Expansions that grow on trees

Arieh Iserles*

1 Linear ordinary differential equations

How to solve linear ordinary differential equations? Like many outstanding mysteries of mathematics, this question has the virtue of simplicity. Linear differential equations are the staple of every mathematical syllabus, familiar to all and sundry, and their investigation has informed much of the development of mathematical analysis in the last three hundred years. Needless to say, we can all provide *partial* answers to this question. Thus, the solution of the scalar equation $y' = a(t)y$, $y(0) = y_0$, is

$$y(t) = \mathrm{e}^{\int_0^t a(\xi)\mathrm{d}\xi} y_0, \quad t \geq 0,$$

while the solution of the vector equation $\boldsymbol{y}' = A\boldsymbol{y}$, $\boldsymbol{y}(0) = \boldsymbol{y}_0 \in \mathbb{R}^N$, where A is a *constant* $N \times N$ matrix, is

$$\boldsymbol{y}(t) = \mathrm{e}^{tA}\boldsymbol{y}_0, \quad t \geq 0.$$

The matrix exponential above is defined by Taylor expansion,

$$\mathrm{e}^{\Omega} = \sum_{m=0}^{\infty} \frac{1}{m!}\Omega^m.$$

It is natural, as the next step in our investigation, to attempt 'mental interpolation' between the last two explicit formulae and check whether the solution of the vector system with variable coefficients, $\boldsymbol{y}' = A(t)\boldsymbol{y}$, $\boldsymbol{y}(0) = \boldsymbol{y}_0$, might perhaps be

$$\boldsymbol{y}(t) = \mathrm{e}^{\int_0^t A(\xi)\mathrm{d}\xi}\boldsymbol{y}_0, \quad t \geq 0, \tag{1}$$

where the integral of the matrix function A is the matrix whose entries are the integrals of the components of A. Unfortunately, this mental leap is in general false.

The general form of $\boldsymbol{y}$, and, indeed, of the fundamental solution Y, where

$$Y' = A(t)Y, \quad t \geq 0, \qquad Y(0) = I, \tag{2}$$

is considerably more intriguing. Its exploration will take us on a tour spanning a hundred years of mathematics and ranging from differential geometry, Lie-algebra theory, and graph theory, all the way to contemporary numerical analysis. The basic idea is to represent the solution of (2) in the form $Y(t) = \mathrm{e}^{\Omega}$ and to rephrase the differential equation in terms of the *logarithm* of Y: the matrix function Ω. The equation for Ω is originally due to Felix Hausdorff [10], and is given by

$$\Omega' = \sum_{m=0}^{\infty} \frac{\mathrm{B}_m}{m!}\mathrm{ad}_{\Omega}^m A, \quad t \geq 0, \qquad \Omega(0) = O. \tag{3}$$

Here $\{\mathrm{B}_m\}_{m\in\mathbb{Z}_+}$ are the *Bernoulli numbers,* defined by

$$\sum_{m=0}^{\infty} \frac{\mathrm{B}_m}{m!} z^m = \frac{z}{\mathrm{e}^z - 1} = 1 - \tfrac{1}{2}z + \tfrac{1}{12}z^2 - \tfrac{1}{720}z^4 + \cdots,$$

while ad_{Ω} is a shorthand for an iterated commutator,

$$\mathrm{ad}_{\Omega}^0 A = A, \quad \mathrm{ad}_{\Omega}^1 A = [\Omega, A], \quad \mathrm{ad}_{\Omega}^2 A = [\Omega, [\Omega, A]]$$

and, in general, $\mathrm{ad}_{\Omega}^{m+1} A = [\Omega, \mathrm{ad}_{\Omega}^m A]$, where $[B, C] = BC - CB$.

On the face of it, we have just taken a large step in rendering the simple equation (2) in a more complex, indeed obscure, manner, as an infinite series of nonlinear terms. Yet, as so often in mathematics, this

*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, United Kingdom, email `ai@damtp.cam.ac.uk`.

is just the price of an ultimate simplification, new insight and superior computational algorithms.

We commence by noting that if $A(t)$ and $\Omega(t)$ commute (which, as we will see soon, is the case whenever $A(t)$ and $\int_0^t A(\xi)\mathrm{d}\xi$ commute), all the terms on the right of (3), except for the first, vanish. We have $\Omega' = A$ and the solution is (1). In other words, the $m \geq 1$ terms and the nonlinearity of (3) are the cost of noncommutativity.

Wilhelm Magnus

The main insight into the equation (3) was provided in a remarkable paper by a remarkable mathematician, Wilhelm Magnus. A topologist, an algebraist, an authority on differential equations and on special functions, a mathematical physicist, Magnus exercised great (and often not sufficiently acknowledged) influence on twentieth century mathematics [1]. Magnus observed in [17] that Ω can be written as a linear combination of multiple integrals. Specifically, employing Picard's iteration,

$$\begin{aligned}\Omega_0(t) &\equiv O,\\ \Omega_{s+1}(t) &= \int_0^t \sum_{m=0}^{\infty} \frac{\mathrm{B}_m}{m!} \mathrm{ad}^m_{\Omega_s(\xi)} A(\xi)\mathrm{d}\xi, \qquad s \in \mathbb{Z}_+,\end{aligned}$$

it is possible to show that

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

Next there are four fourfold integrals, each with three nested commutators, and in general the number of terms in each 'generation' grows exponentially. So does the complexity, which rapidly gets out of hand. As a matter of fact, the situation is even worse, since naive Picard iteration generates a multitude of higher-order terms that need to be excised in each step. Nonetheless, the above *Magnus expansion* became an important tool in numerous application areas, in particular in quantum chemistry and theoretical physics. Many attempts have been made to derive the expansion in explicit form, the most remarkable due to Fomenko and Chakon [8]. This, however, did not eliminate the obvious combinatorial complexity inherent in dealing with multiple integrals and 'multilayered' iterated commutators and, arguably, did not lead to a viable means to generate recursively and analyse the function Ω.

2 On expansions that grow on trees

Several years ago, Syvert Nørsett and I came across the Magnus expansion in the context of *geometric integration,* the emerging discipline concerned with discretising differential equations while respecting their qualitative properties [5]. Blissfully unaware of the substantial research to which Magnus's creation has been subjected, we have developed an alternative approach, using *rooted trees* as a shorthand for expansion terms. This has led to a framework that elucidates the structure of individual terms and their relationship, thus allowing for convenient construction of the expansion (4) by recursion [14].

Examine (4) for a moment: each term is made out of integrals and commutators of the matrix function A. Specifically, each term on the right, except for the first, is of the form

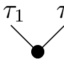
$$\alpha \int_0^t [H_1(\xi), H_2'(\xi)]\mathrm{d}\xi, \tag{5}$$

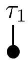
where both H_1 and H_2 have already featured earlier in the expansion, while α is a scalar constant. We thus require terminology to denote repeated integration and commutation in a clear, transparent fashion.

