MAGNUS EXPANSIONS AND PSEUDOSPECTRA OF MASTER EQUATIONS

ARIEH ISERLES* AND SHEV MACNAMARA†

Abstract. New directions in research on master equations are showcased by example. Magnus expansions, time-varying rates, and pseudospectra are highlighted. Exact eigenvalues are found and contrasted with the large errors produced by standard numerical methods in some cases. Isomerisation provides a running example and an illustrative application to chemical kinetics. We also give a brief example of the totally asymmetric exclusion process.

Key words. graph Laplacian, Kirchhoff, Matrix-Tree Theorem, pseudospectra, isomerisation, master equation, Gillespie Stochastic Simulation Algorithm, Magnus expansion, Lie algebra

AMS subject classifications. 60J28, 60H35, 65F40, 65F99, 65L15 , 65FL99, 92C40

1. Introduction.

The term 'master equation' goes back at least as far as the work of Kac in the middle of the twentieth century [26, page 105], and the subject of master equations admits a Feynman–Kac stochastic path integral formulation [43]. 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 and therefore be non-Markovian but the most common interpretation of master equations is as 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 [22]. Choosing the simplest examples, we describe applications to exclusion processes and chemical processes, although the computational methods we present have wider applicability.

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 reaction rate equations 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 [31, 9, 37]. Stochastic differential equations such as the Langevin equation for isomerisation [14] 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.

*Department of Applied Mathematics and Mathematical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK (a.iserles@damtp.cam.ac.uk)

†School of Mathematics and Statistics, University of New South Wales (UNSW) (shev.mac@gmail.com)

The reaction rate equations for this model of isomerisation are the two ordinary differential equations (ODEs)

$$\frac{\mathrm{d}}{\mathrm{d}t}[S_1] = -c_1(t)[S_1] + c_2(t)[S_2], \qquad \frac{\mathrm{d}}{\mathrm{d}t}[S_2] = +c_1(t)[S_1] - c_2(t)[S_2], \tag{1}$$

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

The master equation for this model is a continuous time, discrete state Markov process for which a linear system of ODEs, $p' = Ap$, describes the evolution of the associated probability distribution p. The ith state records the integer number of molecules of each species, and the probability of this state is recorded in the ith entry of the vector p. In a small time $\mathrm{d}t$, 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 nonnegative 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 (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" [20]. In this article we focus on some exceptions. A large class of important and solvable models, including isomerisation, arise when reaction rates are linear as a function of the state [25]. 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 reaction rate equations. (Usually these models agree only approximately.) The exact solution to the CME (2) for our isomerisation example is a binomial distribution, where the time-varying parameter in the binomial distribution comes from the solution to the corresponding RRE (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

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = \Big[A^{[0]} + A^{[1]} f(t)\Big]\boldsymbol{p}, \qquad \boldsymbol{p}(0) = \boldsymbol{p}_0 \in \mathbb{R}^{N+1}, \tag{2}$$

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

$$A^{[0]}_{k,\ell} = \begin{cases} -N, & k = \ell, \\ \ell, & k = \ell - 1, \\ N - \ell, & k = \ell + 1, \\ 0, & \text{otherwise;} \end{cases} \qquad A^{[1]}_{k,\ell} = \begin{cases} N - 2\ell, & k = \ell, \\ \ell, & k = \ell - 1, \\ -N + \ell, & k = \ell + 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

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 $-1 \le f(t) \le 1$ 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 (2) is a master equation, which was originally simulated for the special case $f(t) = \sin t$ [27]. Here, we generalize. It turns out (2) has a truly miraculous structure.

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = \mathbb{A}\boldsymbol{p} \qquad \text{with solution} \qquad \boldsymbol{p}(t) = \exp(t\mathbb{A})\boldsymbol{p}(0). \tag{4}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = \mathbb{A}(t)\boldsymbol{p} \qquad \text{with solution} \qquad \boldsymbol{p}(t) = \exp(\boldsymbol{\Omega}(t))\boldsymbol{p}(0). \tag{5}$$

Here, the Magnus expansion [34] tells us how to find the crucial matrix $\boldsymbol{\Omega}(t)$ as an infinite series, namely

$$\boldsymbol{\Omega}(t) = \int_0^t \mathbb{A}(s)\mathrm{d}s - \frac{1}{2}\int_0^t \left[\int_0^s \mathbb{A}(r)\mathrm{d}r, \mathbb{A}(s)\right] \mathrm{d}s + \dots. \tag{6}$$

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 matrix commutes with itself for all time, i.e. $[\mathbb{A}(t_1), \mathbb{A}(t_2)] \equiv 0$, 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 [24].

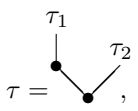
A remarkable correspondence between terms in the Magnus expansion and rooted, binary trees (elucidated in [24, equation (4.10)]) allows (6) to be written as

$$\Omega(t) = \sum_{m=0}^{\infty} \sum_{\tau \in \mathbb{T}_m} \int_0^t \alpha(\tau) G_\tau(x)\,\mathrm{d}x. \tag{7}$$

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.

