A Discrete Spectral Method for the Chemical Master Equation

Stefan Engblom[1] *

June 30, 2006

[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 in the form of a Galerkin spectral method with an inherent natural adaptivity and a very favorable choice of basis functions.

The method is exemplified by the numerical solution of two systems taken from molecular biology. It is shown that the method remains effective and accurate when other traditional solution methods produce less useful results.

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

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

1 Introduction

The celebrated *Markov character* of stochastic processes plays a crucial role in many fields of physics and mathematics. For a certain physical system

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

observed at discrete times $t_1 < t_2 < \cdots t_n$ it states that the conditional probability for the event (y_n, t_n) given the systems history satisfies

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

i.e. that the dependence on past events can be captured by the dependence on the state (y_{n-1}, t_{n-1}) only. Although (1.1) clearly cannot always be true, it is frequently a very accurate and useful approximation. Accuracy results when the discrete time is chosen sufficiently coarse in comparison with the often very short auto-correlation time of the system; usefulness is a consequence of the fact that Markovian systems can be described using only the initial probability $P(y_1, t_1)$ and the *transition probability function* $P(y_s, s|y_t, t)$ [2].

The *master equation* is a formulation of the Markov assumption for discrete variables 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 governs the dynamics of the probability distribution for the system.

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 it are of interest both in research and practice.

Being directly derivable from the master equation, one often considers solving the *reaction-rate* equations instead. This is a set of D ordinary differential equations (ODEs) approximating the expected values of the concentrations of the species in the system. In examples involving few molecules of some of the species, however, this approach can fail to reproduce the actual behavior of the system considered. When biological systems inside living cells are considered the number of molecules is often less than 10^2 [18] and additionally the system is 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.

As an important alternative there are stochastic simulation techniques that offer the ability to follow sample trajectories of the system. *Gillespie's SSA* method [15], or versions of it [14], are the methods in most frequent use. Contrasting to the reaction-rate approach, such methods have in common that they are *exact* in a statistical sense. Although simulating one trajectory can be performed relatively cheaply, high computational cost results when many trajectories need to be simulated for statistical parameters to be accurately determined. Another drawback is the explicit flavor of the method which in particular means that many small time-steps need to be simulated for systems near unstable points or in general when stiff systems are investigated [5].

Recent numerical considerations for the master equation include the numerical solution of the *Fokker-Planck equation* [7, 10] and adaption of the *Sparse grids technique* [19]. Being a representation of fairly general *continuous* stochastic processes [12], numerically solving the Fokker-Planck equation is an interesting subject in itself. When regarded as a continuous approximation to the master equation, however, it is difficult to say *a priori* how good the approximation will be [20]. The sparse grids technique is directly aimed at reducing the computational complexity of high dimensional smooth problems. Its application to the master equation is quite recent and appears to be 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. Efficiency is thus obtained by the spectral compactification 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.

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 two relevant cases from molecular biology. We conclude the paper by discussing the various merits of the method and point 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.

In the context of *Stochastic differential equations* (SDEs) this discretization has been investigated by several authors (see for example [25, 26]), but we have seen no references of spectral methods applied to master equations. The most notable difference between these two types of descriptions lies in the derivation; an SDE is expressed in stochastic variables and is usually arrived at by adding noise to a deterministic description. By contrast, a master equation governs the probability density function of the process and must be determined directly from its statistical properties. Every master equation can be cast into an SDE (in a generalized sense) but going in the opposite direction is in general much more difficult.

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 2.3 and we conclude by devoting Section 2.4 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* $w_r : \mathbf{Z}_+^D \longrightarrow \mathbf{R}_+$. These define the transition probability per unit of time for moving from the state x_r to x;

$$x_r = x + n_r \xrightarrow{w_r(x_r)} x, \tag{2.1}$$

where $n_r \in \mathbf{Z}^D$ is the transition step and is the rth column in the *stoichiometric matrix n*.

The *master equation* [12, 20] 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 to be 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. [20, VII.2] and [12, 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 [20, V.9]. If (p, q) is a pair of not necessarily normalized or positive functions defined over $\mathbf{Z}_+^D$, then as long as both sides make sense (see [8] 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,

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

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 $x \in \mathbf{Z}_+^D$. An additional remark is that 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 $T : \mathbf{Z}_+^D \longrightarrow \mathbf{R}$ a suitable test-function. Using this form of the adjoint, equations for the various moments of X can be formed (see [8] for a numerical investigation of this approach).

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

$$X_{k+1} = X_k + n e_m, \tag{2.6}$$

$$t_{k+1} = t_k + \tau_k, \tag{2.7}$$

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

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

and where the time-step τ_k is determined by selecting an exponentially distributed random variable with parameter α. This formulation is actually directly derived from Gillespie's SSA method [15] and can of course be regarded as a stochastic *difference* equation.

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 the real part of λ must be ≤ 0 so that all eigenvalues of $\mathcal{M}$ share this property. Moreover, we also see that $q \equiv 1$ 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 now prove a certain stability variation of this result.

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

$$\sum_{x \ge 0} |p(x,t)| \le \sum_{x \ge 0} |p(x,0)| \tag{2.10}$$

for any $t \ge 0$.

Proof. It is easier to prove this result by considering the adjoint equation under the dual norm of L^1, which is L^∞. 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 [20, V.9]). Consequently, $\|q\|_{L^\infty}$ cannot increase. In Section 2.3 a second and more direct proof of this result is indicated. □

Remark. Theorem 2.1 guarantees that for any $p \ge 1$, the L^p-norm of the solution stays bounded. However, for $p > 1$ it is frequently the case that the norm is slowly growing in time. □

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

Theorem 2.2 *Let $p(x,0)$ be a given discrete function 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 $t \longrightarrow \infty$. Moreover, if $p(x,0)$ is a discrete probability density, then so is the steady-state solution.*

For a proof and a penetrating discussion we refer to [20, 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

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

Excluding these cases essentially forces $\mathcal{M}$ to define a fully interacting system and thus it is not allowed to consist of several isolated systems.

All these considerations are 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 (eg. the one-dimensional "random walk" on the form $\emptyset \xrightarrow{k} x$). 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 birth-death process [2]

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

that is, 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 Poisson distribution of expectation a_0,

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

For 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 $a(t) = a_0 \exp(-\mu t) + k/\mu \cdot (1 - \exp(-\mu t))$. Independently of the initial data, p approaches a Poisson distribution of expectation k/μ.

