Scientific
Research

On the Stability of Stochastic Jump Kinetics

Stefan Engblom

Division of Scientific Computing, Department of Information Technology, Uppsala University, Uppsala, Sweden
Email: stefane@it.uu.se

Abstract

Motivated by the lack of a suitable constructive framework for analyzing popular stochastic models of Systems Biology, we devise conditions for existence and uniqueness of solutions to certain jump stochastic differential equations (SDEs). Working from simple examples we find *reasonable* and *explicit* assumptions on the driving coefficients for the SDE representation to make sense. By "reasonable" we mean that stronger assumptions generally do not hold for systems of practical interest. In particular, we argue against the traditional use of global Lipschitz conditions and certain common growth restrictions. By "explicit", finally, we like to highlight the fact that the various constants occurring among our assumptions *all can be determined once the model is fixed*. We show how basic long time estimates and some limit results for perturbations can be derived in this setting such that these can be contrasted with the corresponding estimates from deterministic dynamics. The main complication is that the natural path-wise representation is generated by a counting measure with an intensity that depends nonlinearly on the state.

1. Introduction

The observation that detailed modeling of biochemical processes inside living cells is a close to hopeless task is a strong argument in favor of stochastic models. Such models are often thought to be more accurate than conventional rate-diffusion laws, yet remain more manageable than, say, descriptions formed at the level of individual molecules. Indeed, several studies [1]-[3] have showed that noisy models have the ability to capture relevant phenomena and to explain actual, observed dynamics.

In this work we shall consider some "flow" properties of a stochastic dynamical system in the form of a quite general continuous-time Markov chain. Since the pioneering work of Gillespie [4] [5], in the Systems Biology context this type of model is traditionally described in terms of a (chemical) *master equation* (CME). This is the

forward Kolmogorov equation of a certain jump stochastic differential equation (jump SDE for brevity), driven by independent point processes with state-dependent intensities. Despite the popularity of the master equation approach, little analysis on a per trajectory-basis of actual models has been attempted.

In the general literature, when discussing existence/uniqueness and various types of perturbation results, different choices of assumptions with different trade-offs have been made. One finds that the treatment often falls into one of two categories taking either a "mathematical" or a "physical" viewpoint. Both of the conditions are highly general but with subsequently less transparent proofs and resulting in more abstract bounds. Or the conditions are formed out of convenience, say, involving global Lipschitz constants, and classical arguments carry through with only minor modifications.

Protter ([6] Chap. V) offers a nice discussion from the mathematical point of view and in ascending order of generality, including the arguably highly unrestrictive assumption of locally Lipschitz continuous coefficients. Other authors ([7], Chap. 6, [8], Chap. 3-5) also treat the evolution of general jump-diffusion SDEs in *continuous state* spaces.

A study of the flow properties of jump SDEs is found in [9], where the setting is scalar and the state is continuous. In [10] jump stochastic *partial* differential equations are treated, and existence/uniqueness results as well as ergodic results for the case of a multiplicative noise, are found in [11] [12]. Numerical aspects in a similar setting are discussed in [13].

In a more applied context, stability is often thought of as implied from physical premises and the solution is tactically assumed to be confined inside some bounded region ([14], Chap. V). The fundamental issue here is that for *open* systems in a stochastic setting, there is a non-zero probability of reaching *any* finite state and global assumptions must be formed with great care. The analysis of open networks under an *a priori* assumption of boundedness is therefore quite difficult to interpret other than in a qualitative sense. Notable examples in this setting include time discretization strategies [15] [16], time-parallel simulation techniques [17], and parameter perturbations [18].

Evidently, essentially no systems of interest satisfy global Lipschitz assumptions since the fundamental *interaction* almost always takes the form of a quadratic term. Interestingly, for ordinary differential equations, it has been shown [19] that Lipschitz continuous coefficients imply a computationally polynomial-space complete solution; thus providing a kind of explanation for the convenience with this *weak feedback* assumption. It is also known [20], that with SDEs, superlinearly growing coefficients may in fact cause the forward Euler method to diverge.