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 $m \geq 1$, any $\tau \in \mathbb{T}_m$ can be represented in the form

$$\tau = \quad \text{(tree with } \tau_1 \text{ and } \tau_2\text{)}, \qquad \tau_1 \in \mathbb{T}_{m_1},\ \tau_2 \in \mathbb{T}_{m_2}, \quad m_1 + m_2 = m - 1. \tag{8}$$

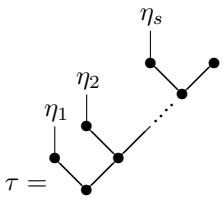


In that case

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

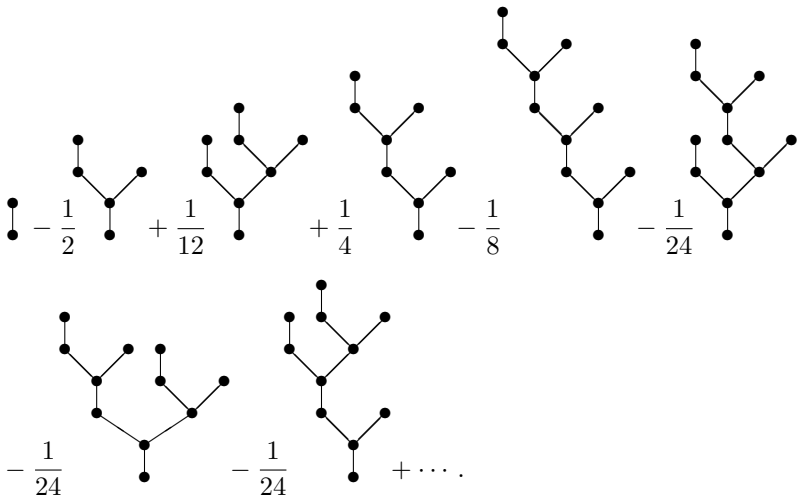
4. $\alpha : \tau \to \mathbb{Q}$ is a mapping from Magnus trees to rational numbers. Specifically,

$\alpha(\bullet) = 1$ and, for any $\tau \in \mathbb{T}_m$ for $m \geq 1$, with B_s denoting Bernoulli numbers,

$\tau =$ 

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

In general, this procedure elegantly expresses the Magnus expansion (7) as

$\Omega(t) \rightsquigarrow$ 

2.1. A special property of isomerisation matrices. Recognising the following special property (confirmed by an easy matrix multiplication)

$$[A^{[0]}, A^{[1]}] = -2A^{[1]} \tag{9}$$

usefully simplifies our Magnus expansion. This simple form of the commutator (9) 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 (9), that $\mathcal{F}$ is

$$\mathcal{F}(A^{[0]}, A^{[1]}) = \mathrm{Span}\,\{A^{[0]}, A^{[1]}\}. \tag{10}$$

In other words, although in general the Magnus expansion of the solution may require many terms, *the Magnus expansion of (2) for isomerisation is simply a linear combination of the form*[1] $\Omega(t) = \sigma_{[0]}(t)A^{[0]} + \sigma_{[1]}(t)A^{[1]}$*!*

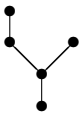
2.2. A Magnus expansion of isomerisation. We now specialize the general form of the expansion (7) to our application of isomerisation (2), for which

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

[1]Indeed, more is true. A Lie algebra $\mathfrak{g}$ is *solvable* if there exists $M \geq 0$ such that $\mathfrak{g}^{[M]} = \{0\}$, where $\mathfrak{g}^{[0]} = \mathfrak{g}$ and $\mathfrak{g}^{[k+1]} = [\mathfrak{g}^{[k]}, \mathfrak{g}^{[k]}]$. By (9), $\dim \mathcal{F}^{[1]} = 1$ so it is a commutative algebra and $\mathcal{F}^{[2]} = \{0\}$. The algebra is solvable!

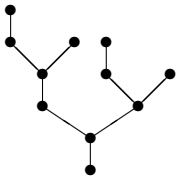
By following the four step algorithm near (8), we find the first few terms in the series (6) and the corresponding trees are

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



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

and so on. Note we made use of (9) for the commutator to simplify the expressions. Moreover, a matrix commutes with itself so some terms are zero, such as



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

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

We already know from (9) and (10) 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,

$\tau =$ 

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

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

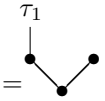
Consider next $m \geq 2$ and (8). If $m_1, m_2 \geq 1$ then, by the induction assumption, both G_{τ_1} and G_{τ_2} are scalar multiples of $A^{[1]}$ and we deduce that $G_\tau \equiv O$. There are two remaining possibilities: either $m_1 = 0$, $m_2 = m-1$ or $m_1 = m-1$, $m_2 = 0$. In the first case

$\tau =$ 

(11)

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

$\tau =$ 

(12)