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

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

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

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

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

2.2 Semi-infinite discrete approximation

In this section 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 polynomial approximation over these discrete spaces is developed. The corresponding results for continuous approximation are of course well-known but the discrete version seems to have been largely overlooked in the literature. We are then able to transfer the results for polynomials into an approximation quality of certain discrete *functions* instead which is crucial for the type of L^1-solutions to the master equation that we wish to represent.

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 additional degree of freedom is discussed towards the end of the section and it proves to be useful as a means of improving the efficiency of the proposed scheme. To this end we shall only consider one-dimensional functions; the corresponding tensor basis is given explicitly in Section 2.4.

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

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

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

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 we shall use the following hierarchy of discrete Sobolev-spaces:

$$H^m(\mathbf{Z}_+) = \left\{ q : \mathbf{Z}_+ \longrightarrow \mathbf{R};\ \sqrt{x^{\underline{k}}} \cdot q(x) \in L^2(\mathbf{Z}_+) \text{ for } 0 \leq k \leq m \right\} \tag{2.24}$$

with 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 $a \in \mathbf{R}_+$ will appear quite naturally.

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

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

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 $p(0) = p(1) \equiv 0$. For simplicity we let the parameter a be equal to 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 the following two basic results concerning the forward- and backward difference operator.

Proposition 2.3 *The map* $\Delta : H^{m+1}_w(\mathbf{Z}_+) \longrightarrow H^m_w(\mathbf{Z}_+)$ *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*

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

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

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

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

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

Summing this for $k = 0 \dots m$ 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. □

In other words, we could equally well use the Sobolev spaces generated by Δ instead, being perhaps more remindful of the usual continuous Sobolev spaces. However, note that the *unweighted* Sobolev spaces cannot be generated by Δ since in fact, $\|\Delta^m p\|_{L^2} \le \sqrt{m} 2^m \|p\|_{L^2}$.

Interestingly, there is no uniform (w.r.t. a) analogue of Proposition 2.3 for the backward difference operator. As a counter-sample we note that the unit pulse at $x = 0$, i.e. $p(x) = 1$ if $x = 0$ and zero otherwise, yields $\|\bar{\nabla} p\|^2_{L^2_w} = (1 + a)\exp(-a)$ whereas $\|p\|^2_{H^m_w} = \exp(-a)$ 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 $F : H^{m+2}_w(\mathbf{Z}_+) \longrightarrow H^m_w(\mathbf{Z}_+)$ defined by $F(q) = x/a \cdot \nabla q$ is continuous. Furthermore, if $a \ge 1$, 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^{m+2}_w and H^m_w, although not necessarily uniformly so. Under the assumption $a \ge 1$ we expand for some constants A_k and B_k, $x^2 = (x-k)(x-k-1) + A_k(x-k) + B_k$. Then

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

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

$$\begin{aligned} a^{-k} \|\sqrt{x^{\underline{k}}}\, x/a \cdot E^{-1} q\|^2_{L^2_w} &= a^{-(k+2)} \sum_{x \ge 0} (x+1)^{\underline{k}} (x+1)^2 q(x)^2 w(x+1) = \\ &= a^{-(k+1)} \sum_{x \ge 0} \left(x^{\underline{k+1}} + A_k x^{\underline{k}} + B_k x^{\underline{k-1}} \right) q(x)^2 w(x). \end{aligned}$$

A similar strategy for $k = 0$ concludes that in fact, $x/a \cdot E^{-1} : H^{m+1}_w(\mathbf{Z}_+) \longrightarrow H^m_w(\mathbf{Z}_+)$ is continuous (and uniformly so if $a \ge 1$). □

We now let $C^a_n(x)$ denote the *normalized* nth degree Charlier polynomial [21] with parameter $a > 0$. These polynomials form an orthonormal set of functions with respect to the L^2_w-product; $(C^a_n, C^a_m)_w = \delta_{nm}$. 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 according to $(\cdot,\cdot)_w$.

The normalized Charlier polynomials satisfy the recurrence

$$\begin{aligned} C^a_0(x) &\equiv 1, \\ C^a_1(x) &\equiv \frac{a-x}{\sqrt{a}}, \\ C^a_{n+1}(x) &= \frac{n+a-x}{\sqrt{a(n+1)}} C^a_n(x) - \sqrt{\frac{n}{n+1}} C^a_{n-1}(x). \end{aligned} \tag{2.29}$$

There is a also a Charlier difference equation,

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

where the Sturm-Liouville operator S can be expanded as

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

Charlier's polynomials combine well with the forward difference operator,

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

and we mention finally also the interesting relation

$$C^a_n(x) = (-1)^n \sqrt{\frac{n!}{a^n}} L^{x-n}_n(a), \tag{2.33}$$

where L^a_n denote *Laguerre polynomials* [21] under 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 2.6 *The operator $S : H^{m+2}_w(\mathbf{Z}_+) \longrightarrow H^m_w(\mathbf{Z}_+)$ is continuous and thus bounded. If $a \geq 1$ is assumed, then the continuity is uniform with respect to this parameter.*

We now need to recall the summation by parts formula on the following form:

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

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 2.7 *Let $p \in H_w^m(\mathbf{Z}_+)$. Then*

$$\begin{aligned} &\textit{(m even)} \quad (p, C_n^a)_w = (a/n)^{m/2} \left(S^{m/2}p, C_n^a\right)_w, & (2.35)\\ &\textit{(m odd)} \quad (p, C_n^a)_w = -(a/n)^{m/2} \left(\Delta S^{(m-1)/2}p, C_{n-1}^a\right)_w. & (2.36)\end{aligned}$$

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 2.8 *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*

$$\|\pi_{N-1}p - p\|_{L_w^2(\mathbf{Z}_+)} \leq C(a/N)^{m/2}\|p\|_{H_w^m(\mathbf{Z}_+)}. \qquad (2.37)$$

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

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

$$\|\pi_{N-1}p - p\|^2_{L^2_w} = \sum_{n \geq N} \bar{p}_n^2,$$

where, provided m is an even integer, we have that

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

by Lemma 2.7. Hence,

$$\|\pi_{N-1}p - p\|^2_{L^2_w} \leq (a/N)^m \sum_{n \geq N} \left(S^{m/2}p, C^a_n\right)^2_w \leq (a/N)^m \|S^{m/2}p\|^2_{L^2_w}.$$

Lemma 2.6 thus concludes the even case. When m is an odd integer we proceed similarly, using instead the odd version of Lemma 2.7. □

Theorem 2.8 is very remindful of results for continuous approximations. See for example Theorem 12.1 in [3] (p. 289) which is a similar result 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 of considering approximation in stronger norms. The following lemma makes this possible although we like to point out that the given bound is very weak and can easily be improved upon. It seems, however, as if such improvements only complicate what follows.

