A Discrete Spectral Method for the Chemical Master Equation

Stefan Engblom[1] *

February 29, 2008

[1] *Div of Scientific Computing, Dept of Information Technology*
Uppsala University, SE-75105 Uppsala, Sweden
email: stefane@it.uu.se

Abstract

As an equivalent formulation of the Markov-assumption of stochastic processes, the master equation of chemical reactions is an accurate description of general systems in chemistry. For D reacting species this is a differential-difference equation in D dimensions, exactly soluble for very simple systems only.

We present and analyze a novel solution strategy based upon a Galerkin spectral method with an inherent natural adaptivity and a very favorable choice of basis functions.

The method is demonstrated by the numerical solution of two model problems followed by two more realistic systems taken from molecular biology. It is shown that the method remains effective and accurate, providing a viable alternative to other solution methods when the dimensionality is not too high.

Keywords: master equation, spectral method, discrete approximation, adaptive basis, unbounded domain, Charlier's polynomials.

AMS subject classification: 65C20, 60H35, 41A10, 41A63.

1 Introduction

The *Markov property* of stochastic processes plays an important role in many descriptions of real-world phenomena. For a physical system observed at

*Financial support has been obtained from the Swedish National Graduate School in Mathematics and Computing.

discrete times $t_1 < t_2 < \cdots t_n$ it states that the probability for the observation (y_n, t_n) conditioned on the systems history satisfies

$$\Pr\left(y_n, t_n | y_1, t_1; \cdots; y_{n-1}, t_{n-1}\right) = \Pr\left(y_n, t_n | y_{n-1}, t_{n-1}\right), \tag{1.1}$$

i.e. that the system depends on the present state only. Although (1.1) is not always fulfilled exactly, it is frequently a very accurate and useful approximation. In particular, the Markov assumption is accurate when the discrete time is chosen sufficiently coarse in comparison with the often very short auto-correlation time of the system. Furthermore, such systems can be described using only the initial probability $\Pr(y_1, t_1)$ and the *transition probability function* $\Pr(y_s, s|y_t, t)$ [2].

The *master equation* is a consequence of the Markov property for a discrete state space in continuous time. In particular, if a chemical system of D reacting species is described by counting the number of molecules of each kind, then the master equation accurately governs the dynamics of the probability distribution for the system. In fact, one can show that the chemical master equation is *exact* under the conditions that the system is well-stirred and in thermal equilibrium [19]

The resulting description is a differential-difference equation in D dimensions and therefore suffers from the well-known "curse of dimensionality"; — each species adds one dimension to the problem leading to a computational complexity that grows exponentially. Only few examples are analytically solvable, and effective numerical methods for solving the master equation are of both practical and theoretical interest.

A common deterministic model is the *reaction-rate* equations. This is a set of D ordinary differential equations (ODEs) approximating the expected values of the concentrations of the species in the system. There are, however, many systems for which the reaction-rate equations fail to reproduce actual behavior [30]. For instance, biological systems inside living cells frequently consists of fewer than 10^2 molecules [22] so that stochastic effects are more pronounced. Additionally, such systems are often driven towards critical points for various biological reasons. Close to such points, small random fluctuations in one variable may slowly "leak" probability mass in a direction that on a longer timescale drastically affects the rest of the system.

An important alternative solution method is the use of stochastic simulation techniques that offer the ability to follow sample trajectories of the system. *Gillespie's SSA* method [18], or versions of it [17], is the method in most frequent use. In contrast to the reaction-rate approach, a common feature of such methods is that they are *exact* in a statistical sense. Although simulating one trajectory can be performed relatively cheaply for many sys-

tems, models in molecular biology are often very stiff and therefore expensive to solve by explicitly simulating the mix of different time-scales [5].

For completeness, we mention that there are analytical techniques for solving the master equation. By using information about a specific system, the Ω-expansion [25, Ch. X] yields an asymptotic expansion of the solution in inverse powers of the size Ω of the system. Another approach, briefly discussed in Section 4 below, is the *Poisson representation* [15, Ch. 7.7], where the solution is written as a superposition of uncorrelated Poisson distributions.

Recent *numerical* considerations for the master equation include numerical solution of the *Fokker-Planck equation* [10, 13] and adaption of the *Sparse grids technique* [24]. When viewed as a master equation for fairly general *continuous* stochastic processes [15, Ch. 7.2], numerical solution of the Fokker-Planck equation is an interesting subject in itself. As a continuous approximation to the master equation, however, it is difficult to say *a priori* how good the approximation will be [25, Ch. VIII]. The sparse grids technique directly aims to reduce the computational complexity of high dimensional smooth problems. Its application to the master equation is quite recent and would appear promising.

In the present paper we will investigate a spectral method for the master equation. The method is unusual in that the basis functions are orthogonal with respect to a *discrete* measure. This suits the discrete character of the solution and avoids the need for continuous approximations to the master operator. The method's efficiency thus depends on the spectral representation of smooth solutions on discrete sets, where "smooth" has to be defined in this new context. Another novel feature of the method is a certain built-in adaptivity of the basis which allows the basis functions to follow the dynamical behavior of the solution. We note that our proposed scheme is reminiscent of an approach considered earlier in [7], and will further comment on this point in Section 4.

The rest of the paper is organized as follows. Section 2 is devoted to a theoretical study of the master equation and the suggested numerical solution method. We define the master equation and briefly touch upon some of its interesting properties. A suitable approximation space is developed and analyzed and the stability of the resulting scheme is discussed. We conclude Section 2 by highlighting some important details that aid in implementing the method efficiently. Section 3 is devoted to numerical experiments and investigates the performance of the method when applied to four different systems: two model problems and two models taken from molecular biology. We conclude the paper by discussing the various merits of the method and pointing to possible future considerations.

2 Concepts and analysis

An efficient and natural solution strategy for any equation of the form $u' = Lu$ with L a linear operator is a spectral Galerkin method. This is in particular true when the natural domain of u and the boundary conditions are "simple" in some sense. Since the spatial domain of the master equation is the set of non-negative integers and since formally, no boundary conditions need to be imposed, it seems reasonable to believe that a spectral method is a favorable numerical tool.

Spectral representations of stochastic processes have been used frequently in the context of *Stochastic differential equations* (SDEs) (see for example [28, 35, 36]), but the application of spectral methods to the master equation seems to be new. Most SDEs in applications are derived by adding noise to a deterministic description as a model of incomplete or inexact data and are expressed in observable variables. By contrast, the master equation directly governs the probability density function of the process and is a consequence of the Markov assumption only.

We now proceed by a discussion of the properties of the master equation. A construction of the approximation spaces then follows where "smoothness" of discrete functions will be defined and where care will be taken to ensure convergence in a relevant norm. The issue of stability is discussed in Section 2.5 and we conclude by devoting Section 2.6 to some important implementation details.

2.1 The master equation

We shall consider the dynamics of a chemical system of D different species under R prescribed reactions. Let $p(x,t)$ be the probability distribution of the states $x \in \mathbf{Z}_+^D = \{0,1,2,\ldots\}^D$ at time t. That is, p simply describes the probability that a certain number of molecules is present at each time.

The reactions are specified as "moves" defined over the states x according to the *reaction propensities* $w_r : \mathbf{Z}_+^D \to \mathbf{R}_+$. These define the transition probability per unit of time for moving from the state x to $x - N_r$;

$$x \xrightarrow{w_r(x)} x - N_r, \tag{2.1}$$

where by convention, $N_r \in \mathbf{Z}^D$ is the transition step and is the rth column in the *stoichiometric matrix* N.

The *master equation* [15, 25] is then given by

$$\begin{aligned}\frac{\partial p(x,t)}{\partial t} &= \sum_{\substack{r=1\\ x+N_r^- \geq 0}}^{R} w_r(x+N_r)p(x+N_r,t) - \sum_{\substack{r=1\\ x-N_r^+ \geq 0}}^{R} w_r(x)p(x,t) \\ &=: \mathcal{M}p, \end{aligned} \tag{2.2}$$

where the transition steps are decomposed into positive and negative parts as $N_r = N_r^+ + N_r^-$.

As indicated, the summations are performed over *feasible* reactions only. In what follows, we shall only consider formulations where $w_r(x) = 0$ whenever $x \not\geq N_r^+$. This assumption is justified as follows (cf. [25, Ch. VII.2] and [15, Ch. 7.5]): let i be such that $N_{ri} > 0$. Then w_r defines a certain reaction for which one or several x_i's are annihilated. Obviously, this reaction cannot occur unless there are sufficiently many x_i's left to annihilate and we therefore postulate that w_r is zero for $x_i \in \{0, 1, \ldots, N_{ri} - 1\}$.

Evidently, under this assumption on the propensities the conditions on x may be removed from the interval of summation in the right sum of (2.2). If additionally p is understood to be zero for negative arguments x, then the left sum may be simplified in the same manner. Regardless of the latter simplification, the adjoint operator $\mathcal{M}^*$ contains no such conditions [25, Ch. V.9]. If (p, q) is a pair of not necessarily normalized or positive functions defined over $\mathbf{Z}_+^D$, then provided both sides make sense (see [11] for a proof),

$$\sum_{x\geq 0} q(x)\mathcal{M}p(x) = \sum_{x\geq 0}\sum_{r=1}^{R}[q(x-N_r) - q(x)]w_r(x)p(x). \tag{2.3}$$

Hence the adjoint operator is given by

$$\mathcal{M}^*q = \sum_{r=1}^{R} w_r(x)[q(x-N_r) - q(x)]. \tag{2.4}$$

Note that both x and N_r in (2.3) and (2.4) are D-dimensional vectors and that the sum in (2.3) runs over $x \in \mathbf{Z}_+^D$. Moreover, if $X = [X_1, \ldots, X_D]$ is the D-dimensional time-dependent stochastic variable for which p is the probability density function, then using this notation (2.3) can be understood as

$$\frac{d}{dt}E[T(X)] = \sum_{r=1}^{R} E\left[\left(T(X-N_r) - T(X)\right) w_r(X)\right] \tag{2.5}$$

for $T : \mathbf{Z}_+^D \to \mathbf{R}$ a suitable test-function. Using this form of the adjoint, equations for the various moments of X can be formed (see [11] for a numerical investigation of this approach).

We also mention that, expressed in the stochastic variable X, the master equation is equivalent to a description in terms of a continuous-time Markov chain:

$$X_{k+1} = X_k + Ne_m, \tag{2.6}$$
$$t_{k+1} = t_k + \tau_k, \tag{2.7}$$

where e_m is the mth unit vector chosen according to the prescription

$$\Pr[m = r] = \alpha w_r(X_k), \tag{2.8}$$

where

$$\alpha \equiv \left(\sum_{r=1}^{R} w_r(X_k) \right)^{-1}, \tag{2.9}$$

and where the time-step τ_k is drawn from an exponential distribution with mean α. This formulation is equivalent to Gillespie's SSA method [18].

Let now (λ, q) be an eigenpair of $\mathcal{M}^*$ normalized so that the largest value of q is positive and real. Then we see from (2.4) that $\Re\lambda \leq 0$ so that all eigenvalues of $\mathcal{M}$ share this property. Moreover, $q \equiv 1$ is an eigenvector corresponding to $\lambda = 0$, a fact that has the natural interpretation that the probability mass of any solution p is conserved by the master equation. We shall need the following related stability result.

Theorem 2.1 *Any solution to the master equation is non-increasing in the l^1-sequence norm. That is,*

$$\sum_{x \geq 0} |p(x,t)| \leq \sum_{x \geq 0} |p(x,0)| \tag{2.10}$$

for any $t \geq 0$.

Note that $p(x, 0)$ in Theorem 2.1 is an arbitrary l^1-measurable function defined on $\mathbf{Z}_+^D$, and not necessarily a probability. The statement follows from the fact that the semigroup corresponding to $\mathcal{M}$ is contractive (see [9, Ch. 1]), but in order to be self-contained we offer the following simple argument.