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 (2) is of the form*

$$\Omega(t) = tA^{[0]} + \sigma(t)A^{[1]} \tag{13}$$

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

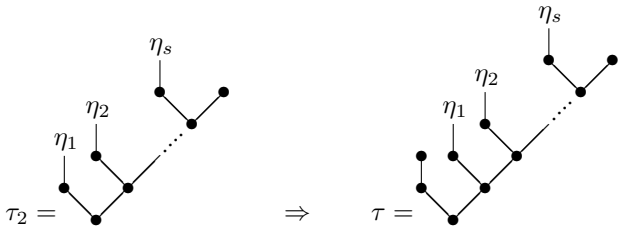
Next, we will explicitly find the function σ of (13) in the Theorem, and thus find the Magnus expansion of isomerisation. We do not present all steps in the derivations to come. Theorem (2.1) and the steps leading to it were deliberately chosen for presentation partly because this quickly gives a good 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 (8). That always works. Often though, particular applications allow simplifications, 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 (11) and of (12).

The situation for (12) 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 (11). There we have



Therefore

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

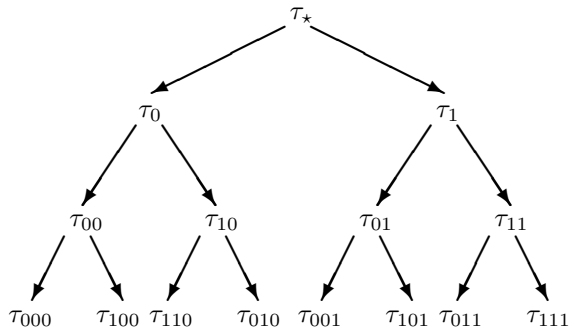
Hence, to summarize

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

This is a moment to comment on the mechanisms giving rise to some of our simplifications. Not all Magnus trees feature — with nonzero coefficients — in the expansion (7). There are two mechanisms that explain this: (i) The coefficient $\alpha(\tau)$ is zero; or (ii) $\sigma_\tau \equiv 0$, 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 (7), 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 $\mathbb{T}_m$, $m \geq 1$ we have 2^{m-1} trees (some with a zero coefficient). What we really have is a binary 'super-tree'



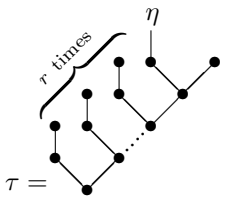
. . . and so on.

The rule is: Each move 'left' (i.e. in the 0 direction – the subscripts are binary strings) corresponds to 'scenario' (11); Each move 'right' corresponds to 'scenario' (12). 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 $\tilde{f}(x) = xf'(x)$,

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

By continuing to find these trees, we see a pattern emerge: For any $\tau \in \mathbb{T}_m$, $m \geq 1$, 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 $\tau \in \mathbb{T}_m$, $m \geq 2$, to work with



(14)

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 (14), that lead to

$$K_\tau(t,x) = 2(-2t)^r \int_x^t K_\eta(y,x)\,\mathrm{d}y, \qquad \alpha(\tau) = \frac{\mathrm{B}_{r+1}}{(r+1)!}\alpha(\eta). \tag{15}$$

We left out one exceptional case, namely $\tau = \tau_{\mathbf{0}}$. In that case the representation (14) 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

$$\frac{t(1-\mathrm{e}^{-2t})}{1-2t-\mathrm{e}^{-2t}}\Theta(t,x) = \int_x^t \Theta(y,x)\,\mathrm{d}y - 1, \tag{16}$$

with solution

$$\Theta(t,x) = -\exp\left(-4\int_x^t \frac{1-y-(1+y)\mathrm{e}^{-2y}}{(1-\mathrm{e}^{-2y})(1-2y-\mathrm{e}^{-2y})}\,\mathrm{d}y\right)\frac{1-2x-\mathrm{e}^{-2x}}{x(1-\mathrm{e}^{-2x})}\,\mathrm{d}\xi. \tag{17}$$

Finally, we integrate the contribution of the individual σ_τs, scaled by $\alpha(\tau)$, from each tree, for all Magnus 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 (16), 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

$$\sigma(t) = tf(t) + \frac{t(1-\mathrm{e}^{-2t})}{1-2t-\mathrm{e}^{-2t}}\int_0^t xf'(x)\Theta(t,x)\,\mathrm{d}x. \tag{18}$$

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 (17), everything is now explicit. Combining σ in (18) with Theorem (2.1), we have now found the (complete!) Magnus expansion of isomerisation.

Note that (18) is bounded for all $t \geq 0$, 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* (13) *for isomerisation converges for every* $t \geq 0$. 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 Moan–Niesen sufficient condition for convergence of the Magnus expansion can be taken while still maintaining good accuracy with Magnus-based numerical methods [27, Figure 1]. That good experience of taking larger time steps 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 [23]. 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.