Lemma 2.9 *For a constant C depending only on m,*

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

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

$$\|\Delta^k C^a_n\|^2_{L^2_w} = \frac{n^{\underline{k}}}{a^k} \leq \frac{n^k}{a^k}.$$

Summing we get $\|C^a_n\|_{H^m_{w,\Delta}} \leq \sqrt{m} \max(1, n/a)^{m/2}$ and the bound follows. □

This ”smoothness” of the basis polynomials yields the following generalization of Theorem 2.8:

Theorem 2.10 *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 2.8 and we proceed by induction, assuming that (2.39) holds for some k. Split the error according to

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

where Theorem 2.8 and the induction hypothesis were used. Evidently, in this norm we have that $\|\Delta p\|_{H_{w,\Delta}^{m-1}} \leq \|p\|_{H_{w,\Delta}^m}$ and so we focus on the last term. Writing as before

$$p = \sum_{n \geq 0} \bar{p}_n C_n^a$$

we readily obtain

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

so that the last term is

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

where Lemma 2.9 was used. By Lemma 2.7, Cauchy-Schwarz's inequality and Lemma 2.6 this becomes

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

finishing the induction step. □

In other words, the cost for measuring the error in the stronger norm $\|\cdot\|_{H_w^k(\mathbf{Z}_+)}$ 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 $\hat{C}_n^a(x) := C_n^a(x) \cdot w(x)^{1/2}$ 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

$$\hat{\pi}_N p = w(x)^{1/2} \pi_N \left(p(x) \cdot w(x)^{-1/2}\right) \tag{2.40}$$

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

Corollary 2.11 *For any nonnegative integers k and m, $k \le m$, there exists a positive constant C depending only on m and a (or only on m provided $a \ge 1$ is given) such that, for any function $p \in H^m(\mathbf{Z}_+)$, the following estimate holds*

$$\|\hat{\pi}_{N-1} p - p\|_{H^k(\mathbf{Z}_+)} \le C(a/N)^{m/2} \max(1, N/a)^{k/2} \|p\|_{H^m(\mathbf{Z}_+)}. \tag{2.41}$$

Corollary 2.11 is again related to similar results for continuous approximation; see for example [24].

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. Firstly, any Galerkin formulation of the master equation in the inner product $(\cdot,\cdot)_w$ will at best lead to convergence in the corresponding norm $\|\cdot\|_{L_w^2}$. Contrary to this, 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. Secondly, solutions in X_N are not probability distributions and statistical functionals of interest, such as the mean and variance, can therefore not be computed.

Unfortunately, the projection $\hat{\pi}_N$ is not sufficiently conservative for our present purposes. The reason is that is 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 $\hat{\pi}_N^0 p = p_N$ which for some Lagrange multiplicator λ satisfies

$$\left.\begin{array}{rcl} (q, p_N - p) + \lambda(f(q), 1) & = & 0 \\ (1, p_N - p) & = & 0 \end{array}\right\} \text{ for } \forall q \in \hat{X}_N, \tag{2.42}$$

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

$$\|\hat{\pi}_N^0 p - p\| \leq \|\hat{\pi}_N^0 p - \hat{\pi}_N p\| + \|\hat{\pi}_N p - p\| \tag{2.43}$$

and that as consequences of (2.42),

$$\hat{\pi}_N^0 p - \hat{\pi}_N p = \sum_{n=0}^{N} (\tilde{p}_n - \bar{p}_n) \hat{C}_n^a = -\lambda \sum_{n=0}^{N} (f(\hat{C}_n^a), 1) \hat{C}_n^a, \tag{2.44}$$

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

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

$$\|\hat{\pi}_N^0 p - \hat{\pi}_N p\|_{L^2}^2 = \lambda^2 \sum_{n=0}^{N} (f(\hat{C}_n^a), 1)^2. \tag{2.46}$$

For the traditional Lagrangian choice $f(q) = q$, (2.44) and (2.45) can be combined into

$$\|\hat{\pi}_N^0 p - \hat{\pi}_N p\|_{L^2}^2 = \frac{(1, \hat{\pi}_N p - p)^2}{\sum_{n=0}^{N} (\hat{C}_n^a, 1)^2}. \tag{2.47}$$

However, the Lagrangian projection turns out to be a somewhat inconvenient choice in a Galerkin formulation of the master equation. It *is* possible to efficiently compute this projection by using the Sherman-Morrison formula [16], but the tendency to spread the loss of probability mass across all frequencies $\hat{C}_n^a$ is slightly more "nervous" than the much simpler choice $f(q) = \hat{\pi}_0 q$. This choice satisfies (from (2.44) and (2.45))

$$\|\hat{\pi}_N^0 p - \hat{\pi}_N p\|_{L^2}^2 = \frac{(1, \hat{\pi}_N p - p)^2}{(\hat{C}_0^a, 1)^2}. \tag{2.48}$$

Here the lowest frequency alone absorbs the lost probability mass and although the error (2.48) is slightly larger than (2.47) 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 (2.47) is replaced by the largest term so that in practice the difference is not critical. Interestingly then, using the "tau-method" [11, 17], 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 (2.48), but with the denominator replaced by $(\hat{C}_N^a, 1)^2$. Since the mass of $\hat{C}_N^a$ is generally smaller than that of $\hat{C}_0^a$ this method performs poorer 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 (2.43), (2.44) and (2.45),

$$\|\hat{\pi}_N^0 p - p\|_{L^1} \leq |(1, \hat{\pi}_N p - p)| + \|\hat{\pi}_N p - p\|_{L^1} \leq 2\|\hat{\pi}_N p - p\|_{L^1}. \qquad (2.49)$$