Proof. It is easier to prove this result by considering the adjoint equation under the dual norm of l^1, which is the l^∞-norm. From the relation $\partial q/\partial t =$

$\mathcal{M}^* q$ and the definition (2.4) we see that the largest positive value of q cannot increase and that the most negative value of q cannot decrease (this observation is due to van Kampen, see the "remarkable exercise" in [25, Ch. V.9]). Consequently, $\|q\|_{l^\infty}$ cannot increase. In Section 2.5 a second and more direct proof of this result is obtained. $\square$

Remark. Theorem 2.1 guarantees that for any $p \geq 1$, the l^p-norm of the solution stays bounded. However, for $p > 1$ it is frequently the case that the norm grows slowly in time. $\square$

We now cite the following important result which follows from slightly more careful considerations of the structure of the master operator.

Theorem 2.2 *Let $p(x,0)$ be an l^1-measurable discrete function defined on $\mathbf{Z}_+^D$ and let $\mathcal{M}$ be neither* decomposable *nor a* splitting *(see below). Then the master equation (2.2) admits a unique steady-state solution as $t \to \infty$. Moreover, if $p(x,0)$ is a discrete probability density, then so is the steady-state solution.*

For a proof and an insightful discussion we refer the reader to [25, Ch. V.3]. A *decomposable* linear operator can be cast in the form (by relabeling the states)

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{11} & 0 \\ 0 & \mathcal{M}_{22} \end{bmatrix}, \tag{2.11}$$

while a *splitting* operator can be written as

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & 0 \\ 0 & \mathcal{M}_{22} & 0 \\ 0 & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}. \tag{2.12}$$

Exclusion of these cases essentially forces $\mathcal{M}$ to define a fully interacting system, one not allowed to consist of several isolated subsystems.

Theorem 2.2 is formally only valid when the number of states is finite. Indeed, there are many examples of master equations for which no steady-state solution exists (e.g. the unbounded one-dimensional "random walk" on the form $\emptyset \xrightarrow{k} X$ where X-molecules arrive indefinitely). For *chemical* master equations in a bounded environment, however, each species must have sufficiently strong "sinks" to match the inflow from the "sources". We therefore expect reasoning based on assuming a finite number of states to be valid for all physically realizable systems.

As a concluding model problem in one dimension we consider the linear birth-death process [2]

$$\left.\begin{array}{ccc} \emptyset & \xrightarrow{k} & X \\ X & \xrightarrow{\mu x} & \emptyset \end{array}\right\}, \tag{2.13}$$

where conventionally we use uppercase letters to denote molecule *names*, while lowercases are used for counting the number of molecules of a certain species. Eq. (2.13) states that X-molecules are created at constant rate and simultaneously destroyed at a rate proportional to the total number of molecules. The master equation for this system can be written in terms of the forward- and backward difference operator $\Delta q(x) = q(x+1) - q(x)$ and $\nabla q(x) = q(x) - q(x-1)$,

$$\frac{\partial p(x,t)}{\partial t} = \mathcal{M}p(x,t) = -k\bar{\nabla}p(x,t) + \mu\Delta[xp(x,t)], \tag{2.14}$$

where we use a bar over ∇ to express the convention that $p(-1) = 0$. This problem can be solved completely if initial data is given in the form of a Poissonian distribution of expectation a_0,

$$p(x,0) = \frac{a_0^x}{x!}e^{-a_0}. \tag{2.15}$$

In this case one easily verifies that the full dynamic solution is given by

$$p(x,t) = \frac{a(t)^x}{x!}e^{-a(t)}, \tag{2.16}$$

where $a(t) = a_0\exp(-\mu t) + k/\mu \cdot (1 - \exp(-\mu t))$. Independent of the initial data, p approaches a Poissonian distribution with expectation k/μ.

This example is actually part of the motivation for the unusual choice of basis functions suggested in the next section. Intuitively, for this example, a perfect choice would be the ansatz

$$p(x,t) = \sum_n b_n B_n(x)\frac{a^x}{x!}e^{-a} \tag{2.17}$$

where B_n are polynomials and (possibly) $a = a(t)$ to make the ansatz "follow" the solution in some way. Unfortunately, there are no known polynomials that are orthogonal with respect to the implied discrete inner product

$$\langle B_n, B_m\rangle_1 = \sum_{x\geq 0} B_n(x)B_m(x)\frac{a^{2x}}{(x!)^2}e^{-2a}, \tag{2.18}$$

as would be required by (2.17). However, there *are* polynomials that are orthogonal under the similar product

$$\langle C_n, C_m \rangle_2 = \sum_{x \geq 0} C_n(x) C_m(x) \frac{a^x}{x!} e^{-a}. \tag{2.19}$$

These are *Charlier's polynomials*, to be defined in the next section. Consequently, we abandon the ansatz (2.17) and focus instead on the form

$$p(x,t) = \sum_n c_n C_n(x) \sqrt{\frac{a^x}{x!} e^{-a}}. \tag{2.20}$$

The properties of this heuristically motivated ansatz are investigated more carefully in the following sections.

2.2 Functions on $\mathbf{Z}_+$

In the following three sections we shall study the approximation of real functions defined over the set of non-negative integers. Suitable spaces of functions are introduced and investigated and a theory for approximation over these discrete spaces is developed. The corresponding results for continuous approximation are fairly well understood but the discrete version seems to have been largely overlooked in the literature.

For clarity, we mention here that the theory will contain a certain parameter a and we make some efforts to obtain uniform results. The precise choice of this parameter is discussed in Section 2.4, and we will later use this degree of freedom as a means of improving the efficiency of the resulting scheme. To this end we shall only consider one-dimensional functions; the corresponding tensor basis is mentioned in Section 2.6.

For $p \in \{1, 2, \infty\}$ we will make use of the ordinary normed $l^p(\mathbf{Z}_+)$-spaces,

$$l^p(\mathbf{Z}_+) = \left\{ q : \mathbf{Z}_+ \to \mathbf{R};\ \|q\|_{l^p(\mathbf{Z}_+)} < \infty \right\}, \tag{2.21}$$

$$\|q\|^p_{l^p(\mathbf{Z}_+)} \equiv \sum_{x \geq 0} |q(x)|^p, \tag{2.22}$$

where the usual sup-norm is to be understood when $p = \infty$. For $p = 2$ we additionally associate the discrete Euclidean inner product,

$$(p, q) \equiv \sum_{x \geq 0} p(x) q(x). \tag{2.23}$$

Define now the falling factorial function by $x^{\underline{m}} = \prod_{i=0}^{m-1}(x-i)$. For reasons that will be clear later on (see Lemma 2.4) we shall use the following hierarchy of discrete Sobolev-spaces:

$$h^m(\mathbf{Z}_+) = \left\{ q : \mathbf{Z}_+ \to \mathbf{R};\ \sqrt{x^{\underline{k}}} \cdot q(x) \in l^2(\mathbf{Z}_+) \text{ for } 0 \le k \le m \right\} \tag{2.24}$$

with the corresponding norm

$$\|q\|^2_{h^m(\mathbf{Z}_+)} \equiv \sum_{k=0}^{m} a^{-k} \|\sqrt{x^{\underline{k}}} \cdot q(x)\|^2_{l^2(\mathbf{Z}_+)}, \tag{2.25}$$

and where the parameter $a \in \mathbf{R}_+$.

We also define an analogous set of *weighted* Sobolev-spaces with weight $w(x) = a^x/x! \cdot e^{-a}$. The weighted inner product is

$$(p,q)_w \equiv \sum_{x \ge 0} p(x)q(x)w(x) \tag{2.26}$$

with generated norm. This yields the weighted space $l^2_w(\mathbf{Z}_+)$ and the definition of each weighted discrete Sobolev-space $h^m_w(\mathbf{Z}_+)$ follows as in (2.24) and (2.25). Since these spaces are less common in analysis and in order to get some feeling for them we consider the following example: set $p(x) = \sqrt{(x-2)!}$ with $p(0) = p(1) \equiv 0$ and let $a = 1$. Clearly, $p \in h^0_w$, and in fact $\|p\|^2_{h^0_w} = \exp(-1)$. However, $p \notin h^1_w$ by the divergence of the harmonic sum. As an easy generalization we note that $p(x)^2 = (x-m-2)!$ is in h^m_w but not in h^{m+1}_w. We now further examine these spaces by proving a basic result concerning the forward difference operator.

Proposition 2.3 *The map $\Delta : h^{m+1}_w(\mathbf{Z}_+) \to h^m_w(\mathbf{Z}_+)$ is continuous uniformly w.r.t. to the parameter a.*

In fact, the following stronger result will be convenient later on and additionally provides us with some insight:

Lemma 2.4 *Define the norm*

$$\|q\|^2_{h^m_{w,\Delta}(\mathbf{Z}_+)} \equiv \sum_{k=0}^{m} \|\Delta^k q\|^2_{l^2_w(\mathbf{Z}_+)}. \tag{2.27}$$

Then the norms $\|\cdot\|_{h^m_{w,\Delta}(\mathbf{Z}_+)}$ and $\|\cdot\|_{h^m_w(\mathbf{Z}_+)}$ are uniformly equivalent. *That is, there are positive constants C_1 and C_2 depending only on m such that*

$$C_1\|q\|_{h^m_w(\mathbf{Z}_+)} \le \|q\|_{h^m_{w,\Delta}(\mathbf{Z}_+)} \le C_2\|q\|_{h^m_w(\mathbf{Z}_+)} \tag{2.28}$$

holds for any function $q \in h^m_w(\mathbf{Z}_+)$.

Proof. Denote the forward shift operator by $Eq(x) = q(x+1)$. We start by noting the useful relation

$$a^{-k}\|\sqrt{x^{\underline{k}}}q\|_{l_w^2}^2 = \|E^k q\|_{l_w^2}^2.$$

Expanding $E^k = (I+\Delta)^k$ in binomial terms yields

$$\begin{aligned}\|E^k q\|_{l_w^2}^2 &= \sum_{x\geq 0}\left(\sum_{j=0}^{k}\binom{k}{j}\Delta^j q(x)\right)^2 w(x) \leq \\ &\leq \sum_{x\geq 0}\sum_{j=0}^{k}\binom{k}{j}^2\sum_{j=0}^{k}\left(\Delta^j q(x)\right)^2 w(x) \leq 4^k\|q\|_{h_{w,\Delta}^k}^2.\end{aligned}$$

Summing this for $k = 0\ldots m$ gives the first bound with (say) $C_1^{-1} = 2^{m+1}$. The second bound with $C_2 = 2^{m+1}$ can be proved in exactly the same way, expanding $\Delta^k = (E-I)^k$ instead. $\square$

The Sobolev-spaces generated by Δ are sometimes more convenient to work with and are perhaps more natural analogues to the standard continuous Sobolev-spaces. In Section 2.1 we defined the modified backward difference operator $\bar{\nabla}p(x) \equiv p(x-1) - p(x)$ with the exception of $\bar{\nabla}p(0) \equiv -p(0)$. Interestingly, there is no uniform (w.r.t. a) analogue of Proposition 2.3 for this operator. As a counter-example we note that the unit pulse at $x = 0$, i.e. $p(x) = 1$ if $x = 0$ and zero otherwise, yields $\|\bar{\nabla}p\|_{l_w^2}^2 = (1+a)\exp(-a)$ whereas $\|p\|_{h_w^m}^2 = \exp(-a)$ for *any* m. Thus, any bound on $\bar{\nabla}$ (or ∇) must be non-uniform with respect to a. In spite of this we prove the following partial result in this direction which will be helpful in order to bound a certain Sturm-Liouville operator to be introduced shortly.

