Euro. Jnl of Applied Mathematics (2019), *vol.* 30, *pp.* 400–425.
doi:10.1017/S0956792518000177

Applications of Magnus expansions and pseudospectra to Markov processes†

A. ISERLES[1] and S. MACNAMARA[2]

[1]*Department of Applied Mathematics and Mathematical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK*
email: a.iserles@damtp.cam.ac.uk
[2]*Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), School of Mathematical and Physical Sciences, University of Technology Sydney, Ultimo, Australia*
email: shev.macnamara@uts.edu.au

(*Received 22 August 2017; revised 18 March 2018; accepted 19 March 2018; first published online 17 April 2018*)

New directions in Markov processes and research on master equations are showcased by example. The utility of Magnus expansions for handling time-varying rates is demonstrated. The useful notion in applied mathematics often turns out to be the pseudospectra and not simply the eigenvalues. We highlight that general principle with our own examples of Markov processes where exact eigenvalues are found and contrasted with the large errors produced by standard numerical methods. As a motivating application, isomerisation provides a running example and an illustration of our approaches to chemical kinetics. We also present a brief example of a totally asymmetric exclusion process.

Key words: 60J28, 60H35, 65F40

1 Introduction

The term ‘master equation’ goes back at least as far as the work of Kac in the middle of the twentieth century [27, page 105], and the subject of master equations admits a Feynman–Kac stochastic path integral formulation [44]. The general principle of a governing equation emerging from ensemble averages goes back much further in the history of statistical mechanics, including the kinetic theories of Boltzmann and, earlier, of Bernoulli in the 1700s. Generalised master equations can cater to some form of memory in non-Markovian models but the most common application of master equations is to *Markov processes*. Perhaps the first application of the eponymous Markov process was Andrei Markov’s model of a poem ‘Eugeny Onegin’ as a Markov chain, which he presented in 1913 in St. Petersburg. Other famous applications include Shannon’s Information Theory and Google’s PageRank to find order in the information on the World Wide Web [23]. Choosing the simplest examples, we describe applications to exclusion processes and chemical processes, although the computational methods we present have wider applicability.

† This research and Shev MacNamara have been partially supported by a David G. Crighton Fellowship to DAMTP, Cambridge.

The purpose of this paper is to elucidate connections between Markov processes and a number of diverse computational issues, as surveyed in Section 4. These seem to have received limited attention in the Markov process literature but are nevertheless significant for computational approaches. Here is a *summary of our major contributions:*

- Theorem 3.1 is a restatement of a known result concerning the spectra of a Kac matrix arising in a Markov process, and we find an original proof. All other theorems and algorithms are original results. Theorem 2.1 and the *explicit* formula that we derive in (2.16) for the Magnus expansion are examples of particular contributions.
- Exact results for the spectra are further extended computationally, leading to especially attractive contributions in Figures 1–3.
- As a primary and important application, isomerization serves as a running example. In particular, and as a major focus of the work, we have demonstrated how to apply the techniques described in (2.5) for the Magnus expansion in (2.6) to *Markov processes with time-varying rates.*

1.1 Models of isomerisation

The same chemical species can sometimes exist in two distinct molecular forms, S_1 and S_2, and can reversibly convert from one form, or isomer, to the other in a process named *isomerisation*: $S_1 \longleftrightarrow S_2$. A mathematical model involves two rate constants (this terminology is common, but in our examples the rate 'constants' are often *time-dependent*), $c_1(t)$ associated with the forward reaction $S_1 \xrightarrow{c_1} S_2$, and $c_2(t)$ for the backward reaction $S_1 \xleftarrow{c_2} S_2$.

A hierarchy of three mathematical frameworks for modelling chemical reactions is provided by the reaction rate equations (RRE), the chemical Langevin equation, and the chemical master equation (CME). Typically, when all species are present in high concentrations, the deterministic RRE are a good model at a macroscopic scale, but if some species are present in small numbers of molecules then often the discrete and stochastic CME is a more appropriate model at a mesoscopic scale [10, 32, 38]. Stochastic differential equations such as the Langevin equation for isomerisation [15] and their corresponding Fokker–Planck partial differential equations provide models at scales that are intermediate between those of the deterministic rate equations and the discrete and stochastic master equations.

The RRE for this model of isomerisation are the two linear ordinary differential equations (ODEs)

$$\frac{\mathrm{d}}{\mathrm{d}t}[S_1] = -c_1(t)[S_1] + c_2(t)[S_2], \qquad \frac{\mathrm{d}}{\mathrm{d}t}[S_2] = +c_1(t)[S_1] - c_2(t)[S_2], \tag{1.1}$$

where $[S_i]$ indicates the concentration (molecules per unit volume) of species i.

These chemical reactions can also be modelled by a continuous time, discrete state Markov process for which a linear system of ODEs known as the master equation, $p' = Ap$, describes the evolution of the associated probability distribution p. (This is also sometimes known as a Chapman–Kolmogorov forward equation, our notation involves a matrix A that is the transpose of the 'Q-matrix' notation that is sometimes encountered,

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

and statements of the master equation for isomerisation in notation without a matrix can be found in [15].) The state records the integer number of molecules of each species, and, after enumeration of all possible states, the probability of the ith state is recorded in the ith entry of the vector p. In a small time dt, the probability mass that flows from state j to a different state i is approximately given by $A_{ij}\mathrm{d}t$. The matrix A has non-negative off-diagonals and zero column sum, and is thus a *graph Laplacian*. As an example, if we start with N molecules of species S_1 and zero molecules of S_2, then there are $N+1$ states, $(i, N-i)$ for $i = 0, \ldots, N$, where state i has i molecules of S_1. If our initial condition has all probability concentrated on state $(0, N)$, then our initial probability vector is $p(0) = (0, 0, \ldots, 1)^\top$. With rates $c_1(t) = 1 + f(t)$ and $c_2(t) = 1 - f(t)$, the probability vector evolves according to the linear ODE (1.2), introduced below, which is the CME for isomerisation.

'Generally, the CME has such extremely high dimension that it cannot be handled analytically or computationally' [21]. In this article, we focus on some exceptions. A large class of important and solvable models, including isomerization, arise when reaction rates are linear as a function of the state [26]. For this special class of models, we have exact agreement between the average value of the stochastic CME model and the solution of the corresponding deterministic RRE. (Usually these models agree only approximately.) The exact solution to the CME (1.2) for our isomerisation example is a binomial distribution, where the time-varying parameter in the binomial distribution originates in the solution to the corresponding RRE (1.1). This makes it an ideal candidate for demonstrating novel applications of Magnus methods, which as we will see, reveal finer structure in the master equations.

1.2 A master equation for isomerisation with explicitly time-varying rates

We are concerned with the linear ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = \left[A^{[0]} + A^{[1]} f(t)\right]\boldsymbol{p}, \qquad \boldsymbol{p}(0) = \boldsymbol{p}_0 \in \mathbb{R}^{N+1}, \tag{1.2}$$

involving two matrices $A^{[0]}$ and $A^{[1]}$ defined by, for $k, \ell = 0, \ldots, N$,

$$A^{[0]}_{k,\ell} = \begin{cases} -N, & k = \ell, \\ \ell, & k = \ell - 1, \\ N - \ell, & k = \ell + 1, \\ 0, & \text{otherwise}; \end{cases} \qquad A^{[1]}_{k,\ell} = \begin{cases} N - 2\ell, & k = \ell, \\ \ell, & k = \ell - 1, \\ -N + \ell, & k = \ell + 1, \\ 0, & \text{otherwise}. \end{cases} \tag{1.3}$$

The $A^{[0]}$ matrix is remarkably close to the `'clement'` matrix in the `MATLAB` gallery, which has a zero main diagonal but is otherwise the same.

If $-1 \leqslant f(t) \leqslant 1$ then $\mathcal{A} = A^{[0]} + A^{[1]} f(t)$ has the usual properties of a graph Laplacian matrix (sometimes called the *infinitesimal generator* of the Markov process). In that case (1.2) is a master equation, which was originally simulated for the special case $f(t) = \sin t$ [28]. Here, we generalize. It turns out (1.2) has a truly miraculous structure.

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

2 The Magnus expansion

The matrix exponential is the solution of a linear ODE when the coefficient matrix is constant, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = \mathbb{A}\boldsymbol{p} \qquad \text{with solution} \qquad \boldsymbol{p}(t) = \exp(t\mathbb{A})\boldsymbol{p}(0). \tag{2.1}$$

When the matrix varies in time, $\mathbb{A} = \mathbb{A}(t)$, the solution is no longer simply the matrix exponential, but it can still be expressed in an exponential form. We write

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = \mathbb{A}(t)\boldsymbol{p} \qquad \text{with solution} \qquad \boldsymbol{p}(t) = \exp(\boldsymbol{\Omega}(t))\boldsymbol{p}(0). \tag{2.2}$$

Here, the Magnus expansion [35] tells us how to find the crucial matrix $\boldsymbol{\Omega}(t)$ as an infinite series, namely

$$\boldsymbol{\Omega}(t) = \int_0^t \mathbb{A}(s)\mathrm{d}s - \frac{1}{2}\int_0^t \left[\int_0^s \mathbb{A}(r)\mathrm{d}r, \mathbb{A}(s)\right] \mathrm{d}s + \dots. \tag{2.3}$$

All higher order terms in the expansion can be generated recursively by integration and commutation, thus involving commutators as a factor. The *commutator* of two matrices is, as usual, $[A, B] \equiv AB - BA$. In the special case that the time-varying matrix commutes with itself at different time points, i.e., $\forall t_1, \forall t_2$, $[\mathbb{A}(t_1), \mathbb{A}(t_2)] \equiv 0$, those commutators are all zero so the expansion simplifies to $\Omega(t) = \int_0^t \mathbb{A}(s)\mathrm{d}s$, agreeing with our intuition from the scalar case. This expansion, which is valid for all sufficiently small times t, was originally motivated by applications in quantum mechanics where it was derived by an analogy with Cauchy–Picard iteration in the 1950s. For a long time it remained merely a theoretical tool, and it was only nearing the turn of the century that it was fashioned into an effective computational tool [25].

