Think globally, act locally: solving highly-oscillatory ordinary differential equations

Arieh Iserles

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

Abstract

In this paper we explore the solution of highly-oscillatory differential equations, with a special reference to the linear oscillator $y'' + g(t)y = 0$, where $g(t) \stackrel{t\to\infty}{\longrightarrow} +\infty$. Commencing from a global-error formula, we explore the accumulation of the error by Runge–Kutta and Magnus methods. Motivated by our analysis, we present a modification of the Magnus method which results in substantially better performance.

Key words: Highly-oscillatory equations, Runge–Kutta methods, Lie-group methods, WKB analysis.

1 On the global error incurred by time-stepping methods

Many ordinary differential equations occurring in applications of mathematics, from molecular dynamics to celestial mechanics, from quantum chemistry to atomic physics and beyond, exhibit highly-oscillatory behaviour. This phenomenon poses significant challenge to the design, analysis and implementation of discretization methods and has spawned a number of general approaches. Occasionally, underlying equations possess attributes like symplecticity or time-reversibility and their retention under discretisation leads to significantly slower error accumulation (Cano & Sanz-Serna, 1997; Petzold et al., 1997). In other cases, it is possible to extract the highly-oscillatory part of the solution by an appropriate projection (Hochbruck & Lubich, 1999; Ziv & Amdursky, 1977). In yet other cases it is possible to reformulate the underlying equations with respect to fast-rotating local coordinates (Scheid, 1983). The latter approach is adopted (albeit, in a very different form) in this paper and in (Iserles, 2001) and it proves particularly effective when combined with a Lie-group method (Iserles et al., 2000).

We commence this paper with a formula for the leading term in the expansion of the global error in powers of the step size h. This expression has been used time and again in numerical literature under different disguises but its earliest presentation is by Henrici (1963). Thus, suppose that the ordinary differential system

$$\boldsymbol{y}' = \boldsymbol{f}(t, \boldsymbol{y}), \quad t \geq 0, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0, \tag{1}$$

is solved by an arbitrary p-order time-stepping method with constant step size $h > 0$. Let $\boldsymbol{\varphi}_{t_1,t_2}$ be the *flow* of (1),

$$\boldsymbol{\varphi}_{t_1,t_2}(\boldsymbol{y}(t_1)) = \boldsymbol{y}(t_2), \qquad t_1, t_2 \geq 0,$$

and let Ψ_t stand for its *variational flow:* the fundamental solution of the matrix equation

$$\Psi_t' = \frac{\partial \boldsymbol{f}(t, \boldsymbol{y}(t))}{\partial \boldsymbol{y}} \Psi_t, \quad t \geq 0, \qquad \Psi_0 = I.$$

Theorem 1 *Let the system (1) be solved by a p-order convergent numerical method with constant step size $h > 0$ and incur local truncation error $\boldsymbol{\ell}(\boldsymbol{y}_n)$ at $t_{n+1} = (n+1)h$,*

$$\boldsymbol{y}_{n+1} - \boldsymbol{x}(t_{n+1}) = h^{p+1} \boldsymbol{\ell}(\boldsymbol{y}_n) + \mathcal{O}\left(h^{p+2}\right),$$

where $\boldsymbol{x}$ is the exact solution of (1) with the initial condition $\boldsymbol{x}(t_n) = \boldsymbol{y}_n$. Then the global error at t_n is

$$\boldsymbol{e}_h(t_n) = h^p \int_0^{t_n} \Psi_{t_n - \tau} \boldsymbol{\ell}(\boldsymbol{\varphi}_{0,\tau}(\boldsymbol{y}_0)) \mathrm{d}\tau + \mathcal{O}\left(h^{p+1}\right). \tag{2}$$

Elementary alternative proofs of Theorem 1 are available in (Iserles, 2001) and (Niesen, 2001).

In our quest for global error, the main tool at our disposal is the availability of local error estimates. Thus, *even if we wish to think globally, we need to act locally!* The estimate (2) demonstrates this point. Upon closer examination, it displays another remarkable feature. Noting that $\Psi_{t_n-\tau} = \Psi_{t_n} \Psi_\tau^{-1}$, we rewrite

(2) in the form

$$\boldsymbol{e}_h(t_n) = h^p \Psi_{t_n} \left[\boldsymbol{c}^* + \int_{t^*}^{t_n} \Psi_\tau^{-1} \boldsymbol{\ell}(\boldsymbol{\varphi}_{0,\tau}(\boldsymbol{y}_0)) \mathrm{d}\tau \right] + \mathcal{O}\left(h^{p+1}\right), \tag{3}$$

where $t^* > 0$ is given and $\boldsymbol{c}^* = \int_0^{t^*} \Psi_\tau^{-1} \boldsymbol{\ell}(\boldsymbol{\varphi}_{0,\tau}(\boldsymbol{y}_0))\mathrm{d}\tau$ is a constant. Suppose now that t^* is large enough, so that the behaviour of Ψ_t and $\boldsymbol{\varphi}_{0,t}$ for $t \geq t^*$ can be described solely in terms of asymptotic expansions. Then, up to the constant vector $\boldsymbol{c}^*$, so can be the global error in (3)!

The formula (3) is an example of an *a priori error estimate:* provided that, as often is the case, the asymptotics of Ψ and $\boldsymbol{\varphi}$ are known for $t \gg 1$, we can deduce the qualitative behaviour of the global error without even computing the solution. This yields a convenient tool to compare the performance of different methods for different classes of initial-value problems. This formula is less useful that *a priori error estimates* in local monitoring of error in the course of numerical solution.

2 Classical and Magnus methods for the linear oscillator

Provided that (1) is linear, it is true that $\boldsymbol{\varphi}_{0,t} = \Psi_t \boldsymbol{y}_0$ and it suffices for the application of the global-error formula (3) to know the asymptotics of Ψ_t. One important instance when such an estimate is widely available is the *linear oscillator*

$$y'' + g(t)y = 0, \tag{4}$$

where $g(t) > 0$ for $t > 0$, $\lim_{t\to\infty} g(t) = +\infty$ (our analysis remains valid if g is just 'very large' for $t \gg 1$) and $g'(t)$ is relatively small, $|g'(t)| = o(g(t))$, $t \to \infty$. Rephrasing (4) as a two-variable linear system, we can use WKB analysis (Hinch, 1991) to argue that

$$\Psi_t \approx \begin{bmatrix} g^{-1/4}(t) & 0 \\ 0 & g^{1/4}(t) \end{bmatrix} \begin{bmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{bmatrix} S, \qquad t \gg 1, \tag{5}$$

where $\theta(t) = \int_0^t g^{1/2}(\tau)\mathrm{d}\tau$ and S is a constant 2×2 matrix. Note that (5) conveniently separates *amplitude* and *frequency,* in the leftmost and the central matrix respectively. Thus, while the amplitude of y decays, the frequency of its oscillation increases.