Proposition 2.5 *The map $F : h_w^{m+2}(\mathbf{Z}_+) \to h_w^m(\mathbf{Z}_+)$ defined by $F(q) = x/a\cdot\nabla q$ is continuous. Furthermore, if $a \geq 1$, then the continuity is uniform with respect to this parameter.*

Proof. Split the operator according to $F(q) = x/a\cdot q - x/a\cdot E^{-1}q$, where E^{-1} is the backward shift operator (note that the convention $q(-1) = 0$ is not needed here). By definition the former map is continuous between h_w^{m+2} and h_w^m, although not necessarily uniformly so. Under the assumption $a \geq 1$, we expand for some constants A_k and B_k, $x^2 = (x-k)(x-k-1) + A_k(x-k) + B_k$. Then

$$a^{-k}\|\sqrt{x^{\underline{k}}}\,x/a\cdot q\|_{l_w^2}^2 = a^{-(k+2)}\|\sqrt{x^{\underline{k+2}} + A_k x^{\underline{k+1}} + B_k x^{\underline{k}}}\cdot q\|_{l_w^2}^2,$$

and the bound is uniform with respect to a. As for the operator $x/a \cdot E^{-1}$, we proceed similarly for $k \geq 1$,

$$a^{-k}\|\sqrt{x^{\underline{k}}}x/a \cdot E^{-1}q\|_{l^2_w}^2 = a^{-(k+2)}\sum_{x\geq 0}(x+1)^{\underline{k}}(x+1)^2 q(x)^2 w(x+1) = \\ = a^{-(k+1)}\sum_{x\geq 0}\left(x^{\underline{k+1}} + A_k x^{\underline{k}} + B_k x^{\underline{k-1}}\right) q(x)^2 w(x).$$

A similar strategy for $k = 0$ establishes that in fact, $x/a \cdot E^{-1} : h_w^{m+1}(\mathbf{Z}_+) \to h_w^m(\mathbf{Z}_+)$ is continuous (and uniformly so if $a \geq 1$). $\square$

In contrast to Propositions 2.3 and 2.5, we note that by expanding $\Delta^m = (E-I)^m$ into binomials and using the triangle inequality we get $\|\Delta^m p\|_{l^2} \leq 2^m\|p\|_{l^2}$. Incidentally, this explains why the *unweighted* Sobolev-spaces h^m cannot be generated by Δ. A slight generalization of this fact will later be useful to bound the regularity of general master operators:

Proposition 2.6 *The maps $\bar{\nabla} : h^m(\mathbf{Z}_+) \to h^m(\mathbf{Z}_+)$ and $\Delta : h^m(\mathbf{Z}_+) \to h^m(\mathbf{Z}_+)$ are continuous.*

Proof. We only consider the result for Δ since the proof of the other case is similar. By writing $\Delta = E - I$ it suffices to prove the boundedness of the forward shift operator E. Define $\bar{q}(x) = q(x)$ except for $\bar{q}(0) \equiv 0$ and note that by inspection,

$$\|\sqrt{x^{\underline{k}}}Eq\|_{l^2} = \|\sqrt{(x-1)^{\underline{k}}}\bar{q}\|_{l^2} = \|\sqrt{x^{\underline{k}} - kx^{\underline{k-1}}}\bar{q}\|_{l^2} \leq \|\sqrt{x^{\underline{k}}}q\|_{l^2},$$

with the exception of the trivial case $k = 0$. $\square$

2.3 Charlier approximation

We now let $C_n^a(x)$ denote the *normalized* nth degree Charlier polynomial [26] with parameter $a > 0$. These polynomials form an orthonormal set of functions with respect to the l_w^2-product; $(C_n^a, C_m^a)_w = \delta_{nm}$. We write X_N for the span of the (Charlier-) polynomials of degree $\leq N$ and define π_N as the orthogonal projection onto X_N associated with $(\cdot,\cdot)_w$.

The normalized Charlier polynomials satisfy the recurrence

$$\begin{aligned} C_0^a(x) &\equiv 1, \\ C_1^a(x) &\equiv \frac{a-x}{\sqrt{a}}, \\ C_{n+1}^a(x) &= \frac{n+a-x}{\sqrt{a(n+1)}}C_n^a(x) - \sqrt{\frac{n}{n+1}}C_{n-1}^a(x). \end{aligned} \tag{2.29}$$

There is a also a Charlier difference equation,

$$SC_n^a(x) := -w^{-1}(x)\nabla\left[w(x)\Delta C_n^a(x)\right] = \frac{n}{a}C_n^a(x), \tag{2.30}$$

where the Sturm-Liouville operator S can be expanded as

$$Sp = \frac{x}{a}\nabla p - \Delta p. \tag{2.31}$$

Charlier's polynomials are compatible with the forward difference operator in the sense that

$$\Delta C_n^a(x) = -\sqrt{n/a}\cdot C_{n-1}^a(x), \tag{2.32}$$

and they also obey the interesting relation

$$C_n^a(x) = (-1)^n\sqrt{\frac{n!}{a^n}}L_n^{x-n}(a), \tag{2.33}$$

where L_n^a denote *Laguerre polynomials* [26] with the usual normalization.

As is well-known in Sturm-Liouville theory, the approximation properties of orthogonal functions depend crucially on the regularity of the corresponding Sturm-Liouville operator. This is the motivation for our interest in Propositions 2.5 and 2.3 since they immediately yield (cf. the definition (2.31) of the operator S),

Lemma 2.7 *The operator $S : h_w^{m+2}(\mathbf{Z}_+) \to h_w^m(\mathbf{Z}_+)$ is continuous and thus bounded. If $a \geq 1$ is assumed, then the continuity is uniform with respect to this parameter.*

Recall now the summation by parts formula in the following form:

$$\sum_{x=0}^{N} p(x)\Delta q(x) = p(N)q(N+1) - p(-1)q(0) - \sum_{x=0}^{N}\nabla p(x)q(x), \tag{2.34}$$

where usually we will have that both boundary terms vanish. The following lemma relates the coefficients of an orthogonal expansion in terms of Charlier polynomials with the Sturm-Liouville operator S.

Lemma 2.8 *Let $p \in h_w^m(\mathbf{Z}_+)$. Then*

$$\textit{(m even)} \quad (p, C_n^a)_w = (a/n)^{m/2}\left(S^{m/2}p, C_n^a\right)_w, \tag{2.35}$$

$$\textit{(m odd)} \quad (p, C_n^a)_w = -(a/n)^{m/2}\left(\Delta S^{(m-1)/2}p, C_{n-1}^a\right)_w. \tag{2.36}$$

Proof. In view of (2.30) we get

$$(p, C_n^a)_w = -\frac{a}{n}\left(p, \nabla\left[w\Delta C_n^a\right]\right).$$

Summation by parts then yields in turn

$$= \frac{a}{n}\left(\Delta p, w\Delta C_n^a\right) = -\frac{a}{n}\left(\nabla\left[w\Delta p\right], C_n^a\right) = \frac{a}{n}\left(Sp, C_n^a\right)_w.$$

If m is even, repeating this procedure a total of $m/2$ times concludes the proof of (2.35). For the odd case we continue from

$$\begin{aligned}(a/n)^{(m-1)/2}\left(S^{(m-1)/2}p, C_n^a\right)_w &= -(a/n)^{(m+1)/2}\left(S^{(m-1)/2}p, \nabla\left[w\Delta C_n^a\right]\right) = \\ &= (a/n)^{(m+1)/2}\left(\Delta S^{(m-1)/2}p, \Delta C_n^a\right)_w.\end{aligned}$$

Using (2.32) now produces (2.36). □

Theorem 2.9 *For any nonnegative integer m, there exists a positive constant C depending only on m and a such that, for any function $p \in h_w^m(\mathbf{Z}_+)$, the following estimate holds*

$$\|\pi_{N-1}p - p\|_{l_w^2(\mathbf{Z}_+)} \le C(a/N)^{m/2}\|p\|_{h_w^m(\mathbf{Z}_+)}. \tag{2.37}$$

If in addition, $a \ge 1$ is assumed, then C depends only on m.

Proof. Expanding any function $p \in l_w^2$ in terms of Charlier polynomials, we readily get

$$\|\pi_{N-1}p - p\|_{l_w^2}^2 = \sum_{n\ge N}\bar{p}_n^2,$$

where, provided m is an even integer, we have by Lemma 2.8 that

$$\bar{p}_n = (p, C_n^a)_w = (a/n)^{m/2}\left(S^{m/2}p, C_n^a\right)_w.$$

Hence,

$$\|\pi_{N-1}p - p\|_{l_w^2}^2 \le (a/N)^m \sum_{n\ge N}\left(S^{m/2}p, C_n^a\right)_w^2 \le (a/N)^m\|S^{m/2}p\|_{l_w^2}^2.$$

Lemma 2.7 thus concludes the even case. When m is an odd integer we proceed similarly, using instead the odd version of Lemma 2.8. □

Theorem 2.9 is reminiscent of results for continuous approximations. See for example Theorem 12.1 in [3] (p. 289) for approximating continuous functions over $\mathbf{R}_+$ by Laguerre polynomials. Worth noting with the continuous theory is the technical need for an additional hierarchy of Sobolev-spaces (cf. equation (12.6) in [3], p. 288). This can be avoided completely in the present case thanks to Lemma 2.4.

We are now in the position to consider approximation in stronger norms. The following lemma makes this possible although we would like to point out that the given bound is very weak and can easily be improved upon. It seems, however, that such improvements only complicate what follows.

Lemma 2.10 *For a constant C depending only on m,*

$$\|C_n^a\|_{h_w^m(\mathbf{Z}_+)} \leq C \max(1, n/a)^{m/2}. \tag{2.38}$$

Proof. It is easier to prove this using the norm $\|\cdot\|_{h_{w,\Delta}^m(\mathbf{Z}_+)}$. From (2.32) we immediately get

$$\|\Delta^k C_n^a\|_{l_w^2}^2 = \frac{n^{\underline{k}}}{a^k} \leq \frac{n^k}{a^k}.$$

Summation yields $\|C_n^a\|_{h_{w,\Delta}^m} \leq \sqrt{m}\max(1, n/a)^{m/2}$ and the bound follows. □

This "smoothness" of the basis polynomials yields the following generalization of Theorem 2.9:

Theorem 2.11 *For any nonnegative integers k and m, $k \leq m$, there exists a positive constant C depending only on m and a such that, for any function $p \in h_w^m(\mathbf{Z}_+)$, the following estimate holds*

$$\|\pi_{N-1}p - p\|_{h_w^k(\mathbf{Z}_+)} \leq C(a/N)^{m/2}\max(1, N/a)^{k/2}\|p\|_{h_w^m(\mathbf{Z}_+)}. \tag{2.39}$$

Again, C depends only on m if $a \geq 1$ is assumed.

Proof. Again it is convenient to construct the proof in the uniformly equivalent norm $\|\cdot\|_{h_{w,\Delta}^m(\mathbf{Z}_+)}$. The case $k = 0$ corresponds to Theorem 2.9 and we proceed by induction, assuming that (2.39) holds for some k. Split the error according to

$$\begin{aligned}\|\pi_{N-1}p - p\|_{h_{w,\Delta}^{k+1}} &\leq \|\pi_{N-1}p - p\|_{l_w^2} + \|\pi_{N-1}\Delta p - \Delta p\|_{h_{w,\Delta}^k} + \\ &+ \|\Delta\pi_{N-1}p - \pi_{N-1}\Delta p\|_{h_{w,\Delta}^k} \leq \\ &\leq C_1(a/N)^{m/2}\|p\|_{h_{w,\Delta}^m} + C_2(a/N)^{(m-1)/2-k/2}\|\Delta p\|_{h_{w,\Delta}^{m-1}} + \\ &+ \|\Delta\pi_{N-1}p - \pi_{N-1}\Delta p\|_{h_{w,\Delta}^k},\end{aligned} \tag{2.40}$$

where Theorem 2.9 and the induction hypothesis have been used. Evidently, in this norm we have that $\|\Delta p\|_{h^{m-1}_{w,\Delta}} \leq \|p\|_{h^m_{w,\Delta}}$, and so we focus on the last term. Writing as before

$$p = \sum_{n\geq 0} \bar{p}_n C^a_n,$$

we readily obtain

$$\Delta\pi_{N-1}p = \sum_{n=0}^{N-1} \bar{p}_n \Delta C^a_n \quad \text{and} \quad \pi_{N-1}\Delta p = \sum_{n=0}^{N} \bar{p}_n \Delta C^a_n.$$

The last term in (2.40) is therefore

$$\|\bar{p}_N \Delta C^a_N\|_{h^k_{w,\Delta}} \leq |\bar{p}_N| \|C^a_N\|_{h^{k+1}_{w,\Delta}} \leq C_3 |(p, C^a_N)_w| \max(1, N/a)^{(k+1)/2},$$

where Lemma 2.10 was used. By Lemma 2.8, Cauchy-Schwarz's inequality, and Lemma 2.7 this becomes

$$\leq C_4 (a/N)^{m/2} \max(1, N/a)^{(k+1)/2} \|p\|_{h^m_{w,\Delta}},$$

finishing the induction step. □

In other words, the cost for measuring the error in the stronger norm $\|\cdot\|_{h^k_w(\mathbf{Z}_+)}$ is determined by the regularity of the basis, a situation that again is encountered in many continuous settings (for the corresponding result for Laguerre polynomials, see Theorem 12.3 in [3], p. 291).