The next higher order contribution in (2.3) is the sum of two terms, namely,

$$\begin{aligned}\frac{1}{12}\int_0^t \left[\int_0^{s_1} \mathbb{A}(s_2)\mathrm{d}s_2, \left[\int_0^{s_1} \mathbb{A}(s_2)\mathrm{d}s_2, \mathbb{A}(s_1)\right]\right] \mathrm{d}s_1 \\ + \frac{1}{4}\int_0^t \left[\int_0^{s_1} \left[\int_0^{s_2} \mathbb{A}(s_3)\mathrm{d}s_3, \mathbb{A}(s_2)\right] \mathrm{d}s_2, \mathbb{A}(s_1)\right] \mathrm{d}s_1.\end{aligned}$$

The next higher order contribution is the sum of four terms. The point is that successively higher order contributions in the expansion involve a growing number of terms, so the expansion quickly becomes difficult to work with directly. Instead, using graph theory, more specifically rooted trees, is a successful way to manage the expansion. A remarkable correspondence between terms in the Magnus expansion and rooted, binary trees (elucidated in [25, equation (4.10)]) allows (2.3) to be written as

$$\Omega(t) = \sum_{m=0}^{\infty} \sum_{\tau \in \mathbb{T}_m} \int_0^t \alpha(\tau) G_\tau(x)\,\mathrm{d}x. \tag{2.4}$$

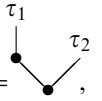
All terms in the expansion are identified with a rooted, binary tree in the set of Magnus trees, denoted $\cup_m \mathbb{T}_m$. In this correspondence, vertical lines correspond to integration and joining trees corresponds to commutation. Here is the four-step recipe.

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

(1) $\mathbb{T}_m$ is the set of Magnus trees with m vertical lines.

(2) The only member of $\mathbb{T}_0$ is $\bullet$.

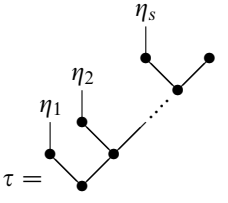
(3) $\tau \to G_\tau$ is a mapping from Magnus trees to matrices. Specifically, $G_\bullet = \mathcal{A}$ and, given $m \geqslant 1$, any $\tau \in \mathbb{T}_m$ can be represented in the form

$\tau =$  $\tau_1 \in \mathbb{T}_{m_1},\ \tau_2 \in \mathbb{T}_{m_2}, \quad m_1 + m_2 = m - 1.$ (2.5)

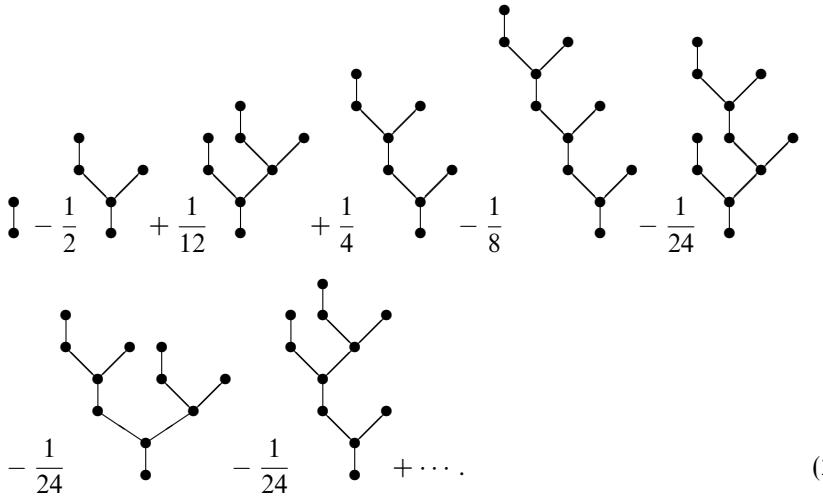
In that case

$$G_\tau(t) = \left[\int_0^t G_{\tau_1}(x)\,\mathrm{d}x, G_{\tau_2}(t)\right].$$

(4) $\alpha : \tau \to \mathbb{Q}$ is a mapping from Magnus trees to rational numbers. Specifically, $\alpha(\bullet) = 1$ and, for any $\tau \in \mathbb{T}_m$ for $m \geqslant 1$, with B_s denoting Bernoulli numbers,

$\tau =$  $\Rightarrow \qquad \alpha(\tau) = \dfrac{\mathrm{B}_s}{s!} \displaystyle\prod_{j=1}^{s} \alpha(\eta_j).$

In general, this procedure elegantly expresses the Magnus expansion (2.4) as

$\Omega(t) \rightsquigarrow$  (2.6)

2.1 A special property of isomerisation matrices

Recognising the following special property (confirmed by an easy matrix multiplication)

$$[A^{[0]}, A^{[1]}] = -2A^{[1]} \tag{2.7}$$

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

usefully simplifies our Magnus expansion. This simple form of the commutator (2.7) is fundamental because the Magnus expansion is constructed as a linear combination of terms that can be obtained from $\mathcal{A}(t) = A^{[0]} + A^{[1]} f(t)$ using only integration and commutation. It thus resides in the *free Lie algebra* $\mathcal{F}$ generated by $A^{[0]}$ and $A^{[1]}$. In light of (2.7), that $\mathcal{F}$ is

$$\mathcal{F}(A^{[0]}, A^{[1]}) = \text{Span}\,\{A^{[0]}, A^{[1]}\}. \tag{2.8}$$

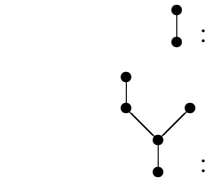
In other words, although in general the Magnus expansion of the solution may require many terms, *the Magnus expansion of (1.2) for isomerisation is simply a linear combination of the form*[1] $\Omega(t) = \sigma_{[0]}(t)A^{[0]} + \sigma_{[1]}(t)A^{[1]}$ *!*

2.2 A Magnus expansion of isomerisation

We now specialize the general form of the expansion (2.4) to our application of isomerisation (1.2), for which

$$\bullet \rightsquigarrow A^{[0]} + f(t)A^{[1]}.$$

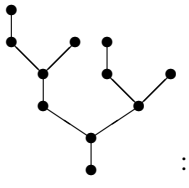
By following the four step algorithm near (2.5), we find the first few terms in the series (2.3) and the corresponding trees are



$$\int_0^t \mathcal{A}(x)\,\mathrm{d}x = tA^{[0]} + \int_0^t f(x)\,\mathrm{d}x A^{[1]},$$

$$\begin{aligned}\int_0^t \int_0^{x_1} [\mathcal{A}(x_2), \mathcal{A}(x_1)]\,\mathrm{d}x_2\,\mathrm{d}x_1 &= \int_0^t \left[x_1 f(x_1) - \int_0^{x_1} f(x_2)\,\mathrm{d}x_2\right] \mathrm{d}x_1 [A^{[0]}, A^{[1]}] \\ &= 2\int_0^t (t - 2x) f(x)\,\mathrm{d}x A^{[1]}.\end{aligned}$$

Note we made use of (2.7) for the commutator to simplify the expressions. Moreover, a matrix commutes with itself so some terms are zero, such as



$$\left[2\int_0^t (t - 2x) f(x)\,\mathrm{d}x A^{[1]}, -2\int_0^t [f(t) - f(x)]\,\mathrm{d}x A^{[1]}\right] = O.$$

We claim that for $\tau \in \mathbb{T}_m$, $m \geqslant 1$, necessarily G_τ is a scalar multiple of $A^{[1]}$, i.e., $G_\tau(t) = \sigma_\tau(t)A^{[1]}$.

[1] Indeed, more is true. A Lie algebra $\mathfrak{g}$ is *solvable* if there exists $M \geqslant 0$ such that $\mathfrak{g}^{[M]} = \{0\}$, where $\mathfrak{g}^{[0]} = \mathfrak{g}$ and $\mathfrak{g}^{[k+1]} = [\mathfrak{g}^{[k]}, \mathfrak{g}^{[k]}]$. By (2.7), $\dim \mathcal{F}^{[1]} = 1$ so it is a commutative algebra and $\mathcal{F}^{[2]} = \{0\}$. The algebra is solvable!

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

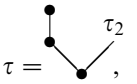
We already know from (2.7) and (2.8) that our Magnus expansion is of the form $\sigma_{[0]}(t)A^{[0]} + \sigma_{[1]}(t)A^{[1]}$. In view of the first few trees above, our claim immediately implies $\sigma_{[0]}(t) = t$. Having now found $\sigma_{[0]}$, it remains only to find $\sigma_{[1]}$, so to simplify notation, we drop the subscript from now on and let $\sigma = \sigma_{[1]}$.

The proof of the claim is by induction. For $m = 1$ there is only one Magnus tree,

$\tau =$  $\quad \Rightarrow \quad G_\tau(t) = -2\int_0^t [f(t) - f(x)]\,\mathrm{d}x A^{[1]}.$

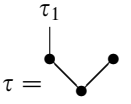
Therefore, $\sigma_\tau(t) = -2\int_0^t [f(t) - f(x)]\,\mathrm{d}x$.

Consider next $m \geqslant 2$ and (2.5). If $m_1, m_2 \geqslant 1$ then, by the induction assumption, both G_{τ_1} and G_{τ_2} are scalar multiples of $A^{[1]}$ and we deduce that $G_\tau \equiv O$. There are two remaining possibilities: either $m_1 = 0$, $m_2 = m - 1$ or $m_1 = m - 1$, $m_2 = 0$. In the first case

$\tau =$  (2.9)

so $G_\tau(t) = \left[tA^{[0]} + \int_0^t f(x)\,\mathrm{d}x A^{[1]}, \sigma_{\tau_2}(t)A^{[1]}\right] = t\sigma_{\tau_2}(t)[A^{[0]}, A^{[1]}] = -2t\sigma_{\tau_2}(t)A^{[1]}$, and $\sigma_\tau(t) = -2t\sigma_{\tau_2}(t)$.

Finally, for $m_1 = m - 1$ and $m_2 = 0$, we have

$\tau =$  (2.10)

for which $G_\tau(t) = \left[\int_0^t \sigma_{\tau_1}(x)\,\mathrm{d}x A^{[1]}, A^{[0]} + f(t)A^{[1]}\right] = -\int_0^t \sigma_{\tau_1}(x)\,\mathrm{d}x[A^{[0]}, A^{[1]}] = 2\int_0^t \sigma_{\tau_1}(x)\,\mathrm{d}x A^{[1]}$ and $\sigma_\tau(t) = 2\int_0^t \sigma_{\tau_1}(x)\,\mathrm{d}x$. This completes the proof of

Theorem 2.1 *The Magnus expansion for isomerisation (1.2) is of the form*

$$\Omega(t) = tA^{[0]} + \sigma(t)A^{[1]} \tag{2.11}$$

for a function σ which has been described above in a recursive manner.

Next, we will explicitly find in the next theorem the function σ of (2.11), thus finding the Magnus expansion of isomerisation. We do not present all steps in the derivations to come. Theorem 2.1 and the steps leading to it were deliberately chosen for presentation partly because this quickly gives a good sense of the style of arguments needed in this area, while still being very accessible. The steps required in our other proofs follow a similar pattern, albeit more detailed.