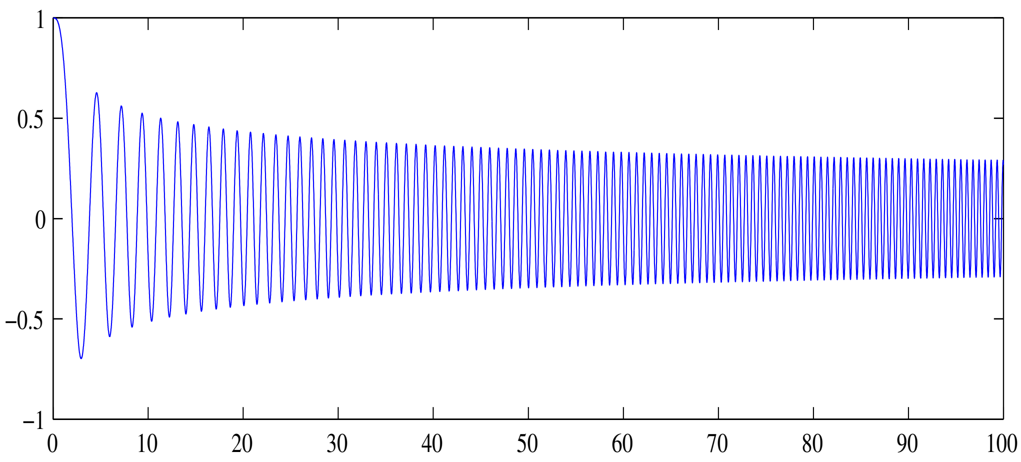


Fig. 1. The exact solution of the Airy equation (6).

An example running through this paper is the *Airy equation*

$$y'' + ty = 0. \tag{6}$$

Setting $g(t) = t$ in (5), we deduce that

$$\Psi_t \approx \begin{bmatrix} t^{-1/4} & 0 \\ 0 & t^{1/4} \end{bmatrix} \begin{bmatrix} \cos\frac{2}{3}t^{3/2} & \sin\frac{2}{3}t^{3/2} \\ -\sin\frac{2}{3}t^{3/2} & \cos\frac{2}{3}t^{3/2} \end{bmatrix} S, \qquad t \gg 1.$$

Solving $\boldsymbol{y}' = \boldsymbol{f}(\boldsymbol{y})$ with a classical p-order numerical method, e.g. Runge–Kutta, multistep scheme, Taylor expansion or general linear method, the local truncation error can be expressed as a linear combination of *elementary differentials:* terms of the form

$$\sum_{j_1,j_2,\ldots,j_p} f^i_{\boldsymbol{k}_1} f^{j_1}_{\boldsymbol{k}_2} \cdots f^{j_{p-1}}_{\boldsymbol{k}_p} f^{j_p},$$

where v^i is the ith component of the vector $\boldsymbol{v}$,

$$f^i_{\boldsymbol{m}} = \frac{\partial^{m_1+m_2+\cdots+m_{p+1}} f^i(\boldsymbol{y})}{\partial y^{m_1} \partial y^{m_2} \cdots \partial y^{m_{p+1}}} \qquad \text{and} \qquad \bigcup_{l=1}^{p} \boldsymbol{k}_l = \boldsymbol{j}.$$

Specialising to (4), it possible to identify the largest elementary differential, thereby demonstrating that, for even order $p = 2s$,

$$\boldsymbol{\ell}(\boldsymbol{y}) \approx c(-g)^s \begin{bmatrix} 0 & 1 \\ -g & 0 \end{bmatrix} \boldsymbol{y},$$

where c is a constant (Iserles, 2001). Substitution in (3) and easy algebra yield the following expression for the leading global-error term.

Theorem 2 (Iserles 2001) *Suppose that the linear oscillator (4) is solved by a classical convergent method of order $2s$. Then*

$$\boldsymbol{e}_h(t) \approx h^{2s} \left[c_1 + (-1)^s c_2 \int_0^t g^{s+1/2}(\tau)\mathrm{d}\tau \right] \begin{bmatrix} g^{-1/4}(t) & 0 \\ 0 & g^{1/4}(t) \end{bmatrix} G(t)\boldsymbol{s}, \qquad (7)$$

where $G(t) = \begin{bmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{bmatrix}$, *$c_1, c_2$ are scalar constants and $\boldsymbol{s}$ is a 2-vector.*

We hasten to point out an important issue. In our analysis we fix (sufficiently large) $t > 0$ and expand in powers of h. In principle, it is more desirable to seek *uniform* estimates, valid for all $t \in [0, \infty)$. Unfortunately, for most methods this makes little sense: once $h > 0$ is fixed, sooner or later accuracy is lost altogether. Stability of a method might keep the numerical solution bounded, except that the latter is no longer a valid approximation to the exact solution of (4). In Section 3 we describe a new method which admits uniform estimates for (4) with 'nice' functions g.

Note an interesting point about Theorem 2: the accumulation of global error is independent of stability! This is in variance with Hamiltonian systems, where symplectic or time-reversible methods display superior error propagation (Cano & Sanz-Serna, 1997).

Figure 2 displays the error committed by the familiar fourth-order explicit Runge–Kutta scheme and by the (A-stable, symplectic) fourth-order Gauss–Legendre–Runge–Kutta method, with Butcher tableaux

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \qquad \text{and} \qquad \begin{array}{c|cc} \frac{1}{2} - \frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{1}{2} + \frac{\sqrt{3}}{6} & \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

as well as by a Magnus method of similar order, which will be described in the next few paragraphs. Note that for both Runge–Kutta methods

$$y_{t/h} - y(t) = \tfrac{2}{7} c h^4 t^{13/4} \chi(t) + \mathcal{O}\left(h^5\right),$$