We commence by assigning to A a trivial tree, consisting of single vertex,

$$\bullet \rightsquigarrow A.$$

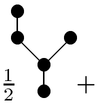
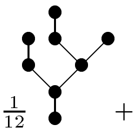
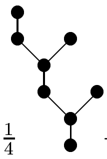
Moreover, suppose that we have already constructed two expansion terms, H_{τ_1} and H_{τ_2}, say, such that $\tau_1 \rightsquigarrow H_{\tau_1}$ and $\tau_2 \rightsquigarrow H_{\tau_2}$. Then



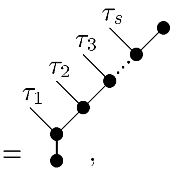
$\rightsquigarrow [H_{\tau_1}, H_{\tau_2}],$ 

$\rightsquigarrow \int_0^t H_{\tau_1}(\xi)\mathrm{d}\xi.$

Given an $N \times N$ matrix function A with integrable entries, and commencing from $\bullet$, we have just defined a map $\tau \to H_\tau$ from $\mathbb{F}$, a subset of binary rooted trees, into $N \times N$ matrix functions. In particular, (4) can be denoted in the shorthand

 $-\frac{1}{2}$  $+\frac{1}{12}$  $+\frac{1}{4}$  $+\cdots$. (6)

Each tree in (6), except for the first, is of the form on the right side of (7),

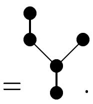
$\tau =$ 

, (7)

where $s \in \mathbb{Z}_+$ and the trees $\tau_1, \tau_2, \ldots, \tau_s$ have already featured earlier in the expansion. For example, in the third tree in (6)

$s = 2, \qquad \tau_1 = \tau_2 =$ ,

while the fourth tree is consistent with

$s = 1, \qquad \tau_1 =$ .

It is possible to deduce from (5) that the form (7) remains valid for subsequent expansion terms, indeed that it characterises all the trees in $\mathbb{F}$. With a little more effort it is also possible to deduce from (7) the explicit form of the constant α in (5) [14]. Set

$\alpha \Big($  $\Big) = 1$

and suppose that $\alpha(\tau_l)$ is known for $l = 1, 2, \ldots, s$. Then

$$\alpha(\tau) = \frac{\mathrm{B}_s}{s!} \prod_{l=1}^{s} \alpha(\tau_l), \qquad s \in \mathbb{N}, \tag{8}$$

where B_s is the sth Bernoulli number.

We have now all the ingredients to describe a general recursive rule for the generation of a Magnus expansion, except that, in our quest for tidiness, we are reluctant to lump all expansion coefficients into a hotchpotch of unclassified terms. In our earlier discussion of (4) we have aggregated terms according to the number of integrals therein: in the shorthand of rooted trees, this corresponds to the number of 'vertical' edges. This, however, is not the best way to display the expansion in the most economical fashion, without including terms which might be redundant once the expansion has been truncated for practical purposes. Instead, we say that a tree τ is of *power* $r \in \mathbb{N}$ if r is the greatest integer such that

$$H_\tau(t) = \mathcal{O}(t^r), \qquad t \to 0,$$

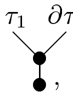
for all sufficiently smooth matrix functions A. Let $\mathbb{F}_r$ be the set of all trees of power r. The Magnus expansion can be written in the form

$$\Omega(t) = \sum_{r=1}^{\infty} \sum_{\tau \in \mathbb{F}_r} \alpha(\tau) H_\tau(t) \tag{9}$$

according to the following rules:

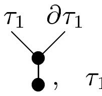
1. Let $\mathbb{F}_1 = \Big\{$  $\Big\}$ and $\alpha \Big($  $\Big) = 1$;

2. Given $r \geq 2$, assemble $\mathbb{F}_r$ by taking all the trees of the form



$$, \quad \tau_1 \in \mathbb{F}_{k_1}, \tau_2 \in \mathbb{F}_{k_2}, \; k_1+k_2 = r-1, \; \tau_1 \neq \tau_2,$$

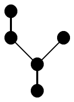
where $\partial\tau$ stands for the tree τ with its root excised, i.e. $\partial\tau \rightsquigarrow \mathrm{d}H_\tau/\mathrm{d}t$. Moreover, if r is even then add to $\mathbb{F}_r$ all trees of the form



$$, \quad \tau_1 \in \mathbb{F}_{k_1}, \; k_1 = r/2 - 1.$$

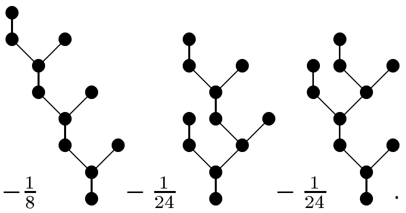
3. Identify each tree τ with the representation (7) and use (8) to evaluate its coefficient $\alpha(\tau)$.

Note that $\mathbb{F}_2 = \emptyset$, since (as follows from the algorithm above but can be easily verified directly) the power of



is three.

All trees of power up to four are displayed in (6). The next contribution to Ω, corresponding to trees in $\mathbb{F}_5$, is



$$-\tfrac{1}{8} \quad - \tfrac{1}{24} \quad - \tfrac{1}{24} \quad . \qquad (10)$$

As we go along, the number of terms in $\mathbb{F}_r$ increases exponentially: it was proved in [15] that $\limsup_{r\to\infty}[\#\mathbb{F}_r]^{1/r} = 3.1167\ldots$. In comparison, the number of terms with r integrals increases as $c(r)4^r$, where c is a slowly-varying function [14]. Even for moderate r the difference is highly significant.

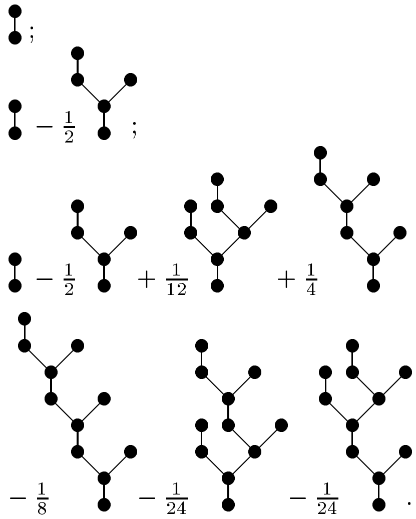
By construction, once we truncate (9),

$$\Omega^{[p]}(t) = \sum_{r=1}^{p} \sum_{\tau\in\mathbb{F}_r} \alpha(\tau) H_\tau(t),$$

we can expect $\Omega^{[p]}(t) = \Omega(t) + \mathcal{O}\big(t^{p+1}\big)$, an order-$p$ approximation. However, the situation is better than this!