1.1. Agenda

Besides its expository material, the purpose of this paper is to devise simple conditions that imply stability for finite and, in certain cases, infinite times, and that, when applied to *systems of practical interest*, yield explicit expressions for the associated stability estimates. As a result the framework developed herein applies in a constructive way to any chemical network, of arbitrary size and topology, formed by any combination of the elementary reactions (2.3) to be presented in Section 2. Additionally, it will be clear how to encompass also other types of nonlinear reactions that typically result from adiabatic simplifications.

As an argument in favor of this bottom-up approach one can note that, for evolutionary reasons, biochemical systems tend to operate close to critical points in phase-space where the efficiency is the highest. Clearly, for such dynamical systems, an analysis *by analogy* might be highly misleading.

We also like to argue that our results are of interest from the modeling point of view. Due to the type of phenomenological arguments often involved, judging the relative effect of the (non-probabilistic) *epistemic uncertainty* is a fundamental issue which has so far not rendered a consistent analysis.

1.2. Outline

The expository material in Section 2 is devoted to formulating the type of processes we are interested in. We state the master equation as well as the corresponding jump SDE and we also look at some simple, yet informative actual examples. Since it is expected that the properties of the stochastic dynamics are somehow similar to those of the deterministic version, we search for a set of minimal assumptions in the latter setting in Section 3. Techniques for finding explicit values of the constants occurring among our assumptions are also devised. The main results of the paper are found in Section 4 where we put our theory together and prove existence and uni-

queness, as well as long time estimates and limit results for perturbations. A concluding discussion is found in Section 5.

2. Stochastic Jump Kinetics

In this section we start with the physicist's traditional viewpoint of pure jump processes and write down the governing *master equations*. These are evolution equations for the probability densities of continuous-time Markov chains over a discrete state space. Although the application considered here is mesoscopic chemical kinetics, identical or very similar stochastic models are also used in Epidemiology [21], Genetics [22] and Sociodynamics [23], to name just a few.

We then proceed with discussing a path-wise representation in terms of a stochastic jump differential equation. The reason the sample path representation is interesting is the possibility to reason about *flow* properties and thus compare functionals of single trajectories. This is generally not possible with the master equation approach.

For later use we conclude the section by looking at some prototypical models. A simple analysis shows, somewhat surprisingly, that an innocent-looking example produces second moments that grow indefinitely.

2.1. Reaction Networks and the Master Equation

We consider a chemical network consisting of D different *chemical species* interacting according to R prescribed *reaction pathways*. At any given time t, the state of the system is an integer vector $X(t) \in \mathbf{Z}_+^D = \{0,1,2,\cdots\}^D$ counting the number of individual molecules of each species. A reaction law is a prescribed change of state with an intensity defined by a *reaction propensity*, $w_r : \mathbf{Z}_+^D \to \mathbf{R}_+$. This is the transition probability per unit of time for moving from the state x to $x - \mathbb{N}_r$:
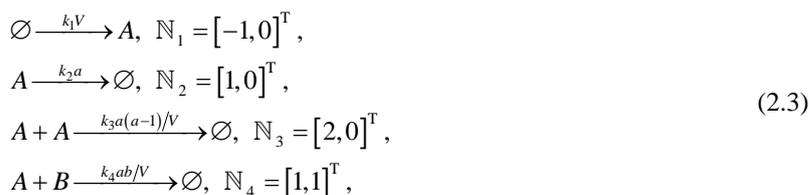
$$\mathbf{P}\left[X(t+dt) = x - \mathbb{N}_r \mid X(t) = x \right] = w_r(x)dt + o(dt). \tag{2.1}$$

where $\mathbb{N}_r \in \mathbf{Z}^D$ is the transition step and is the rth column in the *stoichiometric matrix* $\mathbb{N} \in \mathbf{Z}^{D \times R}$. Informally, for states $x(t) \in \mathbf{R}_+^D$, we can picture (2.1) as a stochastic version of the time-homogeneous ordinary differential equation

$$x'(t) = -\sum_{r=1}^{R} \mathbb{N}_r w_r(x) = -\mathbb{N}w(x) =: F(x), \tag{2.2}$$

where $w(x) \equiv \left[w_1(x), \cdots, w_R(x) \right]^{\mathrm{T}}$ is the column vector of reaction propensities.