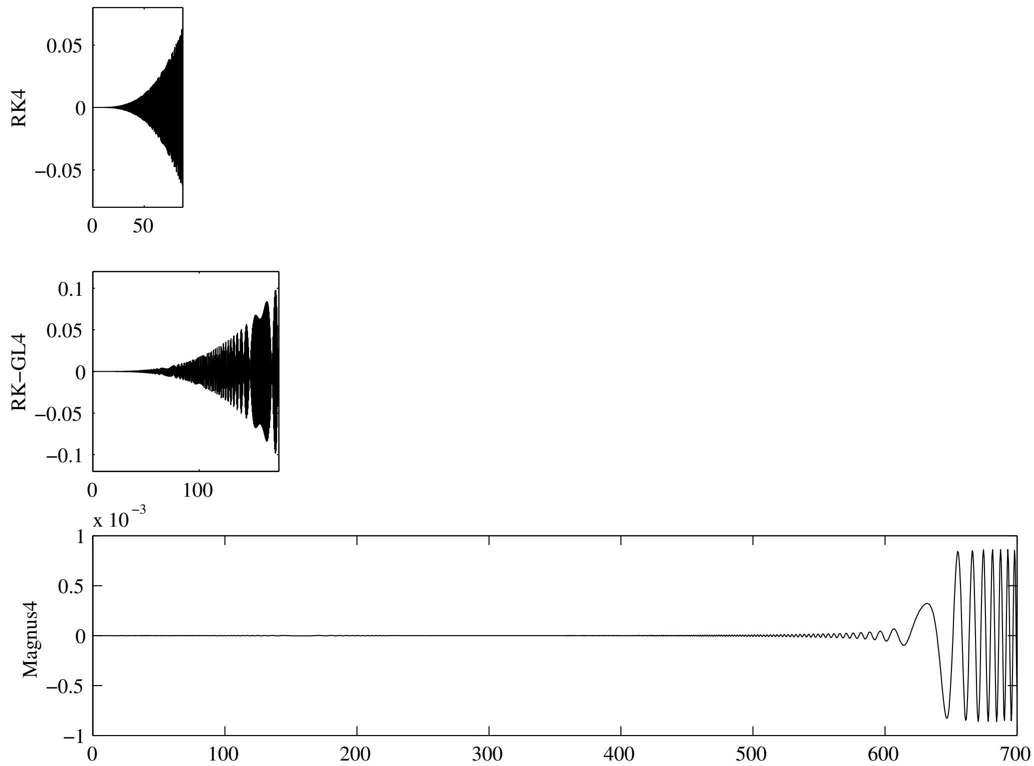


Fig. 2. Global error for the Airy equation with different fourth-order methods and $h = \frac{1}{16}$.

where $\chi(t) \in [-1, 1]$ is a fast-oscillating function. The only difference is that c is smaller for the A-stable method, but this does not change the qualitative behaviour of the global error.

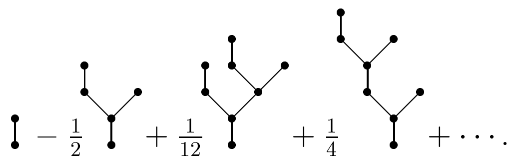
A *Magnus method* is an example of a Lie-group method. An important characteristic of such methods, that have been surveyed in (Iserles et al., 2000), is that, provided that the solution of (1) evolves in a Lie group $\mathcal{G}$, so does the numerical solution. This retention of crucial geometric structure is desirable in many applications, although it is probably only of marginal interest in the present context.

Most Lie group methods solve (1) by mapping it first to the underlying Lie algebra $\mathfrak{g}$, say (a procedure known as 'trivialisation'), solving it there and mapping the solution back to $\mathcal{G}$. Since $\mathfrak{g}$ is a linear space, closed under commutation, its structure is respected by any method which restricts itself to linear-space operations and commutations. In particular, Wilhelm Magnus (1954) used the exponential trivialisation, followed by a Lie-algebra expansions in multivariate integrals of iterated commutators. The method has been formalised, employing graph theory, by Iserles & Nørsett (1999), who have also demonstrated how to approximate the underlying multivariate integrals with surprisingly small number of function evaluations. The theory of Magnus expansions has been further perfected by a number of mathematicians and the reader is referred to (Iserles et al., 2000) for details and further references.

Given a linear matrix system $Y' = A(t)Y$, the Magnus approach is to seek a matrix function Ω such that $Y(t_n + t) = \mathrm{e}^{\Omega(t)} Y(t_n)$. It is possible to show that

$$\begin{aligned}\Omega(t) = & \int_0^t A(\tau)\mathrm{d}\tau - \tfrac{1}{2}\int_0^t\int_0^{\tau_1}[A(\tau_2), A(\tau_1)]\mathrm{d}\tau_2\mathrm{d}\tau_1 \\ & + \tfrac{1}{12}\int_0^t\int_0^{\tau_1}\int_0^{\tau_1}[A(\tau_3), [A(\tau_2), A(\tau_1)]]\mathrm{d}\tau_3\mathrm{d}\tau_2\mathrm{d}\tau_1 \\ & + \tfrac{1}{4}\int_0^t\int_0^{\tau_1}\int_0^{\tau_2}[[A(\tau_3), A(\tau_2)], A(\tau_1)]\mathrm{d}\tau_3\mathrm{d}\tau_2\mathrm{d}\tau_1 + \cdots .\end{aligned} \tag{8}$$

There are further four quadruple integrals and, in general, the number of terms grows exponentially with multiplicity of integrals. Yet, a sense can be made of out this 'integral soup' by identifying each term with a *rooted tree* and demonstrating that more complex trees can be generated from simpler ones by well-defined recursive rules (Iserles & Nørsett, 1999). For example, the expansion (8) above can be written concisely in terms of trees as



To obtain an expansion term from a tree proceed as follows: First affix the matrix function A to every leaf. Then prune the tree according to the following two rules: If a leaf is a single child, integrate, while if its parent has two children, commute the terms corresponding to their leaves.

For the sake of simplicity, we assume that the matrix A can be integrated explicitly, which is the case when its terms are, for example, products and linear combinations of polynomials, exponentials and trigonometric functions. While mentioning in passing that (8) can be discretized very effectively by quadrature, we will mostly ignore this important issue in the present paper.

Considering the application of the Magnus expansion (8) to the linear oscillator (4), it is possible to 'measure' the contribution of each tree. Since the leading term of the error in the Lie algebra $\mathfrak{g}$ (the space where Ω evolves) dominates the global error in $\boldsymbol{y}$, we thereby have enough information to use (3) and determine the rate of global-error accumulation by Magnus methods.

Theorem 3 (Iserles 2001) *Suppose that the linear oscillator (4) is solved by a Magnus method of order $p = 2s$, $s \geq 2$, with exact integration. Then, for fixed $t \gg 1$ and sufficiently small $h > 0$ it is true that*

$$\begin{aligned} \boldsymbol{e}_h(t) \approx h^{2s} \Bigg\{ & c_1 \begin{bmatrix} g^{-1/4}(t) & 0 \\ 0 & g^{1/4}(t) \end{bmatrix} \begin{bmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{bmatrix} \\ & + c_2 g'(t) \begin{bmatrix} g^{s-7/4}(t) & 0 \\ 0 & g^{s-5/4}(t) \end{bmatrix} \begin{bmatrix} -\sin\theta(t) & \cos\theta(t) \\ \cos\theta(t) & \sin\theta(t) \end{bmatrix} \Bigg\} \boldsymbol{s}. \end{aligned} \tag{9}$$

Here c_1 and c_2 are scalar constants, while $\boldsymbol{s}$ is a constant 2-vector. In the case $p = 2$ the second term drops out and the global-error estimate is just

$$\boldsymbol{e}_h(t) \approx h^2 \begin{bmatrix} g^{-1/4}(t) & 0 \\ 0 & g^{1/4}(t) \end{bmatrix} \begin{bmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{bmatrix} \boldsymbol{s}. \tag{10}$$

We note that for second-order Magnus the amplitude of global error matches exactly the amplitude of the solution itself: surely one cannot expect more than this from a numerical method! This is not the case for higher-order methods. Thus, fourth-order Magnus expansion of the Airy oscillator (6) yields

$$y^1_{t/h} - y(t) \approx c_1 h^4 t^{1/4} \chi_1(t), \qquad y^2_{t/h} - y'(t) \approx c_2 h^4 t^{3/4} \chi_2(t),$$

where c_1, c_2 are constants, while χ_1, χ_2 are fast-oscillating functions with unit amplitude. Although not perfect, this is much better than classical fourth-order methods and explains the superior behaviour of Magnus in Figure 2.