In order to conclude this section we shall finally pay attention to the choice of the parameter a which must be decided upon 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 is meant a distribution with one single distinct peak located closely 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",

$$P^{1/2}(x;\, a) = C^{-1}\sqrt{\frac{a^x}{x!}e^{-a}}. \qquad (2.50)$$

Here C is the normalizing constant given by

$$\begin{aligned} C &= \sum_{x \geq 0} \sqrt{\frac{a^x}{x!}e^{-a}} = \\ &= \sum_{x \geq 0} \frac{(a/2)^x}{x!}e^{-a/2} \times \underbrace{\pi^{1/4}\sqrt{\frac{\Gamma(x+1)}{\Gamma(x+1/2)}}\left(1 + \sqrt{\frac{a}{2x+1}}\right)}_{=:f(x)}, \end{aligned} \qquad (2.51)$$

where the last line follows from summing even and odd terms separately and using the duplication formula for the gamma function [1]. To evaluate (2.51), note that the sum is nothing but $E[f(X)]$ for X a Poissonian stochastic variable of expectation $a/2$. Expand f in a formal Taylor series around $a/2$ and assume a to be large so that standard asymptotics [1] for the gamma function apply. Using formulas for the central moments of the Poisson distribution then yields after some work,

$$C \sim 2^{3/4}\pi^{1/4}a^{1/4}\left(1 - \frac{1}{16a} + \mathcal{O}\left(a^{-2}\right)\right). \qquad (2.52)$$

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

$$m \sim a + \frac{1}{2} + \frac{1}{8a} + \mathcal{O}\left(a^{-2}\right), \tag{2.53}$$

$$\sigma^2 \sim 2a - \frac{1}{4a} + \mathcal{O}\left(a^{-2}\right). \tag{2.54}$$

The asymptotical method described above relates loosely to the method of Laplace [23] in the theory of asymptotics 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 (2.53) yields a relative error less than 0.06% even for $a = 2$. Taken together, (2.53) and (2.54) 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 $P^{1/2}(x; m - 1/2)$ in (2.50) closely. As N grows this estimate no longer holds true since the approximating space gets larger. By Corollary 2.11 we see that a should reasonably be small for the error to decrease rapidly. 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 [6] for this fairly complicated issue.

We thus summarize this section: We started by looking at polynomial approximation in $L^2_w(\mathbf{Z}_+)$ according to the Poissonian weight $w(x) = a^x/x! \cdot e^{-a}$. The corresponding system of orthogonal polynomials is Charlier's polynomials and the strongest approximation result for this system was obtained in Theorem 2.10. We then transferred our results in L^2_w to L^2 using the Charlier *functions* as a new basis instead. A mass-preserving version of the projection operator was constructed and analyzed and we concluded by examining the dependence of the projection's quality on the parameter a.

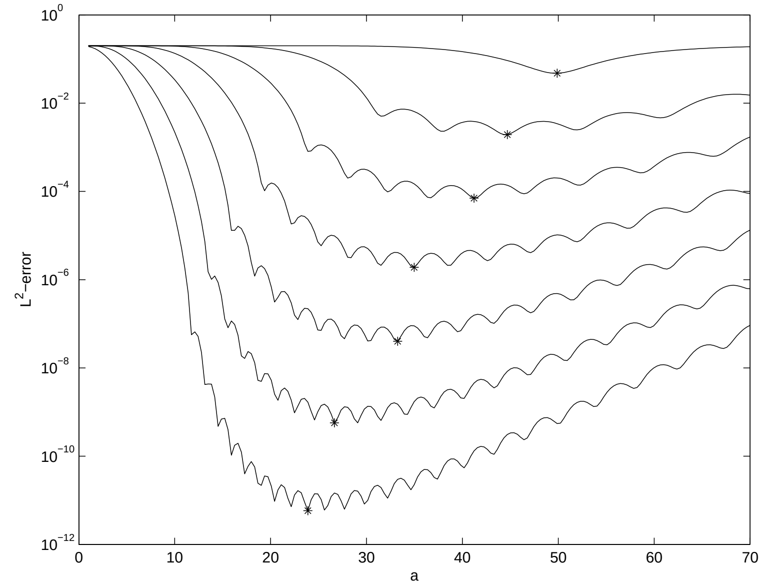

Figure 2.1: A Poisson distribution of expectation value 50 is projected on $\hat{X}_N$ for $N = 0, 5, \ldots, 30$ using several values of a. The L^2-error is determined for each choice of a producing the dependence shown. For each value of N, the asterisk indicates the optimal value of a.

2.3 Stability

The Galerkin approximation to (2.2) that we shall now analyze reads as follows (compare (2.42)): find $p_N \in \hat{X}_N$ such that

$$\left.\begin{array}{rcl} (q, \partial p_N/\partial t) + \lambda(\hat{\pi}_0 q, 1) & = & (q, \mathcal{M} p_N) \\ (1, \partial p_N/\partial t) & = & 0 \end{array}\right\} \quad \text{for } \forall q \in \hat{X}_N. \tag{2.55}$$

Since the master operator in general is unbounded, indefinite and unsymmetric with non-orthogonal eigenvectors, we cannot possibly hope to capture any convergence properties of (2.55) by adhering to standard energy estimates. The partial results we present in this section are instead based on observations due to van Kampen [20] and Theorem 2.1. We will attempt to make it reasonable to believe that (2.55) is stable in the L^1-norm so that convergence follows from the Lax-Richtmyer equivalence theorem.

We first write (2.55) in the equivalent form

$$\frac{\partial p_N}{\partial t} = \hat{\pi}_N^0 \mathcal{M} p_N, \tag{2.56}$$

where the representation

$$p_N(x,t) = \sum_{n=0}^{N} c_n(t)\hat{C}_n^a(x) \tag{2.57}$$

is implicitly understood.

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

$$U_N(t) \equiv \sum_{x \geq 0} p_N(x,t)^+ = \sum_{x \in \xi^+(t)} p_N(x,t), \tag{2.58}$$

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,

$$U_N(t) + V_N(t) = \text{constant}. \tag{2.59}$$

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

$$\begin{aligned} U_N'(t) = \sum_{x \in \xi^+(t)} \hat{\pi}_N^0 \mathcal{M} p_N &= \sum_{x \in \xi^+(t)} \mathcal{M} p_N + \sum_{x \in \xi^+(t)} \left[\hat{\pi}_N^0 \mathcal{M} p_N - \mathcal{M} p_N\right] \\ &=: A_N(t) + B_N(t), \end{aligned} \tag{2.60}$$

where from (2.2),

$$A_N(t) = \sum_{x \in \xi^+(t)} \sum_{r=1}^{R} \left[x \geq -n_r^-\right] w_r(x+n_r) p_N(x+n_r,t) - w_r(x) p_N(x,t), \tag{2.61}$$