The physical premises leading to a description in the form of discrete transition laws (2.1) often imply the existence of a *system size* V (e.g. physical volume or total number of individuals). For instance, in a given volume V the *elementary* chemical reactions can be written using the state vector $x = [a,b]^{\mathrm{T}}$,

$$
\begin{aligned}
\varnothing &\xrightarrow{\;k_1 V\;} A, \quad \mathbb{N}_1 = [-1,0]^{\mathrm{T}}, \\
A &\xrightarrow{\;k_2 a\;} \varnothing, \quad \mathbb{N}_2 = [1,0]^{\mathrm{T}}, \\
A + A &\xrightarrow{\;k_3 a(a-1)/V\;} \varnothing, \quad \mathbb{N}_3 = [2,0]^{\mathrm{T}}, \\
A + B &\xrightarrow{\;k_4 ab/V\;} \varnothing, \quad \mathbb{N}_4 = [1,1]^{\mathrm{T}},
\end{aligned}
\tag{2.3}
$$

with the names of the species in capitals. These propensities are generally scaled such that $w_r(x) = V u_r(x/V)$ for some dimensionless function u_r. Intensities of this form are called *density dependent* and arise naturally in a number of situations ([24], Chap. 11). For the rest of this paper, we conveniently take $V = 1$ and defer system's size analysis to another occasion.

The models we consider here all have states in the positive integer lattice and the assumption that no transition can yield a state outside $\mathbf{Z}_+^D$ is therefore natural. We make this formal as follows ([25], Chap. 8.2.2, Definition 2.4):

Assumption 2.1 (*Conservation and stability*). For all propensities, $w_r(x) = 0$ for any $x \in \mathbf{Z}_+^D$ such that $x - \mathbb{N}_r \notin \mathbf{Z}_+^D$, and we also restrict initial data to $\mathbf{Z}_+^D$. Furthermore, $w_r : \mathbf{Z}_+^D \to \mathbf{R}_+$ such that $w_r(x)$ is finite for all finite arguments x.

To state the chemical master equation (CME), let for brevity $p(x,t) = \mathbf{P}\left(X(t) = x \mid X(0) = x_0 \right)$ be the prob-

ability that a certain number x of molecules is present at time t conditioned upon an initial state x_0. The CME is then given by ([14], Chap. V)

$$\frac{\partial p(x,t)}{\partial t} = \sum_{r=1}^{R} w_r (x + \mathbb{N}_r) p(x + \mathbb{N}_r, t) - w_r (x) p(x,t) =: \mathbb{M}^{\mathrm{T}} p(x,t). \tag{2.4}$$

The convention of the transpose of the operator to the right of (2.4) is the standard mathematical formulation of Kolmogorov's forward differential system ([25], Chap. 8.3) in terms of which $\mathbb{M}$ is the infinitesimal generator of the associated Markov process. This is also the adjoint of the master operator $\mathbb{M}^{\mathrm{T}}$ in the sense that $\left(\mathbb{M}^{\mathrm{T}} p, q \right) = \left(p, \mathbb{M} q \right)$ in the Euclidean inner product over the state space. An explicit representation is

$$\mathbb{M}q(x) = \sum_{r=1}^{R} w_r (x) \left[q(x - \mathbb{N}_r) - q(x) \right], \tag{2.5}$$

such that the propensities in (2.1) can be retrieved,

$$\mathbb{M}(x, x - \mathbb{N}_r) = w_r (x). \tag{2.6}$$

Under assumptions to be prescribed in Section 4.1 it holds that the dynamics of the expected value of some time-independent unknown function f, conditioned upon the initial state x_0, can be written

$$\frac{\mathrm{d}}{\mathrm{d}t} E^{x_0} \left[f(X_t) \right] = \sum_{x \in Z_+^D} \frac{\partial p(x,t)}{\partial t} f(x) = \left(\mathbb{M}^{\mathrm{T}} p, f \right) = \left(p, \mathbb{M}f \right)$$
$$= \sum_{r=1}^{R} E^{x_0} \left[w_r (X_t) \left(f(X_t - \mathbb{N}_r) - f(X_t) \right) \right]. \tag{2.7}$$

We now consider a path-wise representation for the stochastic process X_t.