An operator Ψ_t, which for every t maps (for example) $N \times N$ matrix functions to themselves, is said to be *time symmetric* if $\Psi_{-t} \circ \Psi_t = \mathrm{Id}$. In particular, every solution operator of a differential equation, $\Psi_h Y(t) = Y(t+h)$, is time symmetric, but this is not necessarily the case with approximations to the solution – and the truncation $\Omega^{[p]}(t)$ can be envisaged as a numerical method to compute Ω with step size t. (Much more later about Magnus expansions as numerical methods.) Yet, it is possible to prove that the map $\Psi_t X(t) = \mathrm{e}^{\Omega^{[p]}(t)} X(t)$ is time symmetric [15].

Time symmetry is important because it implies that for odd p, the tail $\Omega(t) - \Omega^{[p]}(t)$ is $\mathcal{O}\big(t^{p+2}\big)$ and we gain an extra unit of order [15]. The first few such truncated Magnus expansions are



order 2: ;

order 4: $\; - \tfrac{1}{2} \;$;

order 6: $\; - \tfrac{1}{2} \; + \tfrac{1}{12} \; + \tfrac{1}{4} \; - \tfrac{1}{8} \; - \tfrac{1}{24} \; - \tfrac{1}{24} \;$.

As a simple example, which the reader (perhaps with the help of a symbolic algebra program) is encouraged to verify, let us consider the *Airy equation*

$$y'' + ty = 0, \quad t \geq 0, \qquad y(0) = 1, \;\; y'(0) = \tfrac{1}{2}, \quad (11)$$

which we rewrite in a vector form as

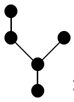
$$\boldsymbol{y}' = \begin{bmatrix} 0 & 1 \\ -t & 0 \end{bmatrix} \boldsymbol{y}, \quad t \geq 0, \qquad \boldsymbol{y}(0) = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}.$$

Although the exact solution of (11) can be written explicitly in terms of Airy functions, it is more illustrative to express it as a series,

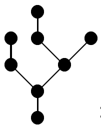
$$y = 1 + \tfrac{1}{2}t - \tfrac{1}{6}t^3 - \tfrac{1}{24}t^4 + \tfrac{1}{180}t^6 + \tfrac{1}{1008}t^7 + \mathcal{O}\big(t^9\big)\,.$$

The leading Magnus trees and corresponding matrix functions are

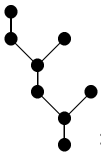
$$: \quad \begin{bmatrix} 0 & t \\ -\frac{1}{2}t^2 & 0 \end{bmatrix};$$



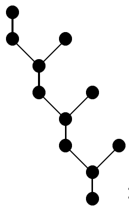
$$: \quad \begin{bmatrix} -\frac{1}{6}t^3 & 0 \\ 0 & \frac{1}{6}t^3 \end{bmatrix};$$



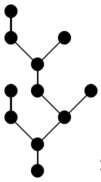
$$: \quad \begin{bmatrix} 0 & \frac{1}{4}t^4 \\ \frac{1}{10}t^5 & 0 \end{bmatrix};$$



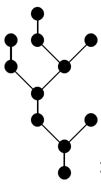
$$: \quad \begin{bmatrix} 0 & -\frac{1}{12}t^4 \\ -\frac{1}{15}t^5 & 0 \end{bmatrix};$$



$$: \quad \begin{bmatrix} \frac{1}{40}t^6 & 0 \\ 0 & -\frac{1}{40}t^6 \end{bmatrix};$$



$$: \quad \begin{bmatrix} -\frac{1}{12}t^6 & 0 \\ 0 & \frac{1}{12}t^6 \end{bmatrix};$$



$$: \quad \begin{bmatrix} -\frac{7}{120}t^6 & 0 \\ 0 & \frac{7}{120}t^6 \end{bmatrix}.$$

We thus deduce that

$$\Omega^{[1]} = \begin{bmatrix} 0 & t \\ -\frac{1}{2}t^2 & 0 \end{bmatrix};$$

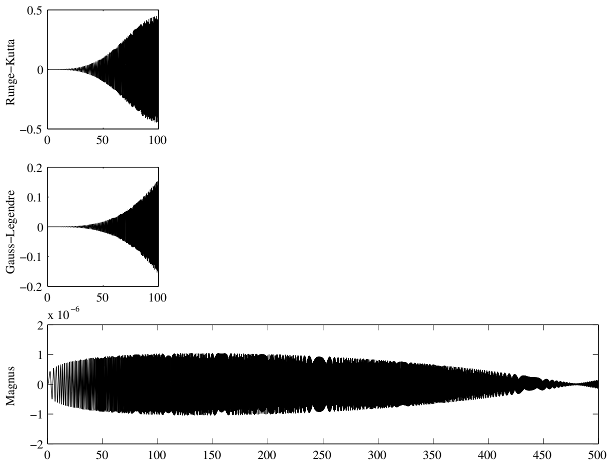


Figure 1: The error in the solution of the Airy equation by three numerical methods.

$$\begin{aligned} \Omega^{[3]} &= \begin{bmatrix} \frac{1}{12}t^3 & t \\ -\frac{1}{2}t^2 & -\frac{1}{12}t^3 \end{bmatrix}; \\ \Omega^{[5]} &= \begin{bmatrix} \frac{1}{12}t^3 + \frac{1}{360}t^6 & t \\ -\frac{1}{2}t^2 - \frac{1}{120}t^5 & -\frac{1}{12}t^3 - \frac{1}{360}t^6 \end{bmatrix}. \end{aligned}$$

Letting $\boldsymbol{x}_k = \mathrm{e}^{\Omega^{[k]}}\boldsymbol{y}(0)$, we obtain

$$\begin{aligned} \boldsymbol{x}_1 &= 1 + \tfrac{1}{2}t^2 - \tfrac{1}{4}t^3 + \mathcal{O}\big(t^4\big)\,, \\ \boldsymbol{x}_3 &= 1 + \tfrac{1}{2}t^2 - \tfrac{1}{6}t^3 - \tfrac{1}{24}t^4 + \tfrac{1}{144}t^6 + \mathcal{O}\big(t^7\big)\,, \\ \boldsymbol{x}_5 &= 1 + \tfrac{1}{2}t^2 - \tfrac{1}{6}t^3 - \tfrac{1}{24}t^4 + \tfrac{1}{180}t^6 + \tfrac{1}{1080}t^7 + \mathcal{O}\big(t^9\big)\,. \end{aligned}$$

Comparison with the Taylor expansion of y confirms that this is consistent with order $k+1$.

Figure 1 displays the pointwise error in the solution of the Airy equation (11) by three fourth-order methods: the classical Runge–Kutta method, the (implicit, A-stable) Gauss–Legendre Runge–Kutta method and a truncation of the Magnus expansion, all time-stepped with constant step size of $\frac{1}{10}$. The exact solution of (11) oscillates with increasing frequency and it behaves asymptotically like $t^{-1/4}\sin\frac{2}{3}t^{3/2}$ as $t \to \infty$. This is sufficient to defeat even the hardiest 'classical' numerical schemes, while Magnus, less impressed by rapid oscillation, delivers significantly superior performance: Note that the vertical scale on the bottom diagram in Figure 1 is six orders of magnitude smaller than in the other two diagrams.