2.3 Constructing the trees

In general, when we want to find the Magnus trees, we can follow the four-step algorithm near (2.5). That always works. Often though, particular applications allow simplifications,

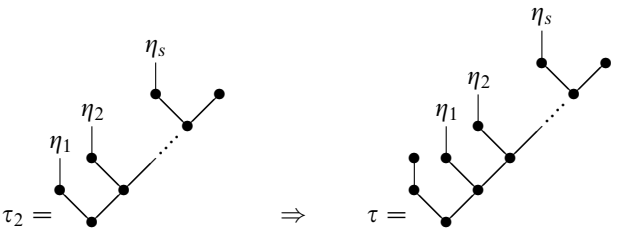
https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

as we now use our application to illustrate. The main question to be answered for this example is how to connect the coefficients $\alpha(\tau)$ to the trees in the situations of (2.9) and of (2.10).

The situation for (2.10) is trivial: since $s = 1$, we have

$$\alpha(\tau) = \frac{\mathrm{B}_1}{1!}\alpha(\tau_1) = -\frac{1}{2}\alpha(\tau_1).$$

It is more complicated in the situation of (2.9). There we have



Therefore,

$$\alpha(\tau_2) = \frac{\mathrm{B}_s}{s!}\prod_{j=1}^{s}\alpha(\eta_j), \qquad \alpha(\tau) = \frac{\mathrm{B}_{s+1}}{(s+1)!}\prod_{j=1}^{s}\alpha(\eta_j).$$

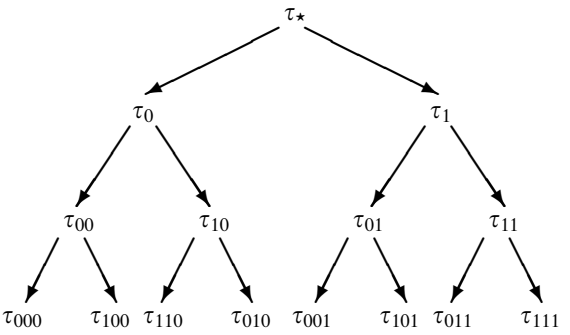
Hence, to summarize

$$\begin{aligned} s = 1: &\quad \alpha(\tau_2) = -\frac{1}{2}\alpha(\eta_1), \quad \alpha(\tau) = \frac{1}{12}\alpha(\eta_1) = -\frac{1}{6}\alpha(\tau_2); \\ s \text{ even}: &\quad \mathrm{B}_{s+1} = 0 \quad \Rightarrow \quad \alpha(\tau) = 0; \\ s \geqslant 3 \text{ odd}: &\quad \mathrm{B}_s = 0 \quad \Rightarrow \quad \alpha(\tau_2) = 0. \end{aligned}$$

This is a moment to comment on the mechanisms giving rise to some of our simplifications. Not all Magnus trees feature — with non-zero coefficients — in the expansion (2.4). There are two mechanisms that explain this: (a) The coefficient $\alpha(\tau)$ is zero; or (b) $\sigma_\tau \equiv 0$, because a matrix commutes with itself and τ originates in trees τ_1 and τ_2 such that $G_{\tau_k}(t) = \sigma_{\tau_k}(t)A^{[1]}$, for $k = 1, 2$. There is an important difference between these two situations. For the first mechanism, while we do not include the tree τ in (2.4), we must retain it for further recursions. In the second mechanism, though, if a tree is zero then all its 'children' are zero too.

The long-and-short is that in every $\mathbb{T}_m$, $m \geqslant 1$ we have 2^{m-1} trees (some with a zero coefficient). What we really have is a binary 'super-tree'

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press



. . . and so on.

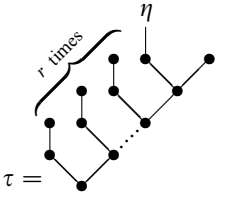
The rule is: Each move 'left' (i.e., in the 0 direction – the subscripts are binary strings) corresponds to 'scenario' (2.9); Each move 'right' corresponds to 'scenario' (2.10). Now that we have simplified our system for dealing with the trees, we are ready to proceed to find σ.

2.4 An explicit formula for σ

As we have seen, except for $\mathbb{T}_0$, every $\tau \in \mathbb{T}_m$ leads to an expression of the form $\sigma_\tau(t)A^{[1]}$. For example, setting $\tilde{f}(x) = xf'(x)$,

$$\mathbb{T}_1 : \qquad \tau_\star = \quad \Rightarrow \quad \sigma_{\tau_\star} = -2\int_0^t \tilde{f}(x)\,\mathrm{d}x, \quad \alpha(\tau_\star) = -\frac{1}{2}.$$

By continuing to find these trees, we see a pattern emerge: For any $\tau \in \mathbb{T}_m$, $m \geqslant 1$, our $\sigma_\tau(t)$ is of the form $\sigma_\tau(t) = \int_0^t K_\tau(t,x)\tilde{f}(x)\,\mathrm{d}x$ for some *kernel* K_τ. To find the kernels, it is convenient for $\tau \in \mathbb{T}_m$, $m \geqslant 2$, to work with



(2.12)

Let $r \in \{0, 1, \ldots, m-2\}$ and $\eta \in \mathbb{T}_{m-r}$. Straightforward computation shows that

$$\eta \rightsquigarrow K_\eta(t,x), \qquad \eta \rightsquigarrow \int_x^t K_\eta(y,x)\,\mathrm{d}y, \qquad \eta \rightsquigarrow 2\int_x^t K_\eta(y,x)\,\mathrm{d}y.$$

This pattern motivates arguments by induction for (2.12), that lead to

$$K_\tau(t,x) = 2(-2t)^r \int_x^t K_\eta(y,x)\,\mathrm{d}y, \qquad \alpha(\tau) = \frac{\mathrm{B}_{r+1}}{(r+1)!}\alpha(\eta). \tag{2.13}$$

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

We left out one exceptional case, namely $\tau = \tau_{\mathbf{0}}$. In that case, the representation (2.12) is still true but $\eta \in \mathbb{T}_0$, so is not associated with a kernel. However, easy computation confirms that $K_{\tau_{\mathbf{0}}}(t,x) = -2(-2t)^{m-1}$, $\alpha(\tau_{\mathbf{0}}) = \frac{\mathrm{B}_m}{m!}$.

Now that we have the kernels, we sum them. Let $\Theta_m(t,x) = \sum_{\tau\in\mathbb{T}_m} \alpha(\tau) K_\tau(t,x)$, for $m \in \mathbb{N}$. For example, $\Theta_1(t,x) \equiv 1$ and $\Theta_2(t,x) = -\frac{2}{3}t + x$. Next, let $\Theta(t,x) = \sum_{m=1}^{\infty} \Theta_m(t,x)$. After some recursion we are led to the Volterra-type equation

$$\frac{t(1-\mathrm{e}^{-2t})}{1-2t-\mathrm{e}^{-2t}}\Theta(t,x) = \int_x^t \Theta(y,x)\,\mathrm{d}y - 1, \tag{2.14}$$

with the solution

$$\Theta(t,x) = -\exp\left(-4\int_x^t \frac{1-y-(1+y)\mathrm{e}^{-2y}}{(1-\mathrm{e}^{-2y})(1-2y-\mathrm{e}^{-2y})}\,\mathrm{d}y\right)\frac{1-2x-\mathrm{e}^{-2x}}{x(1-\mathrm{e}^{-2x})}. \tag{2.15}$$

Now integrate σ_τs, scaled by $\alpha(\tau)$, for all trees: $\sigma(t) = \int_0^t \sum_{m=0}^{\infty}\sum_{\tau\in\mathbb{T}_m} \alpha(\tau)\sigma_\tau(\xi)\,\mathrm{d}\xi = \int_0^t f(x)\,\mathrm{d}x + \int_0^t \sum_{m=1}^{\infty}\sum_{\tau\in\mathbb{T}_m} \alpha(\tau)\int_0^\xi K_\tau(\xi,x)\tilde{f}(x)\,\mathrm{d}x\,\mathrm{d}\xi$. Swapping integration and summation, we have $\sigma(t) = \int_0^t f(x)\,\mathrm{d}x + \int_0^t xf'(x)\int_x^t \Theta(\xi,x)\,\mathrm{d}\xi\,\mathrm{d}x$. Substituting (2.14), we attain our desired goal: $\sigma(t) = \int_0^t f(x)\,\mathrm{d}x + \int_0^t xf'(x)\left[\frac{t(1-\mathrm{e}^{-2t})}{1-2t-\mathrm{e}^{-2t}}\Theta(t,x) + 1\right]\mathrm{d}x$, or

$$\sigma(t) = tf(t) + \frac{t(1-\mathrm{e}^{-2t})}{1-2t-\mathrm{e}^{-2t}}\int_0^t xf'(x)\Theta(t,x)\,\mathrm{d}x. \tag{2.16}$$

Here, we used integration by parts, $\int_0^t xf'(x)\,\mathrm{d}x = tf(t) - \int_0^t f(x)\,\mathrm{d}x$. With (2.15), everything is now explicit. Combining σ in (2.16) with Theorem 2.1, we have now found the (complete!) Magnus expansion of isomerisation.

Note that (2.16) is bounded for all $t \geqslant 0$, because $t(1-\mathrm{e}^{-2t})/(1-2t-\mathrm{e}^{-2t})$ is bounded[2] for all $t \in \mathbb{R}$. As a consequence, *the Magnus series* (2.11) *for isomerisation converges for every* $t \geqslant 0$. That is a significant finding for isomerisation, because in general the Magnus series is only convergent for small times.

There is further significance. Our own exposition of the Magnus expansion here also explains the intriguing numerical evidence appearing in earlier work that time-steps larger than the bound in the Moan–Niesen sufficient condition for convergence of the Magnus expansion can be taken while still maintaining good accuracy with Magnus-based numerical methods [28, Figure 1]. That good experience of taking larger time steps (larger than the bound in the Moan–Niesen condition) with Magnus-based methods has previously been reported in numerous numerical studies in the context of the Schrödinger equation, and was eventually carefully explained by Hochbruck and Lubich [24]. We are also seeing it here in a novel context of master equations, although our explanation via the Magnus expansion shows that same good experience in this novel context is for completely different reasons.

[2] Actually, it is analytic.

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

2.5 A role for automorphisms

Theorems 2.1 and 2.16 tell us, explicitly, the matrix $\boldsymbol{\Omega}(t)$ in the Magnus expansion. Ultimately, we want the solution (2.2). For that, we need the exponential, $\exp(\boldsymbol{\Omega}(t))$. This is an opportunity to show how automorphisms can simplify exponentials arising in master equations.