We now take a look at approximation in the unweighted Sobolev-spaces $h^m(\mathbf{Z}_+)$. Define *Charlier's functions* by $\hat{C}^a_n(x) := C^a_n(x) \cdot w(x)^{1/2}$ along with the space $\hat{X}_N = \{p(x) = q(x) \cdot w(x)^{1/2};\, q \in X_N\}$. Evidently, these functions are orthonormal under the usual l^2-product $(\cdot, \cdot)$ and we use $\hat{\pi}_N$ to denote the corresponding orthogonal projection on $\hat{X}_N$. The relation

$$\hat{\pi}_N p = w(x)^{1/2} \pi_N \left(p(x) \cdot w(x)^{-1/2} \right) \tag{2.41}$$

is immediate and we make the crucial observation that the map $p \to w^{1/2}p$ is an *isomorphism* between h^m_w and h^m. This implies the following result.

Corollary 2.12 *For any nonnegative integers k and m, $k \leq m$, there exists a positive constant C depending only on m and a (or only on m provided $a \geq 1$ is given) such that, for any function $p \in h^m(\mathbf{Z}_+)$, the following estimate holds*

$$\|\hat{\pi}_{N-1}p - p\|_{h^k(\mathbf{Z}_+)} \leq C(a/N)^{m/2} \max(1, N/a)^{k/2} \|p\|_{h^m(\mathbf{Z}_+)}. \tag{2.42}$$

Corollary 2.12 is again related to similar results for continuous approximation; see for example [31].

There are several reasons for preferring to seek approximations to solutions of the master equation in the space $\hat{X}_N$ rather than in X_N. First, any Galerkin formulation of the master equation in the inner product $(\cdot,\cdot)_w$ will at best lead to convergence in the corresponding norm $\|\cdot\|_{l^2_w}$. In contrast, a convergent Galerkin formulation in the l^2-product will of course imply the existence of error estimates in the much stronger l^2-norm. Second, solutions in X_N are not probability distributions and statistical functionals of interest, such as the mean and variance, can therefore not be computed.

2.4 Conservation of probability and adaptivity

Unfortunately, the projection $\hat{\pi}_N$ is not sufficiently conservative for our present purposes. The reason is that it does not preserve the probability mass; in general we have that $(1, \hat{\pi}_N p) \neq (1, p) = 1$. For the projection to be conservative we need to somehow enforce the preservation of total probability. We therefore consider the projection $\hat{\pi}^0_N p = p_N$ which for some Lagrange multiplier λ satisfies

$$\left.\begin{array}{rcl} (q, p_N - p) + \lambda(f(q), 1) & = & 0 \\ (1, p_N - p) & = & 0 \end{array}\right\} \text{ for } \forall q \in \hat{X}_N, \tag{2.43}$$

where f is a suitable nonzero linear function to be determined. To analyze this projection, we first note that, regardless of the norm,

$$\|\hat{\pi}^0_N p - p\| \leq \|\hat{\pi}^0_N p - \hat{\pi}_N p\| + \|\hat{\pi}_N p - p\| \tag{2.44}$$

and that as consequences of (2.43),

$$\hat{\pi}^0_N p - \hat{\pi}_N p = \sum_{n=0}^{N} (\tilde{p}_n - \bar{p}_n)\hat{C}^a_n = -\lambda \sum_{n=0}^{N} (f(\hat{C}^a_n), 1)\hat{C}^a_n, \tag{2.45}$$

$$(1, \hat{\pi}_N p - p) = -(1, \hat{\pi}^0_N p - \hat{\pi}_N p) = \lambda \sum_{n=0}^{N} (f(\hat{C}^a_n), 1)(\hat{C}^a_n, 1), \tag{2.46}$$

where $\tilde{p}_n$ and $\bar{p}_n$ are the coefficients produced by $\hat{\pi}^0_N$ and $\hat{\pi}_N$, respectively. The l^2-deviation between $\hat{\pi}^0_N$ and the orthogonal projection $\hat{\pi}_N$ is therefore generally given by

$$\|\hat{\pi}^0_N p - \hat{\pi}_N p\|^2_{l^2} = \lambda^2 \sum_{n=0}^{N} (f(\hat{C}^a_n), 1)^2. \tag{2.47}$$

For the traditional Lagrangian choice $f(q) = q$, (2.45) and (2.46) can be combined into

$$\|\hat{\pi}_N^0 p - \hat{\pi}_N p\|_{l^2}^2 = \frac{(1, \hat{\pi}_N p - p)^2}{\sum_{n=0}^{N} (\hat{C}_n^a, 1)^2}. \tag{2.48}$$

However, this Lagrangian projection turns out to be somewhat inconvenient in a Galerkin formulation of the master equation. It *can* efficiently be computed by using the Sherman-Morrison formula [20], but modifying all frequencies $\hat{C}_n^a$ is more complicated and does not yield a much smaller error than modifying the lowest frequency $\hat{C}_0^a$ alone. This corresponds to the choice $f(q) = \hat{\pi}_0 q$ and satisfies (from (2.45) and (2.46))

$$\|\hat{\pi}_N^0 p - \hat{\pi}_N p\|_{l^2}^2 = \frac{(1, \hat{\pi}_N p - p)^2}{(\hat{C}_0^a, 1)^2}. \tag{2.49}$$

Although the error (2.49) is slightly larger than (2.48), it is found experimentally that $\hat{C}_0^a$ carries more mass than the rest of the modes. Consequently, the sum in the denominator of (2.48) is replaced by the largest term so that in practice the difference is not critical. Remarkably then, using the "tau-method" [14, 21], or what amounts to the same thing, making the choice $f(q) = (\hat{C}_N^a, q)\hat{C}_N^a$, can not be recommended. The corresponding deviation has the same appearance as (2.49), but with the denominator replaced by $(\hat{C}_N^a, 1)^2$. Since the mass of $\hat{C}_N^a$ is smaller than that of $\hat{C}_0^a$, this method performs worse than the suggested projection.

Another feature of the choice $f(q) = \hat{\pi}_0 q$ is that a reasonably sharp error estimate in the l^1-norm is easily obtained. By inspection $\hat{C}_0^a$ is everywhere positive so that $(1, \hat{C}_0^a) = \|\hat{C}_0^a\|_{l^1}$. Hence from (2.44), (2.45) and (2.46),

$$\|\hat{\pi}_N^0 p - p\|_{l^1} \leq |(1, \hat{\pi}_N p - p)| + \|\hat{\pi}_N p - p\|_{l^1} \leq 2\|\hat{\pi}_N p - p\|_{l^1}. \tag{2.50}$$

In order to conclude this section we shall finally pay attention to the choice of the parameter a which must be chosen prior to forming any projection onto $\hat{X}_N$. We first claim that *if N is small and p is a "one-peak" probability distribution with expectation value m, then $a \approx m - 1/2$ is close to optimal.* By a "one-peak" probability distribution we mean a unimodal distribution with standard deviation relatively small compared to the expectation value.

To motivate this statement we consider the case $N = 0$ which means that p is to be approximated by the "half Poissonian distribution",

$$P^{1/2}(x;\, a) = C^{-1}\sqrt{\frac{a^x}{x!}e^{-a}}. \tag{2.51}$$

Here C is the normalizing constant given by

$$C = \sum_{x\geq 0} \sqrt{\frac{a^x}{x!}e^{-a}} = \\ = \sum_{x\geq 0} \frac{(a/2)^x}{x!}e^{-a/2} \times \underbrace{\pi^{1/4}\sqrt{\frac{\Gamma(x+1)}{\Gamma(x+1/2)}}\left(1+\sqrt{\frac{a}{2x+1}}\right)}_{=:f(x)}, \quad (2.52)$$

where the last line follows from summing even and odd terms separately and using the duplication formula for the gamma function [1]. To evaluate the sum in (2.52), we note that it can be compactly written as $E[f(X)]$ for X a Poissonian stochastic variable of expectation $a/2$. Expand f in a Taylor series around $a/2$ and assume a to be large so that Stirling's expansion [1] for the gamma function apply. Inserting formulas for the central moments of the Poissonian distribution then yields after some work,

$$C \sim 2^{3/4}\pi^{1/4}a^{1/4}\left(1-\frac{1}{16a}+\mathcal{O}\left(a^{-2}\right)\right). \quad (2.53)$$

Proceeding similarly for the expectation value m and the variance σ^2 we obtain

$$m \sim a+\frac{1}{2}+\frac{1}{8a}+\mathcal{O}\left(a^{-2}\right), \quad (2.54)$$

$$\sigma^2 \sim 2a-\frac{1}{4a}+\mathcal{O}\left(a^{-2}\right). \quad (2.55)$$

The method just described for obtaining these expressions is somewhat related to the method of Laplace [29] in the theory of asymptotic expansions of integrals. Interestingly, it can also be shown to be equivalent to a technique due to Ramanujan (see [4], Ch. 3, entry 10). The resulting formulas are surprisingly accurate already for quite small values of a. For example, the indicated three terms in (2.54) yield a relative error less than 0.06% even for $a = 2$. Taken together, (2.54) and (2.55) show that $P^{1/2}$ is situated slightly to the right of a Poisson distribution with the same parameter and is about 41% wider.

In conclusion then, if the probability distribution p is reasonably centered around its expectation value m, then we expect that the optimal approximation in $\hat{X}_0$ is nearly $P^{1/2}(x;\, m-1/2)$ in (2.51). As N grows and the approximating space gets larger, this estimate no longer holds true. By Corollary 2.12 we see that a should reasonably decrease for the error to rapidly become small. We resort to a small informative experiment.

In Figure 2.1 the l^2-error of projections using different a's and N's are shown together with the optimal choice of a thus determined. The behavior of the optimal value a_{opt} is found to agree with the above discussion; for small N we see that a_{opt} is slightly less than the expectation value, while it decreases with increasing order N. We note that the *global* trend of the error near the optimal value is quite flat so that the precise choice is not so important. The oscillating *local* behavior of the error can be explained by considering asymptotic expansions for the Charlier polynomials involving Bessel functions (see [8] for this fairly complicated issue).

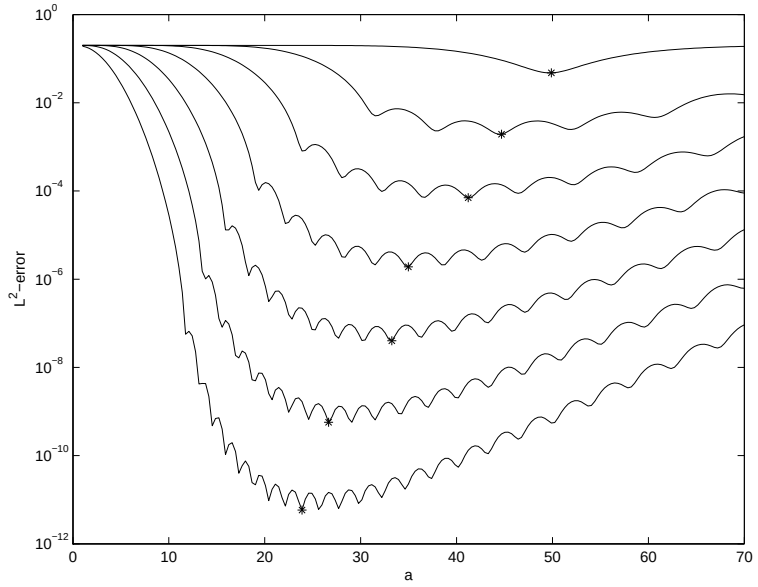

Figure 2.1: A Poisson distribution of expectation value 50 is projected on $\hat{X}_N$ for $N = 0, 5, \ldots, 30$ using several values of a. The l^2-error is determined for each choice of a producing the dependence shown. For each value of N, the asterisk indicates the optimal value of a.