For the record, to obtain similar accuracy of six significant digits in $[0, 500]$ with a Runge–Kutta method would have required a step size of $\approx \frac{1}{1200}$, at a relative cost (in computer time) of more than sixty times that of the Magnus method. More sophisticated methods need not be necessarily more expensive!

3 Lie groups and geometric integration

Like Molière's Monsieur Jourdain, who had been speaking prose all his life and didn't know it, we have just been, perhaps unknowingly, speaking the language of Lie groups.

Since their introduction by Sophus Lie in the late Nineteenth Century, Lie groups have been among our most powerful tools for understanding structure, symmetry and geometry, in particular within the context of differential equations [22].

A *Lie group* $\mathcal{G}$ is a manifold endowed with group structure. Being a manifold, it can be covered by an atlas of overlapping local charts, which are homeomorphic to open subsets of Euclidean space. Being a group, $\mathcal{G}$ is closed under inversion and multiplication. Moreover, we stipulate that these two operations are continuous with respect to the underlying topology of the manifold. In case this sounds mysterious and abstract, let us consider several examples. Firstly, the *orthogonal group* $\mathrm{O}(N)$ of real $N \times N$ orthogonal matrices. Clearly, it is a group: the identity is an orthogonal matrix and if $Q_1, Q_2 \in \mathrm{O}(N)$ then also $Q_1^{-1} = Q_1^\top, Q_1 Q_2 \in \mathrm{O}(N)$. It is also an $\frac{1}{2}(N-1)N$-dimensional manifold (can you prove it?) which can be embedded in the N^2-dimensional space of all real $N \times N$ matrices, while both matrix multiplication and inversion are smooth operations. Another example is the *special linear group* $\mathrm{SL}(N)$ of real $N \times N$ matrices with unit determinant. Yet another is the *symplectic group* $\mathrm{Sp}(N)$, consisting of $(2N) \times (2N)$ real matrices X such that $XJX^\top = J$, where

$$J = \begin{bmatrix} O & I \\ -I & O \end{bmatrix},$$

with $N \times N$ zero and identity matrices O and I respectively. Finally, consider the set of all *rigid motions* (rotations, reflections and translations) of the space $\mathbb{R}^N$, namely $\boldsymbol{x} \mapsto A\boldsymbol{x} + \boldsymbol{b}$, where $A \in \mathrm{O}(N)$ and $\boldsymbol{b} \in \mathbb{R}^N$. Each rigid motion is thus characterised by the pair $(A, \boldsymbol{b})$, and we can endow such pairs with group structure. The outcome is the *Euclidean group* $\mathrm{E}(N)$, with the operations

$$\begin{aligned} (A_2, \boldsymbol{b}_2) \cdot (A_1, \boldsymbol{b}_1) &= (A_2 A_1, A_2 \boldsymbol{b}_1 + \boldsymbol{b}_2), \\ (A, \boldsymbol{b})^{-1} &= (A^\top, -A^\top \boldsymbol{b}) \end{aligned}$$

(verify!).

All our examples are subgroups of matrices, so we should perhaps comment that Lie groups need be neither multiplicative subgroups of matrices nor, indeed, finite dimensional. Glossing over some technical details, an important example of an infinite-dimensional group is SDiff $\mathcal{M}$, the set of all volume-preserving diffeomorphisms from a manifold $\mathcal{M}$ to itself.

Lie groups feature in multitude of applications. Orthogonal and Euclidean groups are crucial in mechanics, computer vision and numerical algebra, the special linear group in volume conservation, the symplectic group in Hamiltonian mechanics and SDiff $\mathcal{M}$ in geometric theories of fluid flow. Lie groups are increasingly used in diverse areas of applied mathematics where it is important to identify and preserve geometric structure, e.g. in control theory, robotics, Kalman filtering, subdivision schemes in computer graphics, geometric mechanics....

Typically, a mathematical model is phrased as a solution of a differential system. Whenever such a system is known to evolve in a Lie group, this represents valuable information since, more often than not, Lie-group structure embodies qualitative features which are central to the underlying natural phenomenon that we are endeavouring to understand.

Once we consider differential equations evolving on a manifold $\mathcal{M}$, we are concerned with *tangents* to points on $\mathcal{M}$, since these are all the possible values that a vector field may attain. Let $\mathrm{T}\mathcal{M}|_x$ be the linear space of all vectors tangent to $\mathcal{M}$ at some $x \in \mathcal{M}$. It is easy to verify that any differential equation evolving in $\mathcal{M}$ can be written in the form

$$y' = f(t, y), \quad t \geq 0, \qquad y(0) \in \mathcal{M},$$

where the (sufficiently smooth) function $f(t, y)$ maps $\mathbb{R}_+ \times \mathcal{M}$ to $\mathrm{T}\mathcal{M}|_y$. Typically, the structure of the *tangent bundle* $\mathrm{T}\mathcal{M}$ of all tangent spaces $\mathrm{T}\mathcal{M}|_x$, $x \in \mathcal{M}$, is complicated. One of the great virtues of Lie groups is that their tangency relations are relatively transparent and endowed with beautiful features. We first note that the group structure implies that $\mathrm{T}\mathcal{G}|_X = \mathrm{T}\mathcal{G}|_I X$ for all $X \in \mathcal{G}$. Therefore, all ordinary differential equations evolving on $\mathcal{G}$ can be brought to the form

$$Y' = A(t, Y)Y, \quad t \geq 0, \qquad Y(0) \in \mathcal{G}, \tag{12}$$

where the (suitably smooth) function A maps $\mathbb{R}_+ \times \mathcal{G}$ to $\mathfrak{g} = \mathrm{T}\mathcal{G}|_I$. Secondly, the set $\mathfrak{g}$ is not just a humble linear space but a *Lie algebra:* it is closed under a bilinear, skew-symmetric operation $[\![\cdot\,,\cdot]\!]$, say. For every $x_1, x_2, x_3 \in \mathfrak{g}$ and $\alpha_1, \alpha_2 \in \mathbb{R}$ (or, for example, in $\mathbb{C}$, but we do not aim here for the greatest generality) it is true that

$$\begin{aligned} [\![x_1, x_2]\!] + [\![x_2, x_1]\!] &= 0, \\ [\![\alpha_1 x_1 + \alpha_2 x_2, x_3]\!] &= \alpha_1 [\![x_1, x_3]\!] + \alpha_2 [\![x_2, x_3]\!] \end{aligned}$$

and

$$[\![x_1, [\![x_2, x_3]\!]]\!] + [\![x_2, [\![x_3, x_1]\!]]\!] + [\![x_3, [\![x_1, x_2]\!]]\!] = 0$$

(the *Jacobi identity.*) Thirdly, it is possible to construct a smooth map $e : \mathfrak{g} \to \mathcal{G}$ which takes zero to the identity and affords us a way back from the algebra to the group.

