On the global error of discretization methods for highly-oscillatory ordinary differential equations

Arieh Iserles [1]

Abstract

Commencing from a global-error formula, originally due to Henrici, we investigate the accumulation of global error in the numerical solution of linear highly-oscillating systems of the form $y'' + g(t)y = 0$, where $g(t) \stackrel{t\to\infty}{\longrightarrow} \infty$. Using WKB analysis we derive an explicit form of the global-error envelope for Runge–Kutta and Magnus methods. Our results are closely matched by numerical experiments.

Motivated by the superior performance of Lie-group methods, we present a modification of the Magnus expansion which displays even better long-term behaviour in the presence of oscillations.

1 Introduction

The starting point to the work of the paper is the observation that methods based upon Magnus expansions display truly remarkable performance when applied to certain highly-oscillating ordinary differential systems: Airy (Iserles & Nørsett 1999), Mathieu (Iserles, Marthinsen & Nørsett 1999) and Bessel-like (Iserles, Munthe-Kaas, Nørsett & Zanna 2000) equations. One may expect from general considerations that methods reproducing Lie-group structure (or, more specifically, SL(2) structure) are likely to yield more precise solution for the linear oscillator $y'' + g(t)y = 0$, say, than classical Runge–Kutta methods. Yet, the degree of improvement in (Iserles & Nørsett 1999) and elsewhere can hardly be explained by Lie-group considerations alone. A plausible suspicion is that methods based upon Magnus expansions are intrinsically better for highly-oscillating systems for reasons that have little to do with the maintenance of Lie-group structure. The goal of this paper is to pursue this train of thought and analyse the different rate of global-error accumulation by classical and Magnus methods.

We commence from an integral formula for the leading term in a global-error expansion of an arbitrary time-stepping method. Although in the form displayed in Section 2 this formula is apparently new, as is its (surprisingly elementary) proof, it features repeatedly, often in a subtly different form, in modern theory of numerical ordinary differential equations. The earliest reference is probably Henrici (1963), who derived a similar expression for multistep methods. This was extended by Hairer & Lubich (1984) to a more general mechanism to expand the global error of a strictly-stable multistep method. More recently, Viswanath (2000) used a similar formula to investigate global error. Moreover, the formula is reminiscent of the classical Alekseev–Gröbner lemma (Hairer, Nørsett & Wanner 1987).

A remarkable feature of the integral formula is that, once the asymptotic behaviour of a differential system and its variational equation is known, it is possible to determine

[1]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, England, email A.Iserles@damtp.cam.ac.uk.

explicitly the asymptotic form of leading global error term for most numerical solvers of interest. In Sections 3 and 4 we use WKB analysis to analyse the performance of Runge–Kutta and Magnus methods for the linear oscillator $y'' + g(t)y = 0$, where $\lim_{t\to\infty} g(t) = +\infty$, while the derivatives $|g^{(l)}(t)|$ are small for $t \to \infty$, $l \in \mathbb{N}$. This analysis puts on a firm mathematical basis the observation that commenced this paper: Magnus methods indeed outperform classical solvers insofar as highly-oscillating systems are concerned and the comparison can be quantified very precisely indeed.

Once the mechanism that explains the superior performance of Magnus has been understood, there opens up an opportunity to make such solvers even better. Thus, in Section 5 we propose a modification of the Magnus method, based on local linearisation, which leads to a method that, in the case of the linear oscillator $y'' + g(t)y = 0$, displays global error attenuation identical to the attenuation of the exact solution, regardless of the frequency of oscillation. When $g(t) \to +\infty$ the frequency progressively grows and, assuming constant step size, so does the number of oscillations per step. Yet, the envelope of the global error goes on decaying!

The paper is accompanied by a number of numerical examples that illustrate our results.

2 An integral expression for the global error

We consider the initial-value ordinary differential system

$$\boldsymbol{y}' = \boldsymbol{f}(t, \boldsymbol{y}), \quad t \geq 0, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0, \tag{2.1}$$

where $\boldsymbol{y}_0 \in \mathbb{R}^m$, while $\boldsymbol{f} : \mathbb{R}_+ \times \mathbb{R}^m \to \mathbb{R}^m$ is a C^1 function. The *flow* of (2.2) will be denoted by Φ_{t_1,t_2}, thus

$$\Phi_{t_1,t_2}(\boldsymbol{y}(t_1)) = \boldsymbol{y}(t_2), \qquad t_1, t_2 \geq 0. \tag{2.2}$$

Alongside (2.1) we consider its *variational equation*

$$\Psi' = J(t)\Psi, \quad t \geq 0, \qquad \Psi(0) = I, \tag{2.3}$$

where

$$J(t) = \frac{\partial \boldsymbol{f}(t, \boldsymbol{y}(t))}{\partial \boldsymbol{y}}, \qquad t \geq 0.$$

The equation (2.1) is being solved by an r-step convergent pth-order numerical method

$$\boldsymbol{y}_{n+1} = R_{n,h}(\boldsymbol{y}_{n-r+1}, \boldsymbol{y}_{n-r+2}, \ldots, \boldsymbol{y}_n), \qquad n \in \mathbb{N}, \tag{2.4}$$

where the starting values $\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{r-1}$ are either exact or provided to an order exceeding p. We assume that (2.4) is applied with a constant step size $h > 0$, therefore $\boldsymbol{y}_n \approx \boldsymbol{y}(t_n)$, where $t_n = nh$, $n \in \mathbb{Z}_+$. The *local error* of (2.4) at t_{n+1} is denoted by

$$h^{p+1}\boldsymbol{\ell}(\boldsymbol{y}_n) = \boldsymbol{y}_{n+1} - \Phi_{t_n,t_{n+1}}(\boldsymbol{y}_n), \qquad n \in \mathbb{N},$$

while its *global error* is

$$\boldsymbol{e}_{n+1} = \boldsymbol{y}_{n+1} - \boldsymbol{y}(t_{n+1}).$$

We commence our global-error analysis with the observation that

$$\begin{aligned}\boldsymbol{e}_{n+1} &= R_{n,h}(\boldsymbol{y}_{n-r+1}, \boldsymbol{y}_{n-r+2}, \ldots, \boldsymbol{y}_n) - \Phi_{t_n,t_{n+1}}(\boldsymbol{y}(t_n)) \\ &= h^{p+1}\boldsymbol{\ell}(\boldsymbol{y}_n) + \Phi_{t_n,t_{n+1}}(\boldsymbol{y}_n) - \Phi_{t_n,t_{n+1}}(\boldsymbol{y}(t_n)) + \mathcal{O}\big(h^{p+2}\big)\,.\end{aligned}$$

The variational equation, as is well known, provides the leading term in perturbation analysis (Hinch 1991): for sufficiently small $\|\boldsymbol{\delta}\|$ it is true that

$$\Phi_{t_n,t_{n+1}}(\boldsymbol{y}(t_n) + \boldsymbol{\delta}) - \Phi_{t_n,t_{n+1}}(\boldsymbol{y}(t_n)) = \Psi(t_{n+1})\Psi^{-1}(t_n)\boldsymbol{\delta} + \mathcal{O}\big(\|\boldsymbol{\delta}\|^2\big)\,.$$

In particular, letting $\boldsymbol{\delta} = \boldsymbol{e}_n$, we have $\boldsymbol{y}(t_n) + \boldsymbol{\delta} = \boldsymbol{y}_n$ and

$$\boldsymbol{e}_{n+1} = h^{p+1}\boldsymbol{\ell}(\boldsymbol{y}(t_n)) + \Psi(t_{n+1})\Psi^{-1}(t_n)\boldsymbol{e}_n + \mathcal{O}\big(h^{p+2}\big)\,. \tag{2.5}$$

Note that we have replaced $\boldsymbol{y}_n$ with $\boldsymbol{y}(t_n)$ in the first term on the right: the difference is subsumed in the $\mathcal{O}\big(h^{p+2}\big)$ term.

Let $T = nh$, We denote by $\boldsymbol{E}(T)$ the global error at T. Although, clearly, $\boldsymbol{E}(T) = \boldsymbol{e}_n$, this notation helps to focus attention on the behaviour of the global error as a function of $T \gg 0$.

Theorem 1 *Let $h > 0$ be sufficiently small. Then*

$$\boldsymbol{E}(T) = h^p\Psi(T)\int_0^T \Psi^{-1}(t)\boldsymbol{\ell}(\Phi_{0,t}(\boldsymbol{y}_0))\mathrm{d}t + \mathcal{O}\big(h^{p+1}\big)\,. \tag{2.6}$$

Proof We commence by observing that, by easy induction, (2.5) implies

$$\boldsymbol{e}_n = h^{p+1}\Psi(t_n)\sum_{k=0}^{n-1}\Psi^{-1}(t_k)\boldsymbol{\ell}(\boldsymbol{y}(t_k)) + \mathcal{O}\big(h^{p+2}\big)\,.$$

Letting $T = nh$ and approximating the Riemann sum with an integral yields the formula (2.6). □

Note that (2.6) can be written in a convolution form,

$$\boldsymbol{E}(T) = h^p\int_0^T \Psi(T-t)\boldsymbol{\ell}(\Phi_{0,t}(\boldsymbol{y}_0))\mathrm{d}t + \mathcal{O}\big(h^{p+1}\big)\,.$$

This, however, obscures what we regard as the most salient observation about this formula: *provided that the asymptotic behaviour of (2.1) and (2.3) is known, we can work out with moderate effort the asymptotic form of (2.6).* Thus, suppose that $T > 0$ is large and choose $0 < T^* < T$ so that both $\Phi_{0,t}$ and $\Psi(t)$ can be approximated well from their asymptotics for $t \ge T^*$. Setting

$$c^* = \int_0^{T^*}\Psi^{-1}(t)\boldsymbol{\ell}(\Phi_{0,t}(\boldsymbol{y}_0))\mathrm{d}t,$$

we deduce at once from (2.6) that

$$\boldsymbol{E}(T) = h^p\Psi(T)\left[c^* + \int_{T^*}^T \Psi^{-1}(t)\boldsymbol{\ell}(\Phi_{0,t}(\boldsymbol{y}_0))\mathrm{d}t\right] + \mathcal{O}\big(h^{p+1}\big)\,. \tag{2.7}$$

Hence, assuming that the explicit form of $\boldsymbol{\ell}$ is known, it is possible to work out the asymptotics of $\boldsymbol{E}(T)$ from asymptotic information for the exact solution and the variational equation. Indeed, it is (2.7), rather than (2.6), which is of a form suitable for the analysis of Sections 3–5.

It is important to emphasise that the scope of Theorem 1 is not unlimited. Given $T > T^*$ and $0 < h \le h^* = h^*(T)$, the formula (2.6) represents the global error well. However, the sensitivity of h^* as a function of T and the influence of higher-order terms are not entirely clear.

An interesting consequence of (2.6), consistent with the results in (Hairer & Lubich 1984) and (Henrici 1963), is that the accumulation of global error is driven by the size of the variational solution (2.3), rather than by the differential flow (2.2). Needless to say, the two can be widely different. For example, for the Riccati equation

$$y' + y^2 = 0, \quad t \ge 0, \qquad y(0) = \alpha, \tag{2.8}$$

the exact solution is $y(t) = \alpha/(1+\alpha t)$. Therefore, $J(t) = -2\alpha/(1+\alpha t)$ and (2.3) yields

$$\Psi(t) = \frac{1}{(1+\alpha t)^2}.$$

Thus, $\Phi_{0,t} = \mathcal{O}(t^{-1})$, while $\Psi(t) = \mathcal{O}(t^{-2})$.

To persuade ourselves in the validity of (2.6) by means of a simple example, let us suppose that (2.8) is solved by the first-order *Euler method*

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h\boldsymbol{f}(t_n, \boldsymbol{y}_n).$$

Therefore $\ell(y_n) = -\frac{1}{2}y''(t_n) + \mathcal{O}(h) = \alpha^3/(1+\alpha t_n)^3 + \mathcal{O}(h)$ and (2.6) yields the estimate

$$E(h) = -\frac{h\alpha^3}{(1+hT)^2}\int_0^T \frac{\mathrm{d}t}{1+\alpha t} + \mathcal{O}(h^2) = -\frac{h\alpha^2}{1+\alpha T}\log(1+\alpha T) + \mathcal{O}(h^2). \tag{2.9}$$

Figure 1 displays at the top the exact solution of (2.8) with $\alpha = \frac{1}{2}$. The middle plot is the global error $e_n = y_n - y(t_n)$ for $h = \frac{1}{10}$ and $n = 0, 1, \ldots, 10^4$, while the bottom plot depicts the leading (i.e. $\mathcal{O}(h)$) term in the estimate (2.9). The fit between the exact global error and the estimate is persuasive enough and, as expected, it is becoming better as T increases and asymptotic behaviour takes over.

The distinction between the exact and the variational equation disappears once we restrict the attention to linear equations: letting $\boldsymbol{f}(t, \boldsymbol{y}) = A(t)\boldsymbol{y}$, we have $J \equiv A$ and $\boldsymbol{y}(t) = \Phi_{0,t}(\boldsymbol{y}_0) = \Psi(t)\boldsymbol{y}_0$. Therefore (2.7) becomes

$$\boldsymbol{E}(T) = h^p\Psi(T)\left[c^* + \int_{T^*}^T \Psi^{-1}(t)\boldsymbol{\ell}(\Psi(t)\boldsymbol{y}_0)\mathrm{d}t\right] + \mathcal{O}(h^{p+1}). \tag{2.10}$$

This form will be used repeatedly in Sections 3–5.

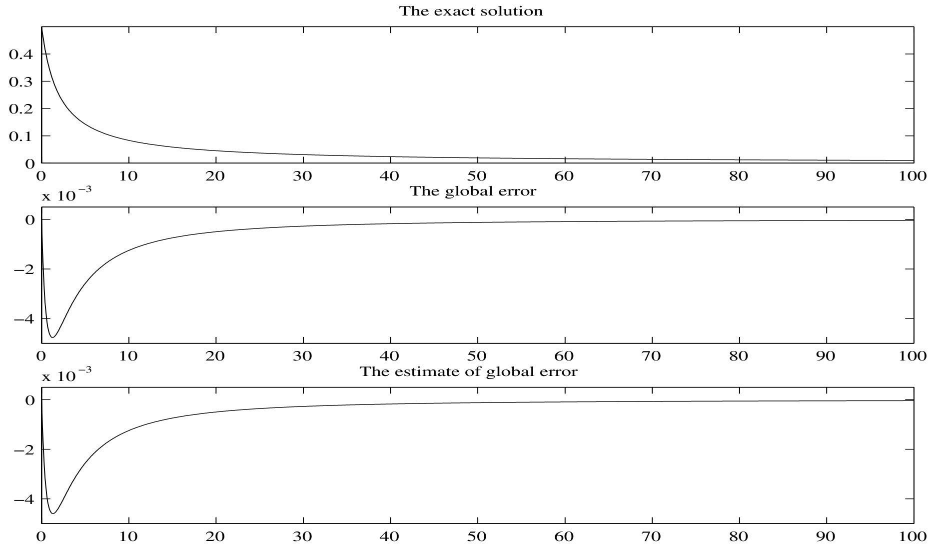


Figure 1: The solution of the Riccati equation (2.8) with Euler's method.