Let P be the $(N+1)\times(N+1)$ *persymmetric identity*: $P_{i,j}=1$ if $j=N-i$, and is zero otherwise (with indexing beginning at 0). Note that $P\in\mathrm{O}(N+1)\cap\mathrm{Sym}(N+1)$ so P is an *orthogonal involution:* $P^{-1}=P^{\top}=P$ and $P^2=I$. Matrix multiplication confirms the useful properties

$$PA^{[0]}P=A^{[0]},\qquad PA^{[1]}P=-A^{[1]}. \tag{2.17}$$

Being an orthogonal involution, P defines an inner automorphism on $\mathfrak{gl}(N+1)$, namely $\imath(B)=PBP$ for $B\in\mathfrak{gl}(N+1)$. Following [37], we let $\mathfrak{k}=\{B\in\mathfrak{gl}(N+1)\,:\,\imath(B)=B\}$ and $\mathfrak{p}=\{B\in\mathfrak{gl}(N+1)\,:\,\imath(B)=-B\}$ be the *fixed points* and *anti-fix points* of the automorphism $\imath$. Here is a list of the three main features of our general strategy. First, in the *Generalised Cartan Decomposition*, $\mathfrak{gl}(N+1)=\mathfrak{k}\oplus\mathfrak{p}$. That is, given $B\in\mathfrak{gl}(N+1)$, we split it into $\frac{1}{2}[B+\imath(B)]\in\mathfrak{k}$ and $\frac{1}{2}[B-\imath(B)]\in\mathfrak{p}$. Second, here $\mathfrak{k}$ is a subalgebra of $\mathfrak{gl}(N+1)$, while $\mathfrak{p}$ is a *Lie triple system:* $[\mathfrak{k},\mathfrak{k}],[\mathfrak{p},\mathfrak{p}]\subseteq\mathfrak{k}$ and $[\mathfrak{k},\mathfrak{p}],[\mathfrak{p},\mathfrak{k}]\in\mathfrak{p}$. Third, letting $B=k+p$, where $k\in\mathfrak{k}$ and $p\in\mathfrak{p}$, we have (and we will apply this form to our example momentarily)

$$\mathrm{e}^{tB}=\mathrm{e}^{X}\mathrm{e}^{Y},$$

where $X\in\mathfrak{k}$, $Y\in\mathfrak{p}$ have the Taylor expansion

$$\begin{aligned}X=\,&tp-\frac{1}{2}t^2[p,k]-\frac{1}{6}t^3[k,[p,k]]+t^4\left(\frac{1}{24}[p.[p,[p.k]]]-\frac{1}{24}[k,[k,[p,k]]]\right)\\&+t^5\left(\frac{7}{360}[k,[p,[p,[p,k]]]]-\frac{1}{120}[k,[k,[k,[p,k]]]]-\frac{1}{180}[[p,k],[p,[p,k]]]\right)\\&+t^6\left(-\frac{1}{240}[p,[p,[p,[p,[p,k]]]]]+\frac{1}{180}[k,[k,[p,[p,[p,k]]]]]\right.\\&-\frac{1}{720}[k,[k,[k,[k,[p,k]]]]]+\frac{1}{720}[[p,k],[k,[p,[p,k]]]]\\&\left.+\frac{1}{180}[[p,[p,k]],[k,[p,k]]]\right)+\mathcal{O}\big(t^7\big)\,,\end{aligned} \tag{2.18}$$

$$\begin{aligned}Y=\,&tk-\frac{1}{12}t^3[p,[p,k]]+t^5\left(\frac{1}{120}[p,[p,[p,[p,k]]]]+\frac{1}{720}[k,[k,[p,[p,k]]]]\right.\\&\left.-\frac{1}{240}[[p,k],[k,[p,k]]]\right)+\mathcal{O}\big(t^7\big)\,.\end{aligned} \tag{2.19}$$

Now, let $k=A^{[0]}$ and $p=A^{[1]}$ so by (2.7), $[p,k]=2p$. Look again at (2.18) and (2.19). Each term necessarily contains the commutator $[p,k]$. Suppose that, except for this commutator, the term contains at least one additional p. Then, necessarily, it is zero. The reason is there must be a sub-term of the form $[p,[k,[k,[\ldots,[k,[p,k]]\cdots]]]]$. Beginning from the inner bracket, we replace $[p,k]$ by $2p$, so $[k,[p,k]]=-4p$, and so on, until we reach the outermost commutator: up to a power of 2, it will be $[p,p]=0$, proving our

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

assertion. We deduce that the only terms surviving in (2.18), except for the first, are of the form (where in this line we are also introducing an adjoint operator notation ad_k^{r+1}, to simplify expressions with nested commutators)

$$\overbrace{[k,[k,\cdots,k}^{r\geqslant 0 \text{ times}},[p,k]]] = -\mathrm{ad}_k^{r+1}p = (-1)^r 2^{r+1}p$$

so

$$X = -\sum_{r=1}^{\infty}\frac{t^r}{r!}\mathrm{ad}_k^{r-1}p = \frac{1-\mathrm{e}^{-2t}}{2}p. \tag{2.20}$$

Insofar as Y is concerned, things are even simpler. While p features an odd number of times in X (because $X \in \mathfrak{k}$), $Y \in \mathfrak{p}$ implies that p features there an even number of times. Except for the leading term, it features $\mathfrak{p}$ at least twice, and each such term must vanish, so

$$Y = tk. \tag{2.21}$$

Of course, what we really need to compute is $\exp(\boldsymbol{\Omega}(t)) = \exp(tA^{[0]} + \sigma(t)A^{[1]}) = \mathrm{e}^{tB} = \mathrm{e}^X\mathrm{e}^Y$. For that, we keep (2.21) intact (hence $Y = tA^{[0]}$), but t in (2.20) need be replaced by $\sigma(t)/t$ (which is not problematic since $\sigma(0) = 0$), i.e.,

$$X = \frac{1}{2}\left[1 - \exp\left(-\frac{2\sigma(t)}{t}\right)\right]A^{[1]}.$$

Thus automorphisms have simplified the required $\exp(tA^{[0]} + \sigma(t)A^{[1]})$ to computing exponentials of $A^{[0]}$ and of $A^{[1]}$ separately. Those come from the spectral decomposition, which we set about finding next.

3 Spectra and pseudospectra of isomerisation matrices

3.1 Spectral decomposition of $A^{[0]}$

We wish first to determine the eigenvalues and eigenvectors of $A^{[0]}$. They are essentially given by [11, Theorem 2.1]. Here, we provide an alternative proof and an explicit formula for the eigenvectors.

Theorem 3.1 *The spectrum of* $A^{[0]}$ *is*

$$\{-2r \,:\, r = 0,1,\ldots,N\}.$$

Moreover, an (unnormalised) eigenvector corresponding to the eigenvalue $-2r$*, for* $r = 0,\ldots,N$*, is*

$$v_m = (-1)^m\binom{r}{m}{}_2F_1\left[\begin{matrix}-N+r,-m;\\ r-m+1;\end{matrix}\,-1\right], \qquad m = 0,\ldots,r, \tag{3.1}$$

$$v_m = (-1)^r\binom{N-r}{m-r}{}_2F_1\left[\begin{matrix}-N+m,-r;\\ m-r+1;\end{matrix}\,-1\right], \qquad m = r,\ldots,N. \tag{3.2}$$

where ${}_kF_\ell$ *is the generalized hypergeometric function.*

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

Proof By definition, λ is an eigenvalue of $A^{[0]}$ and $\boldsymbol{v} \neq \boldsymbol{0}$ a corresponding eigenvector if and only if

$$(N+1-m)v_{m-1} - (N+\lambda)v_m + (m+1)v_{m+1} = 0, \qquad m = 0, \ldots, N, \tag{3.3}$$

with the boundary conditions $v_{-1} = v_{N+1} = 0$. One way to arrive at the theorem is to let

$$\mathcal{V}(t) := \sum_{m=0}^{N} v_m t^m$$

and establish $\mathcal{V} = (1+t)^{N+\lambda/2}(1-t)^{-\lambda/2}$ using (3.3). Then impose conditions on λ to ensure $\mathcal{V}$ is a polynomial of degree N. The exact details of the eigenvectors $\boldsymbol{v}$ can come by expanding $(1+t)^{N+\lambda/2}(1-t)^{-\lambda/2}$. □

Incidentally, (3.1)–(3.2) reveal symmetry. Denoting the eigenvector corresponding to the eigenvalue $-2r$ by $\boldsymbol{v}^{[r]}$, we have: $v^{[r]}_{N-m} = (-1)^{m-r} v^{[N-r]}_m$, $m = 0, \ldots, N$.

What else can we say about the eigenvector matrix $V = [\boldsymbol{v}^{[0]}, \boldsymbol{v}^{[1]}, \ldots, \boldsymbol{v}^{[N]}]$? Computer experiments seem to demonstrate the remarkable result $V^2 = 2^N I$, hence

$$V^{-1} = 2^{-N} V \tag{3.4}$$

and this is true: for brevity we omit the proof. More importantly, having the spectral decomposition and having V^{-1}, we now have the exponential, exactly:

$$\mathrm{e}^{tA^{[0]}} = \frac{1}{2^N} V \Lambda(t) V, \qquad \text{where} \qquad \Lambda(t) = \operatorname{diag}\left(1, \mathrm{e}^{-2t}, \mathrm{e}^{-4t}, \cdots, \mathrm{e}^{-2Nt}\right).$$

It is tempting to compute matrix exponentials via diagonalization. In general, this is not necessarily a good numerical choice, even in situations where the spectral decomposition is cheaply available. An issue is that the condition number of the eigenvector matrix can be very large, as happens here[3] — $\kappa(V)$ grows quickly with N. Also, expressions such as e^{-2Nt} are at risk of underflow error.

3.2 A Jordan form of $A^{[1]}$

Unlike $A^{[0]}$, the matrix $A^{[1]}$ is not diagonalizable. It can still be usefully factorized in

Theorem 3.2 *The Jordan form of* $A^{[1]}$ *is*

$$A^{[1]} = W E W^{-1}, \tag{3.5}$$

[3] In hindsight, such poor conditioning of the eigenvector matrix was to be expected because $A^{[0]}$ exhibits a humongous pseudospectrum. The best case scenario is when eigenvectors form an orthogonal basis (consistent with our intuition from numerical linear algebra that orthogonal matrices have the ideal condition number of 1), as happens in the real symmetric case. Pseudospectra measures the departure of a *nonnormal matrix* from that good orthogonal case. Our example has eigenvectors in Theorem 3.1 that are far from orthogonal.