Recalling our examples of Lie groups, we just list the corresponding Lie algebras. Thus, the Lie algebra of $\mathrm{O}(N)$ is $\mathfrak{so}(N)$, the linear space of $N \times N$ skew-symmetric matrices; to $\mathrm{SL}(N)$ there corresponds the algebra $\mathfrak{sl}(N)$ of zero-trace matrices; the Lie-algebraic counterpart of the symplectic group $\mathrm{Sp}(N)$ is $\mathfrak{sp}(N)$, such that $X \in \mathfrak{sp}(N)$ implies $XJ + JX^\top = O$ and, finally, the Euclidean algebra $\mathfrak{se}(N)$ comprises of all pairs $(X, \boldsymbol{b})$ such that $X \in \mathfrak{so}(N)$ and $\boldsymbol{b} \in \mathbb{R}^N$.

All this may sound unrelated to linear equations. Yet, a powerful connection becomes apparent once we restrict our attention to finite-dimensional groups and algebras. According to a beautiful theorem of Ado, every finite-dimensional Lie algebra is isomorphic to a Lie subalgebra of matrices [22] and, without loss of generality, we may assume that $\mathfrak{g}$ is indeed a subalgebra of $N \times N$ real matrices. In that case $[\![X_1, X_2]\!] = [X_1, X_2]$ and $e(X) = \mathrm{e}^X$, the familiar commutator and exponential functions. Thus, the entire discourse of the previous section can be rendered in the language of Lie: The matrix function $A(t)$ assumes values in a matrix Lie algebra $\mathfrak{g}$, hence the solution of (2) evolves in 'its' Lie group $\mathcal{G}$. (In the most unrestricted case, $\mathfrak{g} = \mathfrak{gl}(N)$, the Lie algebra of all $N \times N$ matrices, and $\mathcal{G} = \mathrm{GL}(N)$, the *general linear group* of all nonsingular $N \times N$ matrices.) Replacing Y with Ω and equation (2) with (3) corresponds to translating the problem from a Lie group to a Lie algebra. Since the latter is closed with respect to linear combinations, it is also closed with respect to integration, which is a limit of Riemann sums. Therefore, both the Magnus expansion (9) and its truncation $\Omega^{[p]}$ evolve in $\mathfrak{g}$ and the exponential restores the solution to the Lie group.

The great virtue of working in a Lie algebra, rather than a Lie group, becomes clear once we consider approximation, whether numerical or perturbative. General Lie groups are nonlinear creatures and classical numerical methods are often useless when the retention of Lie-group structure is at issue. For example, no familiar discretisation method, e.g. Runge–Kutta or multistep or a truncated Taylor expansion, can be assured to evolve in $\mathrm{SL}(N)$ for $N \geq 3$ [11, 16]. On the other hand, $\mathfrak{g}$ is a linear space: as long as we restrict our arsenal to linear combinations and commutation, we cannot do wrong! Thus, Magnus expansions are a prime example of *geometric integrators,* which are guaranteed to respect the underlying geometry of the differential equation [13].

An obvious alternative to Lie-group methods, like the Magnus expansion, is *projection:* use an arbitrary discretisation method and ensure that geometry is modelled correctly by projecting the numerical solution on $\mathcal{G}$. Although naive projection may run foul of instabilities and rapid error accumulation, a more sophisticated approach can be effective [9].

4 Computing Magnus

In principle, the Magnus expansion presents a powerful means to compute numerically the solution of a linear system (2). However, nothing is ever simple or straightforward with computation. The first stumbling block is that the expansion (9) has only a limited radius of convergence: to converge in norm $\|\cdot\|$ we require

$$\int_0^t \|A(\xi)\|\mathrm{d}\xi \le \int_0^{2\pi} \frac{\mathrm{d}\xi}{4+\xi[1-\cot(\xi/2)]} \approx 1.08687$$

[2, 18]. If t is so large that this inequality is violated then we can use time-stepping with, say, constant step $h > 0$. A more substantial obstacle is that the evaluation of $\Omega^{[p]}$ requires the computation of several multivariate integrals, each over a different polytope, at each time step. Now, it is very easy to approximate a univariate integral but most challenging and labour intensive to do so in a multivariate setting [7].

Before we can run, though, we need to learn how to walk. Or, more specifically, how to evaluate univariate integrals. Even if you are familiar with numerical quadrature, do bear with me for, miraculously, the univariate case extends in our setting to multiple integrals. Let $c_1, c_2, \ldots, c_\nu$ be the zeros of the ν-degree Legendre polynomial, shifted to the interval $[0,1]$. It is well known since Gauss's time that there exist weights $b_1, b_2, \ldots, b_\nu > 0$ such that the *quadrature*

$$\int_0^h A(\xi)\mathrm{d}\xi \approx h\sum_{k=1}^{\nu} b_k A(c_k h) \tag{13}$$

is of order 2ν: for sufficiently-smooth functions A it carries an error of $\mathcal{O}(h^{2\nu+1})$. Moreover, no other choice of quadrature nodes $c_1, c_2, \ldots, c_\nu$ and weights can improve (or, indeed, match) this order [7]. It is helpful to rephrase (13). Thus, let the matrices $B_1, B_2, \ldots, B_\nu$ be the solution of the Vandermonde linear system

$$\sum_{l=1}^{\nu}(c_k - \tfrac{1}{2})^{l-1}B_l = hA(c_k h), \quad k = 1, 2, \ldots, \nu.$$

It is possible to show that $B_l = d_l h^l A^{(l-1)}(\frac{1}{2}h) + \mathcal{O}(h^{l+1})$, where the d_ls are nonzero constants, and that the order-(2ν) quadrature (13) is equivalent to

$$\int_0^h A(\xi)\mathrm{d}\xi \approx \sum_{l=0}^{\lfloor(\nu-1)/2\rfloor} \frac{1}{4^l(2l+1)} B_{2l+1}.$$

We will soon see that, when generalisation to 'Magnus integrals' is at issue, the currency of the B_ls is superior to that of the $A(c_k h)$s.

Taking a leaf from [14], we next consider the double integral from (4),

$$\int_0^h \int_0^{\xi_1} [A(\xi_2), A(\xi_1)]\mathrm{d}\xi_2\mathrm{d}\xi_1.$$

Being thrifty by nature, we are loath to abandon the matrices B_l. Instead, in the spirit of environmental awareness, we recycle them. The integrand being a commutator of two values of A, we may attempt to approximate the integral by a linear combination of commutators of the form $[B_k, B_l]$. This works surprisingly well and it is possible to prove that

$$\begin{aligned}&\sum_{k=1}^{\nu-1} \sum_{\substack{l=k+1\\ k+l \text{ odd}}}^{\nu} \frac{(l-k)-(-1)^k(l+k)}{2^{l+k-1}lk(l+k)}[B_k, B_l]\\ &= -\tfrac{1}{6}[B_1,B_2] + \tfrac{1}{120}[B_2,B_3] - \tfrac{1}{40}[B_1,B_4] + \cdots\end{aligned}$$

approximates the double integral to order 2ν, identical to the univariate Gaussian quadrature (13) [13].