3 Linear oscillators I: Runge–Kutta methods

In this section we commence detailed analysis of the global error in the discretization of the linear oscillator

$$y'' + g(t)y = 0, \quad t \geq 0, \qquad y(0) = y_0,\ y'(0) = y_0', \tag{3.1}$$

where $y_0^2 + {y_0'}^2 > 0$ and $g \in \mathrm{C}^1[\mathbb{R}_+]$ obeys the following conditions: $g(t) > 0$, $t \gg 1$, $\limsup_{t\to\infty} g(t) = +\infty$ and all the derivatives of g are comparatively small, $0 \leq |g^{(l)}(t)|/g(t)^{1/l} \ll 1$, $l \in \mathbb{N}$, $t \gg 1$. The exact solution of (3.1) is a highly-oscillating function whose asymptotic behaviour can be easily determined by the classical *Wentzel–Kramers–Brillouin (WKB)* method (Hinch 1991):

$$\begin{bmatrix} y(t) \\ y'(t) \end{bmatrix} \approx \begin{bmatrix} [g(t)]^{-1/4} & 0 \\ 0 & [g(t)]^{1/4} \end{bmatrix} G(t) S_0, \qquad t \gg 1, \tag{3.2}$$

where

$$G(t) = \begin{bmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{bmatrix}, \qquad \theta(t) = \int_0^t [g(x)]^{1/2} \mathrm{d}x,$$

while S_0 is a 2×2 constant matrix. Thus, for large t the solution of $y(t)$ is (to a very high degree of approximation) a fast-oscillating function whose amplitude is $[g(t)]^{-1/4}$. Higher-order terms in the WKB expansion can be derived by classical techniques but they contribute little to present analysis.

We convert (3.1) to the vector equation

$$\boldsymbol{y}' = \begin{bmatrix} 0 & 1 \\ -g(t) & 0 \end{bmatrix} \boldsymbol{y}, \quad t \geq 0, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0, \tag{3.3}$$

and solve it with a p-order *Runge–Kutta method* (Hairer et al. 1987, Iserles 1996). Without loss of generality, we assume that p is even: the case of odd p lends itself to similar treatment. As is well known, the local error $\boldsymbol{\ell}(\boldsymbol{y}(t_n))$ is a linear combination, specific to the method in question, of *elementary differentials* of order $p+1$ (Hairer et al. 1987). Elementary differentials themselves are multilinear forms in derivatives of $\boldsymbol{y}$ at t_n. Fortunately, in the case of linear equations

$$\boldsymbol{y}' = A(t)\boldsymbol{y}, \qquad A(t) = \begin{bmatrix} 0 & 1 \\ -g(t) & 0 \end{bmatrix}, \tag{3.4}$$

it is possible to identify terms that dominate the aforementioned linear combination.

Given an autonomous ordinary differential system

$$\boldsymbol{y}' = \boldsymbol{f}(\boldsymbol{y}), \quad t \geq 0, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0 \in \mathbb{R}^m, \tag{3.5}$$

where $\boldsymbol{f}$ is C^∞, we denote the components of $\boldsymbol{y}$ and $\boldsymbol{f}$ by $y^1, \ldots, y^m$ and $f^1, \ldots, f^m$ respectively and set

$$f^i_{k_1,k_2,\ldots,k_r} = \frac{\partial f^i(\boldsymbol{y})}{\partial y^{k_1} \partial y^{k_2} \cdots \partial y^{k_r}}.$$

Each ith component of an elementary differential is a sum over terms of the form f^j and $f^j_{k_1,k_2,\ldots,k_r}$ for different values of j and $\boldsymbol{k}$ (Hairer et al. 1987). In the case of the *nonautonomuos* linear system (3.4) we need to add a third, dummy variable to render it into the autonomous form (3.5). With greater generality, if $A(t)$ is $(m-1)\times(m-1)$, we let $y_m = t$, whence

$$f^i(\boldsymbol{y}) = \begin{cases} \sum_{j=1}^{m-1} A_{i,j}(y^m)y^j, & i = 1,2,\ldots,m-1, \\ y^m, & i = m. \end{cases}$$

Therefore

$$f^i_k = \begin{cases} A_{i,k}, & i,k = 1,2,\ldots,m-1, \\ \sum_{j=1}^{m-1} A'_{i,j}y^j, & i = 1,2,\ldots,m-1, \ k = m, \\ 0, & i = m, \ k = 1,2,\ldots,m-1, \\ 1, & i = k = m. \end{cases}$$

The pattern soon emerges. Thus, $f^i_{k_1,k_2} = 0$ for $k_1,k_2 = 1,2,\ldots,m-1$ or $i = m$,

$$f^i_{k_1,m} = f^i_{m+1,k_1} = A'_{i,k_1}, \qquad f^i_{m,m} = \sum_{j=1}^{m-1} A''_{i,j}y^j,$$

while for third derivatives the only nonvanishing terms are

$$f^i_{k,m,m} = f^i_{m,k,m} = f^i_{m,m,k} = A''_{i,k} \qquad \text{and} \qquad f^i_{m,m,m} = \sum_{i=1}^{m-1} A'''_{i,j}y^j.$$

Specialising to $\boldsymbol{f} = [y^2, -g(y^3)y^1, 1]^{\mathrm{T}}$, the only nonzero terms are

$$\begin{aligned}
&f^1_2 = 1, \quad f^2_1 = -g, \quad f^2_3 = -g'y^1, \quad f^3_3 = 1,\\
&f^2_{1,3} = f^2_{3,1} = -g', \quad f^2_{3,3} = -g''y^1,\\
&f^2_{1,3,3} = f^2_{3,1,3} = f^2_{3,3,1} = -g'', \quad f^2_{3,3,3} = -g'''y^1\\
&f^2_{1,3,3,3} = f^2_{3,1,3,3} = f^2_{3,3,1,3} = f^2_{3,3,3,1} = -g''', \quad f^2_{3,3,3,3} = -g^{(iv)}y^1
\end{aligned}$$

and so on. The general rule is obvious.

In general, the ith component of an order-r elementary differential is of the form

$$\sum_{j_1,j_2,\ldots,j_{r-1}} f^i_{\boldsymbol{k}_1} f^{j_1}_{\boldsymbol{k}_2} \cdots f^{j_{r-2}}_{\boldsymbol{k}_{r-1}} f^{j_{r-1}} \tag{3.6}$$

(Hairer et al. 1987). Here

$$\bigcup_{l=1}^{r-1} \boldsymbol{k}_l = \boldsymbol{j}$$

and each of the $\boldsymbol{k}_l$s can be empty: we identify $f^i_\emptyset$ with f^i. For example, the components of elementary differentials for $r = 4$ are

$$\sum_{j_1,j_2,j_3} f^i_{j_1,j_2,j_3} f^{j_1} f^{j_2} f^{j_3}, \; \sum_{j_1,j_2,j_3} f^i_{j_1,j_3} f^{j_1}_{j_2} f^{j_2} f^{j_3}, \; \sum_{j_1,j_2,j_3} f^i_{j_1} f^{j_1}_{j_2,j_3} f^{j_2} f^{j_3}, \; \sum_{j_1,j_2,j_3} f^i_{j_1} f^{j_1}_{j_2} f^{j_2}_{j_3} f^{j_3}.$$

Recall that g is large, while its derivatives are small. Letting $r = p+1 = 2s+1$ in (3.6), it is easy to see that the size of the expression is maximised as follows: For $i = 1$ we choose $j_{2l-1} = 2$, $j_{2l} = 1$, $l = 1, 2, \ldots, s$ and $k_l = j_l$, resulting in

$$f^1_2 f^2_1 f^1_2 \cdots f^2_1 f^1 = (f^1_2 f^2_1)^s f_1 = (-g)^s y^2.$$

For $i = 2$ we let $j_{2l-1} = 1$, $j_{2l} = 2$, $k_l = j_l$, whereby

$$f^2_1 f^1_2 f^2_1 \cdots f^1_2 f^2 = (f^2_1 f^1_2)^s f^2 = -(-g)^{s+1} y^1.$$

Finally, for $i = 3$ the only nonzero component is $j_l = k_l = 3$ and it equals one. We deduce that

$$\boldsymbol{\ell}(\boldsymbol{y}) \approx c \begin{bmatrix} (-g)^s y^2 \\ (-g)^{s+1} y^1 \end{bmatrix} = c(-g)^s A\boldsymbol{y},$$

where c is a constant.

Because of (3.2), we have

$$\begin{aligned}
\Psi^{-1}(t)\boldsymbol{\ell}(\boldsymbol{y}(t)) &\approx c(-g)^s S_0^{-1} G^{-1}(t) \begin{bmatrix} g^{1/4} & 0 \\ 0 & g^{-1/4} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -g & 0 \end{bmatrix} \begin{bmatrix} g^{-1/4} & 0 \\ 0 & g^{1/4} \end{bmatrix} G(t) S_0 \\
&= c(-g)^s S_0^{-1} G^{-1}(t) \begin{bmatrix} 0 & g^{1/2} \\ -g^{1/2} & 0 \end{bmatrix} G(t) S_0 \\
&= (-g)^s c g^{s+1/2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},
\end{aligned}$$

since

$$G^{-1}(t) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} G(t) \equiv \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Theorem 2 *The leading global-error term in the discretization of the linear oscillator (3.4) by a Runge–Kutta method of order $p = 2s$ is*

$$\boldsymbol{E}(T) \approx h^{2s} \left\{ c^* + (-1)^s c \int_0^T [g(t)]^{s+1/2} \mathrm{d}t \right\} \left[\begin{array}{cc} [g(T)]^{-1/4} & 0 \\ 0 & [g(T)]^{1/4} \end{array} \right] G(T) S_1, \tag{3.7}$$

where S_1 is a constant 2×2 matrix.

Proof Follows at once by the substitution of $\Psi^{-1}(t)\boldsymbol{\ell}(\boldsymbol{y}(t))$ into (2.10) and another application of (3.2), this time to estimate $\Psi(T)$. □

As our first example we consider the Airy equation

$$y'' + ty = 0, \quad t \geq 0, \qquad y(0) = 1, \ y'(0) = 0. \tag{3.8}$$

namely $g(t) = t$. Substitution in (3.7) yields for sufficiently small $h > 0$

$$y_{T/h} - y(T) \approx \frac{(-1)^s c h^{2s}}{s + \frac{3}{2}} T^{s+5/4} \chi(T), \tag{3.9}$$

where the function χ oscillates fast in $[-1, 1]$. In other words, the envelope of the error is $c_1 h^{2s} T^{s+5/4}$, where $c_1 = |c|/(s + 3/2)$.

In Figure 2 we display the global error in the solution of the Airy equation by the familiar fourth-order explicit Runge–Kutta method with the Butcher tableau

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \tag{3.10}$$

(Hairer et al. 1987). The first row corresponds to the step-size $h = \frac{1}{8}$, the second to $\frac{1}{16}$ and so on. The first column displays the 'true' global error: the solution of (3.8), $y(t) = \pi[\mathrm{Ai}(-t)\mathrm{Bi}'(0) - \mathrm{Ai}'(0)\mathrm{Bi}(-t)]$, can be computed to machine accuracy with, for example, MATLAB or MAPLE. In each case, in the given range the error oscillates rapidly within an increasing envelope. (This behaviour ceases to be true for larger values of T, when the numerical solution loses accuracy altogether and the magnitude of the global error is kept in check by the stability of the method.) According to (3.9), this envelope obeys a power law, asymptotically increasing like $T^{13/4}$. In the second column we have displayed the function

$$\log |y_{t/h} - y(t)| - \tfrac{13}{4} \log t - 4 \log h$$

for integer t/h. Our theory leads us to expect that the above function should oscillate between $-\infty$ and a horizontal line, independent of h. Compliance with this expectation is evident in the figure.

It is important – and perhaps surprising – to observe that the specific nature of a Runge–Kutta method makes no difference to the rate of global error accumulation,

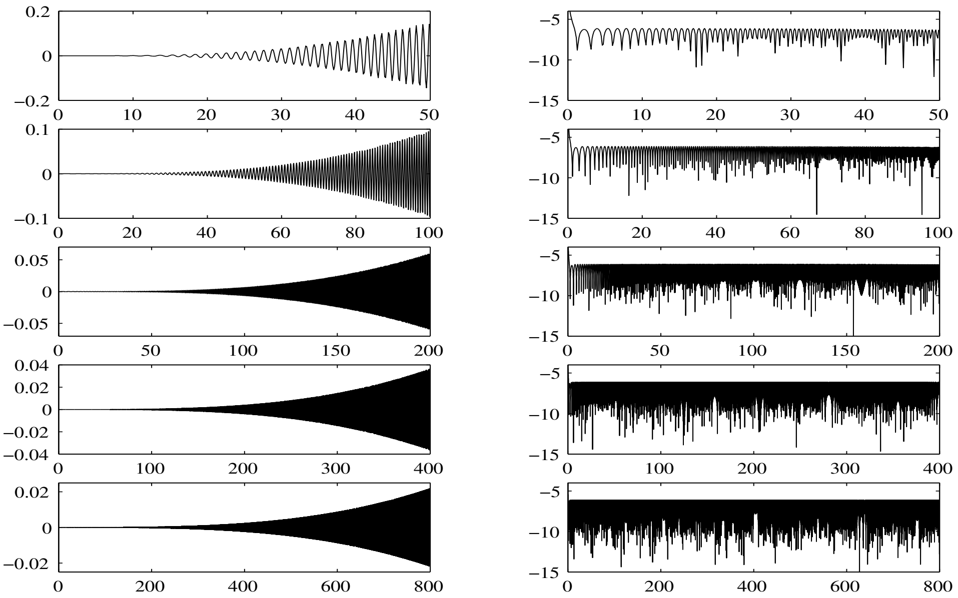


Figure 2: The error in the solution of the Airy equation (3.8) by the fourth-order explicit Runge–Kutta method (3.10).

just its order! The only amendment to (3.9) is in the value of c. Thus, in Figure 3 we display the same information as in Figure 2, arranged in an identical fashion, except that in place of (3.10) we have used the fourth-order Gauss–Legendre Runge–Kutta method

$$\begin{array}{c|cc} \frac{1}{2}-\frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4}-\frac{\sqrt{3}}{6} \\ \frac{1}{2}+\frac{\sqrt{3}}{6} & \frac{1}{4}+\frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}. \tag{3.11}$$

Unlike (3.10), the method (3.11) is implicit. It is also A-stable, algebraically stable and symplectic. Although this might have little relevance to the solution of highly-oscillatory systems, it could have led us to expect different rate of global-error accumulation.[2] It is clear from the figure that this is not the case.