2.2. The Sample Path Representation

In the present context of analyzing models in stochastic chemical kinetics, the path-wise jump SDE representation seems to have been first put to use in [26], and it was later further detailed in [16]. It should be noted, however, that an equivalent representation was used much earlier by Kurtz (see the monograph [24]).

We thus assume the existence of a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with the filtration $\mathcal{F}_{t \geq 0}$ containing R-dimensional Poisson processes. The state of the system $X(t) \in \mathbf{Z}_+^D$ will be constructed from a stochastic integral with respect to suitably chosen Poisson random measures.

The transition probability (2.1) defines a *counting process* $\pi_r (t)$ counting at time t the number of reactions of type r that has occurred since $t = 0$. It follows that these processes fully determine the state $X(t)$,

$$X_t = X_0 - \sum_{r=1}^{R} \mathbb{N}_r \pi_r (t). \tag{2.8}$$

The counting processes are obtained from the transition intensities (cf. (2.1))

$$\mathbf{P} \left[\pi_r (t + \mathrm{d}t) - \pi_r (t) = 1 \big| \mathcal{F}_t \right] = w_r (X_{t-}) \mathrm{d}t + o(\mathrm{d}t), \tag{2.9}$$

where by $X(t-)$ we mean the value of the process prior to any transitions occurring at time t, and where the little-o notation is understood uniformly with respect to the state variable. Alternatively, using Kurtz's *random time change representation* ([24], Chap. 6.2), we can produce the counting process from a standard unit-rate Poisson process Π_r,

$$\pi_r (t) = \Pi_r \left(\int_0^t w_r (X_{s-}) \mathrm{d}s \right). \tag{2.10}$$

The *marked* counting measure ([27], Chap. VIII) $\mu_r (\mathrm{d}t \times \mathrm{d}z; \omega)$ with $\omega \in \Omega$ defines an increasing sequence of arrival times $\tau_i \in \mathbf{R}_+$ with corresponding "marks" $z_i \in I := [0,1]$ according to some probability distribution which we will take to be uniform. The intensity $m_r (\mathrm{d}t \times \mathrm{d}z)$ of $\mu_r (\mathrm{d}t \times \mathrm{d}z)$ is the Lebesgue measure scaled by the corresponding propensity, $m_r (\mathrm{d}t \times \mathrm{d}z) = w_r (X_{t-}) \mathrm{d}t \times \mathrm{d}z$. Using this formalism, (2.8) and (2.10) can be written in the jump SDE form

$$\mathrm{d}X_t = -\int_I \mathbb{N}\boldsymbol{\mu}(\mathrm{d}t \times \mathrm{d}z), \tag{2.11}$$

where $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_R]^{\mathrm{T}}$. Here, the time $\tau - t$ to the arrival of the next reaction of type r is exponentially distributed with intensity $w_r(X_{t-})$. Note that, by virtue of the nature of the propensities, the intensities of the counting processes therefore depend nonlinearly on the state ([27], Chap. II.3).

Using that the point processes are independent and therefore have no common jump times ([25], Chap. 8.1.3), we can obtain a sometimes more transparent notation in terms of a *scalar* counting measure. Define for this purpose and for any state x the cumulative intensities

$$W_r(x) = \sum_{s=1}^{r} w_s(x), \tag{2.12}$$

such that the total intensity is given by $W(x) \equiv W_R(x)$. Let the marks z_i be uniformly distributed on I. Then the frequency of each reaction can be controlled through a set of indicator functions $\hat{w}_r : \mathbf{Z}_+^D \times I \to \{0,1\}$ defined according to

$$\hat{w}_r(x; z) = \begin{cases} 1 & \text{if } W_{r-1}(x) < zW(x) \leq W_r(x), \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

Put $\hat{w}(x) \equiv [\hat{w}_1(x; z), \cdots, \hat{w}_R(x; z)]^{\mathrm{T}}$ and define also for later use the indicator form

$$\hat{F}(x; z) = -\mathbb{N}\hat{w}(x; z), \tag{2.14}$$

such that

$$F(x) = \int_I \hat{F}(x; z)W(x)\mathrm{d}z, \tag{2.15}$$

where $F(x)$ is defined in (2.2).