It is not just the bivariate integral. *All* the integrals in the Magnus expansion can be evaluated to order 2ν using and reusing again the same ν matrices B_l. The general pattern for an s-fold integral is

$$\begin{aligned}&\int_{h\mathcal{S}} \boldsymbol{L}(A(\xi_s), A(\xi_{s-1}), \ldots A(\xi_1))\mathrm{d}\xi_s\mathrm{d}\xi_{s-1}\cdots\mathrm{d}\xi_1\\ &\approx \sum_{\boldsymbol{l}\in \mathrm{C}_s^\nu} \tilde{b}_{\boldsymbol{l}} \boldsymbol{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_s}).\end{aligned} \tag{14}$$

Here $h\mathcal{S} \in \mathbb{R}^s$ is the domain of integration, $\boldsymbol{L}$ is a nested commutator and C_s^ν is the set of all combinations of length s from $\{1, 2, \ldots, \nu\}$. The weights $\tilde{b}_{\boldsymbol{l}}$ are given explicitly by

$$\tilde{b}_{\boldsymbol{l}} = \int_{\mathcal{S}} \prod_{i=1}^{s} (\xi_i - \tfrac{1}{2})^{l_i-1}\mathrm{d}\xi_s\mathrm{d}\xi_{s-1}\cdots\mathrm{d}\xi_1.$$

Regardless of the number of terms in the order-(2ν) approximation $\Omega^{[2\nu-1]}$, their computation requires just ν evaluations of the matrix function A, the same number as we would have needed were the single-integral formula (1) right. Great news? Not exactly, since the *quid pro quo* of this amazing economy in function evaluations is a stupendous price tag in linear algebra. The number of terms in C_s^ν is very large, increasing exponentially with s, and each such term requires the computation of nested commutators. Fortunately, further insight, combined with the magic wand of Lie-algebra theory, has led Hans Munthe-Kaas and Brynjulf Owren to an approach which reduces the cost of algebra down to a significantly more modest size, and allows the computation of high-order Magnus approximations relatively cheaply [20].

5 Graded Lie algebras and optimised Magnus

There are three mechanisms at play that allow us to reduce the number of computations. Firstly, each matrix B_l is $\mathcal{O}(h^l)$, hence

$$\boldsymbol{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_s}) = \mathcal{O}\left(h^{l_1+l_2+\cdots+l_s}\right).$$

We say that $\boldsymbol{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_s})$ is of *grade* $|\boldsymbol{l}| = \sum_{i=1}^{s} l_i$ and note that, whenever $|\boldsymbol{l}| > 2\nu$, we can throw away the underlying nested commutator without affecting the order. Secondly, it is possible to prove that the act of replacing all integrals in $\Omega^{[2\nu-1]}$ with quadratures (14) retains time symmetry [20]. The consequence is that we know in advance that roughly half of all terms come with zero coefficients, hence need not be computed. Specifically, whenever the grade of a term is even, it makes no contribution to the quadrature of $\Omega^{[2\nu-1]}$. It is, however, a third mechanism, that of *graded Lie algebras,* which brings down computational expense to the greatest extent.

It is implicit in our construction that, no matter how complicated the terms $\boldsymbol{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_s})$, they are all 'words' written in an 'alphabet' that consists of just ν separate 'letters', $B_1, B_2, \ldots, B_\nu$. Formally, we say that every such term resides in a *free Lie algebra* $\mathcal{B}$, generated by $\boldsymbol{B} = \{B_1, B_2, \ldots, B_\nu\}$: the linear combination of all possible 'words' that can be created from the elements of $\boldsymbol{B}$ by repeated commutation. We have already seen that each such 'word' has an important attribute, its grade. Formally, we endow each *generator* B_l with the grade $w(B_l) = l$, and propagate this quantity in the recursive generation of nested commutators by

$$Z = [X_1, X_2] \qquad \Rightarrow \qquad w(Z) = w(X_1) + w(X_2).$$

This formula is clearly consistent with our definition of the grade of $\boldsymbol{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_s})$. It is easy to verify that the span, in $\mathcal{B}$, of all the terms of grade m forms a linear space, which we denote by $\mathcal{B}_m$. (In fact, $\mathcal{B}$ is a direct sum of $\mathcal{B}_m$, $m \in \mathbb{N}$.) Remarkably, the *dimension* of $\mathcal{B}_m$ is very small. Let $\lambda_1, \lambda_2, \ldots, \lambda_\nu$ be all the zeros of the polynomial

$$1 - \sum_{i=1}^{\nu} z^i = \frac{1 - 2z + z^{\nu+1}}{1 - z}.$$

It is possible to prove that

$$\dim \mathcal{B}_m = \frac{1}{m} \sum_{d|m} \left(\sum_{i=1}^{\nu} \lambda_i^{-m/d} \right) \mu(d), \qquad (15)$$

where μ is the *Möbius function,*

$$\mu(n) = \begin{cases} 1, & n = 1, \\ (-1)^k, & j_i \equiv 1 \text{ for all } i = 1, 2, \ldots, k, \\ 0, & \text{otherwise,} \end{cases}$$

where $n = p_1^{j_1} p_2^{j_2} \cdots p_k^{j_k}$ is the *prime factorization* of the natural number n and the summation in (15) is carried out over all the integer divisors of m [20]. Moreover, it is also possible to construct (by recursion, how else?) a basis of $\mathcal{B}_m$ and express all the elements in the linear space as linear combinations from the basis. The reason for this truly amazing economy is that a Lie algebra is replete with redundancy, originating in the skew-symmetry of the commutator and in the Jacobi identity. The language of graded free Lie algebras allows us to take advantage of this feature.

To appreciate the savings inherent in this approach, we note that the number of commutators required for an order-6 method drops from 80 to 7 in each step and for an order-8 method goes down from 3304 to 22. For order 10 the number of commutators is reduced from 1256567 in a naive implementation all the way down to 73.

As an example, we herewith present a sixth-order Magnus method, written in a variable-step time-stepping terminology, using quadrature (14) and taking advantage of economies inherent in graded free Lie algebras. In each step n we commence by evaluating $A_k = A(t_n + c_k h_n)$, $k = 1,2,3$, where $h_n = t_{n+1} - t_n$ and

$$c_1 = \tfrac{1}{2} - \tfrac{\sqrt{15}}{10}, \qquad c_2 = \tfrac{1}{2}, \qquad c_3 = \tfrac{1}{2} + \tfrac{\sqrt{15}}{10}.$$

Next, we set

$$\begin{aligned} B_1 &= hA_2, \qquad B_2 = \frac{\sqrt{15}h}{3}(A_3 - A_1), \\ B_3 &= \frac{10h}{3}(A_3 - 2A_2 + A_1). \end{aligned}$$

The method now reads

$$\begin{aligned} \tilde{\Omega}^{[5]} &= B_1 + \tfrac{1}{12}B_3 - \tfrac{1}{12}[B_1, B_2] + \tfrac{1}{240}[B_2, B_3] \\ &\quad + \tfrac{1}{360}[B_1, [B_1, B_3]] - \tfrac{1}{240}[B_2, [B_1, B_2]] \\ &\quad + \tfrac{1}{720}[B_1, [B_1, [B_1, B_2]]] \end{aligned}$$

(note that all the elements are of odd grade!). We conclude each Lie-algebraic step with

$$Y_{n+1} = \mathrm{e}^{\tilde{\Omega}^{[5]}} Y_n,$$

where $Y_n \approx Y(t_n)$. Just three function evaluations, seven commutators and a single exponential, an expense perfectly affordable in a single time step.