2.5 Stability

The Galerkin approximation to (2.2) that we shall now analyze reads as follows (compare (2.43)): find $p_N \in \hat{X}_N$ such that

$$\left.\begin{array}{rcl} (q, \partial p_N/\partial t) + \lambda(\hat{\pi}_0 q, 1) & = & (q, \mathcal{M} p_N) \\ (1, \partial p_N/\partial t) & = & 0 \end{array}\right\} \qquad \text{for } \forall q \in \hat{X}_N. \tag{2.56}$$

Since $\mathcal{M}$ generally is unbounded, indefinite and non-symmetric with non-orthogonal eigenvectors, we cannot possibly hope to capture any convergence

properties of (2.56) by adhering to standard energy estimates. The partial results we present in this section are instead based on observations due to van Kampen [25, Ch. V] and Theorem 2.1. We will attempt to make it reasonable to believe that (2.56) is stable in the l^1-norm so that convergence follows from the Lax-Richtmyer equivalence theorem.

We first write (2.56) in the equivalent form

$$\frac{\partial p_N}{\partial t} = \hat{\pi}_N^0 \mathcal{M} p_N, \tag{2.57}$$

where the representation

$$p_N(x,t) = \sum_{n=0}^{N} c_n(t) \hat{C}_n^a(x) \tag{2.58}$$

is implicitly understood.

For a not necessarily positive or normalized p_N, let $U_N(t)$ denote the sum of the positive elements,

$$U_N(t) \equiv \sum_{x \geq 0} p_N(x,t)^+ = \sum_{x \in \xi^+(t)} p_N(x,t), \tag{2.59}$$

and similarly for $V_N(t)$, the sum of the negative elements. A crucial property of the master operator is that it preserves the probability mass and, since $\hat{\pi}_N^0$ is used rather than $\hat{\pi}_N$, this property holds true for $\hat{\pi}_N^0 \mathcal{M}$ as well. Thus,

$$U_N(t) + V_N(t) = \text{constant}. \tag{2.60}$$

The time derivative of $U_N(t)$ exists between the events when $p_N(x,t)$ changes sign for some x. In such intervals we have that

$$\begin{aligned} U_N'(t) = \sum_{x \in \xi^+(t)} \hat{\pi}_N^0 \mathcal{M} p_N &= \sum_{x \in \xi^+(t)} \mathcal{M} p_N + \sum_{x \in \xi^+(t)} \left[\hat{\pi}_N^0 \mathcal{M} p_N - \mathcal{M} p_N \right] \\ &=: A_N(t) + B_N(t), \end{aligned} \tag{2.61}$$

where from (2.2),

$$A_N(t) = \sum_{x \in \xi^+(t)} \sum_{r=1}^{R} \left[x \geq -N_r^- \right] w_r(x + N_r) p_N(x + N_r, t) - w_r(x) p_N(x,t), \tag{2.62}$$

and where the notation $[f]$, with f a logical expression, is used according to $[f] \equiv 1$ if f and $\equiv 0$ otherwise.

We always have that $A_N(t) \leq 0$. To see why, note that the sum over $x \in \xi^+(t)$ in (2.62) is built up by *some* of the positive elements of $w_r p_N$ minus *all* of its positive elements.

In the analytical case when B_N vanishes this constitutes a second proof of Theorem 2.1, since $\|p\|_{l^1} = U - V$ and, by (2.60), $-V' = U'$. When for some x, $p(x,t)$ changes sign, then although the derivative does not exist, the l^1-norm still depends continuously on time and consequently it cannot increase.

For the numerical investigation to follow, we see that

$$\frac{d}{dt}\|p_N\|_{l^1} = 2A_N(t) + 2B_N(t), \tag{2.63}$$

except for those points in time where a change of sign occurs. We would like to claim that *for* N *sufficiently large,* $|A_N| \geq |B_N|$*, so that the* l^1*-norm of the numerical solution does not increase.*

Although perhaps not a general proof of the claim, the motivation is that $\hat{\pi}_N^0 \mathcal{M} p_N$ approaches $\mathcal{M} p_N$ when N grows so that in principle B_N can be made arbitrarily small. We are then clearly interested in the cases when A_N is zero, since this could induce a growth of the l^1-norm. Three cases are trivial: either one of ξ^+, $\xi^- := \{x;\, p_N(x,t) < 0\}$ or $\xi^0 := \{x;\, p_N(x,t) = 0\}$ is identical to $\mathbf{Z}_+$ so that p_N is either positive, negative or zero. In all these cases B_N vanishes as well since the mass-preserving projection is used — hence the l^1-norm must stay constant. Two cases force a closer examination of the master operator: *(i)* ξ^0 is empty while ξ^+ and ξ^- are not, and, *(ii)* all the sets ξ^+, ξ^- and ξ^0 are non-empty.

Case (i): Relabel the states so that p_N is divided into a positive and a negative part. Then we have

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{++} & 0 \\ 0 & \mathcal{M}_{--} \end{bmatrix}, \tag{2.64}$$

for otherwise A_N cannot be zero (the zero in the lower left corner stems from using $-V' = U'$ and writing down the analogues of (2.61) and (2.62) for V_N). Hence $\mathcal{M}$ is *decomposable* (see (2.11)) and is excluded from the present context since it does not describe a fully interacting chemical system but rather contains two isolated systems. Again, these considerations are formally only valid for a finite number of states (see the discussion in the end of Section 2.1).

Case (ii): Split the states of p_N into a positive, an all-zero, and a negative part, respectively. Accordingly,

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{++} & \mathcal{M}_{+0} & 0 \\ 0 & \mathcal{M}_{00} & 0 \\ 0 & \mathcal{M}_{-0} & \mathcal{M}_{--} \end{bmatrix}. \tag{2.65}$$

That is, $\mathcal{M}$ is a *splitting* (see (2.12)) and can be excluded from the discussion for the same reason as above.

In conclusion then, A_N can only be zero when B_N is simultaneously zero (indicating a constant l^1-norm of p_N). If this is not the case, then we have that $A_N < 0$ and B_N tends to zero with increasing N. To look at the possible dependence of the magnitudes of these two terms, we write out the derivative of the l^1-norm explicitly (by expanding $U_N' - V_N'$ as in (2.61)),

$$\frac{d}{dt}\|p_N\|_{l^1} = \underbrace{\sum_{x\geq 0} \operatorname{sgn}\, p_N\, \mathcal{M}p_N}_{2A_N} + \underbrace{\sum_{x\geq 0} \operatorname{sgn}\, p_N \left[\hat{\pi}_N^0 \mathcal{M}p_N - \mathcal{M}p_N\right]}_{2B_N}, \tag{2.66}$$

where sgn q is zero for $q = 0$. It seems reasonable to believe that there are estimates of the form $2|A_N| \geq \kappa(\mathcal{M})\|\mathcal{M}p_N\|_{l^1}$ and $2|B_N| \leq C_N\|\mathcal{M}p_N\|_{l^1}$, where $\kappa(\mathcal{M})$ is a constant depending on the structure of the master operator and where C_N tends to zero when N increases. Under these assumptions, the derivative of the l^1-norm is $\leq (C_N - \kappa(\mathcal{M}))\|\mathcal{M}p_N\|_{l^1}$ which for N sufficiently large is less than or equal to zero. This argument does not prove l^1-stability unless the indicated estimates are first proved but it does, however, shed some light on the expected stability properties of the Galerkin scheme (2.56). Also, the crucial argument in this discussion is the relation (2.60) which indicates why a mass-preserving projection is a favorable choice.

2.6 Further details

In this section we will describe the suggested numerical scheme in some detail. The assembly process is discussed and we also demonstrate a feasible way to continuously update the parameter a so as to allow the basis functions to capture the dynamics of the solution.

Since the master operator is defined in D dimensions, we need to make use of multi-indices which we denote by small Greek letters. If $\alpha = [\alpha_1, \ldots, \alpha_D]$ and x is a D-dimensional array, then we index x by

$$x_\alpha = x_{\alpha_1,\ldots,\alpha_D}. \tag{2.67}$$

In addition, the following products occur naturally,

$$\beta^\alpha = \beta_1^{\alpha_1} \cdots \beta_D^{\alpha_D}, \tag{2.68}$$

$$\alpha! = \alpha_1! \cdots \alpha_D!, \tag{2.69}$$

$$e^\alpha = e^{\alpha_1} \cdots e^{\alpha_D}. \tag{2.70}$$

The easiest way of constructing a basis in D-dimensions is to use a tensor basis. We thus write

$$\hat{C}_\gamma^a(x) \equiv \prod_j \hat{C}_{\gamma_j}^{a_j}(x_j). \tag{2.71}$$

Evidently, this system of polynomials is orthonormal with respect to the inner product

$$(f, g) \equiv \sum_{x \geq 0} f(x)g(x) \prod_j \frac{a_j^{x_j}}{x_j!} e^{-a_j} = \sum_{x \geq 0} f(x)g(x) \frac{a^x}{x!} e^{-a}, \tag{2.72}$$

where x and a now are vector quantities. The solution to the master equation is thus represented compactly as

$$p(x, t) = \sum_\gamma c_\gamma(t) \hat{C}_\gamma^a(x). \tag{2.73}$$

Multiplying both sides of the master equation (2.2) by $\hat{C}_\delta^a$ and summing over $\mathbf{Z}_+^D$ yields the set of equations

$$\begin{aligned} c_\delta' &= \left(\hat{C}_\delta^a, \mathcal{M}p\right) = \left(\mathcal{M}^* \hat{C}_\delta^a, p\right) = \\ &= \sum_{r=1}^{R} \sum_\gamma \left(C_\delta^a(x - N_r) \cdot \sqrt{a^{-N_r} x!/(x - N_r)!} - C_\delta^a(x), w_r(x) c_\gamma C_\gamma^a(x) \right), \end{aligned} \tag{2.74}$$

where the favorable representation of the adjoint has been used. The use of orthogonality to simplify the above expression is notationally non-trivial but computationally quite simple. What is left is R different sums to be performed over the dimensions involved in each reaction; i.e. the dimensions i such that w_r depends on x_i and/or N_{ri} is non-zero. *The number of dimensions in each sum is almost always bounded by 4.* For example, this is the case with the reaction $x + y \to z$ with propensity $w(x, y, e)$. That is, when two species interact under the influence of an enzyme e.

The sums themselves are computed using an associated *Gauss-Charlier quadrature* [12]. In one dimension it is given by

$$\sum_{x\geq 0} f(x)\frac{a^x}{x!}e^{-a} = \sum_{j=1}^{n} f(x_j)w_j + R_n, \tag{2.75}$$

$$R_n = a^n n!\frac{f^{(2n)}(\xi)}{(2n)!}, \qquad \xi \in (0,\infty). \tag{2.76}$$

The x_j's are the roots of $C_n^a(x)$ and the weights can be computed according to the formula

$$w_j \equiv -\frac{(an)^{-1/2}}{C_{n-1}^a(x_j)\cdot d/dx\, C_n^a(x_j)}. \tag{2.77}$$

Turn now to a discussion of the interesting and novel strategy of dynamically adapting the parameter a. Intuitively, the basis is most "active" in a neighborhood of $x \sim a$, and consequently we would like to adjust a so as to rapidly capture the behavior of the represented solution. A different but related viewpoint is that the quadrature points tend to be more densely populated around a and we would like to ensure that no quadrature points are "wasted".