2.5. A role for automorphisms. Theorem (2.1) and (18) tell us the answer to the question of the matrix $\Omega(t)$ in the Magnus expansion. Ultimately, we want the solution (5). For that, we need the exponential, $\exp(\mathbf{\Omega}(t))$. 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. Note $P \in \mathrm{O}(N+1)\cap\mathrm{Sym}(N+1)$ so P is an *orthogonal involution:* $P^{-1} = P^\top = P$ and $P^2 = I$. Matrix multiplication confirms the useful properties

$$PA^{[0]}P = A^{[0]}, \qquad PA^{[1]}P = -A^{[1]}. \tag{19}$$

[2] Actually, it is analytic.

Being an orthogonal involution, P defines an inner automorphism on $\mathfrak{gl}(N+1)$, namely $\iota(B) = PBP$ for $B \in \mathfrak{gl}(N+1)$. Following [36], we let $\mathfrak{k} = \{B \in \mathfrak{gl}(N+1) : \iota(B) = B\}$ and $\mathfrak{p} = \{B \in \mathfrak{gl}(N+1) : \iota(B) = -B\}$ be the *fixed points* and *anti-fix points* of the automorphism ι. 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 $\frac{1}{2}[B + \iota(B)] \in \mathfrak{k}$ and $\frac{1}{2}[B - \iota(B)] \in \mathfrak{p}$. 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

$$\begin{aligned} X = {} & tp - \frac{1}{2}t^2[p,k] - \frac{1}{6}t^3[k,[p,k]] + t^4\left(\frac{1}{24}[p.[p,[p.k]]] - \frac{1}{24}[k,[k,[p,k]]]\right) \\ & + t^5\left(\frac{7}{360}[k,[p,[p,[p,k]]]] - \frac{1}{120}[k,[k,[k,[p,k]]]] - \frac{1}{180}[[p,k],[p,[p,k]]]\right) \\ & + t^6\left(-\frac{1}{240}[p,[p,[p,[p,[p,k]]]]] + \frac{1}{180}[k,[k,[p,[p,[p,k]]]]]\right. \\ & - \frac{1}{720}[k,[k,[k,[k,[p,k]]]]] + \frac{1}{720}[[p,k],[k,[p,[p,k]]]] \\ & \left. + \frac{1}{180}[[p,[p,k]],[k,[p,k]]]\right) + \mathcal{O}(t^7)\,, \end{aligned} \tag{20}$$

$$\begin{aligned} Y = {} & tk - \frac{1}{12}t^3[p,[p,k]] + t^5\left(\frac{1}{120}[p,[p,[p,[p,k]]]] + \frac{1}{720}[k,[k,[p,[p,k]]]]\right. \\ & \left. - \frac{1}{240}[[p,k],[k,[p,k]]]\right) + \mathcal{O}(t^7)\,. \end{aligned} \tag{21}$$

Now, let $k = A^{[0]}$ and $p = A^{[1]}$ so by (9), $[p,k] = 2p$. Look again at (20) and (21). 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 assertion. We deduce that the only terms surviving in (20), 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)

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

so

$$X = -\sum_{r=1}^{\infty}\frac{t^r}{r!}\mathrm{ad}_k^{r-1}p = \frac{1-\mathrm{e}^{-2t}}{2}p. \tag{22}$$

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 at least twice, and each such term must vanish, so

$$Y = tk. \tag{23}$$

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 (23) intact (hence $Y = tA^{[0]}$), but t in (22) 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 to determine the eigenvalues and eigenvectors of $A^{[0]}$. They are essentially given by [10, Theorem 2.1]. Here we provide an alternative proof and a 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

$$v_m = (-1)^m \binom{r}{m} {}_2F_1\left[\begin{matrix} -N+r, -m; \\ r-m+1; \end{matrix} \; -1\right], \qquad m = 0, \ldots, r, \tag{24}$$

$$v_m = (-1)^r \binom{N-r}{m-r} {}_2F_1\left[\begin{matrix} -N+m, -r; \\ m-r+1; \end{matrix} \; -1\right], \qquad m = r, \ldots, N. \tag{25}$$

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

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

$$(N+1-m)v_{m-1} - (N+\lambda)v_m + (m+1)v_{m+1} = 0, \qquad m = 0, \ldots, N, \tag{26}$$

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 (26). 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, (24)–(25) 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

$$V^{-1} = 2^{-N} V \tag{27}$$

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:

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

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*

$$A^{[1]} = WEW^{-1}, \tag{28}$$

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,*

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

An immediate consequence of this Jordan form (28) is that $A^{[1]}$ is nilpotent.

Proof. The Jordan form (28) 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 $n \ge m+2$ 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 $n = m$

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

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

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

[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.