The jump SDE (2.11) can now be written in terms of a scalar counting random measure μ through a state-dependent *thinning procedure* ([28], Chap. 7.5),

$$\mathrm{d}X_t = -\int_I \mathbb{N}\hat{w}(X_{t-}; z)\mu(\mathrm{d}t \times \mathrm{d}z). \tag{2.16}$$

Equation (2.16) expresses exponentially distributed reaction times that arrive according to a point process of intensity $m(\mathrm{d}t \times \mathrm{d}z) = W(X_{t-})\mathrm{d}t \times \mathrm{d}z$ carrying a mark which is uniformly distributed in I. This mark implies the ignition of one of the reaction channels according to the acceptance-rejection rule (2.13).

One frequently decomposes (2.16) into its "drift" and "jump" parts,

$$\mathrm{d}X_t = -\mathbb{N}w(X_t)\mathrm{d}t - \int_I \mathbb{N}\hat{w}(X_{t-}; z)(\mu - m)(\mathrm{d}t \times \mathrm{d}z). \tag{2.17}$$

The second term in (2.17) is driven by the compensated measure $(\mu - m)$ and is a local martingale provided in essence that the path is absolutely integrable (see [27], Chap. VIII.1, Corollary C4 for details).

2.2.1. Localization: Itô's and Dynkin's Formulas

In analytic work it is often necessary to "tame" the process by deriving results under a stopping time $\tau_P := \inf_{t \geq 0}\{\|X_t\| > P\}$ in some norm. Results for the stopped process $X_{t \wedge \tau_P}$ can then be transferred to the original process by letting $P \to \infty$ under suitable conditions.

Although there are many general versions of Itô's change of variables formula available in the setting of semi-martingales (see for example [8], Chap. 2.7 and [6], Chap. II.7), we shall get around with the following simple version ([7], Chap. 4.4.2). By the properties of the semi-martingale pure jump process we have for $\hat{t} = t \wedge \tau_P$

$$f(X_{\hat{t}}) - f(X_0) = \sum_{0 < s \leq \hat{t}} f(X_s) - f(X_{s-}) = \int_0^{\hat{t}} \int_I f(X_s) - f(X_{s-})\mu(\mathrm{d}s \times \mathrm{d}z), \tag{2.18}$$

where the sum is over jump times $s \in (0, \hat{t}]$. Using that $X_s = X_{s-} - \mathbb{N}\hat{w}(X_{s-}; z)$ we can write this in differential form as

$$\mathrm{d}f(X_t) = \int_I f(X_{t-} - \mathbb{N}\hat{w}(X_{t-}; z)) - f(X_{t-})\mu(\mathrm{d}t \times \mathrm{d}z). \tag{2.19}$$

Alternatively, decomposing (2.18) into drift- and jump parts and taking expectation values we get, since the compensated measure is a local martingale,

$$
\begin{aligned}
Ef(X_{\hat{t}}) - Ef(X_0) &= E\int_0^{\hat{t}}\int_I f\left(X_{s-} - \mathbb{N}\hat{w}(X_{s-};z)\right) - f(X_{s-})m(ds \times dz) \\
&= E\int_0^{\hat{t}}\int_I \left[f\left(X_{s-} - \mathbb{N}\hat{w}(X_{s-};z)\right) - f(X_{s-})\right]W(X_{s-})ds \times dz \\
&= E\int_0^{\hat{t}}\sum_{r=1}^R \left[\left(f(X_s - \mathbb{N}_r) - f(X_s)\right)w_r(X_s)\right]ds.
\end{aligned} \tag{18}
$$

This is Dynkin's formula ([25], Chap. 9.2.2) for the stopped process and we note that (2.7) is just a differential version.