Accordingly, let $a = a(t)$ in (2.73). Then very formally,

$$\frac{\partial p(x,t)}{\partial t} = \sum_\gamma c'_\gamma \hat{C}_\gamma^a(x) + \sum_\gamma \left(\frac{a'}{C_\gamma^a(x)}\frac{d}{da}C_\gamma^a(x) + \frac{x}{2}\frac{a'}{a} - \frac{a'}{2}\right) c_\gamma \hat{C}_\gamma^a(x), \tag{2.78}$$

which is just the product rule for derivatives. From (2.33) and a formula for the derivative of the Laguerre polynomials [1],

$$\frac{d}{dx}L_n^a(x) = -L_{n-1}^{a+1}(x), \tag{2.79}$$

one readily gets in the scalar case,

$$\frac{d}{da}C_n^a(x) = -\frac{n}{2a}C_n^a(x) + \sqrt{\frac{n}{a}}C_{n-1}^a(x). \tag{2.80}$$

Inserting this and using the recurrence (2.29), one can simplify the derivative, thereby finding

$$\frac{\partial p(x,t)}{\partial t} = \sum_\gamma c'_\gamma \hat{C}_\gamma^a(x) + \tag{2.81}$$

$$+ \sum_\gamma c_\gamma \sum_j \hat{C}_{\gamma\setminus\gamma_j}^{a\setminus a_j}(x\setminus x_j)\left(-\frac{1}{2}\sqrt{\frac{\gamma_j+1}{a_j}}\hat{C}_{\gamma_j+1}^{a_j}(x_j) + \frac{1}{2}\sqrt{\frac{\gamma_j}{a_j}}\hat{C}_{\gamma_j-1}^{a_j}(x_j)\right)a'_j.$$

Here we had to be able to remove dimensions from the product — the precise meaning of the above notation is simply

$$\hat{C}^{a\setminus a_j}_{\gamma\setminus\gamma_j}(x \setminus x_j) \equiv \prod_{i\neq j} \hat{C}^{a_i}_{\gamma_i}(x_i). \tag{2.82}$$

It follows that

$$\left(\hat{C}^a_\delta, \frac{\partial p(x,t)}{\partial t}\right) = c'_\delta + \sum_j \left(-\frac{1}{2}\sqrt{\frac{\delta_j}{a_j}} c_{\delta-1_j} + \frac{1}{2}\sqrt{\frac{\delta_j+1}{a_j}} c_{\delta+1_j}\right) a'_j \tag{2.83}$$

where $\delta \pm 1_j$ is just $[\delta_1, \ldots, \delta_j \pm 1, \ldots, \delta_D]$. Once a' has been prescribed, it is straightforward to combine (2.83) and the right side of (2.74) to produce equations for the derivatives of the coefficients.

We have seen in Section 2.4 that it is difficult to exactly find the best value of a to represent the solution p. Once a sufficiently good value has been determined, however, it seems natural from the discussion in Section 2.4 to dynamically update $a(t)$ with the expectation value. Suppose therefore that a good choice of $a(0)$ has been made so that $p(\cdot, 0)$ is efficiently represented. Then we define a' by the derivative of the expectation value and obtain the following algorithm:

1. Compute c'_δ by assembling (2.74).

2. Determine the derivative of the expectation value according to the coefficients just computed and let a' take this value.

3. Account for the dynamic basis by updating c'_δ according to (2.83).

In practice we also enforce $a \geq 1$ since all results in Section 2.3 are uniform under this restriction. The usefulness of this technique is demonstrated in Section 3.3

To conclude this section we finally comment on the mass-preserving issue since we have actually only discussed how to form the l^2-projection $\hat{\pi}_N$. The reason for this is that forming $\hat{\pi}^0_N$ follows as a corollary: simply compute the derivative of the mass under $\hat{\pi}_N$, then update the lowest order derivative c'_0 so that the resulting coefficients carry a stationary mass. In the case of a stationary parameter a this amounts to simply summing the ansatz (2.73) over all the integers using a suitable Gauss-Charlier quadrature. When the parameter is dynamic one proceeds in a similar fashion although this time one has to compute the derivative of the mass according to the slightly more involved expression (2.81).

In summary it is a straightforward (but not trivial) task to write general software using the suggested scheme. Inputs include the reactions (w_r, N_r) and a *reaction topology* to help sorting out the dependence between the dimensions. After forming a suitable initial distribution, any ODE-solver (explicit or implicit) can be used to evolve the coefficients in a time-dependent setting. Alternatively, an iterative linear or nonlinear solver can be used in the case of a steady-state formulation. The dynamic parameter a helps capture solutions which vary over many scales in time, but a static parameter is usually preferable for steady-state solutions.

3 Numerical experiments

We will now demonstrate the feasibility of the proposed method by numerically solving four different models. The first two are one-dimensional model problems with known solutions and are used to demonstrate the application of the theory and the numerical convergence. The third model is four-dimensional with two metabolites and two enzymes, and the task is to find the steady-state distribution. By contrast, the fourth model is dynamic and takes place in two dimensions only. Here, the behavior of the solution is more complicated and the example provides a setting for which the deterministic reaction-rate approach fails. The section concludes with a short discussion of efficiency and a comparison with Monte-Carlo simulations.

3.1 Convergence and application of the theory

As a numerical demonstration of convergence and in order to highlight the application of the theory developed in Sections 2.2 and 2.3 we first consider two one-dimensional models with known solutions. The first is the linear birth-death problem (2.13) with time-dependent solution given by (2.16) and the second is the following set of reactions characterizing spontaneous bimolecular decay:

$$\left.\begin{array}{rcl} \emptyset & \xrightarrow{k} & X \\ X+X & \xrightarrow{\nu x(x-1)} & \emptyset \end{array}\right\}. \tag{3.1}$$

The master equation for this model can be written compactly as

$$\frac{\partial p(x,t)}{\partial t} = -k\bar{\nabla}p(x,t) + \nu\left[\Delta^2 + 2\Delta\right][x(x-1)p(x,t)], \tag{3.2}$$

where again the bar over the backward difference operator ∇ expresses the convention that $p(-1,t) = 0$.

The full time-dependent solution to (3.2) is not known but the steady-state solution can be expressed in terms of the modified Bessel function I_n [25, X.2]. Let $a = \sqrt{k/\nu}$ and define $C = \sqrt{2}I_1(2\sqrt{2}a)$. Then

$$p(x, \infty) = C^{-1}\frac{a^x}{x!}I_{x-1}(2a) \tag{3.3}$$

is the exact stationary solution to (3.2). The expectation value m and the variance v can now be determined from standard asymptotics [1]:

$$m(\infty) \sim \sqrt{\frac{k}{2\nu}} + \frac{1}{8} + \mathcal{O}\left(a^{-1}\right), \tag{3.4}$$

$$v(\infty) \sim \frac{3}{4}\sqrt{\frac{k}{2\nu}} + \frac{1}{16} + \mathcal{O}\left(a^{-1}\right). \tag{3.5}$$

In order to apply the theory of Section 2.2 we first note that the master operator in (2.14) is bounded when regarded as an operator $h^{m+2} \to h^m$. We already know from Proposition 2.6 how the backward- and forward differences behave and it is clear by definition that $p \to xp$ is bounded as a map $h^{m+2} \to h^m$. The situation is similar in (3.2). Here the quadratic factor $x(x-1)$ decreases the regularity by four degrees so that the master operator (3.2) is bounded as a map $h^{m+4} \to h^m$.

The analysis thus conducted is very general. Firstly, note that the reaction steps are always finite so that the master equation can always be written as a linear combination of terms that are finite compositions of difference operators. Secondly, *atomic* propensities that follow from combinatorial arguments are always low order polynomials, easily bounded as shown above. More general non-atomic propensities are also common. For instance, when quasi steady-state approximations are used for the derivation of a reduced model, the result is a set of rational functions involving roots of polynomials. These expressions can generally be majorized by low order polynomials and the resulting master operator can again be bounded as outlined above.

The parts of the analysis that we have to leave in a more incomplete state is the stability of the scheme and the regularity of the solution. The stability was discussed in Section 2.5 and we will simply assume that the scheme is stable. For the present case, the regularity is immediate since we do know the exact solutions. To generally determine conditions on the master operator in D dimensions for a certain given degree of regularity of the solution seems difficult, even in steady-state.

The numerical experiments were conducted as follows. We let the parameters be defined by $[k, \mu] = [1, 10^{-3}]$ in (2.13) and $[k, \nu] = [1, 5 \cdot 10^{-7}]$ in (3.1). This makes the expectation value in steady-state to be $m = 1000$

very closely for both models. For (2.13), we start at $t = 0$ with a Poisson distribution with mean 10 and evolve the system until $t = 10^4$. By comparing the solution thus obtained to the exact solution, the time-averaged error was determined. For the non-linear problem (3.1), since the solution is known at steady-state only, a Poisson distribution of mean 1000 was used as the initial data. Again the system was evolved until $t = 10^4$ and the solution at this point was considered to be in steady-state. In order to be able to compare the two results, a dynamic parameter a was used in both experiments — this is clearly only needed for the linear birth-death problem were the initial data is located far away from steady-state.

In Figure 3.1, errors in various measures are shown and it is clear that the convergence is exponential in the order N of the scheme. The slower rate of convergence for the non-linear problem can be explained by the lesser regularity of both the solution and the corresponding master operator.

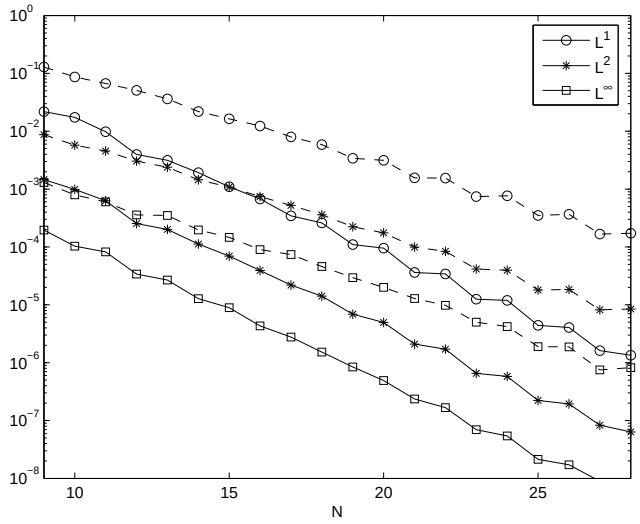

Figure 3.1: Errors of the scheme applied to (2.13) (solid, time-average error) and (3.1) (dashed, error at steady-state) in different norms.

In conclusion all master operators for physically realizable chemical systems can be bounded in a tensor product version of the spaces developed in Section 2.2. In addition, it seems very likely that there are highly general conditions for when the regularity of the solution remains at least at the same degree as the initial data during any interval (finite or infinite) of integration. Under the assumption of stability we then expect exponential convergence of the current scheme and we now proceed to confirm this for two more realistic

models were the analysis would be more difficult.

3.2 Enzyme-control of metabolites

This example comes from [10] and is a model of the synthesis of two metabolites X and Y by two enzymes E_X and E_Y. The reactions are

$$\left.\begin{array}{rcl} \emptyset & \xrightarrow{k_1 e_x/(1+x/k_i)} & X \\ X & \xrightarrow{\mu x} & \emptyset \\ \emptyset & \xrightarrow{k_2 e_y/(1+y/k_i)} & Y \\ Y & \xrightarrow{\mu y} & \emptyset \\ X+Y & \xrightarrow{kxy} & \emptyset \end{array}\right\} \quad \left\{\begin{array}{rcl} \emptyset & \xrightarrow{k_3/(1+x/k_r)} & E_X \\ E_X & \xrightarrow{\mu e_x} & \emptyset \\ \emptyset & \xrightarrow{k_4/(1+y/k_r)} & E_Y \\ E_Y & \xrightarrow{\mu e_y} & \emptyset \end{array}\right\}, \tag{3.6}$$

with parameters $k_1 = k_2 = 0.3$, $k_3 = k_4 = 0.02$, $k = 10^{-3}$, $\mu = 2 \cdot 10^{-3}$, $k_i = 60$ and $k_r = 30$. As it stands, (3.6) is the result of an *adiabatic* [15, Ch. 6.4] simplification of a more complete model. This is generally done by eliminating intermediate products under the assumption that they rapidly reach steady-state.

The steady-state solution as obtained by the scheme is displayed in Figure 3.2. We have tried several different discretizations with the constant value $a = [20, 20, 2, 2]$ as a reasonable parameter for the basis. Steady-state was obtained by explicit time-stepping from initial data in the form of a Poisson distribution with the expectation value a and was reached approximately at $T = 1500$. By comparing each solution to a higher order reference solution, different norms of the error were computed and the result is displayed in Figure 3.3, where the exponential convergence is reasonably explicit. The obtained solution is visually pleasing and free of numerical artifacts already at the quite coarse discretization $[15, 15, 8, 8]$ (degrees per dimension according to the ordering $[x, y, e_x, e_y]$).