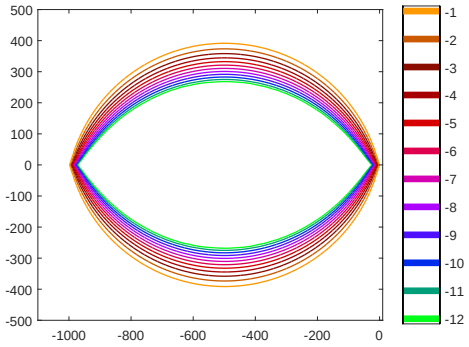


FIG. 1. *An 'almond eye:' Pseudospectrum [42] of a* 500×500 *example of the* $A^{[0]}$ *matrix, defined in* (3)*, as computed by Eigtool [45]. Contours of the minimum singular value,* $s_{min}(zI - A)$*, are displayed on a log scale.*

and we are done. ∎

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

$$Z_{m,n} = \begin{cases} 0, & m \le n-1, \\ (-1)^{m-n}\binom{N-n}{m-n}, & m \ge n, \end{cases} \qquad m, n = 0, \ldots, N.$$

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

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

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

THEOREM 3.3. *The matrix exponential is, in an explicit form,*

$$\mathrm{e}^{tA^{[1]}} = ZC^{-1}\mathrm{e}^{tE}C\tilde{Z}. \tag{29}$$

3.2.1. Evaluating the exponential via (29). 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 $\mathcal{O}\big(N^2\big)$ flops but herewith an algorithm that accomplishes this in just $\mathcal{O}\big(N^2\big)$ *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

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

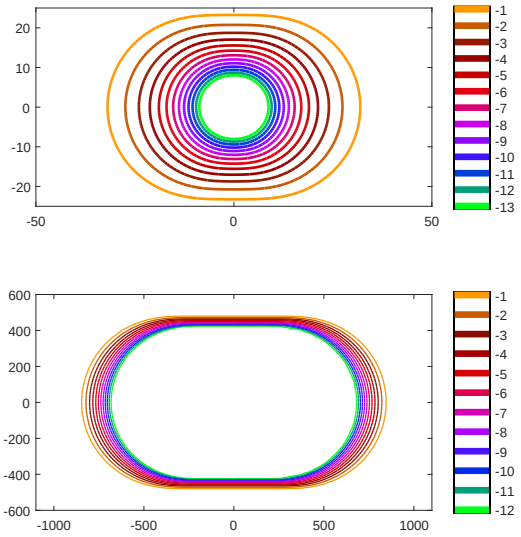


FIG. 2. *The 'athletics track:' Pseudospectrum [42] of the* $A^{[1]}$ *matrix, defined in* (3)*, as computed by Eigtool [45]. Top:* 30×30*. Bottom:* 500×500*.*

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

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

Rewrite this as

$$y_m^{[N]} = y_{m-1}^{[N-1]} + y_m^{[N-1]}, \qquad m = 0, \ldots, N-1 \tag{30}$$

(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 $M = 1, 2, \ldots, N$, add

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

and we are done.

Of course, similar reasoning applies also to a product $\boldsymbol{y} = Z\boldsymbol{u}$. The only difference vis-á-vis (30) is that now $y_m^{[N]} = y_m^{[N-1]} - y_{m-1}^{[N-1]}$, $m = 0, \ldots, N-1$, therefore the recursion 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 (29). We address that now. It is trivial that

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

Therefore (cf. (29))

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

Let us pause to reflect on the exact exponentials that we have just found. We expect the solution to our model of isomerisation to be a binomial distribution [25]. 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 (2) must be a binomial distribution.

As an example, set $\boldsymbol{e}_0 = (1, 0, \dots, 0)^\top$ and compute the leading column, $\mathrm{e}^{qA^{[1]}}\boldsymbol{e}_0 = Z(C^{-1}\mathrm{e}^{qE}C)\tilde{Z}\boldsymbol{e}_0$. Note that $(\tilde{Z}\boldsymbol{e}_0)_m = \tilde{Z}_{m,0} = \binom{N}{m}$. So $[(C^{-1}\mathrm{e}^{qE}C)\tilde{Z}\boldsymbol{e}_0]_m = \binom{N}{m}\sum_{n=0}^{N-m}\binom{N-m}{n}t^n = \binom{N}{m}(1+q)^{N-m}$ 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 [25].

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* [42]. 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.

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 [42] 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 [38], and master equations are closely related to convection-diffusion, suggesting they will also exhibit interesting pseudospectra. Indeed, the matrices that arise in our

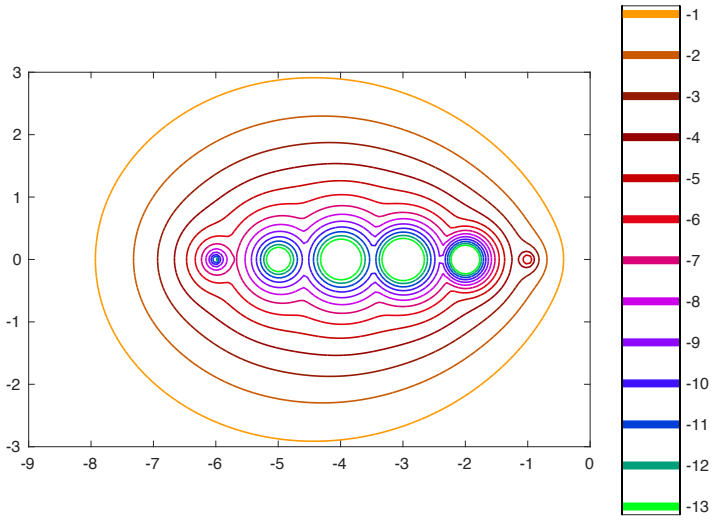


FIG. 3. *The 'seed pod:' Pseudospectrum [42] of a* 1513×1513 *finite section of the singly infinite matrix associated with a totally asymmetric exclusion process (TASEP) with 6 particles beginning in a 'step' initial configuration [8, Figure 9, q=0], as computed by Eigtool [45].*

applications of master equations to isomerizaiton exhibit an humongous pseudospectra. They are examples of the class of *twisted Toeplitz matrices and operators*, which have recently been understood to exhibit a distinctive pseudospectra, captivating more general interest [41].