Our second example is the 'log oscillator'

$$y'' + \log(1+t)y = 0, \quad t \geq 0, \qquad y(0) = 1,\ y'(0) = 0. \tag{3.12}$$

Now $g(t) = \log(1+t)$. It is possible to prove, integrating by parts, that for $T \gg 1$

[2] Although the equation (3.8) is Hamiltonian, it is not autonomous. Therefore symplecticity need not imply linear accumulation of global error, familiar from the computation of autonomous Hamiltonian systems (Hairer & Lubich 1997, Reich 1999)

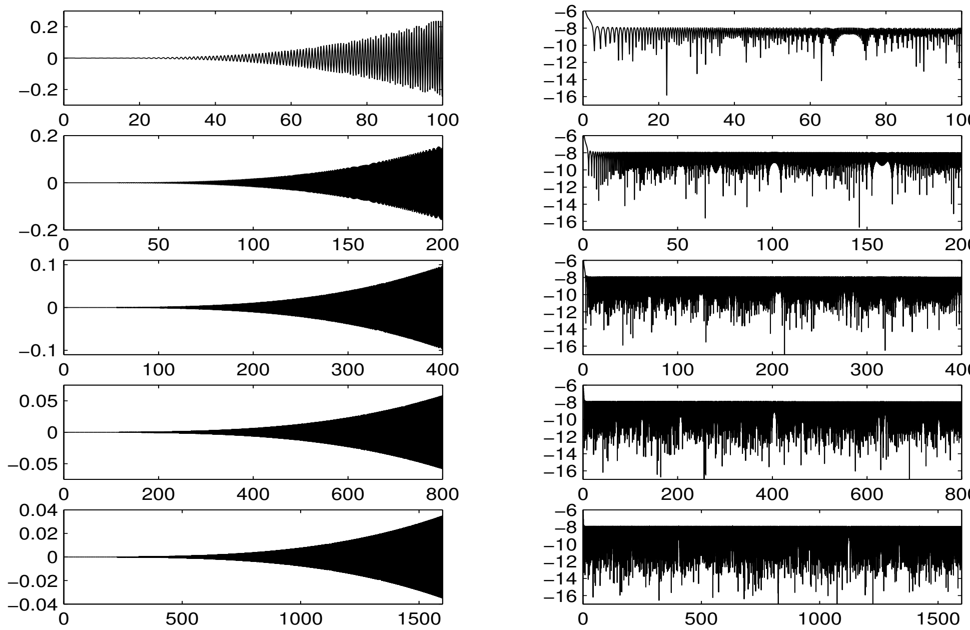


Figure 3: The error in the solution of the Airy equation (3.8) by the fourth-order Gauss–Legendre Runge–Kutta method (3.11).

and $\alpha > 0$

$$\kappa_\alpha(T) = \int_0^T [\log(1+t)]^\alpha \mathrm{d}t \approx \frac{1}{\alpha}(1+T)[\log(1+T)]^\alpha.$$

Therefore, it follows from the WKB approximation (3.2) that asymptotically

$$\begin{aligned} y(t) &\approx \frac{1}{[\log(1+t)]^{1/4}}[d_{1,1}\cos\theta(t) + d_{1,2}\sin\theta(t)], \qquad t \gg 1, \\ y'(t) &\approx [\log(1+t)]^{1/4}[-d_{2,1}\sin\theta(t) + d_{2,2}\cos\theta(t)], \end{aligned}$$

where $\theta(t) \approx 2(1+t)[\log(1+t)]^{1/2}$: the solution oscillates faster as t increases, albeit at a far gentler pace than the Airy equation (3.8), and its amplitude is slowly attenuated. We use (3.7) to estimate the global error, the outcome being

$$\left.\begin{aligned} y^1_{t/h} - y(t) &\approx (-1)^s h^{2s} \frac{\kappa_{s+1/2}(t)}{[\log(1+t)]^{1/4}}\chi_1(t) \\ &\approx (-1)^s \frac{c}{s+\frac{1}{2}} h^{2s}(1+t)[\log(1+t)]^{s+1/4}\chi_1(t), \\ y^2_{t/h} - y'(t) &\approx (-1)^s h^{2s}[\log(1+t)]^{1/4}\kappa_{s+1/2}(t)\chi_2(t) \\ &\approx (-1)^s \frac{c}{s+\frac{1}{2}} h^{2s}(1+t)[\log(1+t)]^{s+3/4}\chi_2(t), \end{aligned}\right\} \quad t \in h\mathbb{Z}_+, \qquad (3.13)$$

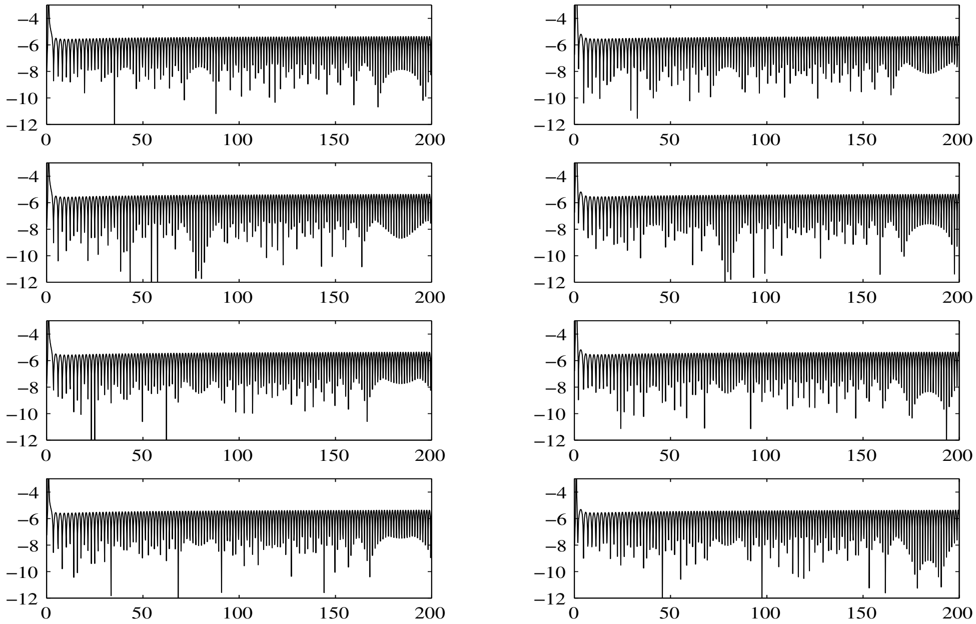


Figure 4: The error in the solution of the 'log oscillator' (3.12) by the fourth-order explicit Runge–Kutta method (3.10).

where χ_1 and χ_2 are highly-oscillating functions.

Figure 4 displays the global error in the solution of (3.12) by the fourth-order Runge–Kutta method (3.10). The rows correspond to decreasing values of the step size, $h = \frac{1}{8}, \frac{1}{16}, \frac{1}{32}$ and $\frac{1}{64}$. The first column displays the quantity

$$\log |y^1_{t/h} - y(t)| - \log(1+t) - \tfrac{9}{4} \log |\log(1+t)| - 4 \log h$$

and the second

$$\log |y^2_{t/h} - y'(t)| - \log(1+t) - \tfrac{11}{4} \log |\log(1+t)| - 4 \log h,$$

both for $t = nh$ in the range $[0, 200]$. According to (3.13), all the plots should oscillate between $-\infty$ and a horizontal line, independent of h. This is confirmed in the figure.

4 Linear oscillators II: Magnus methods

Magnus expansions have been extensively used in many areas of applied mathematics and natural sciences since the pioneering paper of Wilhelm Magnus (1954) and they have recently emerged as an effective computational method for linear (Iserles & Nørsett 1999) and nonlinear (Zanna 1999) ordinary differential systems. An important virtue of Magnus expansions is that they assure that the numerical solution evolves on a Lie group whenever the exact solution does so. We refer the reader to (Iserles et al. 2000) for an extensive survey of Lie-group methods in general and the Magnus expansion in particular. In the remainder of this and the next section we will borrow freely from the terminology, theory and notation in (Iserles et al. 2000).

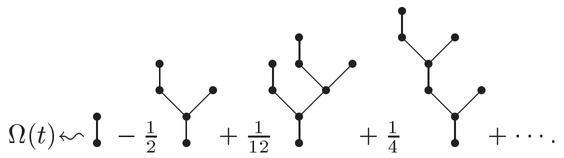
Let $\boldsymbol{y}$ be the solution of

$$\boldsymbol{y}' = A(t)\boldsymbol{y}, \quad t \geq 0, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0. \tag{4.1}$$

We represent it in the form $\boldsymbol{y}(t) = \mathrm{e}^{\Omega(t)}\boldsymbol{y}_0$. A Magnus expansion of the function Ω is

$$\begin{aligned}\Omega(t) = & \int_0^t A(\xi)\mathrm{d}\xi - \tfrac{1}{2}\int_0^t\int_0^{\xi_1}[A(\xi_2), A(\xi_1)]\mathrm{d}\xi_2\mathrm{d}\xi_1 \\ & + \tfrac{1}{12}\int_0^t\int_0^{\xi_1}\int_0^{\xi_1}[A(\xi_3),[A(\xi_2),A(\xi_1)]]\mathrm{d}\xi_3\mathrm{d}\xi_2\mathrm{d}\xi_1 \\ & + \tfrac{1}{4}\int_0^t\int_0^{\xi_1}\int_0^{\xi_2}[[A(\xi_3),A(\xi_2)],A(\xi_1)]\mathrm{d}\xi_3\mathrm{d}\xi_2\mathrm{d}\xi_1 + \cdots.\end{aligned} \tag{4.2}$$

It is very helpful to identify expansion terms with *rooted trees:* the trivial tree corresponds to the matrix $A(t)$, appending a root to a tree denotes integration and joining two tree with a common root corresponds to commutation. Thus, (4.2) can be written in the form



Binary trees in a Magnus expansion and their coefficients can be obtained recursively (Iserles & Nørsett 1999).

Practical implementation of the Magnus expansion requires the discretization of the integrals in (4.2). Surprisingly effective and affordable quadrature methods have been developed in (Blanes, Casas & Ros 1999, Iserles & Nørsett 1999, Munthe-Kaas & Owren 1999) and they rely on the special form of the integrals and the theory of free graded Lie algebras.

Local error consists of the contribution of two phenomena: the truncation of the infinite Magnus expansion and the replacement of integrals by quadrature. In this section we assume, though, that the integrals are evaluated exactly and the only error source is the truncation.

Advancing the Magnus expansion by step $h > 0$ we have

$$\boldsymbol{y}(t_{n+1}) = \mathrm{e}^{\Omega_n(h)}\boldsymbol{y}(t_n), \qquad \boldsymbol{y}_{n+1} = \mathrm{e}^{\tilde{\Omega}_n(h)}\boldsymbol{y}_n,$$

where $\tilde{\Omega}_n$ is an order-p truncation of Ω_n. Note that Ω_n can be expanded in odd powers of h, therefore there is no loss of generality in assuming that the order is even, $p = 2s$ (Iserles, Nørsett & Rasmussen 1998). The local error is thus

$$h^{p+1}\boldsymbol{\ell}(\boldsymbol{y}_n) = \mathrm{e}^{\tilde{\Omega}_n(h)}\boldsymbol{y}_n - \mathrm{e}^{\Omega_n(h)}\boldsymbol{y}_n \approx [\tilde{\Omega}_n(h) - \Omega_n(h)]\boldsymbol{y}_n.$$

Let the Magnus expansion be

$$\Omega_n(x) = \sum_{k=0}^{\infty} \sum_{\tau \in \mathbb{F}_k} \alpha(\tau) \int_{t_n}^{t_n+x} \mathcal{C}_\tau(\xi)\mathrm{d}\xi,$$

where $\mathbb{F}_k$ is the set of all trees such that $\mathcal{C}_\tau(t_n + t) = \mathcal{O}(t^k)$ (cf. (Iserles et al. 2000) for the notation here and in the sequel).[3] Then

$$\tilde{\Omega}_n(x) = \sum_{k=0}^{p-2} \sum_{\tau \in \mathbb{F}_k} \alpha(\tau) \int_{t_n}^{t_n+x} \mathcal{C}_\tau(\xi)\mathrm{d}\xi$$

and the leading term in the local error is

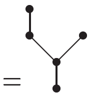
$$h^{p+1}\boldsymbol{\ell}(\boldsymbol{y}_n) \approx \left[\sum_{\tau \in \mathbb{F}_{p-1}} \alpha(\tau) \int_{t_n}^{t_{n+1}} \mathcal{C}_\tau(\xi)\mathrm{d}\xi + \sum_{\tau \in \mathbb{F}_p} \alpha(\tau) \int_{t_n}^{t_{n+1}} \mathcal{C}_\tau(\xi)\mathrm{d}\xi \right] \boldsymbol{y}_n. \tag{4.3}$$

To acquire intuition and illustrate general results, we commence by looking into the size of the terms corresponding to $\tau \in \mathbb{F}_k$ for small $k \in \mathbb{Z}_+$ when a Magnus expansion is applied to the linear oscillator (3.4). Here $t = t_n + x$.

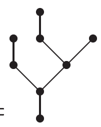
$\mathbb{F}_0$:

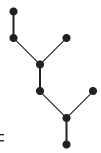
$\tau =$ $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx x \begin{bmatrix} 0 & 1 \\ -g(t_n) & 0 \end{bmatrix},$$

$\mathbb{F}_2$ (the set $\mathbb{F}_1$ is empty):

$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx \tfrac{1}{2}x^3 g'(t_n) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

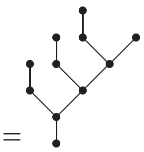
$\mathbb{F}_3$:

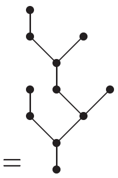
$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx -x^4 g'(t_n) \begin{bmatrix} 0 & 1 \\ g(t_n) & 0 \end{bmatrix},$$

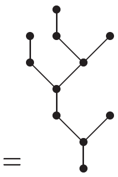
$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx x^4 g'(t_n) \begin{bmatrix} 0 & 1 \\ g(t_n) & 0 \end{bmatrix},$$

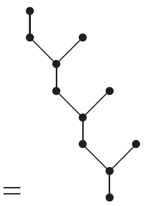
[3] The matrix function $\mathcal{C}_\tau(t)$ can be obtained recursively from the tree τ, pruning the leaves and replacing them by integration or commutation in a transparent manner, except that we do not implement the last integral, associated with the vertical node commencing from the root.