But have we extracted every single ounce of information to drive down the computational cost? No, not yet! Aggregating terms and computing commutators in a clever manner, Sergio Blanes, Fernando Casas and Pepe Ros have recently presented what is probably the best-possible implementation of the numerical Magnus expansion [3]. Revisiting the order-6 case, we compute

$$\begin{aligned} C_1 &= [B_1, B_2], \\ C_2 &= [B_1, 2B_3 + C_1], \\ C_3 &= [-20B_1 - B_3 + C_1, B_2 - \tfrac{1}{60}C_2], \\ \hat{\Omega}^{[5]} &= B_1 + \tfrac{1}{12}B_3 + \tfrac{1}{240}C_3, \\ Y_{n+1} &= \mathrm{e}^{\hat{\Omega}^{[5]}} Y_n, \end{aligned}$$

with just three commutators. Note that $\tilde{\Omega}^{[5]}$ and $\hat{\Omega}^{[5]}$ are not the same, but they differ only in $\mathcal{O}(h_n^7)$ terms, that have no effect on the order.

6 Beyond Magnus

One of main reasons why linear equations are interesting is because they might tell us something about their nonlinear brethren. Indeed, the idea of a Magnus expansion has been generalised by Antonella Zanna to nonlinear Lie-group equations (12) [23]. This, however, leads to implicit methods and nonlinear algebraic equations in every step: other things being equal, it is probably easier to employ explicit Runge–Kutta methods in the Lie algebra, an approach pioneered by Hans Munthe-Kaas [19]. A considerably more significant extension of the scope of Magnus (and other Lie-group) methods is due to Hans Munthe-Kaas and Antonella Zanna, who have applied them to equations evolving in homogeneous manifolds [21].

We say that a Lie group $\mathcal{G}$ *acts* on a manifold $\mathcal{M}$ if there exists a map $\lambda : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$ which is consistent with the group multiplication, $\lambda(x_1, \lambda(x_2, y)) = \lambda(x_1 x_2, y)$ for all $x_1, x_2 \in \mathcal{G}$, $y \in \mathcal{M}$. A group action is said to be *transitive* if for every $y_1, y_2 \in \mathcal{M}$ there exists $x \in \mathcal{G}$ such that $\lambda(x, y_1) = y_2$. If there exists a transitive group action for $\mathcal{M}$, we say that the latter is a *homogeneous manifold* (or a homogeneous space).

An example of a homogeneous manifold is the Euclidean unit sphere $\mathbb{S}_{N-1} \subset \mathbb{R}^N$, which is acted upon transitively by orthogonal matrices $\mathrm{O}(N)$,

$$\|\boldsymbol{x}\|_2 = 1, \quad Q \in \mathrm{O}(N) \qquad \Rightarrow \qquad \|Q\boldsymbol{x}\|_2 = 1.$$

The set $\mathrm{Sym}(N)$ of real symmetric $N \times N$ matrices is subject to an action of $\mathrm{GL}(N)$ by congruence, $S \to VSV^\top$. If we restrict the latter action from $\mathrm{GL}(N)$ to the *special orthogonal group* $\mathrm{SO}(N) = \mathrm{O}(N) \cap \mathrm{SL}(N)$, it acquires an important attribute: the eigenvalues of S and of $VSV^\top$ are the same.

Suppose that a differential equation evolves in a homogeneous manifold $\mathcal{M}$ and we desire to solve it numerically. The conventional objective is thus, given an approximation $\mathcal{M} \ni \boldsymbol{y}_n \approx \boldsymbol{y}(t_n)$, to seek $\boldsymbol{y}_{n+1}$

that approximates $\boldsymbol{y}(t_{n+1})$ to sufficient order of accuracy. Unfortunately, if we employ a 'traditional' numerical method, it is very unlikely that $\boldsymbol{y}_{n+1}$ remains in $\mathcal{M}$. We can instead set a goal more consistent with the underlying geometry of the manifold: given an approximation $\mathcal{M} \ni \boldsymbol{y}_n \approx \boldsymbol{y}(t_n)$, find an element $V_n \in \mathcal{G}$ such that $\boldsymbol{y}_{n+1} = \lambda(V_n, \boldsymbol{y}_n)$ approximates $\boldsymbol{y}(t_{n+1})$ to requisite order of accuracy. Note that this guarantees that we stay in $\mathcal{M}$.

As an example, consider the *isospectral flow*

$$Y' = [A(t,Y), Y], \quad t \geq 0, \quad Y(0) = Y_0 \in \mathrm{Sym}(N), \tag{16}$$

where $A : \mathbb{R}_+ \times \mathrm{Sym}(N) \to \mathfrak{so}(N)$. It is easy to show that the eigenvalues of $Y(t)$ remain constant as the flow evolves by demonstrating that it can be subjected to the (congruent) similarity action by $\mathrm{SO}(N)$, $Y(t) = Q(t)Y_0Q^\top(t)$, where

$$Q' = A(t, QY_0Q^\top)Q, \quad t \geq 0, \qquad Q(0) = I \tag{17}$$

(compare with (12)). This feature is important in many applications, not least in numerical linear algebra [6], but it cannot be retained for $N \geq 3$ by classical numerical methods like Runge–Kutta or multistep. The remedy is to solve (17) by a Lie-group method. Of course, (17) is typically nonlinear, outside the scope of 'plain vanilla' Magnus methods. One remedy is nonlinear Magnus, another a Runge–Kutta method in the algebra. However, in special cases the logic underlying the Magnus expansion can be extended to nonlinear equations. For example, consider the *double bracket equation*

$$Y' = [[M,Y],Y], \quad t \geq 0, \quad Y(0) = Y_0 \in \mathrm{Sym}(N), \tag{18}$$

where $M \in \mathrm{Sym}(N)$ is given. Equation (18) has many interesting applications, which are underpinned by an exciting feature. Suppose that $[M, Y_0] \neq O$. Then $\hat{Y} = \lim_{t\to\infty} Y(t)$ exists, is unique and minimises the distance from M in the *Frobenius norm,* $\|B\|_{\mathsf{F}} = \left(\sum_{k,l=1}^N B_{k,l}^2\right)^{1/2}$, among all symmetric matrices that share the eigenvalues of Y_0 [4].