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

where E is the standard shift matrix, *with $E_{i,j} = 1$ if $j = i+1$ and is zero otherwise, while W is a lower-triangular matrix,*

$$W_{m,n} = \begin{cases} 0, & m \leqslant n-1, \\ \dfrac{(-1)^{m-n}}{n!}\dbinom{N-n}{m-n}, & m \geqslant n, \end{cases} \qquad m,n = 0,\dots,N.$$

An immediate consequence of this Jordan form (3.5) *is that $A^{[1]}$ is nilpotent.*

Proof The Jordan form (3.5) is equivalent to $A^{[1]}W = WE$ and the latter is easier to check. The matrix WE is easy to find because E is the shift matrix: each column of W is shifted rightwards, the Nth column disappears, and the zeroth column is replaced by zeros, so

$$(WE)_{m,n} = \begin{cases} 0, & n = 0, \\ W_{m,n-1}, & n = 1,\dots,N. \end{cases}$$

We proceed to evaluate $A^{[1]}W$ and demonstrate that it is the same.

For every $m,n = 0,\dots,N$ (and with $A^{[1]}_{0,-1} = A^{[1]}_{N,N+1} = 0$) we have

$$(A^{[1]}W)_{m,n} = A^{[1]}_{m,m-1}W_{m-1,n} + A^{[1]}_{m,m}W_{m,n} + A^{[1]}_{m,m+1}W_{m+1,n}.$$

For $n \geqslant m+2$ this obviously vanishes. For $n = m+1$, $A^{[1]}_{m,m+1}W_{m+1,m+1} = \frac{1}{m!} = W_{m,m}$ is all that survives, and for $n = m$

$$A^{[1]}_{m,m}W_{m,m} + A^{[1]}_{m,m+1}W_{m+1,m} = -m\frac{N+1-m}{m!} = \begin{cases} 0, & m = 0, \\ W_{m,m-1}, & m \geqslant 1. \end{cases}$$

Finally, for $n \leqslant m-1$ all three terms are nonzero and their sum is

$$\begin{aligned} &(-N+m-1)\frac{(-1)^{m-1-n}}{n!}\binom{N-n}{m-1-n} + (N-2m)\frac{(-1)^{m-n}}{n!}\binom{N-n}{m-n} \\ &\quad + (m+1)\frac{(m+1-n)}{n!}\binom{N-n}{m+1-n} \\ &= \frac{(-1)^{m-n}n(N-n+1)!}{n!(m-n+1)!(N-m)!} = \begin{cases} 0, & n = 0, \\ W_{m,n-1}, & n \geqslant 1 \end{cases} \end{aligned}$$

and we are done. □

Next, we set about applying our newly found Jordan form to find the matrix exponential. Let $C = \operatorname{diag}(0!, 1!, 2!, \dots, N!)$ be a diagonal matrix and

$$Z_{m,n} = \begin{cases} 0, & m \leqslant n-1, \\ (-1)^{m-n}\dbinom{N-n}{m-n}, & m \geqslant n, \end{cases} \qquad m,n = 0,\dots,N.$$

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

As is trivial to verify, $W = ZC^{-1}$, so $A^{[1]} = ZC^{-1}ECZ^{-1}$. Equally trivial to verify is that Z^{-1} is given by

$$Z^{-1} = \tilde{Z}_{m,n} := \begin{cases} 0, & m \leqslant n-1, \\ \binom{N-n}{m-n}, & m \geqslant n, \end{cases} \qquad m,n = 0,\dots,N.$$

Consequently, $A^{[1]} = ZC^{-1}EC\tilde{Z}$. We have proved

Theorem 3.3 *The matrix exponential is, in an explicit form,*

$$\mathrm{e}^{tA^{[1]}} = ZC^{-1}\mathrm{e}^{tE}C\tilde{Z}. \tag{3.6}$$

3.2.1 *Evaluating the exponential via* (3.6)

Let $\boldsymbol{u} \in \mathbb{R}^{N+1}$ (again, indexed from zero). We wish to compute $\boldsymbol{y} = \tilde{Z}\boldsymbol{u}$. A naïve approach would require $\mathcal{O}(N^2)$ flops but herewith an algorithm that accomplishes this in just $\mathcal{O}(N^2)$ *additions,* without requiring multiplications!

For reasons that become clear, it is useful to indicate N explicitly in the notation, i.e., $\boldsymbol{y}^{[N]} = \tilde{Z}^{[N]}\boldsymbol{u}^{[N]}$. Start by observing that

$$y_m^{[N]} = \sum_{n=0}^{m} \binom{N-n}{m-n} u_n, \qquad m = 0,\dots,N$$

(no need to place superscripts on u_n). Therefore, for $m = 0,\dots,N-1$,

$$\begin{aligned} y_m^{[N]} + y_{m+1}^{[N]} &= \sum_{n=0}^{m} \binom{N-n}{m-n} u_n + \sum_{n=0}^{m+1} \binom{N-n}{m+1-n} u_n = \sum_{n=0}^{m+1} \binom{N+1-n}{m+1-n} u_n \\ &= y_{m+1}^{[N+1]}. \end{aligned}$$

Rewrite this as

$$y_m^{[N]} = y_{m-1}^{[N-1]} + y_m^{[N-1]}, \qquad m = 0,\dots,N-1 \tag{3.7}$$

(in the case $m = 0$ of course $y_0^{[N]} = y_0^{[N-1]} = u_0$, so the above is consistent with $y_{-1}^{[N]} = 0$.) Now proceed from $y_0^{[0]} = u_0$ and then, for $M = 1,2,\dots,N$, add

$$\begin{aligned} y_m^{[M]} &= y_{m-1}^{[M-1]} + y_m^{[M-1]}, \qquad m = 0,\dots,M-1, \\ y_M^{[M]} &= \sum_{n=0}^{M} u_n = y_{M-1}^{[M-1]} + u_M. \end{aligned}$$

and we are done.

Of course, similar reasoning applies also to a product $\boldsymbol{y} = Z\boldsymbol{u}$. The only difference vis-á-vis (3.7) is that now $y_m^{[N]} = y_m^{[N-1]} - y_{m-1}^{[N-1]}$, $m = 0,\dots,N-1$, therefore the recursion

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

steps are

$$y_m^{[M]} = y_m^{[M-1]} - y_{m-1}^{[M-1]}, \qquad m = 0, \ldots, N-1,$$
$$y_M^{[M]} = \sum_{m=0}^{M} (-1)^{M-n} u_n = -y_{M-1}^{[M-1]} + u_N.$$

Having dealt with the $\tilde{Z}\boldsymbol{u}$ and the $Z\boldsymbol{u}$ components, we are left only with the $C^{-1}\mathrm{e}^{tE}C$ portion of (3.6). We address that now. It is trivial that

$$(\mathrm{e}^{tE})_{m,n} = \begin{cases} \dfrac{t^{n-m}}{(n-m)!}, & m = 0, \ldots, n, \\ 0, & m = n+1, \ldots, N. \end{cases}$$

Therefore (*cf.* (3.6))

$$(C^{-1}\mathrm{e}^{tE}C)_{m,n} = \begin{cases} \dbinom{n}{m} t^{n-m} & m = 0, \ldots, n, \\ 0, & m = n+1, \ldots, N. \end{cases}$$

Let us pause to reflect on the exact exponentials that we have just found. We expect the solution to our model of isomerization to be a binomial distribution [26]. In general, that means we expect a linear combination of the columns of the solution matrix $\exp(\Omega(t))$ to be a binomial distribution, when the weights in that linear combination likewise come from a binomial distribution. Perhaps the simplest example is that the first column of the solution of (1.2) must be a binomial distribution.

As an example, set $\boldsymbol{e}_0 = (1, 0, \ldots, 0)^\top$ and compute the leading column, $\mathrm{e}^{qA^{[1]}}\boldsymbol{e}_0 = Z(C^{-1}\mathrm{e}^{qE}C)\tilde{Z}\boldsymbol{e}_0$. Notice that $(\tilde{Z}\boldsymbol{e}_0)_m = \tilde{Z}_{m,0} = \binom{N}{m}$. So we have $[(C^{-1}\mathrm{e}^{qE}C)\tilde{Z}\boldsymbol{e}_0]_m = \binom{N}{m}\sum_{n=0}^{N-m}\binom{N-m}{n}t^n = \binom{N}{m}(1+q)^{N-m}$ and after some simplifications,

$$(\mathrm{e}^{qA^{[1]}}\boldsymbol{e}_0)_m = [Z(C^{-1}\mathrm{e}^{qE}C)\tilde{Z}\boldsymbol{e}_0] = (-1)^m\binom{N}{m}q^m(1+q)^{N-m}.$$

We are seeing on the right that the binomial distribution survives the first term in $X(t) = \mathrm{e}^{tA^{[0]}}\mathrm{e}^{qA^{[1]}}\boldsymbol{e}_0$, where $q = \sigma(t)/t$. Thus, the explicit forms of our exponentials that we have derived allow us to confirm the 'binomial stays binomial' theorem [26].

3.3 Pseudospectra

Having established exact analytic formulæ for spectral decomposition, we are now in a good position to compare exact spectra to numerical estimates of the *pseudospectra* [43]. Two striking contrasts between the numerically computed eigenvalues and the exact eigenvalues are worth pointing out.

First, we proved the matrix $A^{[1]}$ is nilpotent: *exact eigenvalues are precisely zero*. Nonetheless, $A^{[1]}$ has an enormous pseudospectrum, and standard numerical methods lead to wrongly computed non-zero eigenvalues of a large magnitude.

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

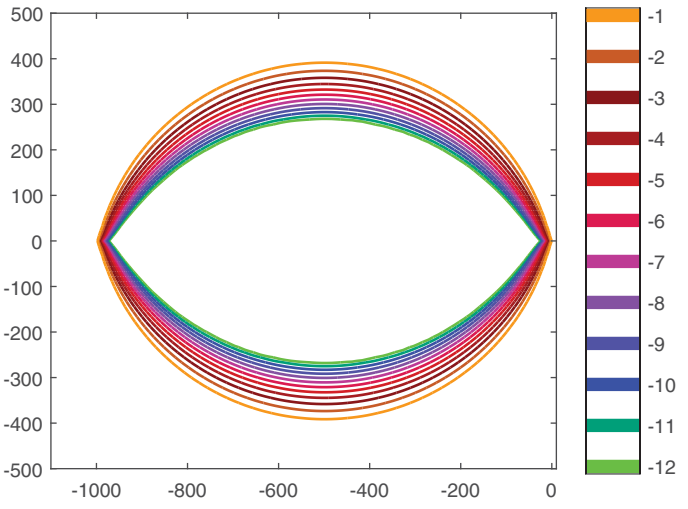


FIGURE 1. An 'almond eye:' Pseudospectrum [43] of a 500×500 example of the $A^{[0]}$ matrix, defined in (1.3), as computed by Eigtool [46]. Contours of the minimum singular value, $s_{\min}(zI - A)$, are displayed on a log scale.