3.3 Bistable toggle switch

A biological *toggle switch* can be formed by two mutually cooperatively repressing gene products X and Y [16]. The relevant equations are

$$\left.\begin{array}{rcl} \emptyset & \xrightarrow{a/(b+y^2)} & X \\ X & \xrightarrow{\mu x} & \emptyset \\ \emptyset & \xrightarrow{c/(d+x^2)} & Y \\ Y & \xrightarrow{\mu y} & \emptyset \end{array}\right\}, \tag{3.7}$$

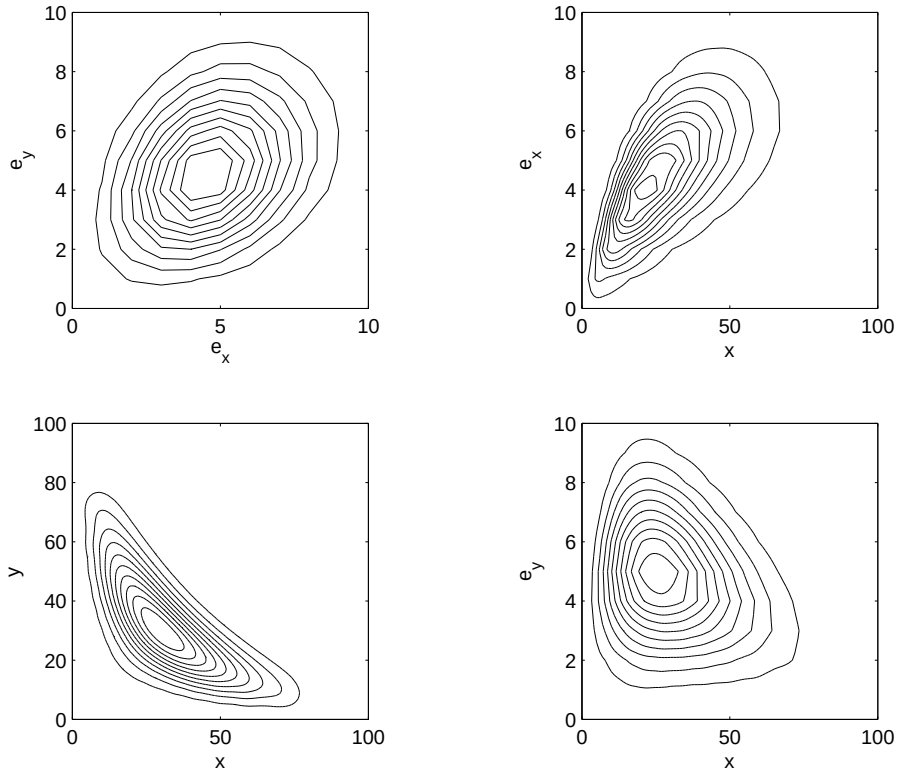

Figure 3.2: Steady-state solution (marginal distributions) to (3.6). The correlation between the various species can be understood from first principles, except perhaps for the somewhat irregular dependence between X and E_Y.

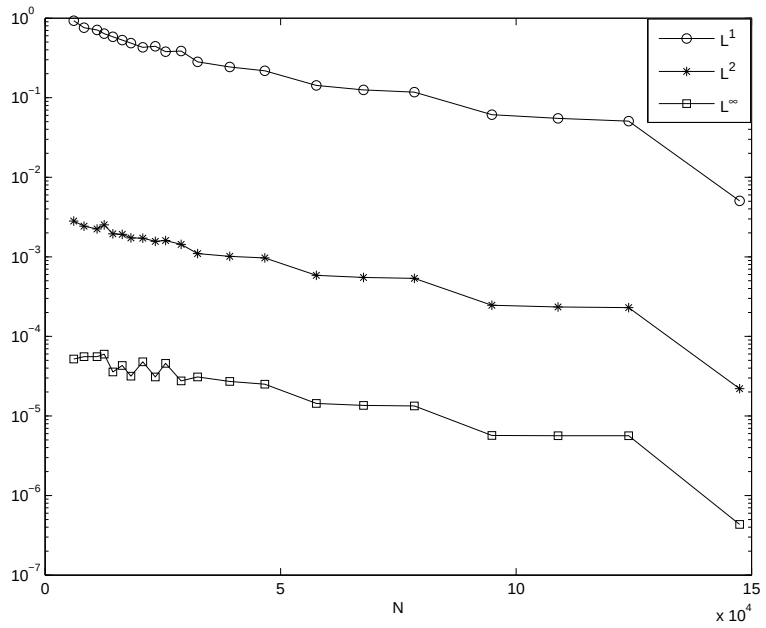

Figure 3.3: Errors of the scheme applied to (3.6) as measured in different norms. Note that the x-axis displays the *total* number of degrees of freedom.

with parameters $a = c = 1000$, $b = d = 6000$ and $\mu = 10^{-3}$. It is easy to get a rough feeling for the behavior of (3.7). Suppose that initially, the number of x-molecules is large and that the number of Y-molecules is small. Then we see that the production of Y-molecules is inhibited so that the system will find a stable state with $x > y$. However, by a certain small probability the stochastic noise can make the number of Y-molecules eventually grow. Due to this 'tunneling' effect, the production of X-molecules will instead be inhibited and the roles of X and Y may suddenly switch. This behavior is explicitly seen in Figure 3.4 where the result of a stochastic simulation with SSA [18] is displayed.

We solved (3.7) using various order N and various initial data. The parameter a was dynamic and followed the expectation value of the solution as explained in Section 2.6. The stiffness of the problem is clearly visible in Figure 3.5, and so an implicit ODE-solver has been our preferred choice (we used `MATLAB`'s `ode15s`).

The error has been estimated using a high order reference solution with stricter tolerances for the time-stepper. In Figure 3.6 several different norms of the error are displayed and the exponential convergence of the method is clearly visible. Figure 3.5 indicates a visually pleasing result already at a quite coarse discretization. Overall, no problems of instabilities were ever

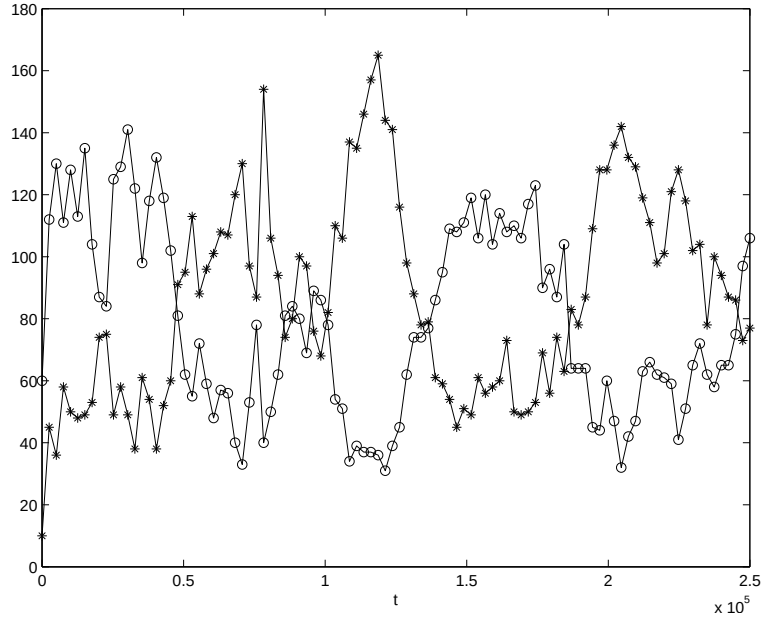

Figure 3.4: One realization of (3.7) obtained by Gillespie's algorithm. In this simulation the system 'switches' three times and we also see a 'near-switch' slightly before $t = 10^5$.

encountered although phase errors were more pronounced for small N.

The formation of two distinct peaks is an interesting feature which makes the toggle switch an example for which the reaction-rate approach must fail. The expectation value obtained using this method comes to rest near one of the two peaks with an additional unstable critical point situated in between them. The 'switching' feature of the system is thus not present and the deterministic solution is of limiting insight.

3.4 Discussion

Although a complete discussion of spectral methods versus Monte-Carlo simulations is beyond the scope here, we offer a few comments of relevance to the examples presented. First, note that *directly* representing the solution in either one of Figure 3.2 or 3.5 is about 100 times more memory consuming than the method proposed.

It is difficult to compare a spectral method to a Monte-Carlo simulation since they produce different information. If we agree to only look at the error in moment, the error ϵ of the spectral method satisfies

$$\epsilon \sim \exp(-N^{1/D}) \tag{3.8}$$

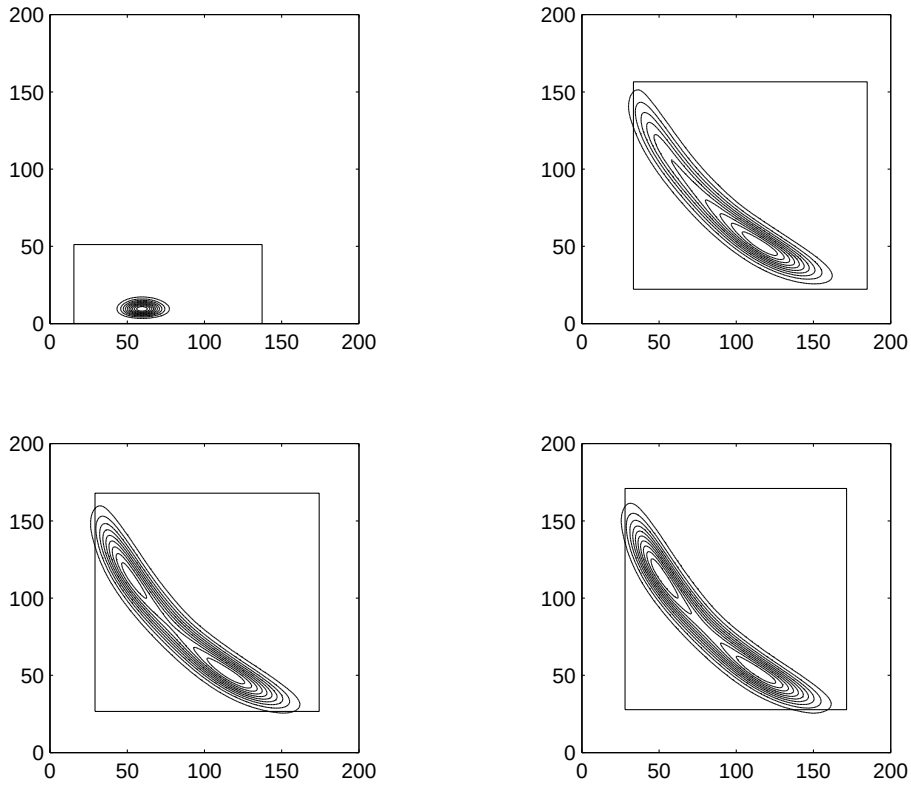

Figure 3.5: The solution to (3.7) at time $t = [0, 1, 1.5, 2.5] \cdot 10^5$ using $N = 19$ (400 coefficients). The simulation starts with a Poissonian solution centered at $(x, y) = (60, 10)$ and ends in equilibrium where two distinct peaks have formed. The indicated bounding box contains all quadrature points and follows the solution quite well. Note the stiffness of the problem: the fast scale is the transport along the line $x - y =$ constant, while the slow distribution along $x \cdot y =$ constant is much more of diffusive character.

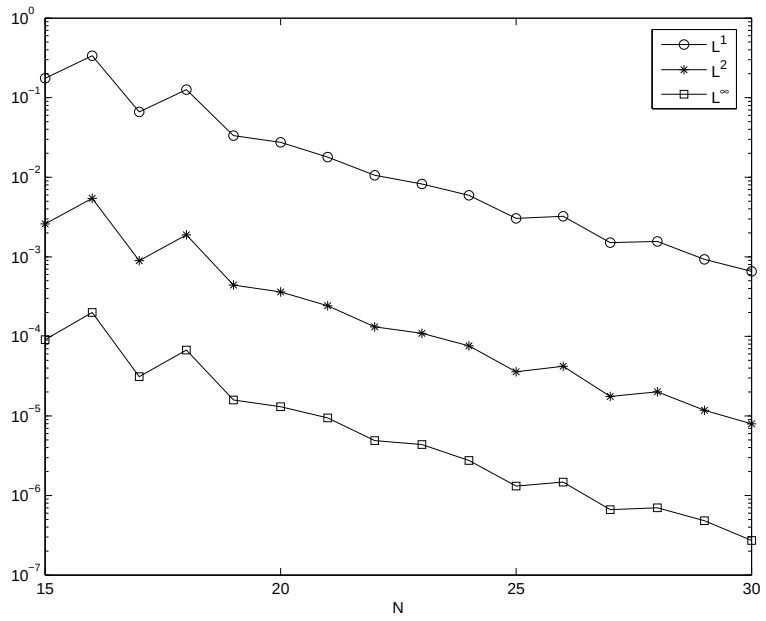

Figure 3.6: Errors of the scheme applied to (3.7) in different norms (time-average) for increasing order N. The exponential convergence of the method is clearly visible.

for N modes in D dimensions. The work for obtaining this error becomes in an idealized case $W \sim N \sim (-\log \epsilon)^D$, whereas for a Monte-Carlo method the work is the familiar $W \sim \epsilon^{-2}$. Thus, for not too high dimensionality and sufficient accuracy demands, the proposed method will be more efficient.