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 (2.61) 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 since, by (2.59), $-V' = U'$. 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 case, following this discussion we see that

$$\frac{d}{dt}\|p_N\|_{L^1} = 2A_N(t) + 2B_N(t) \tag{2.62}$$

except for those points in time when 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.*

While perhaps not leading to a general proof of this claim the motivation is that $\hat{\pi}_N^0 \mathcal{M} p_N$ approaches $\mathcal{M} p_N$ when N grows so that B_N in principle 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 identically 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 forces 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 see that

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{++} & 0 \\ 0 & \mathcal{M}_{--} \end{bmatrix} \tag{2.63}$$

or otherwise A_N cannot be zero (the zero in the lower left corner comes from using $-V' = U'$ and writing down the analogues of (2.60) and (2.61) for V_N). Hence $\mathcal{M}$ is *decomposable* (see (2.11)) and is excluded from the present context since such operators in general does not possess a unique steady-state solution (cf. Theorem 2.2). 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,

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{++} & \mathcal{M}_{+0} & 0 \\ 0 & \mathcal{M}_{00} & 0 \\ 0 & \mathcal{M}_{-0} & \mathcal{M}_{--} \end{bmatrix}. \tag{2.64}$$

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 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 (2.60)),

$$\frac{d}{dt}\|p_N\|_{L^1} = \underbrace{\sum_{x\ge 0} \text{sgn } p_N \, \mathcal{M}p_N}_{2A_N} + \underbrace{\sum_{x\ge 0} \text{sgn } p_N \left[\hat{\pi}_N^0 \mathcal{M}p_N - \mathcal{M}p_N\right]}_{2B_N}, \tag{2.65}$$

where sgn q is zero for $q = 0$. It seems reasonable to believe that there are estimates of the form $2|A_N| \ge \kappa(\mathcal{M})\|\mathcal{M}p_N\|_{L^1}$ and $2|B_N| \le C_N\|\mathcal{M}p_N\|_{L^1}$ 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 $\le (C_N - \kappa(\mathcal{M}))\|\mathcal{M}p_N\|_{L^1}$ 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 proven but it does, however, shed some light over the expected stability properties of the Galerkin scheme (2.55). Also, the crucial argument in this discussion is the relation (2.59) which indicates why a mass-preserving projection is a favorable choice.

2.4 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 for the basis functions to capture the dynamics of the solution.

In order to describe the assembly of the full D-dimensional master equation we need to make use of multi-indices which we shall denote by small Greek letters. In D dimensions with $\alpha = [\alpha_1, \ldots, \alpha_D]$ and x a D-dimensional array we thus index x by

$$x_\alpha = x_{\alpha_1,\ldots,\alpha_D}. \tag{2.66}$$

In addition, the following products occur naturally,

$$\beta^\alpha = \beta_1^{\alpha_1} \cdots \beta_D^{\alpha_D}, \tag{2.67}$$

$$\alpha! = \alpha_1! \cdots \alpha_D!, \tag{2.68}$$

$$e^\alpha = e^{\alpha_1} \cdots e^{\alpha_D}. \tag{2.69}$$

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

$$\hat{C}_\gamma^a(x) \equiv \prod_j \hat{C}_{\gamma_j}^{a_j}(x_j). \tag{2.70}$$

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

$$(f, g) \equiv \sum_{x \geq 0} f(x)g(x) \prod_j \frac{a_j^{x_j}}{x_j!} e^{-a_j} = \sum_{x \geq 0} f(x)g(x) \frac{a^x}{x!} e^{-a}, \tag{2.71}$$

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

$$p(x, t) = \sum_{\gamma} c_{\gamma}(t) \hat{C}_{\gamma}^{a}(x). \tag{2.72}$$

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

$$\begin{aligned} c_{\delta}' &= \left(\hat{C}_{\delta}^{a}, \mathcal{M}p\right) = \left(\mathcal{M}^{*}\hat{C}_{\delta}^{a}, p\right) = \\ &= \sum_{r=1}^{R} \sum_{\gamma} \left(C_{\delta}^{a}(x - n_r) \cdot \sqrt{a^{-n_r} x!/(x - n_r)!} - C_{\delta}^{a}(x), w_r(x) c_{\gamma} C_{\gamma}^{a}(x)\right), \end{aligned} \tag{2.73}$$

where the favorable representation of the adjoint was used. Using 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 $x + y \longrightarrow z$ 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* [9]. In one dimension it is given by

$$\sum_{x \geq 0} f(x) \frac{a^x}{x!} e^{-a} = \sum_{j=1}^{n} f(x_j) w_j + R_n, \tag{2.74}$$

$$R_n = a^n n! \frac{f^{(2n)}(\xi)}{(2n)!}, \qquad \xi \in (0, \infty). \tag{2.75}$$

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

$$w_j \equiv -\frac{(an)^{-1/2}}{C_{n-1}^{a}(x_j) \cdot d/dx\, C_n^a(x_j)}. \tag{2.76}$$

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 be able 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 (2.72). Then very formally,

$$\frac{\partial p(x,t)}{\partial t} = \sum_{\gamma} c'_{\gamma} \hat{C}^{a}_{\gamma}(x) + \sum_{\gamma} \left(\frac{a'}{C^{a}_{\gamma}(x)} \frac{d}{da} C^{a}_{\gamma}(x) + \frac{x}{2}\frac{a'}{a} - \frac{a'}{2} \right) c_{\gamma} \hat{C}^{a}_{\gamma}(x), \tag{2.77}$$

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

$$\frac{d}{dx} L^{a}_{n}(x) = -L^{a+1}_{n-1}(x), \tag{2.78}$$

one readily gets in the scalar case,

$$\frac{d}{da} C^{a}_{n}(x) = -\frac{n}{2a} C^{a}_{n}(x) + \sqrt{\frac{n}{a}} C^{a}_{n-1}(x). \tag{2.79}$$

Inserting this and using the recurrence (2.29) one can simplify the derivative into

$$\begin{aligned} \frac{\partial p(x,t)}{\partial t} &= \sum_{\gamma} c'_{\gamma} \hat{C}^{a}_{\gamma}(x) + \qquad (2.80) \\ &+ \sum_{\gamma} c_{\gamma} \sum_{j} \hat{C}^{a \setminus a_j}_{\gamma \setminus \gamma_j}(x \setminus x_j) \left(-\frac{1}{2}\sqrt{\frac{\gamma_j + 1}{a_j}} \hat{C}^{a_j}_{\gamma_j+1}(x_j) + \frac{1}{2}\sqrt{\frac{\gamma_j}{a_j}} \hat{C}^{a_j}_{\gamma_j-1}(x_j) \right) a'_j. \end{aligned}$$