Second, we found the eigenvalues of $A^{[0]}$ in Theorem 3.1, and they are *purely real*. (Indeed, the same ideas described by Trefethen and Embree [43] also show our $A^{[0]}$ is similar to a real symmetric matrix, so even before Theorem 3.1, we knew eigenvalues had to be real.) However, standard numerical methods to compute the eigenvalues wrongly produce complex numbers (!) with very large imaginary parts.

The reason for the numerical errors in computing the eigenvalues is that the eigenvalues of these matrices are very sensitive to small perturbations. That phenomenal sensitivity is often characterised by the pseudospectra. For $\epsilon > 0$, the ϵ-pseudospectrum is the region of the complex plane, $z \in \mathbb{C}$, where the norm of the *resolvent* is large: $||(zI - A)^{-1}|| > 1/\epsilon$. In the 2-norm, this is equivalent to the region where the minimum singular value, $s_{\min}$, is small: $s_{\min}(zI - A) < \epsilon$.

The pseudospectrum of the convection-diffusion operator is known to be significant [39], and master equations are closely related to convection-diffusion, suggesting they will also exhibit interesting pseudospectra. Indeed, the matrices that arise in our applications of master equations to isomerizaiton exhibit truly humongous pseudospectra. They are examples of the class of *twisted Toeplitz matrices and operators*, which have recently been understood to exhibit distinctive pseudospectra, captivating more general interest [42].

Figure 1 displays the pseudospectrum for $A^{[0]}$ and Figure 2 displays the pseudospectrum for $A^{[1]}$. These are numerical estimates based on the algorithms underlying *eigtool*. In future work it may be possible to analytically bound the region of the complex plane where the pseudospectra are large. For example, the pseudospectrum of the convection-diffusion operator has been shown to be approximately bounded by a parabola [39], and such knowledge of this bounded region has recently been exploited to develop effective contour integral methods based on inverse Laplace transform techniques. Usually the idea of such methods is to choose a contour that stays away from the eigenvalues

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

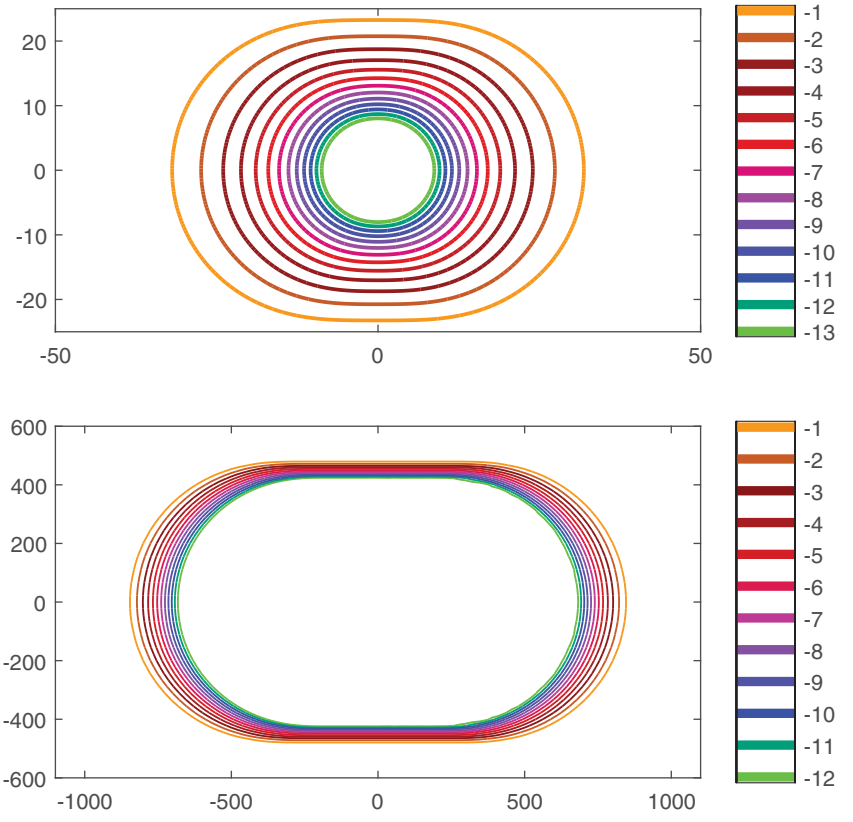


FIGURE 2. The 'athletics track:' Pseudospectrum [43] of the $A^{[1]}$ matrix, defined in (1.3), as computed by Eigtool [46]. Top: 30×30. Bottom: 500×500.

[31, Figure 1]. That works well for real symmetric matrices. But if the operator has a significant pseudospectrum, then more is required: the contour must stay safely away from regions where the resolvent $||(zI - A)^{-1}||$ is large. The figures here show some diversity in pseudospectra. This might inspire research into a computational method that is *adaptive*: instead of requiring detailed knowledge of the pseudospectrum in advance, we require computational methods that adapt the contour of integration so as to control $||(zI - A)^{-1}||$ to be, say, $\mathcal{O}(1)$.

4 Discussion

Master equations and increasingly their applications continue to bring new directions in scientific computation [34]. Monte Carlo approaches, such as Gillespie-like algorithms, are attractive alternatives to directly solving master equations, especially when confronting the grand challenge of high dimensions. To demonstrate Magnus methods, we only considered the matrix case in this article, not the case of operators, but the operator case is important for many models with unbounded state spaces. Anderson & Kurtz give special attention to 'first order systems' [4] where the analysis is more tractable, and our own success here is certainly in large part because the isomerization model is an example of that class [26].

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

Here is an incomplete list of contemporary topics where activity is growing fast.

4.1 Matrix functions of graph Laplacians

A general framework for models of biochemical kinetics has recently been elucidated in terms of graph Laplacians [17]. A simple example of a graph Laplacian on a line of nodes appears in [40], and, like the matrix exponential, it has been shown that a *Mittag-Leffler function* [16] of a graph Laplacian matrix is also a stochastic matrix [33]. All of this suggests research into non-Markovian generalisations of Gillespie-like stochastic simulation algorithms allowing waiting times not exclusively drawn from an exponential distribution [31].

It is known that if we generalise (2.1) to a Caputo fractional derivative of order $0 < \alpha < 1$, $\mathrm{d}^\alpha/\mathrm{d}t^\alpha$, then the matrix exponential is generalised to the Mittag–Leffler function E_α, so that (2.1) becomes $\mathrm{d}^\alpha \boldsymbol{p}/\mathrm{d}t^\alpha = \mathbb{A}\boldsymbol{p}$ with solution $\boldsymbol{p}(t) = E_\alpha(t^\alpha \mathbb{A})\boldsymbol{p}(0)$. This is assuming the coefficient matrix is constant. However, if we allow a time-varying matrix, $\mathbb{A} = \mathbb{A}(t)$, and generalise (2.2) to $\mathrm{d}^\alpha \boldsymbol{p}/\mathrm{d}t^\alpha = \mathbb{A}(t)\boldsymbol{p}$, then an important open question arises: how do we generalise the Magnus expansion of the solution? There is certainly some work in the literature on discrete constructions of continuous-time random walks and their generalised master equations aimed at accommodating time-varying rates. Nevertheless, the authors are not aware of a *fractional generalisation of the Magnus expansion*. Given the current interest in fractional processes and processes with memory, such a generalisation of the Magnus expansion would seem a timely contribution, and would presumably also suggest a fractional generalisation of the Baker–Campbell–Hausdorff formula as a special case.

4.2 Products of matrix exponentials

When matrices commute, a product of exponentials has an especially simple form. Evans, Sturmfels & Uhler recently showed how to successfully exploit this property for master equations governing birth-death processes [13].

This computational approach has the potential for wider applications to master equations where tensor structures involving shift operators often arise. So, let us revisit (1.2) to find, *explicitly,* solutions (without Wilhelm Magnus and without Sophus Lie) in a way that generalises and suggests connections to products of exponentials. To generalise (1.2), consider linearly independent matrices, A and B, such that

$$[A, B] = aA + bB \tag{4.1}$$

for some $a, b \in \mathbb{R}$, not both zero, and the differential equation

$$X' = [\alpha(t)A + \beta(t)B]X, \quad t \geqslant 0, \qquad X(0) = I. \tag{4.2}$$

Here α and β are given scalar functions.

We wish to prove the solution of (4.2) can be expressed in the form

$$X(t) = \mathrm{e}^{\rho_A(t)A}\mathrm{e}^{\rho_B(t)B}, \tag{4.3}$$

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

where ρ_A and ρ_B are scalar functions obeying a certain ODE. Obviously, $\rho_A(0) = \rho_B(0) = 0$.

Assume (without loss of generality) that $b \neq 0$. Differentiating (4.3) and substituting into (4.2), we have $X' = \mathrm{e}^{\rho_A A}(\rho_A' A + \rho_B' B)\mathrm{e}^{\rho_B B} = (\alpha A + \beta B)\mathrm{e}^{\rho_A A}\mathrm{e}^{\rho_B B}$ and, multiplying on the right by $\mathrm{e}^{-\rho_B B}$, we have

$$(\rho_A' - \alpha)A\mathrm{e}^{\rho_A A} + \rho_B' \mathrm{e}^{\rho_A A} B - \beta B \mathrm{e}^{\rho_A A} = O. \tag{4.4}$$

A proof by induction using (4.1) shows that

$$BA^m = (A + bI)^m B - \frac{a}{b} A[A^m - (A + bI)^m], \qquad m \in \mathbb{Z}_+. \tag{4.5}$$

Consequently, $B\mathrm{e}^{\rho_A A} = \sum_{m=0}^{\infty} \frac{\rho_A^m}{m!} BA^m = \sum_{m=0}^{\infty} \frac{\rho_A^m}{m!}(A + bI)^m B - \frac{a}{b} A \sum_{m=0}^{\infty} \frac{\rho_A^m}{m!}[A^m - (A + bI)^m] = \mathrm{e}^{b\rho_A}\mathrm{e}^{tA} B - \frac{a}{b}(1 - \mathrm{e}^{b\rho_A})A\mathrm{e}^{\rho_A A}$. Now substitute into (4.4), $(\rho_A' - \alpha)A\mathrm{e}^{\rho_A A} + \rho_B' \mathrm{e}^{\rho_A A} B - \beta \mathrm{e}^{b\rho_A}\mathrm{e}^{tA} B + \frac{a}{b}\beta(1 - \mathrm{e}^{b\rho_A})A\mathrm{e}^{tA}$. Separating between $A\mathrm{e}^{\rho_A A}$ and $\mathrm{e}^{\rho_A A} B$ above, we obtain two ODEs for ρ_A and ρ_B,

$$\rho_A' = \alpha - \frac{a}{b}\beta(1 - \mathrm{e}^{b\rho_A}), \qquad \rho_A(0) = 0, \tag{4.6}$$

$$\rho_B' = \beta \mathrm{e}^{b\rho_A}, \qquad \rho_B(0) = 0, \tag{4.7}$$

reducing the computation of ρ_A to a scalar ODE and of ρ_B to quadrature.