2.2.2. Coupled Processes

When considering stability properties we will need to compare different trajectories with respect to the same noise. The details of this coupling are not defined in either (2.11) or (2.16) and must in fact be chosen explicitly. Since this equality is easy to inspect for a unit-rate Poisson process, the viewpoint of local time expressed in (2.10) provides an answer; two processes X_t and Y_t may be regarded as coupled if and only if they are evolved using identical Poisson processes Π_r, $r = 1, \cdots, R$ in (2.8) and (2.10). This approach was first used by Kurtz [29] in the context of the random time change representation. Algorithmically it implies the *Common Reaction Path* (CRP) method for simulating coupled processes [30] (see also [17]).

A refinement of this construction was devised, also by Kurtz, in ([31], see Equations (2.2), (2.3)). In turn, this approach implies the *Coupled Finite Difference* method [18] (but see also [16] [26]), and is more amenable to analysis. This is also the construction formalized below under our current framework.

To obtain such a coupled version of (2.16) we will have to make the thinning dependent on both trajectories. This is achieved by firstly replacing the cumulative intensities in (2.12) with the *base* (or *minimal*) intensities

$$
W_r^{(0)}(x,y) = \sum_{s=1}^r w_s(x) \wedge w_s(y), \tag{2.21}
$$

and use the new total base intensity $W^{(0)}(x,y) \equiv W_R^{(0)}(x,y)$ as the intensity of the counting measure μ_0; $m_0(dt \times dz) = W^{(0)}(X_{t-}, Y_{t-})dt \times dz$. We also modify (2.13) accordingly,

$$
\hat{w}_r^0(x,y;z) = \begin{cases} 1 & \text{if } W_{r-1}^{(0)}(x,y) < zW^{(0)}(x,y) \le W_r^{(0)}(x,y), \\ 0 & \text{otherwise.} \end{cases} \tag{2.22}
$$

Secondly, we also define the *remainder* intensity,

$$
W_r^{(\delta)}(x,y) = \sum_{s=1}^r w_s(x) \vee w_s(y) - W_r^{(0)}(x,y) = \sum_{s=1}^r |w_s(x) - w_s(y)|. \tag{2.23}
$$

In analogy with the previous construction we have the associated total intensity $W^{(\delta)}(x,y) \equiv W_R^{(\delta)}(x,y)$ and counting measure μ_δ; $m_\delta(dt \times dz) = W^{(\delta)}(X_{t-}, Y_{t-})dt \times dz$. This time the thinning procedure is non-symmetric in its two first arguments,

$$
\hat{w}_r^\delta(x,y;z) = \begin{cases} 1 & \text{if } W_{r-1}^{(\delta)}(x,y) < zW^{(\delta)}(x,y) \le W_{r-1}^{(\delta)}(x,y) + d(x,y), \\ 0 & \text{otherwise,} \end{cases} \tag{2.24}
$$

with the non-symmetricity due to

$$
d(x,y) = w_r(x) - w_r(x) \wedge w_r(y). \tag{2.25}
$$

As a concrete example of how this comparative thinning might be used, consider the following variant of (2.19),

$$
df(X_t - Y_t) = \int_I f\left(X_{t-} - \mathbb{N}\hat{w}^\delta(X_{t-}, Y_{t-};z) - Y_{t-} + \mathbb{N}\hat{w}^\delta(Y_{t-}, X_{t-};z)\right) - f(X_{t-} - Y_{t-})\mu_\delta(dt \times dz). \tag{2.26}
$$

For this specific example, the terms governed by the base counting measure μ_0 cancel out altogether.

We mention also that an equivalent construction, but one that leads to different algorithms, can be obtained

via a thinning of a single measure [16] [26]. Defining instead

$$W_r^{(+)}(x,y) = \sum_{s=1}^{r} w_s(x) \vee w_s(y),$$ (2.27)

implying the total intensity $W^{(+)}(x,y) \equiv W_R^{(+)}(x,y)$ and associated counting measure μ_+; $m_+(dt \times dz) = W^{(+)}(X_{t-}, Y_{t-})dt \times dz$. By construction the indicator functions are now non-symmetric in their first two arguments,

$$\hat{w}_r^+(x,y;z) = \begin{cases} 1 & \text{if } W_{r-1}^{(+)}(x,y) < zW^{(+)}(x,y) \le W_{r-1}^{(+)}(x,y) + w_r(x), \\ 0 & \text{otherwise.} \end{cases}$$ (2.28)