3 The modified Magnus method

Neither classical methods nor, indeed, Lie-group methods are designed to integrate highly-oscillatory systems. Thus, it is not the poor performance of the first but the remarkably good behaviour of the latter that should come as a surprise. In this section we debate a modification of the Magnus method (8), designed to integrate highly-oscillatory systems. First let us examine in Figure 3 the global error incurred by the modification of fourth-order Magnus, when applied to the Airy equation (6) with constant step of $h = \frac{1}{4}$ (that is, four times that of Figure 2).

Suppose again that we are solving a highly-oscillatory linear system $\boldsymbol{y}' = A(t)\boldsymbol{y}$. The main idea is to perform in each time step a local change of variables

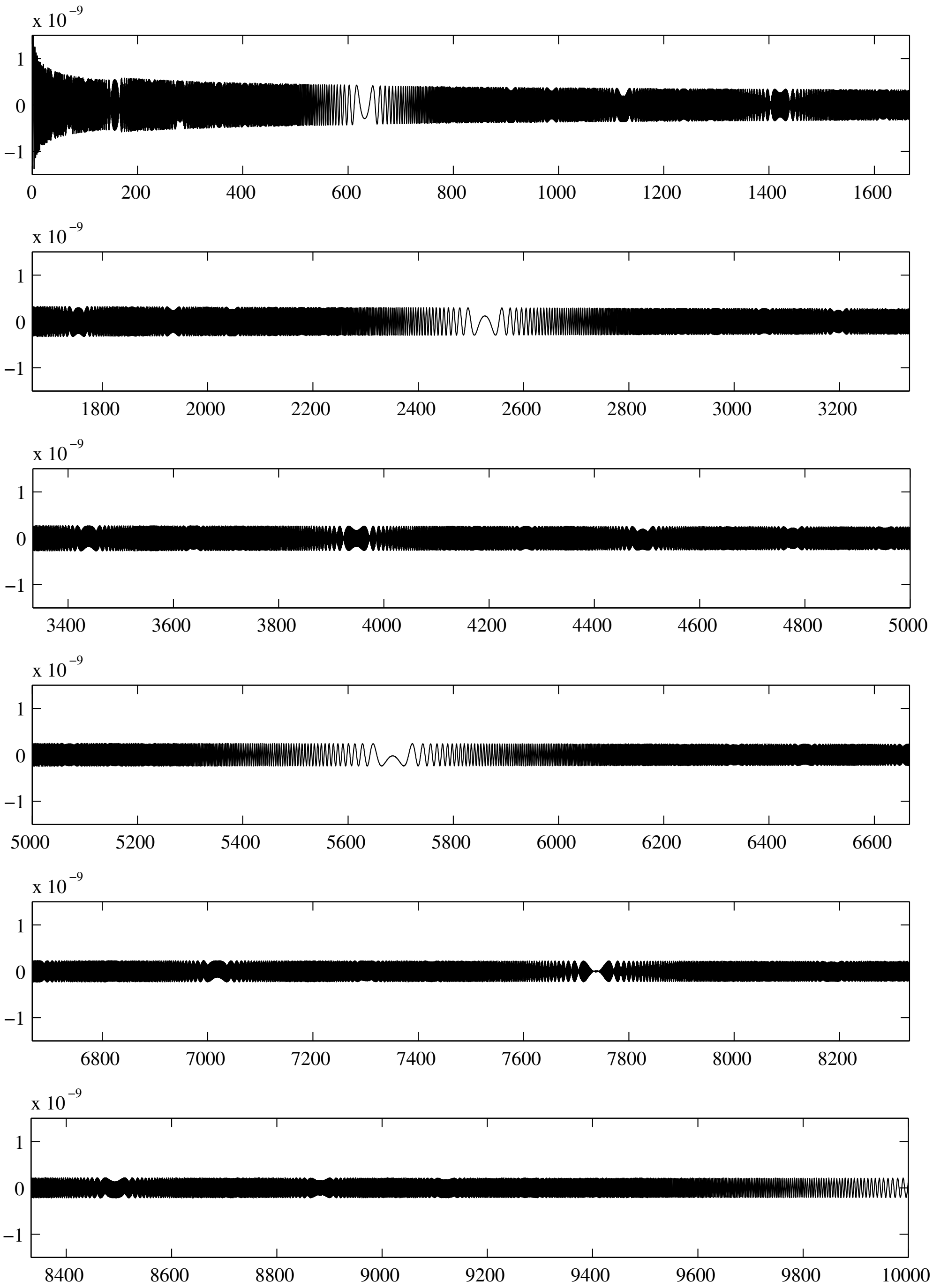


Fig. 3. Global error for the Airy equation with modified Magnus and $h = \frac{1}{4}$.

with respect to a fast-rotating frame of reference induced by a nearby constant-coefficients problem. Thus, we let

$$\boldsymbol{y}(t) = \mathrm{e}^{(t-t_n)A(t_{n+1/2})}\boldsymbol{x}(t-t_n), \qquad t \geq t_n,$$

where $t_{n+1/2} = t_n + \frac{1}{2}h$, and use a Magnus method to solve for the function $\boldsymbol{x}$.

It is easy to verify that the latter obeys the linear equation

$$\boldsymbol{x}' = \mathrm{e}^{-tA(t_{n+1/2})}[A(t_n+t) - A(t_{n+1/2})]\mathrm{e}^{tA(t_{n+1/2})}\boldsymbol{x}, \quad t \geq 0, \tag{11}$$

with the initial condition $\boldsymbol{x}(0) = \boldsymbol{y}_n$. Once (11) is solved, $\boldsymbol{x}(t) \approx \mathrm{e}^{\Omega_n(t)}\boldsymbol{y}_n$, we let

$$\boldsymbol{y}_{n+1} = \mathrm{e}^{hA(t_{n+1/2})}\mathrm{e}^{\Omega_n(t)}\boldsymbol{y}_n.$$

The linear system (11) can be often simplified. In particular, in the case of the linear oscillator (4), we have

$$\boldsymbol{x}' = [g(t_n+t) - \omega]\begin{bmatrix} \frac{1}{2}\omega^{-1}\sin 2\omega t & \omega^{-2}\sin^2\omega t \\ -\cos^2\omega t & -\frac{1}{2}\omega^{-1}\sin 2\omega t \end{bmatrix}\boldsymbol{x},$$

where $\omega = \sqrt{g(t_{n+1/2})}$. Moreover, in that case

$$\mathrm{e}^{tA(t_{n+1/2})} = \begin{bmatrix} \cos\omega t & \omega^{-1}\sin\omega t \\ -\omega\sin\omega t & \cos\omega t \end{bmatrix}.$$

The matrix in (11) is $\mathcal{O}(t)$ and this helps to boost the order of the method from $2s$ to $3s+1$. This, however, falls short of explaining the apparently uniform decay of the global error in Figure 3. Complicated algebra in (Iserles, 2001) determines the order of magnitude of Magnus-expansion terms, as applied to (11), thereby leading to a global-error estimate which should be compared with Theorems 2 and 3.