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 is large. For example, the pseudospectra of the convection-diffusion operator has been shown to be approximately bounded by a parabola [38], 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. 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 especially their applications will continue to occupy new directions in scientific computation for some time [33]. There is always the challenge of *high dimensions*, for instance. 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 [16]. A simple example of a graph Laplacian on a line of nodes appears in [39], and, like the matrix exponential, it has been shown that a *Mittag-Leffler function* [15] of a

graph Laplacian matrix is also a stochastic matrix [32]. 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 [30].

It is known that if we generalise (4) 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 (4) 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 (5) 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 [12].

This computational approach has the potential for wider applications to master equations where tensor structures involving shift operators often arise. So let us revisit (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 (2), consider linearly independent matrices, A and B, such that

$$[A, B] = aA + bB \tag{31}$$

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

$$X' = [\alpha(t)A + \beta(t)B]X, \quad t \geq 0, \qquad X(0) = I. \tag{32}$$

Here α and β are given scalar functions.

We wish to prove the solution of (32) can be expressed in the form

$$X(t) = \mathrm{e}^{\rho_A(t)A}\mathrm{e}^{\rho_B(t)B}, \tag{33}$$

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 (33) and substituting into (32), 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

$$(\rho_A' - \alpha)A\mathrm{e}^{\rho_A A} + \rho_B' \mathrm{e}^{\rho_A A}B - \beta B\mathrm{e}^{\rho_A A} = O. \tag{34}$$

A proof by induction using (31) shows

$$BA^m = (A + bI)^m B - \frac{a}{b}A[A^m - (A + bI)^m], \qquad m \in \mathbb{Z}_+. \tag{35}$$

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 (34), $(\rho_A' - \alpha)A\mathrm{e}^{\rho_A A} +$

$\rho_B' \mathrm{e}^{\rho_A A} B - \beta \mathrm{e}^{b\rho_A} \mathrm{e}^{tA} B + \frac{a}{b}\beta(1 - \mathrm{e}^{b\rho_A}) A \mathrm{e}^{tA}$. 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,

$$\rho_A' = \alpha - \frac{a}{b}\beta(1 - \mathrm{e}^{b\rho_A}), \qquad \rho_A(0) = 0, \tag{36}$$

$$\rho_B' = \beta \mathrm{e}^{b\rho_A}, \qquad \rho_B(0) = 0, \tag{37}$$

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 (36) becomes $\rho_A(t) = t$ and $\rho_B(t) = \int_0^t \mathrm{e}^{-2\tau} f(\tau)\,\mathrm{d}\tau$. Putting (37) in (36), 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* [44]. Specifically, let $A : \mathbb{R}_+ \to \mathfrak{g}$, where $\mathfrak{g}$ is a Lie algebra, $\dim \mathfrak{g} = d$, and consider the ODE

$$X' = A(t)X, \quad t \geq 0, \qquad X(0) = I. \tag{38}$$

Let $\mathcal{P} = \{P_1, P_2, \ldots, P_d\}$ be a basis of $\mathfrak{g}$. Wei & Norman [44] prove that for sufficiently small $t > 0$ there exist functions $g_1, g_2, \ldots, g_d$ such that

$$X(t) = \mathrm{e}^{g_1(t)P_1} \mathrm{e}^{g_2(t)P_2} \cdots \mathrm{e}^{g_d(t)P_d}. \tag{39}$$

This is the situation we have in (2) or, with greater generality, in (32): $P_1 = A$, $P_2 = B$ and, because of (31), the dimension of the free Lie algebra spanned by A and B is $d = 2$. Interestingly enough, this example does not feature in [44].

Coordinates of the second kind have been used extensively in the theory of Lie-group integrators [24] 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 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) [6].

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 [30, 33]. 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. $S + E \leftrightarrow C \to P + E$. 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 $C \to P + E$ [17].

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) [8, 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 numerical picture here in Figure 3 suggests it will turn out to be worthwhile pursuing. 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 [7, 18].

Random Matrix Theory [11] 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 [7]. Although less in the sense of the chemical master equation (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 [40]. 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 $\alpha_f\left(\boldsymbol{x}(s),s\right)=c_1(s)n_1$ and the backward rate by $\alpha_b\left(\boldsymbol{x}(s),s\right)=c_2(s)n_2$. Here n_1 and n_2 are the number of molecules of S_1 and S_2, respectively. The Kurtz random time-change representation [28] of the sample paths corresponding to our master equation (2) with initial state $\boldsymbol{x}(0)$ is

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

At absolute 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*, so methods that can accommodate time-varying rates, such as Magnus expansions described here, may find wider applications [19, 21]. 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 [11].