In analogy to (2.26) we get

$$df(X_t - Y_t) = \int_I f\left(X_{t-} - \mathbb{N}\hat{w}^+(X_{t-}, Y_{t-};z) - Y_{t-} + \mathbb{N}\hat{w}^+(Y_{t-}, X_{t-};z)\right) - f(X_{t-} - Y_{t-})\mu_+(dt \times dz).$$ (2.29)

This time, however, the intensity of the counting measure is generally larger and the equivalence is obtained as a result of the thinning procedure.

2.2.3. The Validity of the Master Equation
With this much formalism developed, we may conveniently quote the following result:

Theorem 2.1 ([25], *Chap.* 8.3.2, *Theorem* 3.3). *Under Assumption* 2.1, *and if additionally, for* $t \in [0,T]$ *it holds that*

$$EW(X_t) < \infty,$$ (2.30)

then (2.7) is valid for $t \in [0,T]$.

Since the governing Equation (2.7) for the expected value of $f(X_t)$ is a direct consequence of (2.4), we can similarly conclude the following:

Corollary 2.2 *Under the assumptions of Theorem* 2.1, *and if, moreover, in an arbitrary norm* $\|\cdot\|$,

$$E\|f(X_t)\| < \infty, \quad t \in [0,T],$$ (2.31)

then (2.7) is valid for $t \in [0,T]$.

In stating these results we have suppressed the conditional dependency on the initial state which we for simplicity consider to be some non-random state x_0.

2.3. Concrete Examples

Consider the bi-molecular birth-death system,

$$\left.\begin{array}{c} \varnothing \xrightarrow{k_1} A \\ \varnothing \xrightarrow{k_1} B \\ A+B \xrightarrow{k_2 ab} \varnothing \end{array}\right\},$$ (2.32)

that is, the system is in contact with a large reservoir such that A- and B-molecules are emitted at a constant rate k_1. Additionally, a *decay* reaction happens with probability k_2 per unit of time whenever two molecules meet. For this example we have the stoichiometric matrix

$$\mathbb{N} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$$

and the vector propensity function

$$w(x) = [k_1, k_1, k_2 ab]^T,$$

where $x = [x_1, x_2]^T = [a,b]^T$.

For a state $X_t = [A_t, B_t]^T$, define $U_t = A_t - B_t \in \mathbf{Z}$. Itô's formula (2.19) with $f(x) = x_1 - x_2$ yields

$$dU_t = df\left(X_t\right) = \int_I -[-1,1,0]\,\hat{w}\left(X_{t-};z\right)\mu\left(dt \times dz\right), \tag{2.33}$$

which upon a moments consideration is just the same thing as the model

$$\varnothing \underset{k_1}{\overset{k_1}{\rightleftharpoons}} U, \tag{2.34}$$

that is, a constant intensity discrete random walk process. An explicit solution is the difference between two independent Poisson distributions,

$$U_t = U_0 + \Pi_1\left(k_1 t\right) - \Pi_2\left(k_1 t\right) \sim \mathcal{N}\left(U_0, 2k_1 t\right), \quad \text{as } t \rightarrow \infty, \tag{2.35}$$

where $\mathcal{N}$ is a normally distributed random variable of the indicated mean and variance. Hence U_t fluctuates between arbitrarily large and small values as $t \rightarrow \infty$.

Reversible Versions

From time to time below we shall be concerned with the following *closed* version of (2.32), consisting of a single reversible reaction,

$$A + B \underset{k_2 c}{\overset{k_1 ab}{\rightleftharpoons}} C \tag{2.36a}$$

This is clearly a finite system since the number $a + b + 2c$ is always preserved. An *open* version of the same system is

$$A + B \underset{k_2 c}{\overset{k_1 ab}{\rightleftharpoons}} C \underset{k_4}{\overset{k_3 c}{\rightleftharpoons}} \varnothing, \tag{2.36b}$$

and will prove to be a useful example in the stochastic setting since formally, all states in $\mathbf{Z}_+^3$ are reachable. For (2.36a) we have

$$\mathbb{N} \equiv \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad w\left(x\right) \equiv \left[k_1 ab, k_2 c\right]^{\mathrm{T}}, \tag{2.37}$$

while (2.36b) is represented by

$$\mathbb{N} \equiv \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ -1 & 1 & 1 & -1 \end{bmatrix}, \quad w\left(x\right) \equiv \left[k_1 ab, k_2 c, k_3 c, k_4\right]^{\mathrm{T}}. \tag{2.38}$$

These examples, while very simple to deal with, will provide good counterexamples in both Sections 3 and 4.