Theorem 4 (Iserles 2001) *Suppose that the linear oscillator (4) is solved by applying an order-$(2s)$ Magnus method to the equation (11). Then, for fixed $t \gg 1$ and sufficiently small $h > 0$ it is true that*

$$\begin{aligned} \boldsymbol{e}_h(t) \approx h^{3s+1} &\begin{bmatrix} g^{-1/4}(t) & 0 \\ 0 & g^{1/4}(t) \end{bmatrix}\begin{bmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{bmatrix} \\ &\times \left\{ c_1 I + \int_{t^*}^{t} \frac{g'^{s+1}(\tau)}{g^{1/2}(\tau)} \begin{bmatrix} -\sin 2\theta(\tau) & -1+\cos 2\theta(\tau) \\ 1+\cos 2\theta(\tau) & \sin 2\theta(\tau) \end{bmatrix} \mathrm{d}\theta \right\} \boldsymbol{s}, \end{aligned} \tag{12}$$

where c_1 and $t^* \gg 1$ are scalar constants and $\boldsymbol{s}$ is a scalar 2-vector.

Applying (12) to the Airy oscillator is disappointing: regardless of $s \geq 1$, we have

$$t_{t/h} - y(t) \approx ch^{3s+1} t^{1/4} \chi(t),$$

where c is a constant and $\chi \in [-1, 1]$ oscillates fast. This is *worse* than 'plain' second-order Magnus and definitely cannot account for the superior performance of the modified Magnus method in Figure 3. The reason is that *we have been measuring the wrong quantity!* To account for the superior performance of modified Magnus, we need to consider *uniform* error estimates, equally valid for all $t \in [0, \infty)$ and sufficiently small $h > 0$. In that case, it is proved in (Iserles, 2001) that the order decreases, yet the global-error estimate is more favourable.

Theorem 5 *Suppose that the linear oscillator (4) is solved by applying an order-$(2s)$ Magnus method to the equation (11). Then, for all $t \in [t^*, \infty)$ and sufficiently small $h > 0$ it is true that*

$$\begin{aligned} \boldsymbol{e}_h(t) \approx h^{3s+1} & \begin{bmatrix} g^{-1/4}(t) & 0 \\ 0 & g^{1/4}(t) \end{bmatrix} \begin{bmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{bmatrix} \Bigg\{ c_1 I \\ & + h^{-\delta(s)} \int_{t^*}^{t} \frac{g'^{s+1}(\tau)}{g^{\delta(s)/2+1}(\tau)} \begin{bmatrix} -\sin 2\theta(\tau) & -1+\cos 2\theta(\tau) \\ 1+\cos 2\theta(\tau) & \sin 2\theta(\tau) \end{bmatrix} \mathrm{d}\tau \Bigg\} \boldsymbol{s}, \end{aligned} \tag{13}$$

where c_1, t^* and $\boldsymbol{s}$ are as in Theorem 4, while $\delta(1) = 2$, $\delta(2) = \delta(3) = 4$.

Proof Straightforward generalisation of the proof for the Airy equation in (Iserles, 2001) □

Corollary 6 *Let*

$$\iota(t) := \int_{\varepsilon}^{t} \frac{|g'(\tau)|^{s+1}}{g^{\delta(s)/2+1}(\tau)} \mathrm{d}\tau,$$

where $\varepsilon \geq 0$ is arbitrary (to avoid integration from the origin since g might vanish there). If $\iota(\infty) < \infty$ then, applying the modified Magnus method to the linear oscillator (4), it is true that

$$y_{t/h} - y(t) \approx ch^{3s+1-\delta(s)} g^{-1/4}(t) \chi(t), \qquad t > 0,$$

and the global error decays at the same speed as the amplitude of the exact

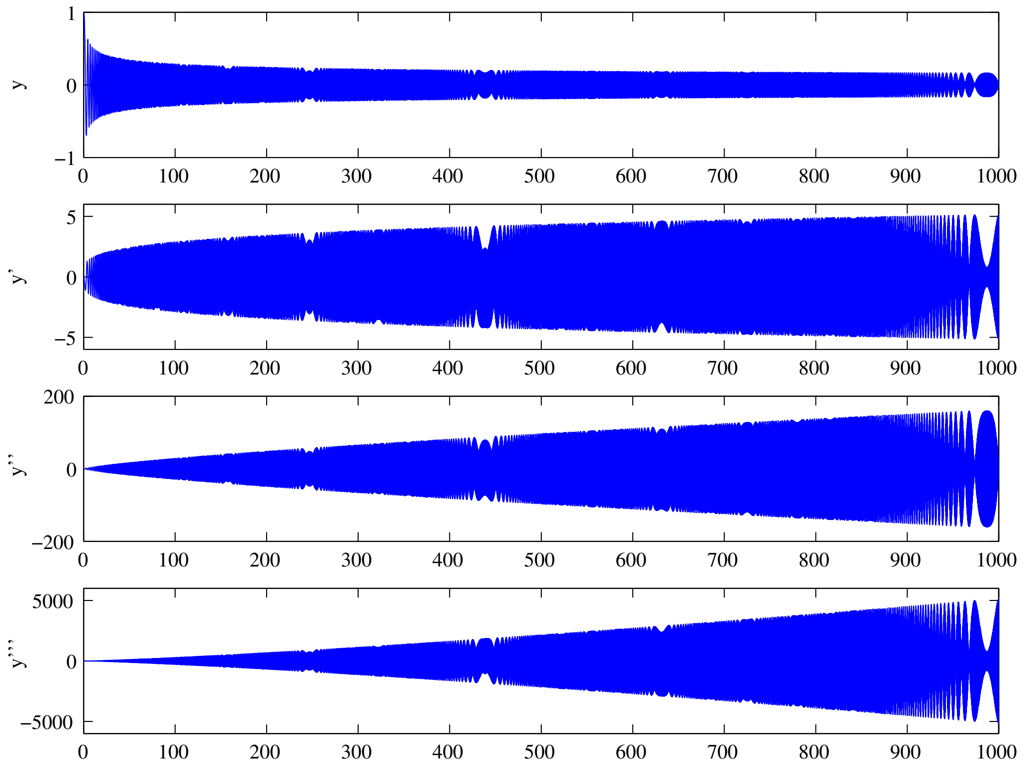


Fig. 4. The solution of the Airy oscillator (6) (with $y(0) = 1$, $y'(0) = 0$) and its derivatives.

solution. If $\iota(\infty) = +\infty$ *then*

$$|y_{t/h} - y(t)| \le ch^{3s+1-\delta(s)} g^{-1/4}(t)\iota(t), \qquad t > 0.$$

In both cases c *is a constant, while* χ *oscillates in* $[-1, 1]$.