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 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'* [13] are likewise adaptable to the discrete setting of master equations [27].

4.5. Preserving positivity. Moler and Van Loan discuss more than nineteen dubious ways for computing the *matrix exponential* [35]. 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 nonnegative and indeed a stochastic matrix — a fundamental question is: do these numerical methods preserve nonnegativity? 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 [5]. 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 [29].

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 [4]. 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 (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" [10]. These matrices have served us well as wonderful running examples to illustrate new directions in master equation research. Kac did not foresee our applications to isomerisation, nor the way those isomerisation master equations are so naturally amenable to Magnus expansions. 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. This research and Shev MacNamara have been partially supported by a David G. Crighton Fellowship to DAMTP, Cambridge. 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] A. H. Al-Mohy and N. J. Higham, *A new scaling and squaring algorithm for the matrix exponential*, SIAM J. Matrix Anal. Appl., (2009), pp. 970–989.
[2] A. H. Al-Mohy and N. J. Higham, *Computing the action of the matrix exponential, with an application to exponential integrators*, SIAM J. Sci. Comp., 33 (2011), pp. 488–511, https://doi.org/10.1137/100788860.
[3] D. Anderson, *An efficient finite difference method for parameter sensitivities of continuous time markov chains*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 2237–2258.
[4] J. R. Barker, T. L. Nguyen, J. F. Stanton, M. C. C. Aieta, F. Gabas, T. J. D. Kumar, C. G. L. Li, L. L. Lohr, A. Maranzana, N. F. Ortiz, J. M. Preses, and P. J. Stimac, *Multiwell-2016 software suite*, tech. report, University of Michigan, Ann Arbor, Michigan, USA, 2016, http://clasp-research.engin.umich.edu/multiwell/.
[5] C. Bolley and M. Crouzeix, *Conservation de la positivité lors de la discrétisation des problèmes d'évolution paraboliques*, RAIRO Anal. Numér., 12 (1978), pp. 237–245, iv.
[6] E. Celledoni and A. Iserles, *Methods for the approximation of the matrix exponential in a Lie-algebraic setting*, IMA J. Numer. Anal., 21 (2001), pp. 463–488, https://doi.org/10.1093/imanum/21.2.463, http://dx.doi.org/10.1093/imanum/21.2.463.
[7] I. Corwin, *Macdonald processes, quantum integrable systems and the Kardar–Parisi–Zhang universality class*, in Proceedings of the International Congress of Mathematicians, arXiv:1403.6877, 2014.
[8] I. Corwin, *Kardar–Parisi–Zhang Universality*, Notices of the AMS, 63 (2016).
[9] B. Drawert, M. Trogdon, S. Toor, L. Petzold, and A. Hellander, *Molns: A cloud platform for interactive, reproducible, and scalable spatial stochastic computational experiments in systems biology using pyurdme*, SIAM Journal on Scientific Computing, 38 (2016), pp. C179–C202, https://doi.org/10.1137/15M1014784.
[10] A. Edelman and E. Kostlan, *The road from Kac's matrix to Kac's random polynomials*, tech. report, University of California, Berkeley, 1994.
[11] A. Edelman and N. R. Rao, *Random matrix theory*, Acta Numerica, (2005), pp. 1–65.
[12] S. N. Evans, B. Sturmfels, and C. Uhler, *Commuting birth-and-death processes*, The Annals of Applied Probability, 20 (2010), pp. 238–266.
[13] M. Giles and P. Glasserman, *Smoking adjoints: fast Monte Carlo Greeks*, Risk, (2006), p. 88.
[14] D. T. Gillespie, *The chemical Langevin and Fokker–Planck equations for the reversible isomerization reaction*, The Journal of Physical Chemistry A, 106 (2002), pp. 5063–5071, https://doi.org/10.1021/jp0128832.
[15] R. Gorenflo, A. Kilbas, F. Mainardi, and S. Rogosin, *Mittag-Leffler Functions, Related Topics and Applications*, Springer, 2014.
[16] J. Gunawardena, *A linear framework for time-scale separation in nonlinear biochemical systems*, PLoS One, 7 (2012), p. e36321, https://doi.org/10.1371/journal.pone.0036321.
[17] J. Gunawardena, *Time-scale separation: Michaelis and Menten's old idea, still bearing fruit*, FEBS J., 281 (2014), pp. 473–488.
[18] M. Hairer, *Singular stochastic PDEs*, Proceedings of the International Congress of Mathematicians, (2014).
[19] A. Hellander, J. Klosa, P. Lötstedt, and S. MacNamara, *Robustness analysis of spatiotemporal models in the presence of extrinsic fluctuations*, submitted, SIAM Journal on Applied Mathematics, arXiv:1610.01323, (2015).
[20] D. J. Higham, *Modeling and simulating chemical reactions*, SIAM Review, 50 (2008), pp. 347–368, https://doi.org/10.1137/060666457, http://dx.doi.org/10.1137/060666457.
[21] A. Hilfinger and J. Paulsson, *Separating intrinsic from extrinsic fluctuations in dynamic biological systems*, Proc. Acad. Natl. Sci., 109 (2011), pp. 12167–72, https://doi.org/10.1073/pnas.1018832108.
[22] P. V. Hilgers and A. N. Langville, *The five greatest applications of Markov chains*, in