$\mathbb{F}_4$:

$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx -2x^5 g'(t_n)g(t_n) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

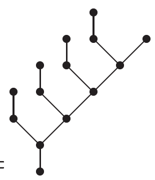
$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx 2x^5 g'(t_n)g(t_n) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx -2x^5 g'(t_n)g(t_n) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

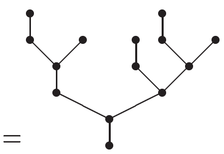
$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx 2x^5 g'(t_n)g(t_n) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

and so on.

In identifying a pattern, we should not jump too hastily to conclusions. It is tempting, for example, to assume that all the terms in the set $\mathbb{F}_k$ correspond to the same matrix function, scaled by a different scalar constant. This is false, since two among the elements of $\mathbb{F}_5$ are

$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx 4x^6 g'(t_n)g(t_n) \begin{bmatrix} 0 & 1 \\ g(t_n) & 0 \end{bmatrix}$$

and

$\tau =$  $\Rightarrow$ $$\int_{t_n}^{t_n+x} \mathcal{C}(\xi)\mathrm{d}\xi \approx -x^6 [g'(t_n)]^2 \begin{bmatrix} 0 & 1 \\ -g(t_n) & 0 \end{bmatrix}.$$

However, since $g(t) \gg |g'(t)|$ for $t \gg 1$, we may disregard the second term. In general, we need to estimate only the *dominant terms,* which exhibit the largest power of $g(t)$.

Proposition 3 *Let the expansion of $g \in \mathrm{C}^2[\mathbb{R}_+]$ about t_n be*

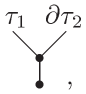
$$g(t) = g_0 + g_1 x + \tfrac{1}{2} g_2 x^2 + o(x^2),$$

where $x = t - t_n$. The dominant terms in the expansion of $\mathcal{C}_\tau(t_n + x)$ are

$$\begin{aligned}
\tau \in \mathbb{F}_0 : \qquad & \begin{bmatrix} 0 & 1 \\ -g_0 - g_1 x & 0 \end{bmatrix}, \\
\tau \in \mathbb{F}_{2k},\ k \in \mathbb{N} : \qquad & (c_1 g_0^{k-1} g_1 x^{2k} + c_2 g_0^{k-1} g_2 x^{2k+1}) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad (4.4)\\
\tau \in \mathbb{F}_{2k+1},\ k \in \mathbb{Z}_+ : \qquad & (d_1 g_0^{k-1} g_1 x^{2k+1} + d_2 g_0^{k-1} g_2 x^{2k+2}) \begin{bmatrix} 0 & 1 \\ g_0 & 0 \end{bmatrix}, \qquad (4.5)
\end{aligned}$$

where c_0, c_1, d_0, d_1 are constants.

Proof By induction. Let

$\tau =$ 

,

where $\partial\tau_2$ is the tree τ_2 with its root removed. According to Iserles et al. (1998), given $\tau_l \in \mathbb{F}_{k_i}$, $i = 1, 2$, it is true that $\tau \in \mathbb{F}_k$, where

$$k = \begin{cases} k_1 + k + 2 + 1, & \tau_1 \neq \tau_2, \\ k_1 + k_2 + 2, & \tau_1 = \tau_2. \end{cases}$$

If $\tau \in \mathbb{F}_{2k}$, $k \in \mathbb{N}$, then there are exactly five possibilities:

1. $\tau_1 \in \mathbb{F}_0$, $\tau_2 \in \mathbb{F}_{2k-1}$: Substitution of (4.5) for τ_1 yields

$$(2d_1 g_0^{k-1} g_1 x^{2k} + 2d_2 g_0^{k-1} g_2 x^{2k+1}) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

2. $\tau_1 \in \mathbb{F}_{2k-1}$, $\tau_2 \in \mathbb{F}_0$: identical to the previous case.

3. $\tau_1 \in \mathbb{F}_{2l}$, $\tau_2 \in \mathbb{F}_{2(k-l)-1}$, $1 \le l \le k-1$: Using (4.4) for τ_{2l} and (4.5) for $\tau_{2(k-l)-1}$, we obtain

$$\left[\frac{2}{2l+1} c_1 d_1 g_0^{k-3} g_1^2 x^{2k} + 2\left(\frac{c_1 d_2}{2l+1} + \frac{c_2 d_1}{2l+2}\right) g_0^{k-3} g_1 g_2 x^{2k+1}\right] \begin{bmatrix} 0 & 1 \\ -g_0 & 0 \end{bmatrix}.$$

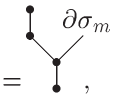
   None of these terms is dominant.

4. $\tau_1 \in \mathbb{F}_{2(k-l)-1}$, $\tau_2 \in \mathbb{F}_{2l}$, $l = 1, 2, \ldots, k-1$: identically to the previous case, no dominant terms survive.

5. $\tau_1 = \tau_2 \in \mathbb{F}_{k-1}$: In that case all the leading terms vanish and, again, there are no dominant terms.

Similar analysis applies when $\tau \in \mathbb{F}_{2k+1}$. Again, there are five cases and the only two that allow for dominant terms are when either $\tau_1 \in \mathbb{F}_0$, $\tau_2 \in \mathbb{F}_{2k}$, or the other way around. $\square$

Note that both (4.4) and (4.5) are attainable. This can be verified easily by considering the trees

$\sigma_0 =$, $\sigma_{m+1} =$  , $m \in \mathbb{Z}_+$.

We substitute the integrals of (4.4) and (4.5) into the expression for local error (4.3), noting that the h^p terms are eliminated automatically, since p is even (Iserles et al. 1998). Recalling that $p = 2s$, the dominant term in the contribution of each $\tau \in \mathbb{F}_{p-1}$ is

$$\frac{d_2}{p+1} g_0^{s-2} g_2 h^{p+1} \begin{bmatrix} 0 & 1 \\ g_0 & 0 \end{bmatrix},$$

while the dominant term of $\tau \in \mathbb{F}_p$ is

$$\frac{c_1}{p+1} g_0^{s-1} g_1 h^{p+1} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Therefore, the $\mathbb{F}_p$ terms are larger by an order of magnitude and

$$\boldsymbol{\ell}(\boldsymbol{y}_n) \approx c g(t_n)^{s-1} g'(t_n) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad n \gg 1. \tag{4.6}$$

Theorem 4 *The leading global-error term in the discretization of the linear oscillator (3.4) by the Magnus expansion of order $p = 2s$ (with exact integrals) is*

$$\begin{aligned} \boldsymbol{E}(T) \approx h^{2s} \Bigg\{ & c^* \begin{bmatrix} [g(T)]^{-1/4} & 0 \\ 0 & [g(T)]^{1/4} \end{bmatrix} G(T) \\ & + \frac{c}{2} g'(T) \begin{bmatrix} [g(T)]^{s-7/4} & 0 \\ 0 & [g(T)]^{s-5/4} \end{bmatrix} \begin{bmatrix} \sin 3\theta(T) & -\cos 3\theta(T) \\ \cos 3\theta(T) & \sin 3\theta(T) \end{bmatrix} \Bigg\} \boldsymbol{v}, \end{aligned} \tag{4.7}$$

for $T \gg 1$, where $\boldsymbol{v} \in \mathbb{R}^2$ is constant, except that for $p = 2$ we have

$$\boldsymbol{E}(T) \approx h^2 c^* \begin{bmatrix} [g(T)]^{-1/4} & 0 \\ 0 & [g(T)]^{1/4} \end{bmatrix} G(T)\boldsymbol{v}, \qquad T \gg 1. \tag{4.8}$$

Proof We substitute (4.6) and the WKB approximation (3.2) into the global-error estimate (2.10), whereby

$$\boldsymbol{E}(T) \approx h^p \Psi(T) \left[c^* + \int_{T^*}^{T} \Psi^{-1}(x) \boldsymbol{\ell}(\Phi(x)\boldsymbol{y}_0) \right]$$

$$
\begin{aligned}
&\approx h^p \begin{bmatrix} [g(T)]^{-1/4} & 0 \\ 0 & [g(T)]^{1/4} \end{bmatrix} G(T) S_0 \left\{ c^* + c \int_{T^*}^{T} g'(x)[g(x)]^{s-1} S_0^{-1} G^{-1}(x) \right. \\
&\quad \left. \times \begin{bmatrix} [g(x)]^{1/4} & 0 \\ 0 & [g(x)]^{-1/4} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} [g(x)^{-1/4} & 0 \\ 0 & [g(x)]^{1/4} \end{bmatrix} G(x) S_0 \mathrm{d}x \right\} \boldsymbol{y}_0 \\
&= h^p \begin{bmatrix} [g(T)]^{-1/4} & 0 \\ 0 & [g(T)]^{1/4} \end{bmatrix} G(T) \left\{ c^* + c \int_{T^*}^{T} g'(x)[g(x)]^{s-1} G^{-1}(x) \right. \\
&\quad \left. \times \begin{bmatrix} [g(x)]^{1/4} & 0 \\ 0 & [g(x)]^{-1/4} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} [g(x)^{-1/4} & 0 \\ 0 & [g(x)]^{1/4} \end{bmatrix} G(x) \mathrm{d}x \right\} \boldsymbol{v},
\end{aligned}
$$

where $\boldsymbol{v} = S_0 \boldsymbol{y}_0$. We observe that

$$
\begin{aligned}
&G^{-1}(x) \begin{bmatrix} [g(x)]^{1/4} & 0 \\ 0 & [g(x)]^{-1/4} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} [g(x)]^{-1/4} & 0 \\ 0 & [g(x)]^{1/4} \end{bmatrix} G(x) \\
&= \begin{bmatrix} \cos\theta(x) & -\sin\theta(x) \\ \sin\theta(x) & \cos\theta(x) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\theta(x) & \sin\theta(x) \\ -\sin\theta(x) & \cos\theta(x) \end{bmatrix} \\
&= \begin{bmatrix} \cos 2\theta(x) & \sin 2\theta(x) \\ -\sin 2\theta(x) & \cos 2\theta(x) \end{bmatrix}, \qquad \text{where} \qquad \theta(x) = \int_0^x [g(\xi)]^{1/2} \mathrm{d}\xi.
\end{aligned}
$$

Therefore, we need to estimate the real and imaginary parts of the function

$$
r_T(\alpha) = \int_{T^*}^{T} g'(x)[g(x)]^{\alpha} \exp\left\{ 2\mathrm{i} \int_0^x [g(\xi)]^{1/2} \mathrm{d}\xi \right\} \mathrm{d}x.
$$

Integrating by parts,

$$
\begin{aligned}
r_T(\alpha) &= -\tfrac{1}{2}\mathrm{i} \int_{T^*}^{T} g'(x)[g(x)]^{\alpha-1/2} \frac{\mathrm{d}}{\mathrm{d}x} \exp\left\{ 2\mathrm{i} \int_0^x [g(\xi)]^{1/2} \mathrm{d}\xi \right\} \mathrm{d}x \\
&= -\tfrac{1}{2}\mathrm{i} \left(g'(T)[g(T)]^{\alpha-1/2} \exp\left\{ 2\mathrm{i} \int_0^T [g(\xi)]^{1/2} \mathrm{d}\xi \right\} \right. \\
&\qquad \left. - g'(T^*)[g(T^*)]^{\alpha-1/2} \exp\left\{ 2\mathrm{i} \int_0^{T^*} [g(\xi)]^{1/2} \mathrm{d}\xi \right\} \right) \\
&\qquad + \tfrac{1}{2}\mathrm{i} \int_{T^*}^{T} \{ g''(x)[g(x)]^{\alpha-1/2} + (\alpha - \tfrac{1}{2})[g'(x)]^2 [g(x)]^{\alpha-3/2} \} \\
&\qquad \times \exp\left\{ 2\mathrm{i} \int_0^x [g(\xi)]^{1/2} \mathrm{d}\xi \right\} \mathrm{d}x.
\end{aligned}
$$

Since $g'' g^{\alpha-1/2}$ are $(g')^2 g^{\alpha-3/2}$ are small in comparison with g^{α} and $g(T) \gg g(T^*)$ for sufficiently large T, we deduce that

$$
r_T(\alpha) \approx -\tfrac{1}{2}\mathrm{i} g'(T)[g(T)]^{\alpha-1/2} \exp\left\{ 2\mathrm{i} \int_0^T [g(\xi)]^{1/2} \mathrm{d}\xi \right\}
$$

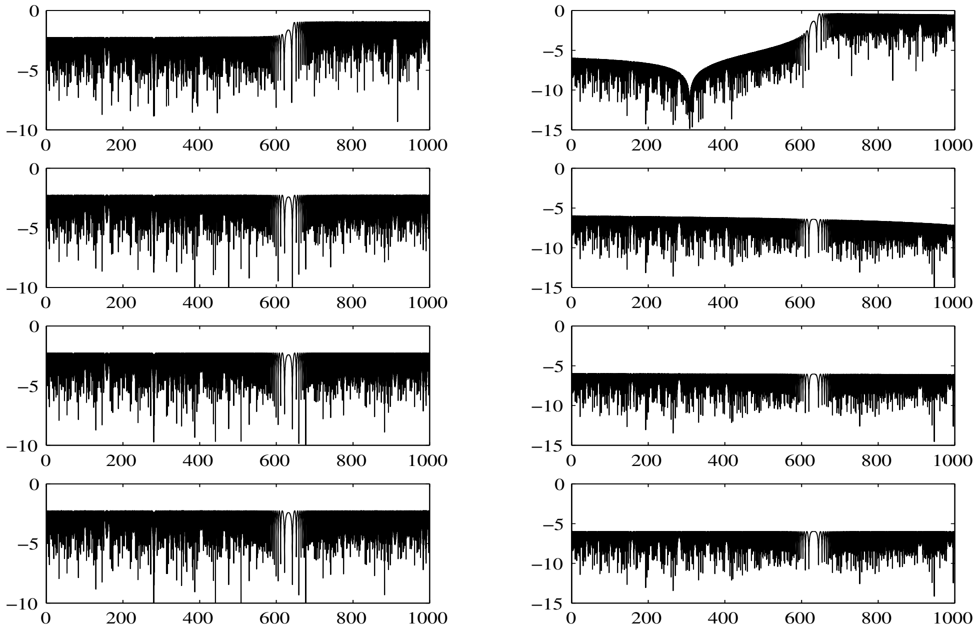


Figure 5: The error in the solution of the Airy equation (3.8) by second-order (left column) and fourth-order (right column) Magnus expansion.

and, substituting into the expression for the global error we have

$$\begin{aligned}\boldsymbol{E}(T) &\approx h^p \begin{bmatrix} [g(T)]^{-1/4} & 0 \\ 0 & [g(T)]^{1/4} \end{bmatrix} G(T) \left\{ c^* + c \begin{bmatrix} \operatorname{Re} r_T(s-1) & \operatorname{Im} r_T(s-1) \\ -\operatorname{Im} r_T(s-1) & \operatorname{Re} r_T(s-1) \end{bmatrix} \right\} \boldsymbol{v} \\ &\approx h^p \begin{bmatrix} [g(T)]^{-1/4} & 0 \\ 0 & [g(T)]^{1/4} \end{bmatrix} G(T) \{ c^* \\ &\quad + \frac{c}{2} g'(T)[g(T)]^{s-3/2} \begin{bmatrix} \sin 2\theta(T) & -\cos 2\theta(T) \\ \cos 2\theta(T) & \sin 2\theta(T) \end{bmatrix} \Bigg\} \boldsymbol{v}.\end{aligned}$$

If $p = 2$, hence $s = 1$, the expression inside the braces is dominated by c^* and we deduce (4.8). If $p \geq 4$, however, either term within the braces might dominate. Since

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sin 2\theta & -\cos 2\theta \\ \cos 2\theta & \sin 2\theta \end{bmatrix} = \begin{bmatrix} \sin 3\theta & -\cos 3\theta \\ \cos 3\theta & \sin 3\theta \end{bmatrix},$$

we obtain the global-error estimate (4.7) and conclude the proof. □

To illustrate Theorem 4 we revisit the two examples from Section 3. For the Airy equation (3.8) we obtain the global error estimate

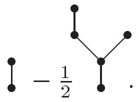
$$y_{T/h} - y(T) \approx \frac{c}{2} h^{2s} T^{\max\{s-7/4,-1/4\}} \chi(T), \qquad 0 < h \ll 1,\ T \gg 1, \qquad (4.9)$$

where χ oscillates fast in $[-1, 1]$. Note that the envelope of the leading term in the global-error expansion for the second-order method is attenuated at exactly the rate of decay of the envelope of the exact solution of (3.8). This is no longer true when the order is higher. Yet, comparison of (4.9) with (3.9) confirms that in that case the error in the Magnus method is smaller by roughly T^3, in comparison with Runge–Kutta methods of equal order.