Specialising to master equations, $\alpha \equiv 1$, $\beta = f$, $a = 0$ and $b = -2$, so (4.6) becomes $\rho_A(t) = t$ and $\rho_B(t) = \int_0^t \mathrm{e}^{-2\tau} f(\tau)\,\mathrm{d}\tau$. Putting (4.7) in (4.6), we obtain $\rho_A' = \alpha - \frac{a}{b}\beta + \frac{a}{b}\rho_B'$. Multiplication by b and integration implies the integral $b\rho_A(t) - a\sigma(t) = b\int_0^t \alpha(\tau)\,\mathrm{d}\tau - a\int_0^t \beta(\tau)\,\mathrm{d}\tau$.

Can all this be (further) generalised, beyond two exponentials? We now suggest the answer to this question is affirmative although applications form the subject of ongoing research. Indeed what we have done thus far is to exemplify *precisely* the Wei–Norman approach of expressing the solution of a linear ODE using *canonical coordinates of the second kind* [45]. Specifically, let $A : \mathbb{R}_+ \to \mathfrak{g}$, where $\mathfrak{g}$ is a Lie algebra, $\dim \mathfrak{g} = d$, and consider the ODE

$$X' = A(t)X, \quad t \geqslant 0, \qquad X(0) = I. \tag{4.8}$$

Let $\mathcal{P} = \{P_1, P_2, \ldots, P_d\}$ be a basis of $\mathfrak{g}$. Wei & Norman [45] prove that for sufficiently small $t > 0$ there exist functions $g_1, g_2, \ldots, g_d$ such that

$$X(t) = \mathrm{e}^{g_1(t)P_1}\mathrm{e}^{g_2(t)P_2} \cdots \mathrm{e}^{g_d(t)P_d}. \tag{4.9}$$

This is the situation we have in (1.2) or, with greater generality, in (4.2): $P_1 = A$, $P_2 = B$ and, because of (4.1), the dimension of the free Lie algebra spanned by A and B is $d = 2$. Interestingly enough, this example does not feature in [45].

Coordinates of the second kind have been used extensively in the theory of Lie-group integrators [25] where it always followed an organising principle that also shows promise for master equations. Specifically, the assumption was – unlike our simple $d = 2$ example – that d is large (e.g., that $\mathfrak{g}$ is the special orthogonal group of matrices $\mathrm{SO}(n)$, say, or

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

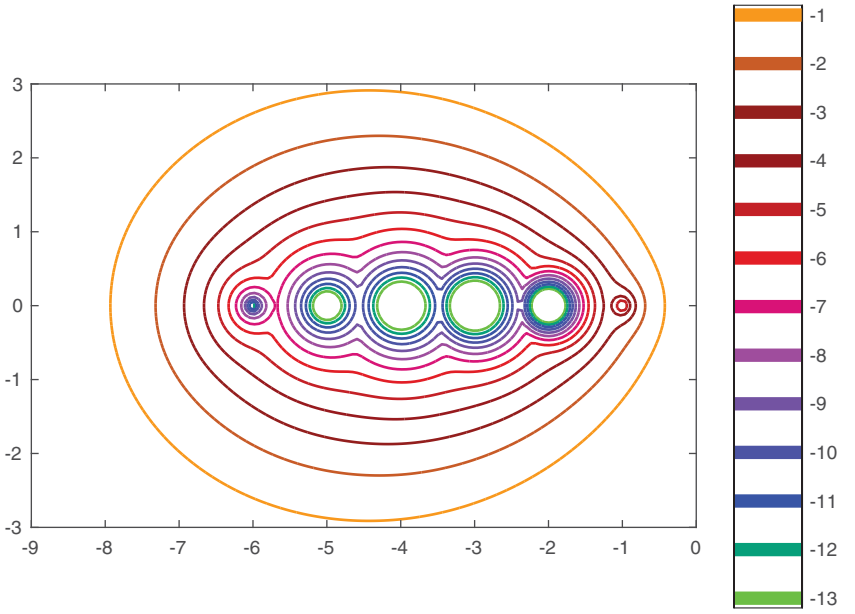


FIGURE 3. The 'seed pod:' Pseudospectrum [43] of a 1513×1513 finite section of the singly infinite matrix associated with a totally asymmetric exclusion process (TASEP) with six particles beginning in a 'step' initial configuration [9, Figure 9, q=0], as computed by Eigtool [46].

the special linear group of matrices $\mathrm{SL}(n)$) and the basis $\mathcal{P}$ selected so that it is easy to evaluate the exponentials $\exp(g_k P_k)$ (e.g., using root space decomposition) [7].

4.3 Pseudospectra of master equations

This is a subject worthy of more attention. For example, we have shown here that even simple isomerisation models exhibit a highly non-trivial pseudospectra. We conjecture that *Michaelis–Menten enzyme kinetics* and a whole host of other important models in biology also exhibit significant pseudospectra [31, 34]. In the usual model of Michaelis–Menten kinetics, a catalytic enzyme E reversibly forms a complex intermediate C with a substrate S, that is eventually irreversibly converted to a product P, viz. $S+E \leftrightarrow C \rightarrow P+E$. There is a need for visualisations of the pseudospectrum of such Michaelis–Menten kinetics, for example. Another open question is how the pseudospectrum of the usual model compares to the pseudospectrum of a more reasonable model suggested by Gunawardena to repent for the 'Original Thermodynamic Sin' of including the irreversible reaction $C \rightarrow P + E$ [18].

As a demonstration of this topic going far beyond merely the isomerisation examples that we have studied here, we have also computed here in Figure 3 the pseudospectrum of the *totally asymmetric exclusion process* (TASEP) [9, Figure 9]. If all that is observed in the picture of the pseudospectrum is merely some '$\epsilon-$balls', centred around each eigenvalue, and well-separated, then the situation is not interesting. For that is simply the picture we would expect for a well-behaved real symmetric matrix anyway. To be interesting, more complex behaviour is required. It is too early to tell for the TASEP, but our preliminary

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

numerical picture here in Figure 3 suggests this is likely to be worthwhile pursuing further. The figure depicts the case with six particles and we can already discern the beginnings of some interesting interactions emerging. Such examples of TASEP models have found applications to single molecule studies of RNA polymerase and protein synthesis. More generally, exclusion processes have witnessed a renaissance of mathematical interest, partly in relation to exactly integrable probabilistic systems, the Kardar–Parisi–Zhang (KPZ) universality class, and the KPZ stochastic partial differential equation [8, 19].

Random Matrix Theory [12] connects to master equations. For example, an important limiting distribution associated with the TASEP master equation is the famous Tracy–Widom distribution for the biggest eigenvalue of a large, random Hermitian matrix [8]. Although less in the sense of the CME (at least so far but that could change) and more in the physicists' sense of Wigner and Freeman Dyson, random matrix theory is also playing a role in recent studies of random graph Laplacians. The resulting distributions are very similar to the standard Gaussian ensembles but the special algebraic properties of graph Laplacians do lead to peculiar discrepancies that persist for large matrix dimension N [41]. Interestingly, the *Matrix-Tree Theorem*, which gives a formula for the stationary distribution (and confirmation of positivity) of such master equations in terms of sums of positive off-diagonal entries, seems yet to be exploited in this random matrix context.

4.4 The Magnus expansion and Kurtz's random time-change representation

Denote the forward rate by $\alpha_f(\boldsymbol{x}(s), s) = c_1(s)n_1$ and the backward rate by $\alpha_b(\boldsymbol{x}(s), s) = c_2(s)n_2$. Here, n_1 and n_2 are the number of molecules of S_1 and S_2, respectively. The Kurtz random time-change representation [29] of the sample paths corresponding to our master equation (1.2) with initial state $\boldsymbol{x}(0)$ is

$$\boldsymbol{x}(t) = \boldsymbol{x}(0) + \begin{pmatrix} -1 \\ +1 \end{pmatrix} Y_1\left(\int_0^t \alpha_f(\boldsymbol{x}(s), s)\,\mathrm{d}s\right) + \begin{pmatrix} +1 \\ -1 \end{pmatrix} Y_2\left(\int_0^t \alpha_b(\boldsymbol{x}(s), s)\,\mathrm{d}s\right).$$

At time t, this stochastic equation has two *internal time frames*: $T_j = \int_0^t \alpha_j(\boldsymbol{x}(s), s)ds$, $j = 1, 2$. Here, Y_1 and Y_2 are independent, unit-rate Poisson processes but dependencies arise through the rates in these internal time-frames. Thus Kurtz and Magnus offer two different representations of the same solution when rates are time-varying. Although much work has appeared on each representation separately, there has been almost no work exploring connections. Such connections would perhaps allow probabilistic interpretations of the Magnus expansion.

More generally, time-varying rates are one way to model *extrinsic noise* (and there are also other approaches not discussed here such as the Nosé–Hoover thermostat), so methods that can accommodate time-varying rates, such as Magnus expansions described here, may find wider applications [20, 22]. Exploring the robustness of master equations to perturbations, including time-varying perturbations, might bring together methods from Magnus-like approaches, pseudospectral studies, and perhaps even stochastic operator approaches [12].

Kurtz's representation has also inspired multi-level Monte Carlo (MLMC) methods to be adapted from the setting of SDEs to the setting of master equations, and in turn this

https://doi.org/10.1017/S0956792518000177 Published online by Cambridge University Press

has led to MLMC methods for estimating the *sensitivity* [3]. It will be interesting to see if *adjoint methods* for sensitivity estimates in the setting of continuous SDEs such as the methods for which Giles and Glasserman won *Risk 'Quant-of-the-Year'* [14] are likewise adaptable to the discrete setting of master equations [28].

4.5 Preserving positivity

Moler and Van Loan discuss more than 19 dubious ways for computing the *matrix exponential* [36]. When such methods are applied to the important class of graph Laplacian matrices — as arise in all master equations and Markov processes, and for which the matrix exponential is provably non-negative and indeed a stochastic matrix — a fundamental question is: do these numerical methods preserve non-negativity? For example, does `MATLAB`'s `expm` function preserve positivity when applied to a graph Laplacian matrix? This question seems especially ripe for research in relation to Krylov-like approximations, Padé-like approximations with scaling and squaring, and recent methods of Al-Mohy and Higham (which are currently the basis of `expm` in `MATLAB`) [1, 2].