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

$$\hat{C}^{a \setminus a_j}_{\gamma \setminus \gamma_j}(x \setminus x_j) \equiv \prod_{i \neq j} \hat{C}^{a_i}_{\gamma_i}(x_i). \tag{2.81}$$

It follows that

$$\left(\hat{C}^{a}_{\delta}, \frac{\partial p(x,t)}{\partial t} \right) = c'_{\delta} + \sum_{j} \left(-\frac{1}{2}\sqrt{\frac{\delta_j}{a_j}} c_{\delta-1_j} + \frac{1}{2}\sqrt{\frac{\delta_j+1}{a_j}} c_{\delta+1_j} \right) a'_j \tag{2.82}$$

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 (2.82) and the right side of (2.73) to produce equations for the derivatives of the coefficients.

Suppose that a good choice of $a(0)$ has been made so that $p(\cdot, 0)$ is efficiently represented. Then the discussion towards the end of Section 2.2 suggest the intuitive idea to define a' by the derivative of the expectation value. This leads to the following simple algorithm:

1. Compute c'_δ by assembling (2.73).
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 (2.82).

In practise we also enforce $a \geq 1$ since all results in Section 2.2 are uniform under this restriction. The usefulness of this technique is demonstrated in Section 3.2

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 $\hat{\pi}_N^0$ 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 set of coefficients implies a stationary mass. In the case of a stationary parameter a this amounts to simply summing the ansatz (2.72) 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 (2.80).

In summary we have seen that it is a straightforward (but not trivial) task to write a 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 feature of a dynamic parameter a helps capturing a solution that varies 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 two different models from molecular biology. The first model is a four-dimensional example involving two metabolites and two enzymes and the task is to find the steady-state distribution. By contrast, the second 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 reaction-rate approach fails.

3.1 Enzyme-control of metabolites

This example is found in [7] and is a model of the synthesis of two metabolites x and y by two enzymes e_x and e_y. The reactions are

$$\left.\begin{array}{ccc} \emptyset & \xrightarrow{k_1 e_x/(1+x/k_i)} & x \\ x & \xrightarrow{\mu x} & \emptyset \\ \emptyset & \xrightarrow{k_2 e_y/(1+y/k_i)} & y \\ y & \xrightarrow{\mu y} & \emptyset \\ x+y & \xrightarrow{kxy} & \emptyset \end{array}\right\} \left.\begin{array}{ccc} \emptyset & \xrightarrow{k_3/(1+x/k_r)} & e_x \\ e_x & \xrightarrow{\mu e_x} & \emptyset \\ \emptyset & \xrightarrow{k_4/(1+y/k_r)} & e_y \\ e_y & \xrightarrow{\mu e_y} & \emptyset \end{array}\right\}, \tag{3.1}$$

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, (3.1) is the result of an *adiabatic* [12] 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 3.1. We 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 was computed and the result is displayed in Figure 3.2 where the exponential convergence is reasonably explicit. The obtained solution is visually pleasing and free of numerical artefacts away from the center of the probability distribution already at the quite coarse discretization $[15, 15, 8, 8]$ (degrees per dimension according to the ordering $[x, y, e_x, e_y]$).

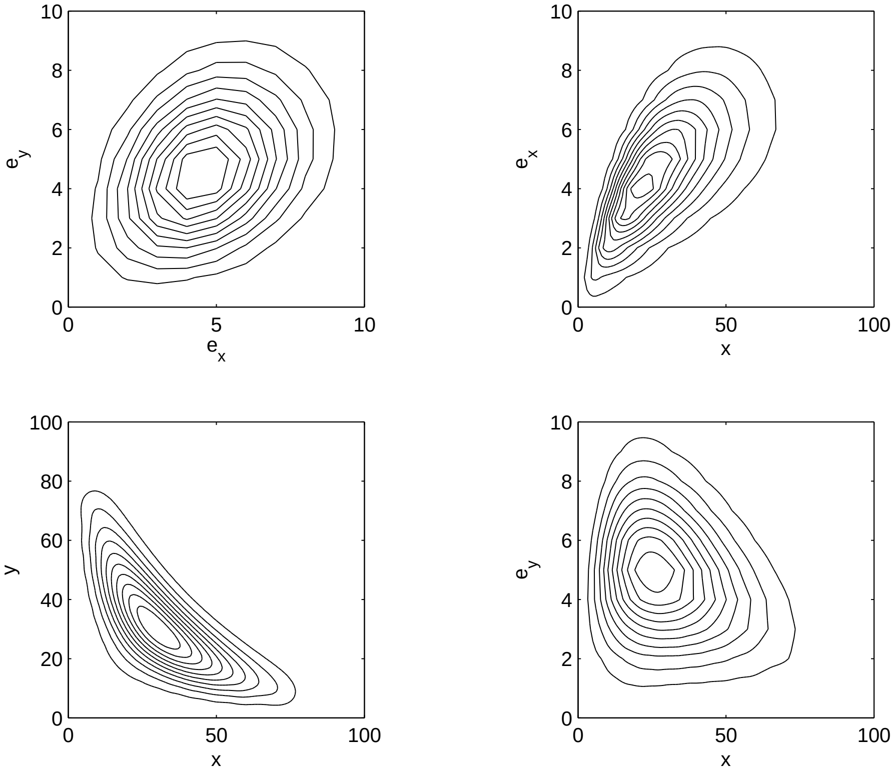

Figure 3.1: Steady-state solution (marginal distributions) to (3.1). The correlation between the various species can be understood from first principles, except perhaps for the somewhat irregular dependence between x and e_y.

