+2}).$$

Thus, since $\mathfrak{O}_0 = O$, it follows by induction on n that $\mathfrak{O}_{2n+2} = \mathcal{O}(h^{p+1})$, thereby concluding the proof. □

Finally, we consider the issue of computing the solution of the isospectral flow (3.2) by alternating between (2.2),(2.3) and (3.4). As before, we restrict our attention to autonomous flows. We assume that $B : \mathrm{M}_d[\mathbb{R}] \to \mathfrak{o}_d[\mathbb{R}]$. An easy means to extend the domain of the function B from $\mathrm{S}_d[\mathbb{R}]$ to $\mathrm{M}_d[\mathbb{R}]$ while keeping the range in $\mathfrak{o}_d[\mathbb{R}]$ is by replacing it with

$$\tilde{B}(Y) := B(\tfrac{1}{2}(Y + Y^T)),$$

a procedure which retains the order of the underlying discretization. It is important to accompany this construction with a rejoinder. Numerical experiments provide compelling evidence that retaining the original function B results in a smaller error. This curious phenomenon is being currently investigated.

Proposition 7 *Suppose that the local orthogonality error of a Runge–Kutta method, as applied to* (2.2)*, is* $\mathfrak{O}_1^{\rm loc} = E_1 h^{\tilde{p}+1} + \mathcal{O}(h^{\tilde{p}+2})$*. Then*

$$\mathfrak{S}_{n+1}^{\rm loc} = \mathfrak{S}_n^{\rm loc} + [E_1^T, Y_{n+1}]h^{\tilde{p}+1} + \mathcal{O}(h^{\tilde{p}+2}), \qquad n \in \mathbb{N}. \tag{3.12}$$

Proof We commence by assuming that $Y_n \in \mathrm{S}_d[\mathbb{R}]$. Therefore

$$\mathfrak{S}_{n+1}^{\rm loc} = U_1 Y_n U_1^{-1} - U_1^{-T} Y_n U_1^T.$$

Hence

$$\begin{aligned} U_1^T \mathfrak{S}_{n+1}^{\rm loc} U_1 &= (U_1^T U_1) Y_n - Y_n (U_1^T U_1) \\ &= [I + (\mathfrak{O}_1^{\rm loc})^T] Y_n - Y_n [I + (\mathfrak{O}_1^{\rm loc})^T] = [(\mathfrak{O}_1^{\rm loc})^T, Y_n]. \end{aligned}$$

Since $U_1 = I + \mathcal{O}(h)$, the estimate (3.12) is true.

Next, we drop the assumption that $Y_n \in \mathrm{S}_d[\mathbb{R}]$. Substituting $Y_n^T = Y_n - \mathfrak{S}_n^{\rm loc}$, we obtain

$$\mathfrak{S}_{n+1}^{\rm loc} = U_1 Y_n U_1^{-1} - U_1^{-T} Y_n U_1^T + U_1^{-T} \mathfrak{S}_n^{\rm loc} U_1^T.$$

Consequently, $U_1 = I + \mathcal{O}(h)$ implies (3.12). □

Combining Theorem 6 and Proposition 7 furnishes a proof of our last result.

Theorem 9 *Let us suppose that* $B : \mathrm{M}_d[\mathbb{R}] \to \mathfrak{o}_d[\mathbb{R}]$, $Y_0 \in \mathrm{S}_d[\mathbb{R}]$ *and that the isospectral flow is solved by a p-order Runge–Kutta method by alternating between* (2.2),(2.3) *and* (3.4) *in even and odd steps, respectively. Then*

$$\mathfrak{S}_{2n} = \mathcal{O}(h^{p+1}), \qquad n \in \mathbb{N}.$$

□

Incidentally, we mention that Theorem 9 generalizes to variable stepsize strategies, as long as we allow h_{2n+1} to vary and let $h_{2n+2} = h_{2n+1}$, $n \in \mathbb{Z}^+$.

We conclude that, by alternating between the orthogonal flow and its adjoint, we can discretize isospectrally, while producing symmetry error substantially smaller than the error in the numerical method itself. In particular, this approach opens up the possibility of combining the computational affordability of explicit Runge–Kutta methods with retention of eigenvalues, while producing small symmetry error.

References

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

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

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

[4] M.P. Calvo, A. Iserles & A. Zanna, "Conservative methods for the Toda lattice equations", in preparation.

[5] M.P. Calvo, A. Iserles & A. Zanna, "Numerical methods on differentiable manifolds", in preparation.

[6] M.T. Chu, "A list of matrix flows with applications", in *Hamiltonian and Gradients Flows, Algorithms and Control* (A. Bloch, ed.), Fields Institute Communications, American Math. Soc. (1994), 87–97.

[7] G.J. Cooper, "Stability of Runge–Kutta methods for trajectory problems", *IMA J. Num. Anal.* **7** (1987), 1–13.

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

[9] A.T. Fomenko, *Symplectic Geometry,* Gordon & Breach, New York (1988).

[10] E. Hairer, S.P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems* (2nd ed.), Springer-Verlag, Berlin (1993).

[11] R.A. Horn and C.R. Johnson, *Matrix Analysis,* Cambridge University Press, Cambridge (1986).

[12] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations,* Cambridge University Press, Cambridge (1996).

[13] A. Iserles and A. Zanna, "Qualitative numerical analysis of ordinary differential equations", DAMTP Tech. Rep. *1995/NA5,* to appear in *Lectures in Applied Mathematics* (J. Renegar, M. Shub & S. Smale, eds), American Mathematical Society, Providence RI.

[14] P.J. Olver, *Applications of Lie Groups to Differential Equations* (2nd ed.), Springer-Verlag, New York, 1991.

[15] A.M. Perelomov, *Integrable Systems of Classical Mechanics and Lie Algebras,* vol. 1, Birkhäuser Verlag, Basel (1990).

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

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

[18] D.S. Watkins, "Isospectral flows", *SIAM Rev.* **26** (1984), 379–391.