We found the complete Magnus expansion for our isomerisation model. Being the full and exact Magnus expansion, it respects the original properties of the system, such as maintaining positivity. Numerical methods in other contexts are often derived by truncation of the Magnus expansion, to a certain prescribed order. In general, *truncation* of the Magnus expansion does not result in the same properties as a graph Laplacian, so positivity is no longer guaranteed. (Although if we are willing to settle for second-order accuracy, then it is possible to truncate so as to maintain these desirable properties.) The issue is that the commutator of two graph Laplacians is not in general a graph Laplacian; it may have negative off-diagonal entries. This observation is motivating ongoing research whose roots are in geometric numerical integration — a subject usually concerned with maintaining equalities — to allow the preservation of *inequalities*, such as preserving positivity.

More generally it has been known for a long time in the context of ODEs that standard numerical methods such as Runge–Kutta methods, usually do not preserve positivity unless they are of first-order accuracy [6]. This also presents a contemporary challenge for Monte Carlo simulation of the sample paths of master equations: the widely used *tau-leap methods* and other analogues of the Euler method or of the Euler–Maruyama method, cannot be guaranteed to preserve positivity. This challenge is motivating much current research appearing on approximations that are able to maintain positivity in these settings, as exemplified in the Kolmogorov Lecture at the most recent World Congress In Probability and Statistics [30].

5 Conclusions

Pafnuty Chebyshev was an academic parent of Markov and today the world has come full circle with Chebyshev polynomials being a useful basis for numerical solvers of Markovian master equations in the quantum world [5]. Here, the adjective 'master' is not used in the sense of an overlord; rather it is in the sense of an ensemble averaging principle that emerges at larger scales from the collective behaviour of the mob of microscopic particles,

each following their own random walk. Edelman and Kostlan take such a walk on 'the road from Kac's matrix to Kac's polynomials' and our own matrix examples $A^{[0]}$ and $A^{[1]}$ of (1.3) also lie at the end of that road, being almost the '*Kac matrix*' (as named by Olga Taussky and John Todd) and 'anti-Kac matrix' [11]. These matrices have served us well as wonderful running examples to illustrate new directions in research for Markov processes. Kac did not foresee our applications to isomerisation, nor the way those isomerisation master equations are so naturally amenable to Magnus expansions. This motivated us to derive a number of original results, as outlined in the introduction. Additionally, we believe the methods in the four step procedure described around (2.5) for the Magnus expansion in (2.6) will find wider applicability to Markov processes with time-varying rates. Similarly, these and other applications that we have surveyed, such as the inchoate subject of the pseudospectra of master equations, no doubt have a bright future that we have yet to fully imagine.

Acknowledgments

Arieh Iserles presented some of the results of this article in a workshop at the Oberwolfach (`https://na.math.kit.edu/marlis/research/meetings/16-oberwolfach/Iserles.pdf`) and thanks participants for interest in the geometry of master equations. He also acknowledges a fruitful discussion with Nick Trefethen on the pseudospectrum of $A[1]$.

References

[1] Al-Mohy, A. H. & Higham, N. J. (2009) A new scaling and squaring algorithm for the matrix exponential. *SIAM J. Matrix Anal. Appl.* **31**(3), 970–989.

[2] Al-Mohy, A. H. & Higham, N. J. (2011) Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM J. Sci. Comp.* **33**, 488–511.

[3] Anderson, D. (2012) An efficient finite difference method for parameter sensitivities of continuous time Markov chains. *SIAM J. Numer. Anal.* **50**, 2237–2258.

[4] Anderson, D. & Kurtz, T. (2011) Continuous time Markov chain models for chemical reaction networks. In: H. Koeppl, G. Setti, M. di Bernardo, D. Densmore (editors), *Design and Analysis of Biomolecular Circuits*, New York: Springer.

[5] Barker, J. R., Nguyen, T. L., Stanton, J. F., Aieta, M. C. C., Gabas, F., Kumar, T. J. D., Li, C. G. L., Lohr, L. L., Maranzana, A., Ortiz, N. F., Preses, J. M. & Stimac, P. J. (2016) *Multiwell-2016 Software Suite*, Technical Report, University of Michigan, Ann Arbor, Michigan, USA.

[6] Bolley, C. & Crouzeix, M. (1978) Conservation de la positivité lors de la discrétisation des problèmes d'évolution paraboliques. *RAIRO Anal. Numér.* **12**, 237–245.

[7] Celledoni, E. & Iserles, A. (2001) Methods for the approximation of the matrix exponential in a Lie-algebraic setting. *IMA J. Numer. Anal.* **21**, 463–488.

[8] Corwin, I. (2014) Macdonald processes, quantum integrable systems and the Kardar–Parisi–Zhang universality class. In: *Proceedings of the International Congress of Mathematicians.*

[9] Corwin, I. (2016) Kardar–Parisi–Zhang Universality. Notices of the American Mathematical Society, March 2016 *Not. AMS* **63**.

[10] Drawert, B., Trogdon, M., Toor, S., Petzold, L. & Hellander, A. (2016) Molns: A cloud platform for interactive, reproducible, and scalable spatial stochastic computational experiments in systems biology using pyurdme. *SIAM J. Sci. Comput.* **38**, C179–C202.

[11] Edelman, A. & Kostlan, E. (1994) *The Road from Kac's Matrix to Kac's Random Polynomials*, Technical Report, University of California, Berkeley.
[12] Edelman, A. & Rao, N. R. (2005) Random matrix theory. *Acta Numer.* **14**, 233–297.
[13] Evans, S. N., Sturmfels, B. & Uhler, C. (2010) Commuting birth-and-death processes. *Ann. Appl. Probab.* **20**, 238–266.
[14] Giles, M. & Glasserman, P. (2006) Smoking adjoints: Fast Monte Carlo Greeks. *Risk* **19**(1), 88–92.
[15] Gillespie, D. T. (2002) The chemical Langevin and Fokker–Planck equations for the reversible isomerization reaction. *J. Phys. Chem. A* **106**, 5063–5071.
[16] Gorenflo, R., Kilbas, A., Mainardi, F. & Rogosin, S. (2014) *Mittag-Leffler Functions, Related Topics and Applications*, New York: Springer.
[17] Gunawardena, J. (2012) A linear framework for time-scale separation in nonlinear biochemical systems. *PLoS One* **7**, e36321.
[18] Gunawardena, J. (2014) Time-scale separation: Michaelis and Menten's old idea, still bearing fruit. *FEBS J.* **281**, 473–488.
[19] Hairer, M. (2014) Singular stochastic PDEs. In: *Proceedings of the International Congress of Mathematicians.*
[20] Hellander, A., Klosa, J., Lötstedt, P. & MacNamara, S. (2017) Robustness analysis of spatiotemporal models in the presence of extrinsic fluctuations. *SIAM J. Appl. Math.* **77**(4), 1157–1183.
[21] Higham, D. J. (2008) Modeling and simulating chemical reactions. *SIAM Rev.* **50**, 347–368.
[22] Hilfinger, A. & Paulsson, J. (2011) Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc. Acad. Natl. Sci.* **109**, 12167–72.
[23] Hilgers, P. V. & Langville, A. N. (2006) The five greatest applications of Markov chains. In: *Proceedings of the Markov Anniversary Meeting*, Boston Press, Boston, MA.
[24] Hochbruck, M. & Lubich, C. (2003) On Magnus integrators for time-dependent Schrödinger equations. *SIAM J. Numer. Anal.* **41**, 945–963.
[25] Iserles, A., Munthe-Kaas, H. Z., Nørsett, S. P. & Zanna, A. (2000) Lie-group methods. *Acta Numer.* **9**, 215–365.
[26] Jahnke, T. & Huisinga, W. (2007) Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Biol.* **54**, 1–26. cited By 97.
[27] Kac, M. (1957) *Probability and Related Topics in Physical Sciences, Summer Seminar in Applied Mathematics*, American Mathematical Society, Boulder, Colorado.
[28] Kormann, K. & MacNamara, S. (2016) Error control for exponential integration of the master equation, Technical Report.
[29] Kurtz, T. (1980) Representations of Markov processes as multiparameter time changes. *Ann. Probab.* **8**, 682–715.
[30] Leite, S. C. & Williams, R. J. (2016) A constrained Langevin approximation for chemical reaction networks. *Kolmogorov Lecture, 9th World Congress In Probability and Statistics*, Toronto, `http://www.math.ucsd.edu/ williams/biochem/biochem.pdf`, (2016).
[31] Macnamara, S. (2015) Cauchy integrals for computational solutions of master equations. *ANZIAM J.* **56**, 32–51.
[32] MacNamara, S., Burrage, K. & Sidje, R. (2008) Multiscale modeling of chemical kinetics via the master equation. *SIAM Multiscale Model. Sim.* **6**, 1146–1168.
[33] MacNamara, S., Henry, B. I. & McLean, W. (2016) Fractional Euler limits and their applications. *SIAM J. Appl. Math.*
[34] MacNamara, S. & Strang, G. (2015) Master equations in 'Essays on New Directions in Numerical Computation', `http://tobydriscoll.net/newdirections2015/`, `http://tobydriscoll.net/newdirections2015/LNT60Essays.pdf`.
[35] Magnus, W. (1954) On the exponential solution of differential equations for a linear operator. *Comm. Pure Appl. Math.* **7**, 649–673.

[36] Moler, C. & Loan, C. V. (2003) Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.* **45**, 3–49.
[37] Munthe-Kaas, H. Z., Quispel, G. R. W. & Zanna, A. (2001) Generalized polar decompositions on Lie groups with involutive automorphisms. *Found. Comput. Math.* **1**, 297–324.
[38] Pavliotis, G. A. & Stuart, A. (2008) *Multiscale Methods: Averaging and Homogenization*, New York: Springer.
[39] Reddy, S. C. & Trefethen, L. N. (1994) Pseudospectra of the convection-diffusion operator. *SIAM J. Appl. Math.*
[40] Strang, G. & MacNamara, S. (2014) Functions of difference matrices are Toeplitz plus Hankel. *SIAM Rev.* **56**, 525–546.
[41] Timm, C. (2009) Random transition-rate matrices for the master equation. *Phys. Rev. E* **80**, 021140, New Jersey.
[42] Trefethen, L. N. & Chapman, S. J. (2004) Wave packet pseudomodes of twisted Toeplitz matrices. *Comm. Pure Appl. Math.* **57**, 1233–1264.
[43] Trefethen, L. N. & Embree, M. (2005) *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press.
[44] Weber, M. F. & Frey, E. (2016) Master equations and the theory of stochastic path integrals, *Rep Prog Phys.* 2017 Apr; **80**(4):046601. doi: 10.1088/1361-6633/aa5ae2.
[45] Wei, J. & Norman, E. (1964) On global representations of the solutions of linear differential equations as a product of exponentials. *Proc. Amer. Math. Soc.* **15**, 327–334.
[46] Wright, T. G. (2002) *Eigtool*, `http://www.comlab.ox.ac.uk/pseudospectra/eigtool/`, 2002.