More precise reasoning has to take the constants into account and as an example of this we consider the bistable system with parameters chosen as in [32] (see Figure 3.7). For this particular example, a single simulated trajectory is just about as expensive to compute as a highly accurate spectral representation of the solution to the master equation.

We finally wish to emphasize that the proposed method is not only an alternative to Monte-Carlo simulations. Methods for simulating realizations of a system and computing its probability density are not mutually exclusive and are effective in answering different kinds of questions.

4 Conclusions

The master equation is a stochastic description of general dynamical systems described by discrete coordinates and is equivalent to the Markov assumption. This description applies in particular to chemical reactions where the

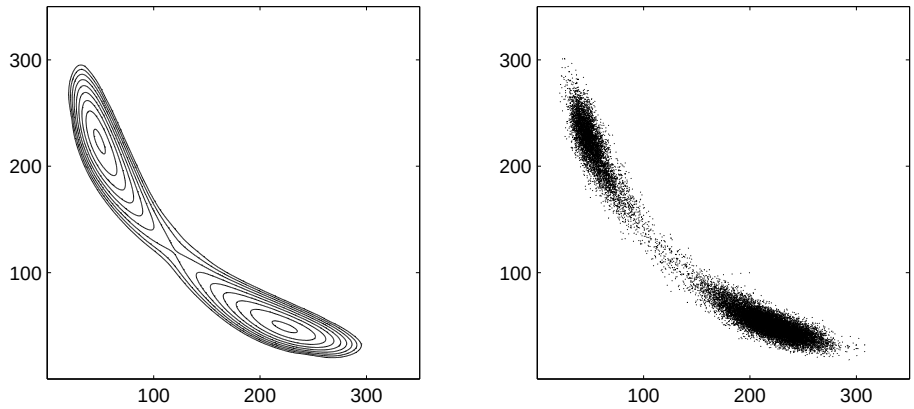

Figure 3.7: Steady-state solution to (3.7) with parameters $a = c = 3000$, $b = d = 11000$ and $\mu = 10^{-3}$ [32]. The problem starts at $t = 0$ in the lower right state and steady-state is approximately reached at $t = 5 \cdot 10^6$. Left: the result of the spectral method of order 39 (1600 coefficients), contour levels $4 \cdot 10^{-4} \cdot 2^{-i}$ for $i = 0, \ldots, 8$. Right: one single trajectory simulated using SSA. Note that the system switches only two times during the simulation and that the computational cost for obtaining the solutions displayed is about the same.

system is described by counting the number of molecules of each kind. If the number of molecules is large, then an effective and usually very accurate description in terms of deterministic ODEs for the expectation values can be formed. However, stochastic descriptions are preferred for many systems of interest; examples can be found inside living cells where the effects of stochasticity are important.

Monte-Carlo simulations such as Gillespie's SSA are effective in computing single trajectories, but the solution obtained in terms of the full probability density function can provide additional insight. Statistical parameters can be accurately determined, certain inverse problems are made feasible or can be solved to greater accuracy, and the exact nature of the processes involved can be studied more closely.

We have presented a novel spectral method for the master equation based on Charlier functions. Key features include high accuracy even at a relatively low resolution per dimension, convergence in the full semi-infinite discrete state-space, and a strategy for dynamically keeping the basis functions adjusted to the solution they represent. The numerical experiments suggest that the scheme is effective when the dimensionality of the problem is not too high.

After this work was completed, the author became aware of a similar

method for polyreaction kinetics devised by Deuflhard and Wulkow [7, 33, 34]. Their original scheme uses the basis $\{\tilde{C}_n^a(x)\} := \{C_n^a(x) \cdot w(x)\}$ and determines the coefficients by taking the l^2-product with $\{C_n^a(x)\}$ in a Petrov-Galerkin formulation. The setting is one-dimensional and the model problem treated is reminiscent of the birth-death problem (2.13). However, when applied to the master equation of dimensionality higher than one, their scheme does not seem to be stable with the method of lines discretization. A related recent work directly aimed at the master equation is found in [6].

We can also relate the method to the *Poisson representation* [15, Ch. 7.7] which assumes that the solution to the master equation can be written as a superposition of multivariate uncorrelated Poisson distributions:

$$p(x,t) = \int f(a,t) \frac{a^x}{x!} e^{-a}\, da. \tag{4.1}$$

It is possible to cast the master equation into an explicit equation for the new unknown density f, thereby mapping the master equation in (x,t)-space into a partial differential equation in (a,t)-space. Note, however, that the relation (4.1) may well imply an arbitrarily peaky and discontinuous f from a fairly smooth p (e.g. f is a Dirac-function for the simple model-problem (2.13)). This observation suggests that perhaps the Poisson representation is better thought of as a tool for deriving various analytical results rather than as a numerical method.

We would also like to mention some possible improvements to the proposed method. First, for certain problems the solution may become a very flat distribution. The proposed basis is really only an effective representation when the essential support of the solution is clustered around the expected value m as $m \pm \mathcal{O}(\sqrt{m})$. If this condition is violated so that the solution is very non-Poissonian, a very high order N is needed in order to resolve the problem. A cure is to *scale* the solution appropriately which for a discrete solution amounts to incorporating *aggregation*. Aggregation of continuous-time Markov chains has been described in [23] and in the setting of the sparse grids technique for the master equation in [24].

Another improvement would stem from coupling the described method to the reaction-rate equations. One is frequently interested in the precise behavior of the solution in a few dimensions only and the representation in terms of expectation values might well suffice for the major part. A drastic efficiency gain is thus possible, making really high dimensional problems tractable. Some steps in this direction for the Fokker-Planck equation have been taken in [27].

Acknowledgment

Many comments and suggestions by Per Lötstedt have improved the content of this paper. Various inputs from Jan Hesthaven and Stephen Lau were helpful during the final stages of writing and Henrik Hult helped in sorting out the dependence of Theorem 2.1 on the contractivity of the corresponding semigroup. The author is also indebted to Hermann Matthies for a long and valuable discussion during the IHP-EU Workshop "Breaking Complexity" in Bad Honnef, Germany. References [6, 7, 33, 34] was kindly provided to the author by Peter Deuflhard and Michael Wulkow.

References

[1] M. Abramovitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1970.

[2] W. J. Anderson. *Continuous-Time Markov Chains*. Springer Series in Statistics. Springer-Verlag, New York, 1991.

[3] C. Bernardi and Y. Maday. Spectral methods. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis*, volume V, pages 209–487. North-Holland, Amsterdam, 1997.

[4] B. C. Berndt. *Ramanujan's Notebooks, Part I*. Springer-Verlag, New York, 1985.

[5] Y. Cao, D. Gillespie, and L. Petzold. Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *J. Comput. Phys.*, 206:395–411, 2005. doi:10.1016/j.jcp.2004.12.014.

[6] P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow. Adaptive discrete Galerkin methods applied to the chemical master equation. Technical Report 07-04, Zuse Institute Berlin, 2007. Available at http://www.zib.de/bib/pub.

[7] P. Deuflhard and M. Wulkow. Computational treatment of polyreaction kinetics by orthogonal polynomials of a discrete variable. *IMPACT Comp. Sci. Eng.*, 1(3):269–301, 1989. doi:10.1016/0899-8248(89)90013-X.

[8] T. M. Dunster. Uniform asymptotic expansions for Charlier polynomials. *J. Approx. Theory*, 112:93–133, 2001.

[9] E. B. Dynkin. *Markov processes*, volume I. Academic Press, 1965.

[10] J. Elf, P. Lötstedt, and P. Sjöberg. Problems of high dimension in molecular biology. In W. Hackbusch, editor, *Proceedings of the 19th GAMM-Seminar in Leipzig "High dimensional problems - Numerical Treatement and Applications"*, pages 21–30, 2003.

[11] S. Engblom. Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comput.*, 180(2):498–515, 2006. doi:10.1016/j.amc.2005.12.032.

[12] S. Engblom. Gaussian quadratures with respect to discrete measures. Technical Report 2006-007, Dept of Information Technology, Uppsala University, Uppsala, Sweden, 2006. Available at http://www.it.uu.se/research.

[13] L. Ferm, P. Lötstedt, and P. Sjöberg. Conservative solution of the Fokker-Planck equation for stochastic chemical reactions. *BIT*, 46:S61–S83, 2006. doi:10.1007/s10543-006-0082-z.

[14] B. Fornberg. *A Practical Guide to Pseudospectral Methods*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 1996.

[15] C. W. Gardiner. *Handbook of Stochastic Methods*. Springer Series in Synergetics. Springer-Verlag, Berlin, 3rd edition, 2004.

[16] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403:339–342, 2000.

[17] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, 104(9):1876–1889, 2000. doi:10.1021/jp993732q.

[18] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4):403–434, 1976. doi:10.1016/0021-9991(76)90041-3.

[19] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1992. doi:10.1016/0378-4371(92)90283-V.

[20] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

[21] D. Gottlieb and S. A. Orszag. *Numerical Analysis of Spectral Methods: Theory and Applications*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, 1977.

[22] P. Guptasarama. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays*, 17(11):987–997, 1995. doi:10.1002/bies.950171112.

[23] M. Haviv. Aggregation/disaggregation method for computing the stationary distribution of a Markov chain. *SIAM J. Numer. Anal.*, 24(4):952–966, 1987.

[24] M. Hegland, C. Burden, L. Santoso, S. MacNamara, and H. Booth. A solver for the stochastic master equation applied to gene regulatory networks. *J. Comput. Appl. Math.*, 205(2):708–724, 2007. doi:10.1016/j.cam.2006.02.053.

[25] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, 5th edition, 2004.

[26] R. Koekoek and R. F. Swarttouw. The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue. Technical Report 98-17, Delft University of Technology, Faculty of Information Technology and Systems, Department of Technical Mathematics and Informatics, 1998. Available at http://aw.twi.tudelft.nl/~koekoek/askey.html.

[27] P. Lötstedt and L. Ferm. Dimensional reduction of the Fokker-Planck equation for stochastic chemical reactions. *Multiscale Meth. Simul.*, 5:593–614, 2006.

[28] H. G. Matthies and A. Keese. Galerkin methods for linear and nonlinear stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 194(12–16):1295–1331, 2005. doi:10.1016/j.cma.2004.05.027.

[29] F. W. J. Olver. *Asymptotics and Special Functions*. Academic Press, New York, 1974.

[30] M. S. Samoilov and A. P. Arkin. Deviant effects in molecular reaction pathways. *Nature Biotech.*, 24:1235–1240, 2006. doi:10.1038/nbt1253.

[31] J. Shen. Stable and efficient spectral methods in unbounded domains using Laguerre functions. *SIAM J. Numer. Anal.*, 38(4):1113–1133, 2000.

[32] P. Sjöberg, P. Lötstedt, and J. Elf. Fokker-Planck approximation of the master equation. Technical Report 2005-044, Dept of Information Technology, Uppsala University, Uppsala, Sweden, 2005. Available at `http://www.it.uu.se/research`. To appear in *Comp. Vis. Sci.*

[33] M. Wulkow. Adaptive treatment of polyreactions in weighted sequence spaces. *IMPACT Comp. Sci. Eng.*, 4(2):153–193, 1992. `doi:10.1016/0899-8248(92)90020-9`.

[34] M. Wulkow. The simulation of molecular weight distributions in polyreaction kinetics by discrete Galerkin methods. *Macromol. Theory. Simul.*, 5(3):393–416, 1996. `doi:10.1002/mats.1996.040050303`.

[35] D. Xiu and J. S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005. `doi:10.1137/040615201`.

[36] D. Xiu and G. E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644, 2002. `doi:10.1137/S1064827501387826`.