Proceedings of the Markov Anniversary Meeting, Boston Press, Boston, MA., 2006.
[23] M. Hochbruck and C. Lubich, *On Magnus integrators for time-dependent Schrödinger equations*, SIAM J. Numer. Anal., 41 (2003), pp. 945–963, https://doi.org/10.1137/S0036142902403875.
[24] A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna, *Lie-group methods*, Acta Numer., 9 (2000), pp. 215–365, https://doi.org/10.1017/S0962492900002154, http://dx.doi.org/10.1017/S0962492900002154.
[25] T. Jahnke and W. Huisinga, *Solving the chemical master equation for monomolecular reaction systems analytically*, Journal of Mathematical Biology, 54 (2007), pp. 1–26, https://doi.org/10.1007/s00285-006-0034-x, http://www.scopus.com/inward/record.url?eid=2-s2.0-33845629747&partnerID=40&md5=c947b5e7b11c3810334b2232d40169e6. cited By 97.
[26] M. Kac, *Probability and Related Topics in Physical Sciences, Summer Seminar in Applied Mathematics, Boulder, Colorado*, American Mathematical Society, 1957.
[27] K. Kormann and S. MacNamara, *Error control for exponential integration of the master equation*, arXiv:1610.03232, (2016).
[28] T. Kurtz, *Representations of Markov processes as multiparameter time changes*, Ann. Probab., 8 (1980), pp. 682–715.
[29] S. C. Leite and R. J. Williams, *A constrained Langevin approximation for chemical reaction networks*, Kolmogorov Lecture, Ninth World Congress In Probability and Statistics, Toronto, (2016), http://www.math.ucsd.edu/~williams/biochem/biochem.pdf.
[30] S. Macnamara, *Cauchy integrals for computational solutions of master equations*, ANZIAM Journal, 56 (2015), pp. 32–51, https://doi.org/10.21914/anziamj.v56i0.9345.
[31] S. MacNamara, K. Burrage, and R. Sidje, *Multiscale modeling of chemical kinetics via the master equation*, SIAM Multiscale Model. & Sim., 6 (2008), pp. 1146–1168.
[32] S. MacNamara, B. I. Henry, and W. McLean, *Fractional Euler limits and their applications*, SIAM Journal on Applied Mathematics, (2016).
[33] S. MacNamara and G. Strang, *Master equations in 'Essays on New Directions in Numerical Computation'*, 2015, http://tobydriscoll.net/newdirections2015/.
[34] W. Magnus, *On the exponential solution of differential equations for a linear operator*, Comm. Pure Appl. Math., 7 (1954), pp. 649–673.
[35] C. Moler and C. V. Loan, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*, SIAM Rev., 45 (2003), pp. 3–49, https://doi.org/10.1137/S00361445024180.
[36] H. Z. Munthe-Kaas, G. R. W. Quispel, and A. Zanna, *Generalized polar decompositions on Lie groups with involutive automorphisms*, Found. Comput. Math., 1 (2001), pp. 297–324, https://doi.org/10.1007/s102080010012, http://dx.doi.org/10.1007/s102080010012.
[37] G. A. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization*, Springer, 2008.
[38] S. C. Reddy and L. N. Trefethen, *Pseudospectra of the convection-diffusion operator*, SIAM J. Appl. Math, (1994).
[39] G. Strang and S. MacNamara, *Functions of difference matrices are Toeplitz plus Hankel*, SIAM Review, 56 (2014), pp. 525–546, https://doi.org/10.1137/120897572.
[40] C. Timm, *Random transition-rate matrices for the master equation*, Phys. Rev. E, 80 (2009), p. 021140.
[41] L. N. Trefethen and S. J. Chapman, *Wave packet pseudomodes of twisted Toeplitz matrices*, Comm. Pure Appl. Math., 57 (2004), pp. 1233–1264, https://doi.org/10.1002/cpa.20034, http://dx.doi.org/10.1002/cpa.20034.
[42] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, 2005.
[43] M. F. Weber and E. Frey, *Master equations and the theory of stochastic path integrals*, arXiv:1609.02849v1, (2016).
[44] J. Wei and E. Norman, *On global representations of the solutions of linear differential equations as a product of exponentials*, Proc. Amer. Math. Soc., 15 (1964), pp. 327–334.
[45] T. G. Wright, *Eigtool*, 2002, http://www.comlab.ox.ac.uk/pseudospectra/eigtool/.