Proof Follows immediately from Theorem 5. □

If g is an mth degree polynomial then it follows readily that $\iota(\infty) < \infty$ if and only if $m[\frac{1}{2}\delta(s) - s + 1] + s > 1$. For $s = 1, 2, 3$ this is satisfied for *all* $m \ge 1$. Likewise, if $g(t) = \mathrm{e}^{p(t)}$, where p is a polynomial, then $s = 1, 2$ yields $\iota(\infty) < \infty$ but $\iota(\infty) = \infty$ for $s = 3$. This implies that modified Magnus displays optimal propagation of global error (i.e., the latter decays at the same speed as the amplitude of the exact solution) for a wide range of functions g. Of course, although $\iota(\infty) < \infty$ is sufficient for this behaviour, it is entirely possible that this desirable feature, which has been observed in many numerical experiments, occurs for other (or even all) g which are amenable to the WKB asymptotics (5).

What is the reason for this remarkable behaviour of modified Magnus? To ex-

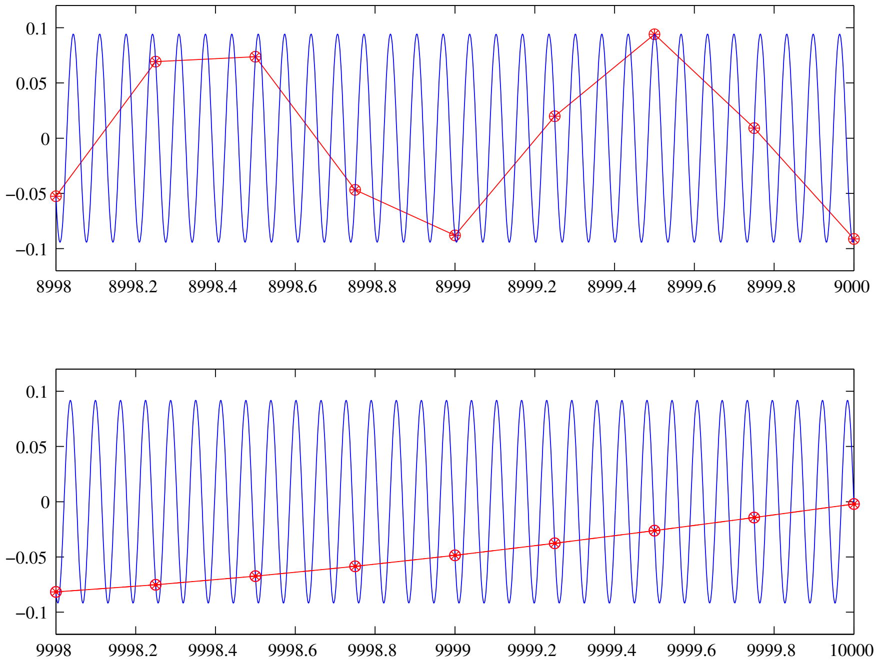


Fig. 5. The exact solution (solid oscillating line) of the Airy oscillator and its numerical solution (solid piecewise-linear line, joining solution at step points with step size $h = \frac{1}{4}$) for large values of t.

amine this phenomenon, we need to go back to basics and query why classical numerical methods do so poorly. The reason is simple: once a function oscillates, so do its derivatives, but with much larger amplitude. Figure 4 revisits the Airy equation that has been already displayed in Figure 1, but it also shows the first few derivatives.

The truncation error of multistep methods and Taylor expansions is always of the form $ch^{p+1}\boldsymbol{y}^{(p+1)}(t_n) + \mathcal{O}(h^{p+2})$, therefore its magnitude is determined by a high-order derivative. The situation is not much better for Runge–Kutta and general linear methods, since the local error can be expanded in elementary differentials. Even for the Magnus method the amplitude of local error is determined by derivatives, except that their order is lower, hence their size is smaller. The situation is radically different for modified Magnus. Recall what the Magnus method is all about: repeated commutation and integration of the vector field. Likewise, the leading term in the truncation of the infinite modified Magnus series, precisely the term that dominates the local error, can be expressed as a multivariate integral over repeated commutators. Note, however, that the vector field in (11) is itself a fast-oscillating matrix! Therefore its integration, the very opposite of differentiation, is a smoothing operator: the amplitude is decreased once the argument is integrated. Moreover, the higher the oscillation, the faster this decay in local error! Like in Oriental martial arts, modified Magnus turns the strength of its opponent into its own.

Figure 5 zooms on the solution of the Airy equation (6) with modified fourth-order Magnus method for large time values. On the face of it, it defies the accepted wisdom, namely that no accuracy can be expected unless 5–8 steps are taken in each period. But this is precisely the point about modified Magnus: even with varying g, *locally* the equation is a moderate perturbation of the harmonic oscillator $y'' + g(t_{n+1/2})y = 0$. (Indeed, it is difficult to notice from the figure that the exact solution is not periodic.) Changing locally the frame of reference allows this feature to precondition the numerical solution, reduce the error and reform altogether the qualitative behaviour of the solution.

4 Further discussion

The analysis of this paper is predicated upon the assumption that the step size h is constant. It is not difficult to generalise Theorem 1 to variable steps, pro-