Figure 5 illustrates (4.9). The values of h are halved progressively: the first row corresponds to $h = \frac{1}{8}$, the second to $h = \frac{1}{16}$ and so on. The left column displays

$$\log|y_{t/h} - y(t)| + \tfrac{1}{4}\log t - 2\log h, \qquad t \in h\mathbb{Z}_+,$$

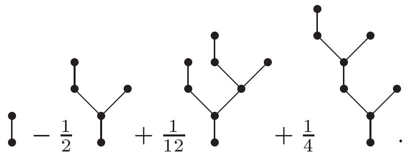
where y_n has been produced by the second-order expansion



The right column depicts

$$\log|y_{t/h} - y(t)| - \tfrac{1}{4}\log t - 4\log h, \qquad t \in h\mathbb{Z}_+,$$

y_n having been obtained from the fourth-order Magnus method



Besides confirming the estimates of Theorem 4, Figure 5 emphasises an important point: our global-error estimates are valid for any given $T > 0$ only when $h > 0$ is small enough. Thus, the second-order method is consistent with (4.8) for $h = \frac{1}{8}$ only until $t \approx 580$, while the error of the fourth-order method is well explained by (4.7) throughout the range $0 \le t \le 10^3$ only for $h \le \frac{1}{32}$.

We next consider the global error in the solution of the 'log oscillator' (3.12). Since

$$g'(T)[g(T)]^{s-3/2} = \frac{[\log(1+T)]^{s-3/2}}{1+T} \to 0, \qquad T \to \infty,$$

the first term in (4.7) is always dominant. Therefore, the envelope of the global error in all Magnus expansions is attenuated at exactly the same speed as the solution itself, namely $[\log(1+T)]^{-1/4}$. This is illustrated in Figure 6. Again, the first column is devoted to the second-order expansion and it displays

$$\log|y_{t/h} - y(t)| + \tfrac{1}{4}\log|\log(1+t)| - 2\log h, \qquad t \in h\mathbb{Z}_+.$$

The second column displays the outcome of the fourth-order method, namely

$$\log|y_{t/h} - y(t)| + \tfrac{1}{4}\log|\log(1+t)| - 4\log h, \qquad t \in h\mathbb{Z}_+.$$

The step size is halved in each consecutive row, commencing from $\frac{1}{8}$. The fit with the theoretical estimates from Theorem 4, expressed by the horizontal straight line in each plot, is remarkable. Note that, unsurprisingly, the plots indicate that the leading error constant for the fourth-order method is substantially smaller.

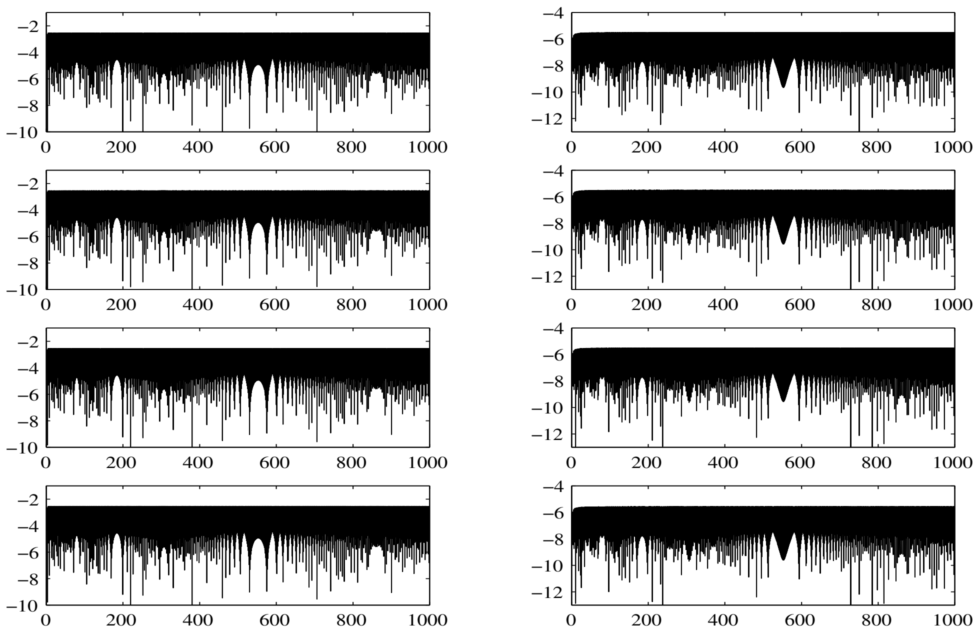


Figure 6: The error in the solution of the 'log oscillator' (3.12) by second-order (left column) and fourth-order (right column) Magnus expansion.

5 Linear oscillators III: Modified Magnus methods

In Sections 3 and 4 we have provided theoretical justification to the observation that motivated this paper, namely the remarkable performance of Magnus methods in the computation of highly-oscillatory initial-value problems. In this section we go a step further and propose a method, in the spirit of Magnus expansions, that is designed explicitly for the integration of highly-oscillatory systems. We commence from a linear problem,

$$\boldsymbol{y}' = A(t)\boldsymbol{y}, \quad t \geq 0, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0, \tag{5.1}$$

but mention an implementation of our ideas in a nonlinear setting in Section 6. Our assumption is that for $t \gg 1$ the matrix $A(t)$ possesses one or more pairs of eigenvalues with large imaginary parts. Although this is not necessary for high oscillation, this state of affairs occurs often enough in practice to justify consideration.

To advance the algorithm from t_n to $t_{n+1} = t_n + h$, where $h = h_n$ is allowed to vary between steps, we set

$$\boldsymbol{y}(t) = \mathrm{e}^{(t-t_n)A(t_{n+1/2})}\boldsymbol{x}(t), \qquad t \geq t_n \tag{5.2}$$

where $t_{n+1/2} = t_n + \frac{1}{2}h$. The function $\boldsymbol{x}$ obeys the differential equation

$$\boldsymbol{x}' = B(t)\boldsymbol{x}, \quad t \geq t_n, \qquad \boldsymbol{x}(t_n) = \boldsymbol{y}(t_n), \tag{5.3}$$

where

$$B(t) = \mathrm{e}^{-(t-t_n)A(t_{n+1/2})}[A(t) - A(t_{n+1/2})]\mathrm{e}^{(t-t_n)A(t_{n+1/2})}, \qquad t \geq t_n. \tag{5.4}$$

The idea is now to solve (5.3) by a standard Magnus method, first replacing $\boldsymbol{y}(t_n)$ by $\boldsymbol{y}_n$. Thus, if the truncated Magnus solution of (5.3) is $\boldsymbol{x}(t) = \mathrm{e}^{\tilde{\Omega}_n(t)}\boldsymbol{y}_n$, we let

$$\boldsymbol{y}_{n+1} = \mathrm{e}^{hA(t_{n+1/2})}\mathrm{e}^{\tilde{\Omega}_n(h)}\boldsymbol{y}_n, \qquad n \in \mathbb{Z}_+. \tag{5.5}$$

The intuition behind the *modified Magnus algorithm* is that the oscillatory behaviour of (5.1) is modelled well *locally* by a linear equation with constant coefficients, namely $\tilde{\boldsymbol{y}}' = A(t_{n+1/2})\tilde{\boldsymbol{y}}$. The matrix B has two features that render our intuition into a more tangible mathematical statement. Firstly, according to (5.4) $B(t) = \mathcal{O}\big((t - t_{n+1/2})\big)$, hence it is small. Therefore (5.3) represents a higher-order correction to the solution. It is the second feature, however, which is more subtle and, arguably, more important in motivating the algorithm: The entries of $B(t)$ are likely to be highly-oscillatory, behaviour inherited from the entries of $\exp[(t-t_n)A(t_{n+1/2})]$. This feature, which will be made explicit in the sequel for the important special case (3.4), is crucial! The reason why the performance of most time-stepping methods, whether 'classical' (Runge–Kutta, multistep, ...) or less standard (Magnus, Runge–Kutta/Munthe-Kaas, ...) deteriorates in the presence of high oscillations is that their error is described locally in the form $h^{p+1}\boldsymbol{w}_n$, where the vector $\boldsymbol{w}_n$ is either a scaled derivative of $\boldsymbol{y}$ or a linear combination of elementary differentials, composed of derivatives. Now, as $\boldsymbol{y}$ oscillates fast, so do its derivatives, *with an increasing amplitude:* the larger the oscillation of $\boldsymbol{y}$, the larger will be the amplitude of its derivatives! Not so for the modified Magnus expansion (5.5), though! The vector field $B(t)$ is itself a highly-oscillating function. Forming a Magnus expansion of (5.3) involves repeated integration of this vector field – but integration is the exact opposite of differentiation, a smoothing operator! Each consecutive integral *lowers* the amplitude. As the Magnus expansion is truncated, the excised terms are themselves iterated integrals, hence the highly-oscillatory nature of $B(t)$ is likely to render them small.

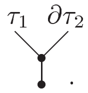
Before proceeding any further, we need to debate how to truncate the Magnus expansion $\tilde{\Omega}$ to obtain given order p. Standard theory of Magnus expansions assumes $A(t) = \mathcal{O}(1)$, but, as discussed in (Iserles & Nørsett 1999), fewer terms are needed when $A(t) = \mathcal{O}(t^q)$ for some $q \in \mathbb{N}$. In our case, replacing A with B, (5.4) implies $B(t) = \mathcal{O}(t)$, but we present analysis for general $q \in \mathbb{N}$. Recall that $\mathbb{F}_k$ is the set of all the trees τ such that $\mathcal{C}_\tau(t) = \mathcal{O}\big(t^k\big)$ for all sufficiently smooth functions A (inclusive of the case $A(t) = \mathcal{O}(1)$) and that $\mathbb{T}_k$ is the set of all trees with k 'bifurcations' (i.e., where $\mathcal{C}_\tau(t)$ has k commutators), $k \in \mathbb{Z}_+$ (Iserles et al. 2000).

We say that a tree τ has a *power* $\boldsymbol{p}(\tau, q) = s$, $s, q \in \mathbb{Z}_+$ if $\mathcal{C}_\tau(t) = \mathcal{O}(t^s)$ for all sufficiently smooth matrix functions A such that $A(t) = \mathcal{O}(t^q)$. Let $\mathbb{S}_{k,r} = \mathbb{T}_k \cap \mathbb{F}_r$, $k, r \in \mathbb{Z}_+$.

Proposition 5 *If* $\tau \in \mathbb{S}_{k,r}$ *then* $\boldsymbol{p}(\tau, q) = r + (k+1)q$.

Proof We assume, for ease of notation, that $t_n = 0$. The proposition is certainly true for $k = 0$, hence we might assume that $k \in \mathbb{N}$. Therefore there exist τ_1 and τ_2

such that

$$\tau = $$  .

Let $\tau_i \in \mathbb{S}_{k_i,r_i}$, $i = 1, 2$. Then, by the definition of the sets $\mathbb{T}_j$, $k = k_1 + k_2 + 1$, . Moreover,

$$\tau \in \begin{cases} \mathbb{F}_{r_1+r_2+1}, & \tau_1 \neq \tau_2, \\ \mathbb{F}_{r_1+r_2+2}, & \tau_1 = \tau_2 \end{cases}$$

(Iserles et al. 1998).