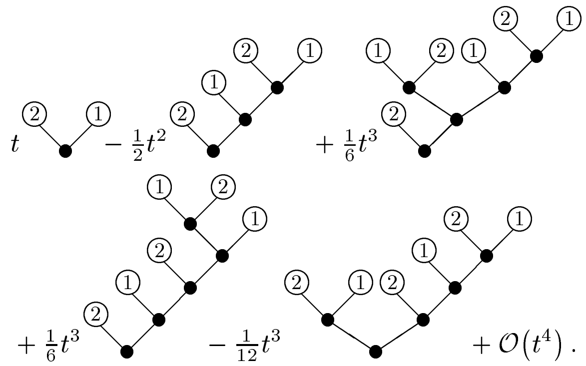
Have a look at (18). No matter how complicated the solution (whether the function Y itself or its Lie-group and Lie-algebraic counterparts), it must be expressible in an 'alphabet' comprising just two 'letters', the matrices Y_0 and M. Since we are keen to stay on the *isospectral orbit,* we express the solution in the form

$$Y(t) = \mathrm{e}^{\Omega(t)} Y_0 \mathrm{e}^{-\Omega(t)}, \qquad t \geq 0,$$

where Ω evolves in $\mathfrak{so}(N)$. Although the equation for Ω is highly nonlinear, it is possible to express this function using our terminology of binary rooted trees, except that we paint their leaves in two colours: we let

$$① \rightsquigarrow Y_0, \qquad ② \rightsquigarrow M.$$

Then the Taylor expansion of Ω can be represented in an arboreal fashion [12],



As before, there exist recursive rules that allow us to approximate Ω to arbitrarily-high order. Double-bracket equations (18) are just one example from a large (and largely unexplored) menagerie of Lie-group and homogeneous-manifold equations that can be written in a finite 'alphabet', thereby lending themselves to similar treatment.

The tale of Magnus expansions and geometric integrators is by no means complete. And the moral of this tale? Two, really: Firstly, the importance of reconciling qualitative information – geometry, structure, invariants – with computation and approximation. It is not a zero-sum game where you must opt for either quality or quantity, but a complicated interactive procedure with huge scope for synergy. Secondly, modern mathematical computation is not about discretising everything in sight by elementary means and massive number crunching. It is about employing every suitable pure-mathematical tool, in our case rooted trees and graded Lie algebras, to render our computations more accurate and affordable. Mathematics can lead us to better computational algorithms and expansions can indeed grow on trees.

Acknowledgements

Local thanks to all my colleagues who have read various versions of this paper, offered their comments and saved me from errors, misprints and self-inflicted embarrassment. In particular, my gratitude is due to Brad Baxter, Harold Boas, Chris Budd, Martin Buhmann, Fernando Casas, Dganit Iserles, Joe Keller and Peter Olver.

No mathematical research, least of all *this* mathematical research, is a creation of a single mind or of several isolated minds. It evolves in the shared intellectual space spanned by the expertise, excitement and sense of fun and discovery of many individuals. Thus, global thanks to the 'Lie group' of colleagues, friends and collaborators who have created in the last few years the new subject area of geometric integration and of computational methods in a Lie-group setting.

References

[1] W. Abikoff, J. S. Birman, and K. Kuiken, editors. *Wilhelm Magnus: Groups, Geometry and Special Functions*, volume 169 of *Contemp. Math.*, Providence, RI, 1994. American Mathematical Society.

[2] S. Blanes, F. Casas, J. A. Oteo, and J. Ros. Magnus and Fer expansions for matrix differential equations: The convergence problem. *J. Phys A*, 31:259–268, 1998.

[3] S. Blanes, F. Casas, and J. Ros. High order optimized geometric integrators for linear differential equations. Technical Report NA07, DAMTP, University of Cambridge, 2000.

[4] R. W. Brockett. Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems. *Linear Algebra Appl.*, 146:79–91, 1991.

[5] C. J. Budd and M. D. Piggott. Geometric integration and its applications. In P. G. Ciarlet and F. Cucker, editors, *Handbook of Numerical Analysis*. North Holland, Amsterdam, 2002. To appear.

[6] M. T. Chu. A list of matrix flows with applications. In *Hamiltonian and Gradient Flows, Algorithms and Control*, volume 3 of *Fields Inst. Commun.*, pages 87–97. Amer. Math. Soc., Providence, RI, 1994.

[7] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, Orlando, FL, 2nd edition, 1984.

[8] A. T. Fomenko and R. V. Chakon. Recurrence formulas for homogeneous terms of a convergent series that represents a logarithm of a multiplicative integral on Lie groups. *Funct. Anal. Applcs*, 1:41–49, 1990.

[9] E. Hairer. Symmetric projection methods for differential equations on manifolds. *BIT*, 40:726–734, 2000.

[10] F. Hausdorff. Die symbolische Exponentialformel in der Gruppentheorie. *Berichte der Sächsischen Akademie der Wißenschaften (Math. Phys. Klasse*, 58:19–48, 1906.

[11] A. Iserles. Multistep methods on manifolds. *IMA J. Num. Anal.*, 21:407–419, 2001.

[12] A. Iserles. On the discretization of double-bracket flows. Technical Report NA08, DAMTP, University of Cambridge, 2001. To appear in *Found. Comp. Maths.*

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

[14] A. Iserles and S. P. Nørsett. On the solution of linear differential equations in Lie groups. *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.*, 357:983–1020, 1999.

[15] A. Iserles, S. P. Nørsett, and A. F. Rasmussen. Time symmetry and high-order Magnus methods. *Appld Num. Maths*, 39:379–401, 2001.

[16] A. Iserles and A. Zanna. Preserving algebraic invariants with Runge–Kutta methods. *J. Comp. Appld Maths*, 125:69–81, 2000.

[17] W. Magnus. On the exponential solution of differential equations for a linear operator. *Comm. Pure Appl. Math.*, 7:649–673, 1954.

[18] P. C. Moan. Efficient approximation of Sturm–Liouville problems using Lie-group methods. Technical Report NA11, DAMTP, University of Cambridge, 1998.

[19] H. Munthe-Kaas. Runge–Kutta methods on Lie groups. *BIT*, 38:92–111, 1998.

[20] H. Munthe-Kaas and B. Owren. Computations in a free Lie algebra. *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.*, 357:957–982, 1999.

[21] H. Munthe-Kaas and A. Zanna. Numerical integration of differential equations on homogeneous manifolds. In F. Cucker and M. Shub, editors, *Foundations of Computational Mathematics*, pages 305–315. Springer Verlag, 1997.

[22] V. S. Varadarajan. *Lie Groups, Lie Algebras, and Their Representations*. GTM 102. Springer-Verlag, 1984.

[23] A. Zanna. Collocation and relaxed collocation for the Fer and the Magnus expansions. *SIAM J. Num. Anal.*, 36:1145–1182, 1999.