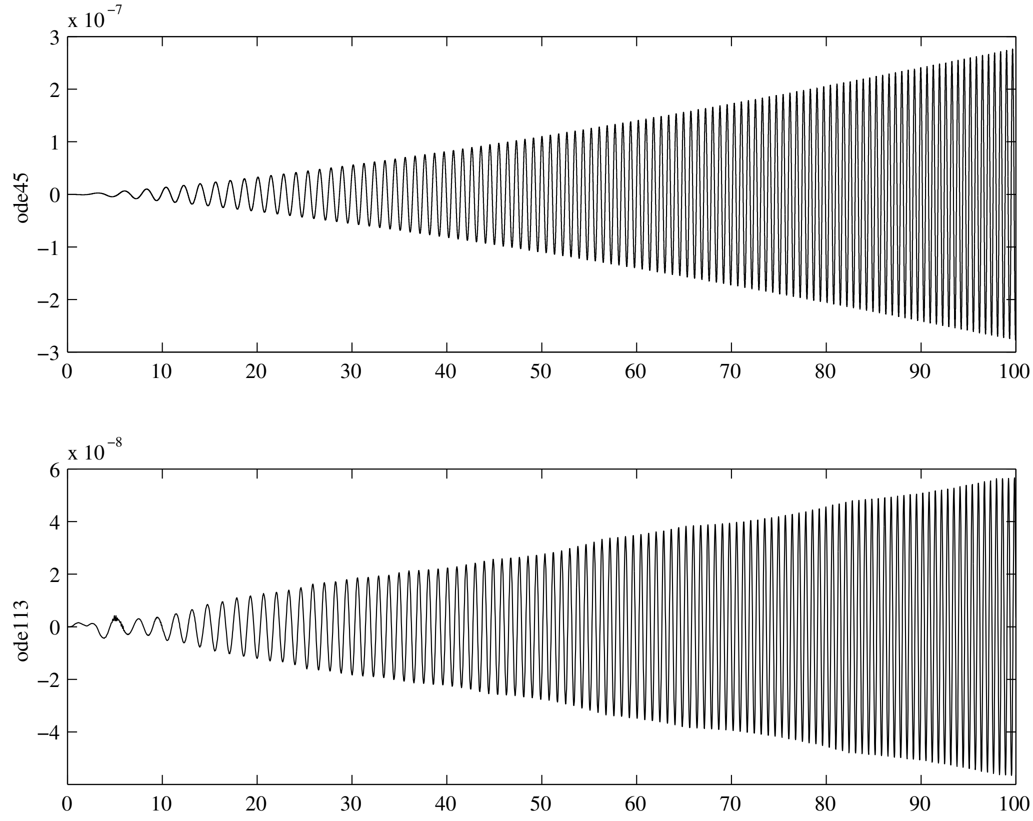


Fig. 6. Global error committed by the MATLAB routines `ode45` and `ode113`, solving the Airy oscillator (6) with $y(0) = 1$, $y'(0) = 0$, `AbsTol=[1e-9,1e-9]` and `RelTol=1e-9`.

vided that they are known *a priori.* This, of course, is not very realistic in the design of competitive numerical software, which amends the step size *a posteriori,* in response to indications from local error-control mechanisms. Having said this, the consideration and optimization of *a priori* step size sequences, like the analysis of the constant-step case, has much virtue in understanding how numerical methods behave and in motivating the development of new methods.

Having said this, the received wisdom is that numerical time-stepping methods should always employ error control and variable step size. (Symplectic solution of Hamiltonian problems is an important exception.) It is thus legitimate to query how standard (variable step) ODE software is likely to cope with the Airy oscillator (6). Can the pessimistic results pertaining to Runge–Kutta methods, say, be reversed by employing error control?

Figure 6 displays the global error in the solution of the Airy equation with standard MATLAB routines, `ode45` (embedded Runge–Kutta–Fehlberg) and `ode113` (variable-order Adams–Bashforth/Adams–Moulton). Although the requested (local) error tolerance, both absolute and relative, was 10^{-9}, global error accumulates much faster and at $t = 100$ exceeds the tolerance by roughly two orders of magnitude. What is not clear from the figure is the number of time steps: 233037 and 4420, respectively. In particular, `ode45` required in excess of two million function evaluations to produce seven significant digits in $[0, 100]$ while, for the record, modified Magnus took just 40000 function evaluations to deliver nine digits of accuracy in $[0, 10000]$ (cf. Figure 3). Given that the rate of oscillation of the exact solution steadily increases, the number of function evaluations with MATLAB routines is likely to be larger by many orders of magnitude in $[0, 10000]$: daring souls with fast computers and unlimited patience are welcome to this computation.

But can the poor behaviour of `ode45` and `ode113` be explained as a consequence of using 'nonstiff' software for an equation which might be described (for large t) as 'stiff'? To explore further this point, Figure 7 displays global error for the stiff MATLAB solvers `ode115` (quasi-constant step backward-differentiation Klopfenstein–Shampine schemes) and `ode23s` (a Rosenbrock 2/3 pair). The results, if at all, are substantially worse than for nonstiff solvers and the global error overshoots the tolerance by even larger margin. The number of time steps is 17305 and 290511, respectively. To illustrate the dependence on the length of the interval, it is enough to report that, computing in $[0, 400]$ under identical conditions with the 'better' method, `ode115s`, yields error of $\approx 2 \times 10^{-5}$ after 137843 time steps.

The second obvious query is to which extent do we need to employ the Magnus method in tandem with the local change of variables that is the starting point for the algorithm from Section 3. Can we replace it with a Runge–Kutta

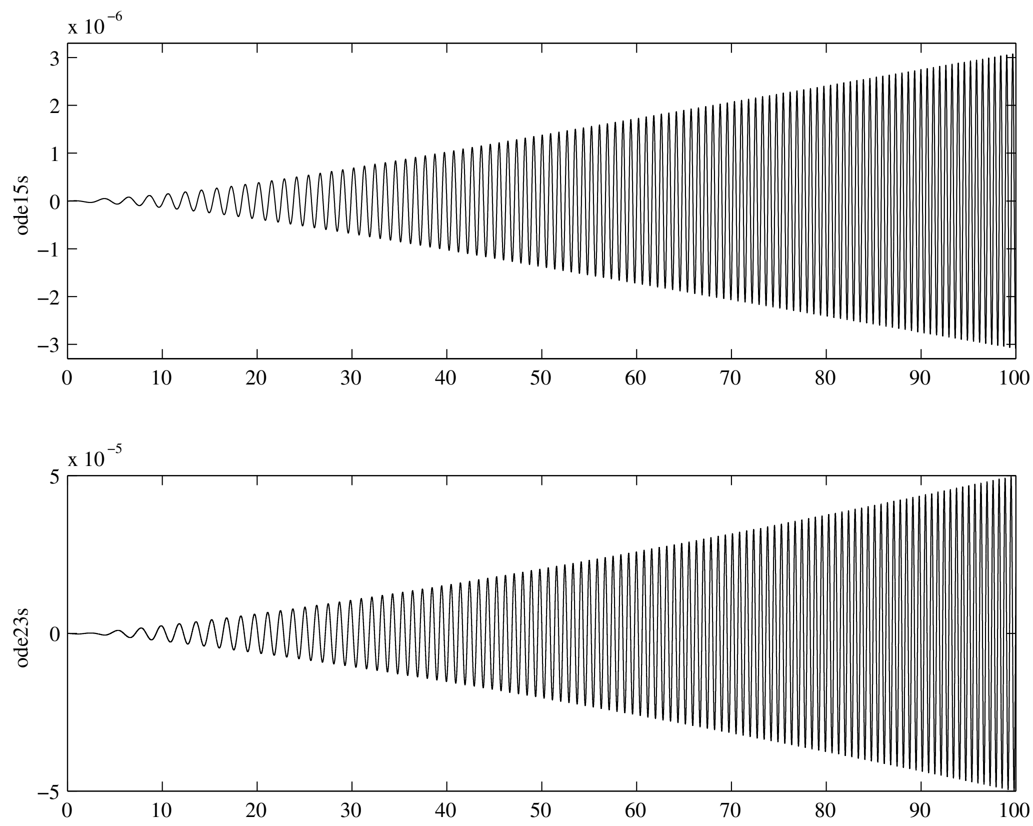


Fig. 7. Global error committed by the MATLAB routines `ode15s` and `ode23s`, solving the Airy oscillator (6) with $y(0) = 1$, $y'(0) = 0$, `AbsTol=[1e-9,1e-9]` and `RelTol=1e-9`.

scheme, say, or with different Lie-group method? Although of little importance for the linear oscillator (4), the implementation of the Magnus method, entailing the evaluation of commutators and exponentials, becomes more expansive in higher dimensions and significantly more complicated for nonlinear equations (Iserles et al., 2000). The discussion toward the end of Section 3 would make it clear that, unfortunately, once Magnus is replaced by another scheme, most of the improvement reported in our paper is likely to be lost. This is also the case when the exponential change of variables is replaced by a Cayley transformation, although the latter is significantly cheaper to evaluate (Diele, 2001).