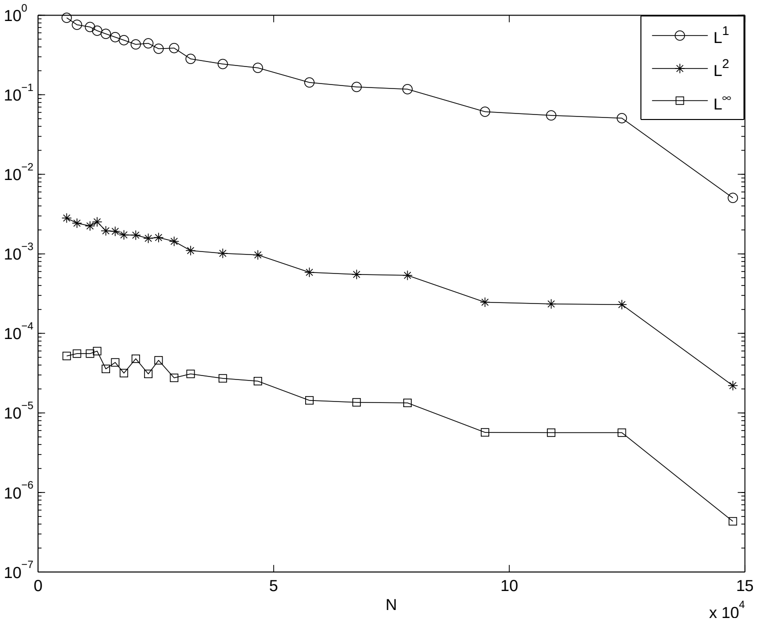

Figure 3.2: Errors of the scheme applied to (3.1) as measured in different norms. Note that the x-axis displays the *total* number of degrees of freedom.

3.2 Bistable toggle switch

A biological *toggle switch* can be formed by two mutually cooperatively repressing products x and y [13]. The relevant equations are

$$\left.\begin{array}{ccc} \emptyset & \xrightarrow{a/(b+y^2)} & x \\ x & \xrightarrow{\mu x} & \emptyset \\ \emptyset & \xrightarrow{c/(d+x^2)} & y \\ y & \xrightarrow{\mu y} & \emptyset \end{array}\right\}, \tag{3.2}$$

with parameters $a = c = 1000$, $b = d = 6000$ and $\mu = 10^{-3}$. The intuitive behavior of (3.2) is easy to grasp. 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. Under 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 3.3 where the result of a stochastic simulation with SSA [15] is displayed.

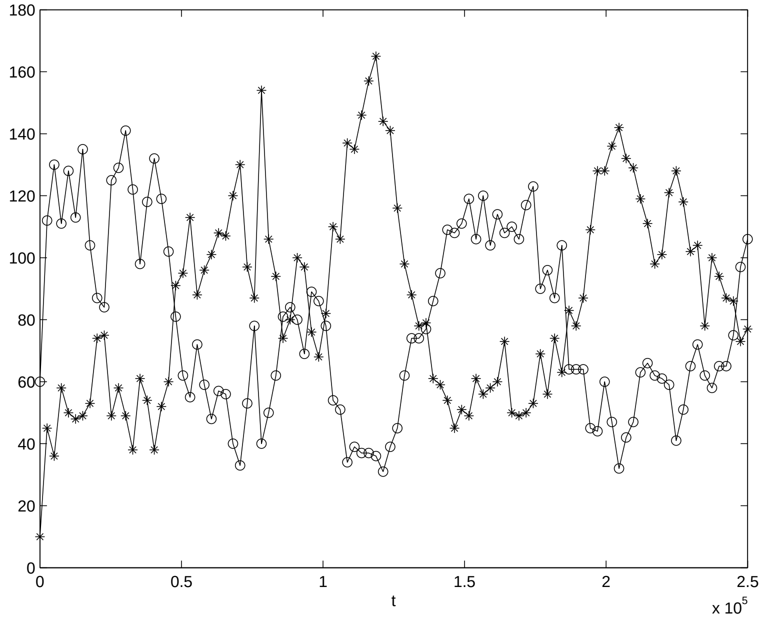

Figure 3.3: One sample trajectory of (3.2) obtained by Gillespie's algorithm. In this simulation the system 'switches' three times and we also see a 'near-switch' slightly before $t = 10^5$.

We solved (3.2) 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 2.4. The stiffness of the problem is clearly visible in Figure 3.4 and so an implicit ODE-solver was preferred (we used `MATLAB`'s `ode15s`).

The error was estimated using a high order reference solution with stricter tolerances for the time-stepper. In Figure 3.4 several different norms of the error are displayed and the exponential convergence of the method is clearly visible. Figure 3.4 indicates a visually pleasing result already at a quite coarse discretization. Overall, no problems of instabilities were ever encountered although phase errors were more pronounced for small N.

The interesting feature of forming two distinct peaks makes the toggle switch an example for which the reaction-rate approach must fail. The solution obtained using this method comes to rest near one of the two peaks with an additional unstable critical point situated in between them. Clearly, anyone of these solutions does not tell the whole story.

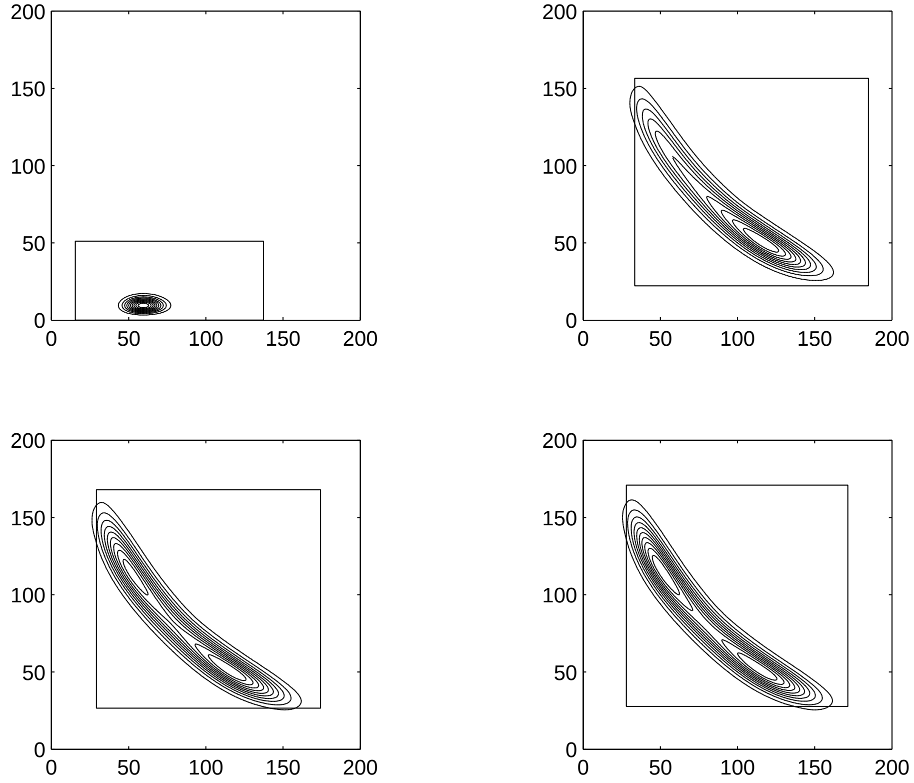

Figure 3.4: The solution to (3.2) at time $t = [0, 1, 1.5, 2.5] \cdot 10^5$ using $N = 19$ (400 coefficients). The simulation starts with a Poissonian solution centered at $(x, y) = (60, 10)$ and ends in equilibrium where two distinct peaks have formed. The indicated bounding box contains all quadrature 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.