By the induction assumption $\boldsymbol{p}(\tau_i, q) = r_i + (k_i + 1)q = s_i$, $i = 1, 2$, say. Therefore there exist matrices $A_{i,j}$ such that

$$\mathcal{C}_{\tau_i}(t) = A_{i,s_i} t^{s_i} + A_{i,s_i+1} t^{s_i+1} + \cdots .$$

Consequently

$$\begin{aligned}
\mathcal{C}_\tau(t) &= \left[\int_0^t \mathcal{C}_{\tau_1}(\xi)\mathrm{d}\xi, \mathcal{C}_{\tau_2}(t) \right] \\
&= [\tfrac{1}{s_1+1} A_{1,s_1} t^{s_1+1} + \tfrac{1}{s_1+2} A_{1,s_1+1} t^{s_1+2} + \cdots, A_{2,s_2} t^{s_2} + A_{2,s_2+1} t^{s_2+1} + \cdots] \\
&= \tfrac{1}{s_1+1} [A_{1,s_1}, A_{2,s_2}] t^{s_1+s_2+1} \\
&\quad + (\tfrac{1}{s_1+2} [A_{1,s_1+1}, A_{2,s_2}] + \tfrac{1}{s_1+1} [A_{1,s_1}, A_{2,s_2+1}]) t^{s_1+s_2+2} + \cdots .
\end{aligned}$$

If $\tau_1 \neq \tau_2$ then there exists $A(t) = \mathcal{O}(t^q)$ such that A_{1,s_1} and A_{2,s_2} do not commute, therefore we deduce that

$$\boldsymbol{p}(\tau, q) = s_1 + s_2 + 1 = (r_1 + r_2 + 1) + (k_1 + k_2 + 2)q = r + (k+1)q,$$

consistently with the statement of the proposition. Likewise, when $\tau_1 = \tau_2$ then the first commutator vanishes,

$$\mathcal{C}_\tau(t) = \tfrac{1}{(s_1+1)(s_1+2)} [A_{1,s_1}, A_{1,s_1+1}] t^{2s_1+2} + \cdots$$

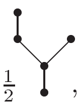
and, again, $\boldsymbol{p}(\tau, q)$ complies with the statement of the proposition. □

Seeking an order-p approximation $\tilde{\Omega}_n$, we discard all trees τ such that $\boldsymbol{p}(\tau, 1) \geq p$. The outcome is

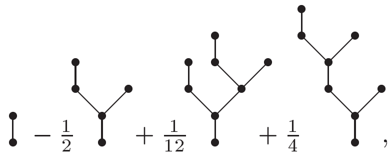
$$\tilde{\Omega}_n(h) = \sum_{k=0}^{p-2} \sum_{r=0}^{p-2-k} \sum_{\tau \in \mathbb{S}_{k,r}} \alpha(\tau) \int_{t_n}^{t_{n+1}} \mathcal{C}_\tau(\xi)\mathrm{d}\xi. \tag{5.6}$$

It is now possible, with little effort, to obtain Magnus expansions of various orders for the case $A(t) = \mathcal{O}(t)$ explicitly:

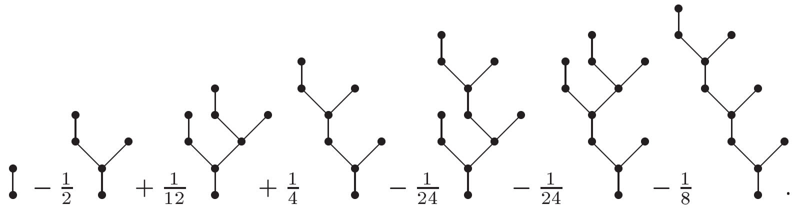
$p = 4:$,

$p = 6:$ $- \frac{1}{2}$  ,

$p = 8:$



$p = 10:$



Twelfth-order method requires 21 trees.

In general, we need far less trees than for standard (i.e., with $A(t) = \mathcal{O}(1)$) Magnus expansions: one, rather than two, for order four and two, instead of seven, for order six. This, needless to say, does not mean that the modified Magnus method is cheaper, far from it: each function evaluation requires the computation of a matrix exponential.

Practical implementation of modified Magnus solvers requires an evaluation of integrals. Since the terms are more complicated than in the case of standard Magnus expansions, we cannot assume without drastic loss of generality that the integrals can be evaluated explicitly, hence we need to replace them with quadrature. An obvious course of action is to use the very powerful set of techniques that have been developed to discretise integrals of standard Magnus expansions (Iserles et al. 2000). Thus, for example, the following MATLAB subroutine codes modified Magnus method of order four with Gaussian quadrature of matching order.

```
function [t, Y] = MMagnus4(N, t0, tN, y0, h);

% Modified Magnus of order 4 for linear equations
% Input:
%  N:   the spacing of solution points (typically N=1)
%  t0:  initial point
%  tN:  endpoint
%  y0:  initial value
%  h:   step size
% Output:
%  t = [t0 : N*h * tN]
%  Y = [y(t0), y(t0+Nh), y(t0+2Nh), ... , y(tN)]

c1 = (1 / 2 - sqrt(3) / 6) * h;
c2 = (1 / 2 + sqrt(3) / 6) * h;
h2 = h / 2;

t = t0;
Y = y0;
M = floor((tN - t0) / h + eps);
```

```
for n = 1 : M
  A0 = Afun(t0 + h2);
  Th = expm(h * A0);
  T1 = expm(c1 * A0);
  T2 = expm(c2 * A0);
  A1 = Afun(t0 + c1);
  A2 = Afun(t0 + c2);
  U1 = T1 \ (A1 - A0) * T1;
  U2 = T2 \ (A2 - A0) * T2;
  Theta = h2 * (U1 + U2);
  y0 = Th * expm(Theta) * y0;
  t0 = n * h;
  if floor(n / N) * N == n
    Y = [Y, y0];
    t = [t, t0];
  end
end
```

where the matrix $A(t)$ is produced by the function `Afun.m`: for the Airy equation (3.8) it reads

```
function A = Afun(t);
A = [0, 1; -t, 0];
```

Note that neither `MMagnus2lin.m` nor `Afun.m` have been optimised for the oscillator (3.4) and they are equally suitable for any linear system (5.1).

Interestingly enough, a sixth-order method requires little extra effort, in comparison with the fourth-order one. The reason is that one of the three nodes of Gauss–Legendre sixth-order quadrature in $[0, h]$ falls at $\frac{1}{2}h$, where the matrix B vanishes. For completeness, herewith a MATLAB sixth-order subroutine,

```
function [t, Y] = MMagnus6(N, t0, tN, y0, h);

% Modified Magnus 6 for linear equations
% Input:
%  N:   the spacing of solution points (typically N=1)
%  t0:  initial point
%  tN:  endpoint
%  y0:  initial value
%  h:   step size
% Output:
%  t = [t0 : N*h * tN]
%  Y = [y(t0), y(t0+Nh), y(t0+2Nh), ... , y(tN)]

c1 = (1 / 2 - sqrt(15) / 10) * h;
c2 = h / 2;
c3 = (1 / 2 + sqrt(15) / 10) * h;
b1 = 5 * h / 18;
```

```
b21 = sqrt(15) / 108 * h^2;

t = t0;
Y = y0;
M = floor((tN - t0) / h + eps);

for n = 1 : M
  A0 = Afun(t0 + c2);
  Th = expm(h * A0);
  T1 = expm(c1 * A0);
  T3 = expm(c3 * A0);
  A1 = Afun(t0 + c1);
  A3 = Afun(t0 + c3);
  U1 = T1 \ (A1 - A0) * T1;
  U3 = T3 \ (A3 - A0) * T3;
  Theta = b1 * (U1 + U3) + b21 * (U3 * U1 - U1 * U3);
  y0 = Th * expm(Theta) * y0;
  t0 = n * h;
  if floor(n / N) * N == n
    Y = [Y, y0];
    t = [t, t0];
  end
end
```

Gauss–Legendre quadrature, however, is not necessarily the best means to discretise integrals with highly-oscillatory integrands. Superior performance can be expected, for example, from Filon quadrature (Davis & Rabinowitz 1984) or other techniques for highly-oscillatory integrands. This is a subject of ongoing research.

Modified Magnus expansion has a particularly simple and appealing form when applied to the linear oscillator (3.4) and the exponential can be evaluated explicitly. Letting $\eta = g(t_{n+1/2})$, we have

$$\exp\left(x\begin{bmatrix} 0 & 1 \\ -\eta & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(\eta^{1/2}x) & \eta^{-1/2}\sin(\eta^{1/2}x) \\ -\eta^{1/2}\sin(\eta^{1/2}x) & \cos(\eta^{1/2}x) \end{bmatrix},$$

whence (5.4) yields

$$B(t_n+x) = [g(t_n+x)-\eta]\begin{bmatrix} \frac{1}{2}\eta^{-1/2}\sin(2\eta^{1/2}x) & \eta^{-1}\sin^2(\eta^{1/2}x) \\ -\cos^2(\eta^{1/2}x) & -\frac{1}{2}\eta^{-1/2}\sin(2\eta^{1/2}x) \end{bmatrix}. \tag{5.7}$$

Note that B is a rank-one nilpotent matrix. A fourth-order modified Magnus solver is

$$\boldsymbol{y}_{n+1} = \begin{bmatrix} \cos\psi & \eta^{-1/2}\sin\psi \\ -\eta^{1/2}\sin\psi & \cos\psi \end{bmatrix} \mathrm{e}^{\tilde{\Omega}_n(h)}\boldsymbol{y}_n, \tag{5.8}$$

where

$$\eta = g(t_{n+1/2}), \qquad \psi = \eta^{1/2}h,$$

$$\tilde{\Omega}_n(h) = \int_0^h [g(t_n+x)-\eta]\begin{bmatrix} \frac{1}{2}\eta^{-1/2}\sin(2\eta^{1/2}x) & \eta^{-1}\sin^2(\eta^{1/2}x) \\ -\cos^2(\eta^{1/2}x) & -\frac{1}{2}\eta^{-1/2}\sin(2\eta^{1/2}x) \end{bmatrix}\mathrm{d}x.$$

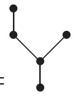
Instead of relying on Proposition 5, we examine the size of the terms $\mathcal{C}_\tau(t)$ directly for (5.7), since the special structure of this matrix is important in expressing the local error in an explicit form. Let

$$g(t_n + x) = g(t_{n+1/2}) + (x - \tfrac{1}{2}h)(\alpha + \beta x + \cdots).$$

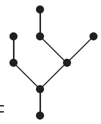
Then

$\tau =$:

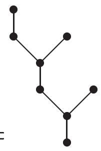
$$\mathcal{C}_\tau(t_n + x) \approx \alpha(x - \tfrac{1}{2}h)\begin{bmatrix} x & x^2 \\ -1 & -x \end{bmatrix},$$

$\tau =$  :

$$\mathcal{C}_\tau(t_n + x) \approx \alpha^2 h(t - \tfrac{1}{2}h)\begin{bmatrix} -\frac{1}{3}x^3 & -\frac{1}{6}x^4 \\ \frac{1}{2}x^2 & \frac{1}{3}x^3 \end{bmatrix},$$

$\tau =$  :

$$\mathcal{C}_\tau(t_n + x) \approx \alpha^3 h^2(t - \tfrac{1}{2}h)\begin{bmatrix} \frac{1}{144}(2\frac{t}{h} - \frac{\beta}{\alpha})x^5 & -\frac{1}{36}x^6 \\ -\frac{1}{12}x^4 & -\frac{1}{144}(2\frac{t}{h} - \frac{\beta}{\alpha})x^5 \end{bmatrix},$$

$\tau =$  :

$$\mathcal{C}_\tau(t_n + x) \approx \alpha^3 h^2(t - \tfrac{1}{2}h)\begin{bmatrix} \frac{1}{15}x^5 & \frac{1}{20}x^6 \\ -\frac{1}{12}x^4 & -\frac{1}{15}x^5 \end{bmatrix}.$$

Proposition 6 *Let $\tau \in \mathbb{T}_k$, $k \in \mathbb{Z}_+$. There exist functions c_1, c_2, c_3 of x/h such that*

$$\mathcal{C}_\tau(t_n + x) = \alpha^{k+1}h^k(x - \tfrac{1}{2}h)\begin{bmatrix} c_1 x^{2k+1} + \cdots & c_2 x^{2k+2} + \cdots \\ c_3 x^{2k} + \cdots & -c_1 x^{2k+1} + \cdots \end{bmatrix}. \tag{5.9}$$

Proof By induction on $k \in \mathbb{Z}_+$, similarly to the proof of Proposition 5. □

Since $x \in [0, h]$, it follows from (5.9) that

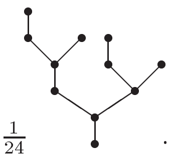
$$\int_{t_n}^{t_{n+1}} \mathcal{C}_\tau(\xi)\mathrm{d}\xi = \alpha^{k+1}\begin{bmatrix} \tilde{c}_1 h^{3k+3} + \cdots & \tilde{c}_2 h^{3k+4} + \cdots \\ \tilde{c}_3 h^{3k+2} + \cdots & -\tilde{c}_1 h^{3k+3} + \cdots \end{bmatrix}, \qquad \tau \in \mathbb{T}_k,\ k \in \mathbb{Z}_+.$$

We deduce that

$$\tilde{\Omega}_{n,r}(h) = \sum_{k=0}^{r-1}\sum_{\tau \in \mathbb{T}_k} \alpha(\tau)\int_{t_n}^{t_{n+1}} \mathcal{C}_\tau(\xi)\mathrm{d}\xi = \Omega_n(h) + \mathcal{O}\big(h^{3r+2}\big), \tag{5.10}$$

hence it is an order-$(3r+1)$ approximation.

Comparison with (5.6) confirms that exploitation of the special structure of (5.7) boosts the order: from six to seven and from eight to ten, while the tenth-order method becomes of order thirteen upon the addition of

$-\frac{1}{24}$  .

Since $\alpha = g'(t_{n+1/2})$, it follows from (5.9) that the local error is

$$\boldsymbol{\ell}(\boldsymbol{y}_n) = \mathrm{e}^{hA(t_{n+1})}[\mathrm{e}^{\tilde{\Omega}_{n,r}(h)} - \mathrm{e}^{\Omega_n(h)}]\boldsymbol{y}_n \approx \tilde{c}_3[g'(t_{n+1/2})]^{r+1}h^{3r+2}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\boldsymbol{y}_n.$$

Using again the WKB approximation (3.2), we have

$$\begin{aligned}\Phi^{-1}\boldsymbol{\ell}(\Phi\boldsymbol{y}_0) &\approx \tilde{c}_3 {g'}^{r+1} S_0^{-1}G^{-1}\begin{bmatrix} g^{1/4} & 0 \\ 0 & g^{-1/4} \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} g^{-1/4} & 0 \\ 0 & g^{1/4} \end{bmatrix} G S_0 \boldsymbol{y}_0 \\ &= \frac{\tilde{c}_3}{2}{g'}^{r+1}g^{-1/2}S_0^{-1}\begin{bmatrix} -\sin 2\theta & -1 + cos 2\theta \\ 1+\cos 2\theta & \sin 2\theta \end{bmatrix} S_0\boldsymbol{y}_0.\end{aligned}$$

Substitution into (2.10) produces the leading term in the global error expansion.

Theorem 7 *The leading global error term in the discretization of the linear oscillator (3.4) by the modified Magnus expansion of order* $p = 3r+1$ *(with exact integrals) is*

$$\begin{aligned}\boldsymbol{E}(T) \approx h^{3r+1}&\begin{bmatrix} [g(T)]^{-1/4} & 0 \\ 0 & [g(T)]^{1/4} \end{bmatrix} G(T)\,\Big\{c^* \\ &+ \int_{T^*}^{T} \frac{[g'(x)]^{r+1}}{[g(x)]^{1/2}} \begin{bmatrix} -\sin 2\theta(x) & -1+\cos 2\theta(x) \\ 1+\cos 2\theta(x) & \sin 2\theta(x) \end{bmatrix} \mathrm{d}x \Big\}\boldsymbol{v},\end{aligned} \tag{5.11}$$

where $\boldsymbol{v} \in \mathbb{R}^2$ *is constant and*

$$\theta(x) = \int_0^x [g(\xi)]^{1/2}\mathrm{d}\xi.$$

We return to the two examples that have already featured in Sections 3 and 4: the Airy equation (3.8) and the 'log oscillator' (3.12). The discussion of the second equation is much simpler: since $g(t) = \log(1+t)$, it is true that

$$\int_{T^*}^{T} \frac{[g'(x)]^{r+1}}{[g(x)]^{1/2}}\mathrm{d}x = \int_{T^*}^{T} \frac{\mathrm{d}x}{(1+x)^{r+1}[\log(1+x)]^{1/2}} = \mathcal{O}(1)\,,$$

therefore, exactly like for standard Magnus expansion, the global-error envelope is attenuated like $[\log(1+t)]^{-1/4}$.