This is perhaps the place to raise another interesting point. Local preconditioning with respect to the solution of a 'locally frozen' system $\boldsymbol{y}' = A(t_{n+1/2})\boldsymbol{y}$ is nothing new in asymptotic theory (Hinch, 1991) and has been occasionally used in numerical analysis, e.g. by Lawson (1967) and Scheid (1983). The main difference with our approach is that the conventional preconditioning is *additive,* namely

$$\boldsymbol{y}(t) = \mathrm{e}^{(t-t_n)A(t_{n+1/2})}\boldsymbol{y}_n + \boldsymbol{x}(t),$$

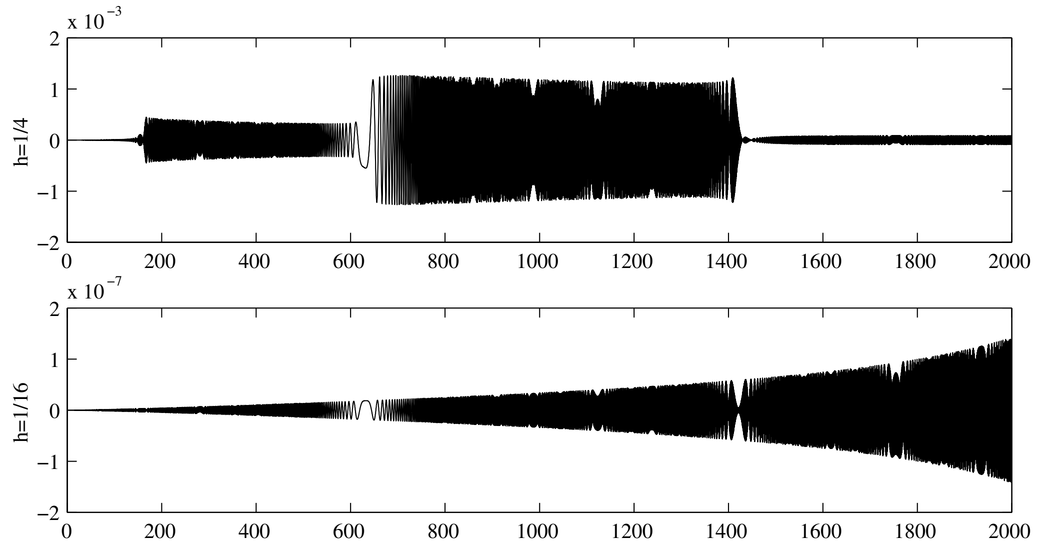


Fig. 8. Global error in the solution of the Airy oscillator (6) by modified Magnus with sixth-order quadrature.

whereas in our case the preconditioning is *multiplicative,*

$$\boldsymbol{y}(t) = \mathrm{e}^{(t-t_n)A(t_{n+1/2})}\boldsymbol{x}(t).$$

Next point to highlight is the dependence of our implementation of the modified Magnus method on exact calculation of integrals. In Figure 8 we report the global error in the solution of the Airy equation by modified Magnus, except that we have used the three-point Gaussian quadrature in the discretization of integrals in the Magnus algorithm (c.f. (Iserles et al., 2000) for exact formulae) for two choices of constant step size h. Comparison with Figure 3 shows that six significant figures are lost for $h = \frac{1}{4}$, in comparison with exact integration! Yet, comparing with the 'plain' Magnus method in Figure 2, we can see that even this, suboptimal implementation of modified Magnus (with $h = \frac{1}{16}$) delivers extra four significant digits.

The performance of modified Magnus with discretized integrals might improve a great deal once Gaussian quadrature is replaced by one of the special quadrature methods for highly-oscillating integrands, e.g. the *Filon rule.* This is an issue for future exploration.

Finally, let us mention the matter of *nonlinear* highly-oscillatory equations. Using the global-error formula (2) becomes considerably more complicated, c.f. the analysis of the Lane–Emden equation by Niesen (2001), both because exact and variational flows and their asymptotics are considerably more obscure and since higher-order terms must be sometimes taken into consideration. Yet, research into well-defined, yet useful classes of nonlinear oscillators is certainly possible. Different issue is a generalisation of modified Magnus to nonlinear setting. It is possible to extend ‘plain’ Magnus to nonlinear Lie-group equations (Iserles et al., 2000), and this can be combined with rotating frame of reference (Iserles, 2001). Another possibility, which has not been yet explored in any great detail, is to combine the modification with explicit Runge–Kutta–Munthe-Kaas schemes (Iserles et al., 2000), arguably the current method of choice for nonlinear Lie-group equations.

References

B. Cano and J.M. Sanz-Serna, Error growth in the numerical integration of periodic orbits, with application to Hamiltonian and reversible systems, *SIAM J. Num. Anal.* **34** (1997), 1391–1417.

F. Diele, personal communication (2001).

P. Henrici, *Error Propagation for Difference Methods,* Wiley (1963), New York.

E.J. Hinch, *Perturbation Methods,* Cambridge University Press (1991), Cambridge.

M. Hochbruck and C. Lubich, Exponential integrators for quantum-classical molecular dynamics, *BIT* **39** (1999), 620–645.

A. Iserles, On the global error of discretization methods for highly-oscillatory ordinary differential equations, to appear in *BIT* (2001).

A. Iserles, H.Z. Munthe-Kaas, S.P. Nørsett and A. Zanna, Lie-group methods, *Acta Numerica* **9** (2000), 215–365.

A. Iserles and S.P. Nørsett, On the solution of linear differential equations in Lie groups, *Philosophical Trans. Royal Soc. A* **357** (1999), 983–1019.

J.D. Lawson, “Generalized Runge–Kutta processes for stable systems with large Lipschitz constant”, *SIAM J. Num. Anal.* **4** (1967), 372–380.

W. Magnus, On the exponential solution of differential equations for a linear operator, *Comm. Pure and Appl. Math* **7** (1954), 649–673.

J. Niesen, “A priori estimates for the global error committed by Runge–Kutta methods for a nonlinear oscillator”, (2001), to appear.

L. Petzold, L.O. Jay and Jeng Yen, Numerical solution of highly oscillatory ordinary differential equations, *Acta Numerica* **9** (1997), 437–483.

R.E. Scheid, The accurate numerical solution of highly oscillatory ordinary differential equations, *Maths Comput.* **41** (1983), 487–509.

A. Ziv and V. Amdursky, On the numerical solution of stiff linear systems of the oscillatory type, *SIAM J. Appl. Maths* **33** (1977), 593–606.