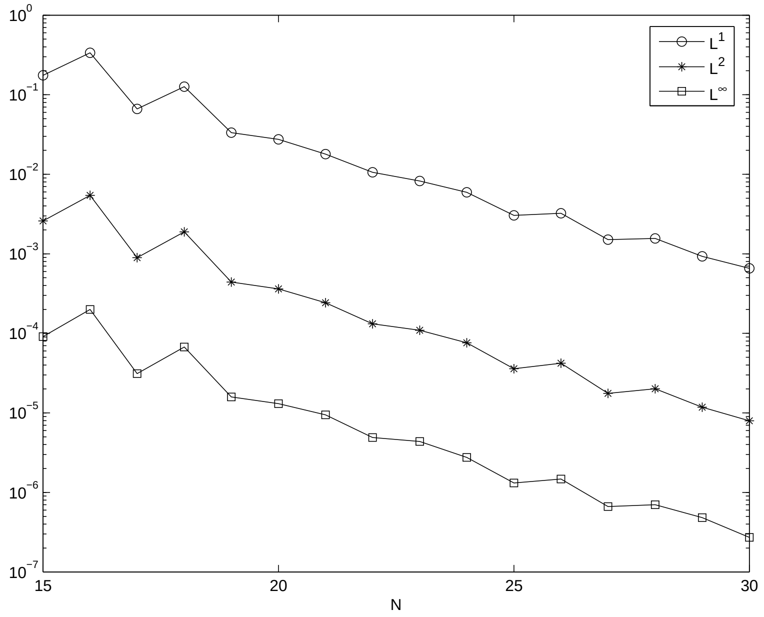

Figure 3.5: Errors of the scheme applied to (3.2) in different norms (time-average) for increasing order N. The exponential convergence of the method is clearly visible.

4 Conclusions

For highly general physical systems described by discrete coordinates, the master equation is a direct consequence of the Markov assumption on the nature of the underlying stochastic process. This description is particularly intuitive for chemical reactions where the system is described by counting the number of molecules of each kind. If the number of participating molecules is large an effective and usually very accurate description in terms of ODEs for the expectation values can be formed.

For a more complete understanding of processes involving few molecules, however, the full master equation represents a viable approach when the number of different species is sufficiently small. Such systems can be found inside living cells where the stochastic nature of the process can be very important.

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 properties 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 an effective, accurate and stable alternative to traditional solution methods when the dimensionality of the problem is sufficiently small. The solution obtained in terms of the full probability density function is useful in determining statistical parameters, improving the accuracy of certain inverse problems and leading to a fuller understanding of the exact nature of the processes involved.

There seem to be very few schemes that are similar to the proposed method. One exception is the *Poisson representation* [12] 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 nice looking distribution p (e.g. f is a Dirac-function for the simple model-problem (2.13)). This observation suggests that the Poissonian 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. Firstly, for certain special problems the solution may become a very wide probability distribution. The (half) Poissonian distribution on which the basis is built is really only an effective representation when the essential support of the solution is clustered around the expected value m as $m \pm \mathcal{O}(\sqrt{m})$ (i.e. the standard deviation should be of order $\sqrt{m}$). If this condition is violated a very high order N is needed in order to capture the solution and the scheme becomes less efficient. A cure is to *scale* the solution appropriately which for a discrete solution amounts to incorporating *aggregation*. If, say, every discrete coordinate is understood to represent 4 points of the solution, then in effect the induced basis functions become 4 times wider. In this case, an additional assumption of smoothness of the coefficients on the form $w_r(x \pm 2) \approx w_r(x)$ is needed in order to close the scheme efficiently. Aggregation of the solution has been described in the setting of the sparse grids technique for the master equation in [19].

Another type of improvement is to couple 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 of the dimensions. If this is the case a drastic efficency gain is possible, making really high dimensional problems tractable. Some steps in this direction have been presented for the Fokker-Planck equation in [22].

Although there is room for some other minor improvements as well, the author believes that the method is already mature enough to serve as an attractive solution method for many interesting applications.

Acknowledgment

Many comments and suggestions by Per Lötstedt have improved the contents of this paper. Various inputs from Jan Hesthaven were also helpful during the final stages of writing. Finally, the author is indebted to Hermann Matthies for a long and valuable discussion during the IHP-EU Workshop "Breaking Complexity" in Bad Honnef, Germany.

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.

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

[7] 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.

[8] S. Engblom. Computing the moments of high dimensional solutions of the master equation. Technical Report 2005-020, Dept of Information Technology, Uppsala University, Uppsala, Sweden, 2005. Available at `http://www.it.uu.se/research`. To appear in *Appl. Math. Comput.*

[9] 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`.

[10] L. Ferm, P. Lötstedt, and P. Sjöberg. Conservative solution of the Fokker-Planck equation for stochastic chemical reactions. Technical Report 2004-054, Dept of Information Technology, Uppsala University, Uppsala, Sweden, 2004. Available at `http://www.it.uu.se/research`. To appear in *BIT*.

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

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

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

[14] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, 104:1876–1889, 2000.

[15] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.

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

[17] 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.

[18] P. Guptasarama. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays*, 17:987–997, 1995.

[19] M. Hegland, C. Burden, L. Santoso, S. MacNamara, and H. Booth. A solver for the stochastic master equation applied to gene regulatory networks. Submitted, e-mail: `markus.hegland@anu.edu.au`.

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

[21] 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`.

[22] P. Lötstedt and L. Ferm. Dimensional reduction of the Fokker-Planck equation for stochastic chemical reactions. Technical Report 2005-023, Dept of Information Technology, Uppsala University, Uppsala, Sweden, 2005. Available at `http://www.it.uu.se/research`. To appear in *Multiscale Meth. Simul.*

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

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

[25] 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.

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