For the Airy equation we have $g(t) = t$, therefore

$$\begin{aligned}\boldsymbol{E}(T) \approx h^{3r+1}&\begin{bmatrix} T^{-1/4} & 0 \\ 0 & T^{1/4} \end{bmatrix} G(T)\,\Big\{c^* \\ &+ \int_{T^*}^{T} \frac{1}{x^{1/2}} \begin{bmatrix} -\sin 2\theta(x) & -1+\cos 2\theta(x) \\ 1+\cos 2\theta(x) & \sin 2\theta(x) \end{bmatrix} \mathrm{d}x \Big\}\boldsymbol{v},\end{aligned}$$

where $\theta(x) = \frac{2}{3}x^{3/2}$. It is easy to verify that

$$\int_{T^*}^{T} \frac{1}{x^{1/2}}\mathrm{e}^{\frac{2}{3}\mathrm{i}x^{3/2}}\mathrm{d}x = \mathcal{O}(1)\,, \qquad 0 \ll T^* < T,$$

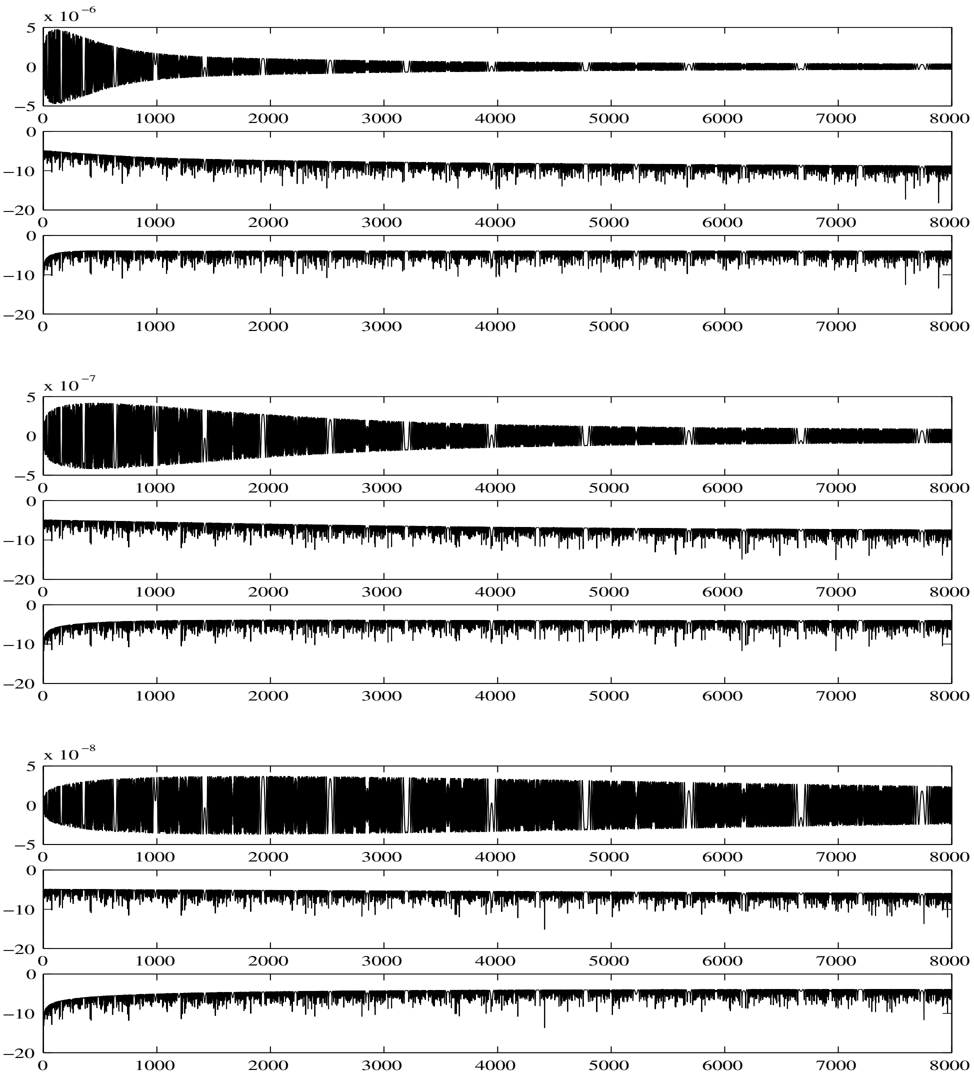


Figure 7: The error in the solution of the Airy equation (3.8) by fourth-order modified Magnus expansion with exact integrals. The first three plots correspond to $h = \frac{1}{8}$, middle three to $h = \frac{1}{16}$ and the three bottom plots to $h = \frac{1}{32}$.

while, of course, $\int_{T^*}^{T} x^{-1/2}\mathrm{d}x = 2(T^{1/2} - T^{*1/2})$. It is now easy to confirm that

$$y_{t/h} - y(t) \approx h^{3r+1} T^{1/4} \chi(T), \qquad t \in h\mathbb{Z}_+ \tag{5.12}$$

where χ oscillates fast in $[-1, 1]$.

Compared with (4.9), the expression (5.12) is somewhat disappointing: After all, the global-error envelope of second-order Magnus expansion decays like $T^{-1/4}$. Al-

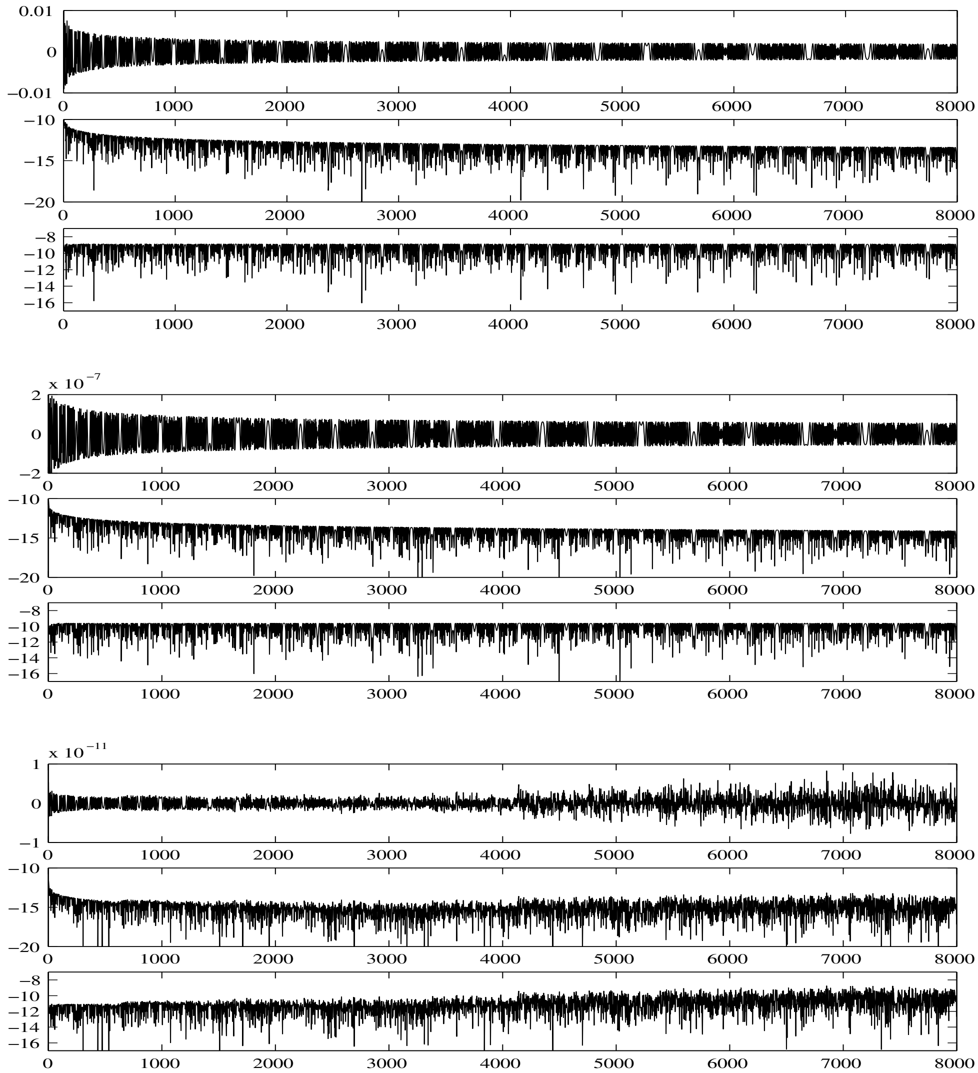


Figure 8: The error in the solution of the Airy equation (3.8) by seventh-order modified Magnus expansion with exact integrals. The first three plots correspond to $h = 2$, middle three to $h = \frac{1}{2}$ and the three bottom plots to $h = \frac{1}{8}$.

though (5.12) is better for high-order methods, on the face of it this result should cause worry. However, brief examination of Figures 7 and 8 demonstrates that, as a matter of fact, the behaviour of the global error is substantially better than predicted by (5.12).

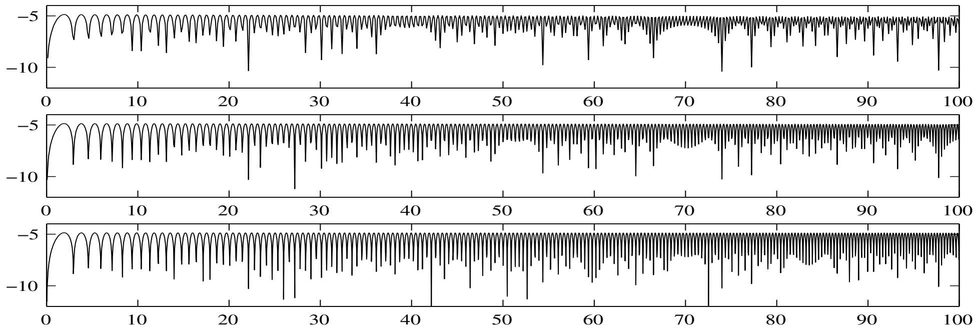


Figure 9: The quantity $\log|y_{t/h} - y(t)| - \frac{1}{4}\log t - 4\log h$ in the solution of the Airy equation (3.8) by fourth-order modified Magnus expansion with exact integrals at the beginning of the integration interval for $h = \frac{1}{8}, \frac{1}{16}$ and $\frac{1}{32}$.

In Figure 7 we display for $h = \frac{1}{8}, \frac{1}{16}$ and $\frac{1}{32}$ the quantities

$$\begin{aligned}
&y_{t/h} - y(t),\\
&\log|y_{t/h} - y(t)| - \tfrac{1}{4}\log t - 4\log h,\\
&\log|y_{t/h} - y(t)| + \tfrac{3}{4}\log t - 2\log h
\end{aligned}$$

for $t = hm$ in the range $[0, 8000]$. Quite evidently, the envelope decays: the best fit is obtained by $t^{-3/4}$, faster than the decay of the exact solution, although, as h decreases, the estimate (5.12) becomes somewhat less misconceived. Note further for future reference that the 'too good to be true' decay of the global-error envelope is accompanied by a decrease in order.

Similarly, in Figure 8 we consider the seventh-order method for $h = 2, \frac{1}{2}, \frac{1}{8}$, displaying

$$\begin{aligned}
&y_{t/h} - y(t),\\
&\log|y_{t/h} - y(t)| - \tfrac{1}{4}\log t - 7\log h,\\
&\log|y_{t/h} - y(t)| + \tfrac{1}{4}\log t - 7\log h
\end{aligned}$$

– again, the second guess, $T^{-1/4}$, seems to provide a better fit with the true decay of the envelope, except for $h = \frac{1}{8}$, where the error is clearly contaminated by roundoff and close to machine accuracy.

Before debating further the seeming contradiction between (5.12) and the two figures, we observe that the precision of seventh-order modified Magnus expansion is truly remarkable: a step of $h = \frac{1}{8}$ yields balance between truncation and roundoff errors in $[0, 8000]$ and even the very crude step $h = 2$ yields two digits of accuracy: toward the end of the range, the function oscillates ≈ 28 times between two grid points, but this does not impede accuracy!

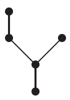
The 'contradiction' between (5.12) and our results is resolved in Figure 9, which zooms on the beginning of the integration interval. The results match (5.12) perfectly!

The reason is as follows: in our analysis we fix $T > 0$ (which might be large) and let $h > 0$ become sufficiently small for the asymptotic behaviour to take place. Provided that $T > 0$ is small enough (yet sufficiently large for the WKB analysis to be valid, although this caveat is not apparent in Figure 9), values of h in the figure fall within this regime. However, for larger T we require values of h which are too small to avoid contamination by roundoff error using standard IEEE computer arithmetic.[4]

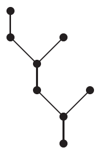
To understand the behaviour for large T we need different approach. Thus, we fix small $h > 0$ and let $T \gg 1$ grow. Our proof focuses on the Airy equation (3.8), but is easy to generalise to linear oscillators (3.4). Moreover, we focus just on the first few modified Magnus methods. The main idea is to examine the size of $\mathcal{C}_\tau(t)$ not just with regard to powers of h, but also of the 'local amplitude' $\nu = h[g(t_{n+1/2})]^{1/2} = h(t_n + \frac{1}{2}h)^{1/2}$.

The lowest-power terms in *both* h and ν^{-1} for trees in $\mathbb{T}_k$, $k = 0, 1, 2, 3$, can be obtained with a symbolic algebra package. Here $C = \cos 2\nu$ and $S = \sin 2\nu$.

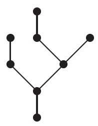
$$: \quad \tfrac{1}{8}h^2 \begin{bmatrix} -h\nu^{-2}(1+C) & -h^2\nu^{-3}S \\ -\nu^{-1}S & h\nu^{-2}(1+C) \end{bmatrix},$$



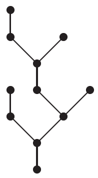
$$: \quad \tfrac{1}{48}h^5 \begin{bmatrix} 3h\nu^{-4}(1-C) & h^2\nu^{-4} \\ -\nu^{-2} & -3h\nu^{-4}(1-C) \end{bmatrix},$$



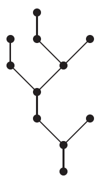
$$: \quad \tfrac{1}{384}h^8 \begin{bmatrix} -2h\nu^{-5}S & h^2\nu^{-6}(4C-3) \\ \nu^{-4}(4C-3) & 2h\nu^{-5}S \end{bmatrix}$$



$$: \quad \tfrac{1}{256}h^8 \begin{bmatrix} -h\nu^{-6}(9-8C+C^2) & -2h^2\nu^{-6} \\ -2\nu^{-4} & h\nu^{-6}(9-8C+C^2) \end{bmatrix},$$

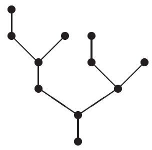


$$: \quad \tfrac{1}{1536}h^{11} \begin{bmatrix} -2h\nu^{-7}S & 5h^2\nu^{-8} \\ \nu^{-6}(4C-1) & 2h\nu^{-7}S \end{bmatrix},$$

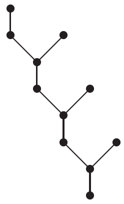


$$: \quad \tfrac{1}{26880}h^{11} \begin{bmatrix} 7h\nu^{-6} & -2h^2\nu^{-6} \\ 2\nu^{-4} & -7h\nu^{-6} \end{bmatrix},$$

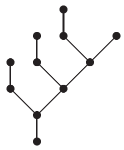
[4] An attempt to reproduce Figure 9 for seventh-order modified Magnus failed, because of contamination by roundoff error.



$$: \quad \frac{1}{26880} h^{11} \begin{bmatrix} -7h\nu^{-6} & 2h^2\nu^{-6} \\ -2\nu^{-4} & 7h\nu^{-6} \end{bmatrix},$$



$$: \quad \frac{1}{1536} h^{11} \begin{bmatrix} -2h\nu^{-7}S & h^2\nu^{-8}(6+C) \\ \nu^{-6}(3C-3) & 2h\nu^{-7}S \end{bmatrix},$$



$$: \quad \frac{1}{26880} h^{11} \begin{bmatrix} 7h\nu^{-6} & -2h^2\nu^{-6} \\ 2\nu^{-4} & -7h\nu^{-6} \end{bmatrix}.$$

We thus observe that for an order-$(3r+1)$ method the local error is

$$\boldsymbol{\ell}(\boldsymbol{y}_n) \approx \tilde{c}_3 h^{3r+2} \frac{1}{\nu_n^{\delta(r)}} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

where $\delta(1) = 2$, $\delta(2) = \delta(3) = 4$. Since $\nu_n = h[g(t_n)]^{1/2}$, we deduce, in place of the global-error formula (5.11), the estimate

$$\begin{aligned} \boldsymbol{E}(T) \approx h^{3r+1} & \begin{bmatrix} T^{-1/4} & 0 \\ 0 & T^{1/4} \end{bmatrix} G(T) \Big\{ c^* \\ & + h^{-\delta(r)} \int_{T^*}^{T} \frac{1}{x^{(\delta(r)+1)/2}} \begin{bmatrix} -\sin 2\theta(x) & -1+\cos 2\theta(x) \\ 1+\cos 2\theta(x) & \sin 2\theta(x) \end{bmatrix} \mathrm{d}x \Big\} \boldsymbol{v}, \end{aligned} \qquad T^* \gg 1. \tag{5.13}$$

Therefore, although the "$h \to 0$, $T \gg 1$ fixed" asymptotics of the Airy equation, as given by (5.12), remain perfectly valid, the "$T \to \infty$, $0 < h \ll 1$ fixed" asymptotic behaviour is vastly different and we obtain from (5.13) the estimate

$$\boldsymbol{E}(T) \approx \boldsymbol{E}(T^*) + c h^{3r+1-\delta(r)} T^{-\delta(r)/2+1/4} \chi(T), \tag{5.14}$$

where χ oscillates in $[-1, 1]$. This is completely consistent with Figure 7: the attenuation of global-error envelope indeed scales up as $T^{-3/4}$ for large T (much faster than the envelope of the exact solution!), but this is accompanied by a drop in order. Results are more equivocal for the order-seven method from Fig. 8, but this can be expected: values of $h, T^{-1} > 0$ small enough to render asymptotics valid run foul of roundoff error, because of the very high precision of the method.

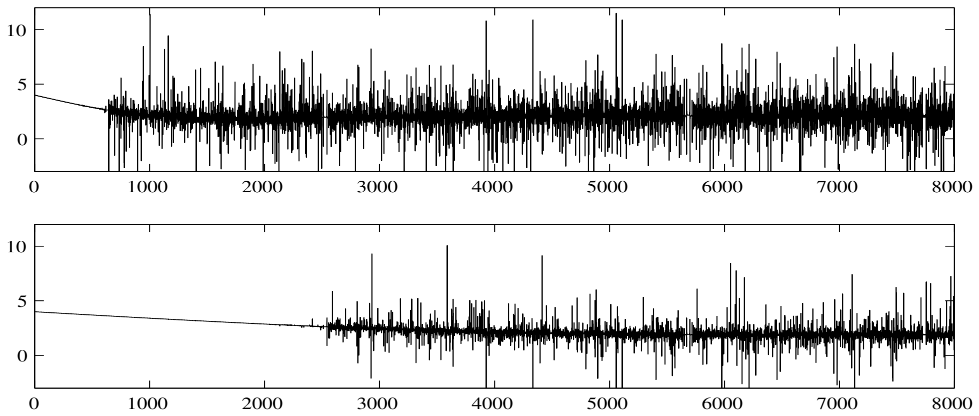


Figure 10: The 'effective order' $p_h(T)$ of the fourth-order modified Magnus method for the Airy equation with $h = \frac{1}{8}$ and $h = \frac{1}{16}$.

The order reduction is genuine. Let $\boldsymbol{E}_h(T)$ be the global error for constant step-size $h > 0$ and assume that $\boldsymbol{E}_h(T) \approx ch^p$. Then

$$p \approx p_h(T) = \frac{\log|\boldsymbol{E}_h(T)| - \log|\boldsymbol{E}_{h/2}(T)|}{\log 2}.$$

Figure 10 displays $p_h(T)$ for fourth-order modified Magnus for two values of h. The function p_h starts at four, decreasing progressively (and linearly). After a while it settles down, oscillating about the mean value of two, as predicted by (5.14). Arguably, for these values the concept of order is meaningless: the error is attenuated because of the smoothing effect of integration, being divided by a power of ν.

6 Conclusions and pointers for further research

This paper hinges upon three developments. Firstly, in Section 2 we have proved and reinterpreted the global-error formula (2.6), which has been originally discussed by Henrici (1963) in a different form. Secondly, in Sections 3 and 4 we have used WKB asymptotic estimates to investigate the accumulation of global error for Runge–Kutta and Magnus methods, respectively, for the model linear oscillator $y'' + g(t)y = 0$. This has motivated the introduction of the modified Magnus method in Section 5.

Each of the three developments begs legitimate questions, which often might lead to further research and improved understanding of the issue at hand.

How valid is the global-error estimate (2.6)? It represents the leading term in an expansion, hence must be correct in an asymptotic sense, when $h \downarrow 0$ and $T > 0$ is fixed. Yet, what is the influence of higher-order terms? What happens when $h > 0$ is small and $T \to \infty$? We have already seen in Section 5 that the latter question might have an unexpected answer, which sheds light on long-term error accumulation, but clearly much more remains to be done. Perhaps most importantly, the question of error accumulation with constant step size is mostly irrelevant to the understanding of contemporary software for numerical ordinary differential equations. Such software usually employs variable step size and error-control strategies. (An exception are Hamiltonian systems, when conservation of symplecticity motivates constant-step strategy (Sanz–Serna & Calvo 1994).) Can (2.6) be generalised to cater for variable steps?

In a very interesting recent paper, Viswanath (2000) has used an approach vaguely similar to (2.6), coupled with Lyapunov's stability theory, to investigate global-error accumulation for a variety of hyperbolic dynamical systems. Clearly, there is room for further research on the overlap of Viswanath's and this paper's ideas.

Equally ripe for further discussion and generalisation is our asymptotic analysis. For a general nonlinear system (2.1) we need to work out the asymptotics of both the exact solution and the (linear) variational flow. The task is much simplified for linear systems and, insofar as the linear oscillator (3.1) is concerned, the WKB technique is a suitable tool. It might be interesting, though, to employ other asymptotic techniques to a wider range of problems, linear as well as nonlinear.

The modified Magnus method (5.5) can be extended to the realm of nonlinear equations by adopting an approach pioneered by Zanna (1999) in the context of standard Magnus expansions. We commence from the highly-oscillatory system

$$\boldsymbol{y}' = A(t,\boldsymbol{y})\boldsymbol{y}, \quad t \geq 0, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0. \tag{6.1}$$

We next let

$$\boldsymbol{y}(t) = \mathrm{e}^{(t-t_n)A(t_n,\boldsymbol{y}_n)}\boldsymbol{x}(t), \qquad t \geq t_n, \tag{6.2}$$

whence

$$\boldsymbol{x}' = F^{-1}(t)[A(t,F(t)\boldsymbol{x}) - A(t_n,\boldsymbol{y}_n)]F(t)\boldsymbol{x}, \quad t \geq 0, \qquad \boldsymbol{x}(0) = \boldsymbol{y}_0, \tag{6.3}$$

where $F(t) = \exp[(t-t_n)A(t_n,\boldsymbol{y}_n)]$. Note that, even if $A(t,\boldsymbol{y}) = A(t)$ and (6.1) is linear, the method (6.2) does not reduce to modified Magnus from Section 5 since we are linearising about t_n, rather than $t_{n+1/2}$.

(6.3) is a nonlinear equation, which we solve with a modification of a second-order method (which, in our context, becomes fourth-order) from (Zanna 1999, Iserles et al. 2000). We let

$$\begin{aligned}
c_1 &= (\tfrac{1}{2} - \tfrac{\sqrt{3}}{6})h, & c_2 &= (\tfrac{1}{2} + \tfrac{\sqrt{3}}{6})h,\\
a_{1,1} &= \tfrac{1}{4}h, & a_{1,2} &= (\tfrac{1}{4} - \tfrac{\sqrt{3}}{6})h,\\
a_{2,1} &= (\tfrac{1}{4} + \tfrac{\sqrt{3}}{6})h, & a_{2,2} &= \tfrac{1}{4}h,\\
\tilde{a}_1 &= (\tfrac{5}{144} - \tfrac{\sqrt{3}}{48})h^2, & \tilde{a}_2 &= -(\tfrac{5}{144} + \tfrac{\sqrt{3}}{48})h^2\\
\hat{a} &= -\tfrac{\sqrt{3}}{12}h^2
\end{aligned}$$

and

$$A_0 = A(t_n,\boldsymbol{y}_n), \qquad F_k = \mathrm{e}^{c_k A_0}, \quad k = 1,2, \qquad \boldsymbol{x}_1 = \boldsymbol{x}_2 = \boldsymbol{y}_0,$$

whence a single iteration is

$$\begin{aligned}
A_k &= A(t_n + hc_k, F_k \boldsymbol{x}_k), \quad U_k = F_k^{-1}(A_k - A_0)F_k, \qquad k = 1,2,\\
U_{1,2} &= [U_1, U_2],\\
G_k &= a_{k,1}U_1 + a_{k,2}U_2 + \tilde{a}_k U_{1,2}, \quad \boldsymbol{x}_k = \mathrm{e}^{G_k}, \qquad k = 1,2,
\end{aligned}$$

Upon convergence, we let

$$\boldsymbol{y}_{n+1} = \mathrm{e}^{hA_0}\mathrm{e}^{\frac{1}{2}h(U_1+U_2)+\hat{a}U_{1,2}}\boldsymbol{y}_n.$$

Note that, strictly speaking, we may remove $U_{1,2}$ altogether from the algorithm without any damage to formal order. Yet, the incorporation of this term adds only marginally to the cost while, as numerical experimentation affirms, improving accuracy.

It is far from clear whether this is the most suitable form of 'nonlinear modified Magnus': our present goal is simply to demonstrate that modified Magnus can be extended to nonlinear equations (6.1). However, any such an extension is likely to be prohibitively expensive for large systems. In that case it might well be advantageous to split the vector field into a linear oscillatory part and a nonoscillatory remainder, in a manner familiar from the solution of Hamiltonian systems (Sanz–Serna & Calvo 1994), and use 'linear modified Magnus' (5.5) for the highly oscillating part.

Acknowledgements

The author wishes to acknowledge numerous discussions with colleagues, in particular with Sergio Blanes, Chris Budd, John Hinch, David Levin, Per Christian Moan, Jitse Niesen, Syvert Nørsett and Divakar Viswanath. The research leading to this paper commenced during the author's visit to Norwegian University of Science and Technology, Trondheim, as the Onsager Professor, and he wishes to thank the Lars Onsager Fund for its assistance.

References

Blanes, S., Casas, F. & Ros, J. (1999), Improved high order integrators based on Magnus expansion, Technical Report NA1999/08, DAMTP, University of Cambridge.

Davis, P. J. & Rabinowitz, P. (1984), *Methods of Numerical Integration*, 2nd edn, Academic Press, Orlando, FL.

Hairer, E. & Lubich, C. (1984), 'Asymptotic expansions of the global error of fixed-stepsize methods', *Numer. Math.* **45**, 345–360.

Hairer, E. & Lubich, C. (1997), 'The life-span of backward error analysis for numerical integrators', *Numer. Math.* **76**, 441–462.

Hairer, E., Nørsett, S. P. & Wanner, G. (1987), *Solving Ordinary Differential Equations I: Nonstiff Problems*, Springer-Verlag, Berlin.

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

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

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

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

Iserles, A., Marthinsen, A. & Nørsett, S. P. (1999), 'On the implementation of the method of Magnus series for linear differential equations', *BIT* **39**, 281–304.

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

Iserles, A., Nørsett, S. P. & Rasmussen, A. F. (1998), Time-symmetry and high-order Magnus methods, Technical Report 1998/NA06, DAMTP, University of Cambridge.

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

Munthe-Kaas, H. & Owren, B. (1999), 'Computations in a free Lie algebra', *Phil. Trans Royal Society A* **357**, 957–982.

Reich, S. (1999), 'Backward error analysis for numerical integrators', *SIAM J. Num. Ana.* **36**, 1549–1570.

Sanz–Serna, J. & Calvo, M. (1994), *Numerical Hamiltonian Problems*, Chapman & Hall, London.

Viswanath, D. (2000), 'Global errors of numerical ODE solvers and Lyapunov's theory of stability', *IMA J. Num. Anal.* to appear.

Zanna, A. (1999), 'Collocation and relaxed collocation for the Fer and the Magnus expansions', *SIAM J. Num. Anal.* **36**, 1145–1182.