3. Deterministic Stability

In this section we shall be concerned with the deterministic drift part of the dynamics (2.17). We are interested in techniques for judging the stability of the time-homogeneous ODE (2.2), the so-called reaction rate equations implied by the rates (2.1). Stability and continuity with respect to initial data are considered in Sections 3.1 and 3.2. The main motivation for this discussion stems from the observation that assumptions that *do not* hold in this very basic setting are unlikely to hold in the stochastic case. In Section 3.3, techniques for explicitly obtaining all our postulated constants are discussed. A good point in favor of taking the time to describe these techniques is that we have not found such a discussion elsewhere.

Initially we will consider states $x \in \mathbf{R}^D$, but we will soon find it convenient to restrict the treatment to $x \in \mathbf{R}_+^D$. In order to remain valid also in the discrete stochastic setting, however, constructed counterexamples will remain relevant also when restricted to $\mathbf{Z}_+^D$.

3.1. Stability

Many stability proofs can be thought of as comparisons with relevant linear cases. This is the motivation for the

well-known Grönwall's inequality which we state in the following two versions.

Lemma 3.1. *Suppose that* $u'(t) \leq A + \alpha u(t)$ *for* $t \geq 0$. *Then*

$$u(t) \leq u(0) e^{\alpha t} + \frac{A}{\alpha} \left(e^{\alpha t} - 1 \right).$$ (3.1)

The same conclusion holds irrespective of the differentiability of u *but with* $\alpha \geq 0$ *and under the weaker integral condition*

$$u(t) \leq u(0) + \int_0^t A + \alpha u(s) \, ds.$$ (3.2)

The most immediate way of comparing the growth of solutions to the ODE (2.2) to those of a linear ODE is to require that the norm of the driving function is bounded in terms of its argument;

$$\|F(x)\| \leq A + \alpha \|x\|,$$ (3.3)

since then by the triangle inequality,

$$\|x(t)\| \leq \|x_0\| + \int_0^t \|F(x)\| \, dt \leq \|x_0\| + \int_0^t A + \alpha \|x(t)\| \, dt,$$ (3.4)

where Grönwall's inequality applies. Unfortunately, (3.3) is a too strict requirement for our applications.

Proposition 3.2 *The bi-molecular birth-death system* (2.32) *does not satisfy* (3.3).

Proof. We compute $\|F(x)\| = \|-\mathbb{N}w(x)\| = \sqrt{2}|k_1 - k_2 ab|$ for a state $x = [a,b]^{\mathrm{T}}$. Hence for $a = b = N = 0$, $1, \cdots$ we have for N large enough that $\|F(x)\| = \sqrt{2}k_2 \cdot N^2 - \sqrt{2}k_1$, which can clearly never be bounded linearly in $\|x\| = \sqrt{2}N$. $\square$

The problem with the simple condition (3.3) is that it does not take the direction of growth into account; the offending quadratic propensity in (2.32) actually *decreases* the number of molecules. To deal with this, let $x \in \mathbf{R}^D$ be an arbitrary vector defining an "outward" direction. The length of the component of the driving function along this direction is $(x, F(x))$ and in order not to have x driven too strongly out along this ray we may, in view of Grönwall's inequality, naturally require that $(x/\|x\|, F(x)) \leq \text{constant} \times \|x\|$ for $\|x\|$ sufficiently large. Equivalently, for any *x*,

$$(x, F(x)) \leq A + \alpha \|x\|^2,$$ (3.5)

from which one deduces

$$\frac{d}{dt} \frac{\|x\|^2}{2} = (x, F(x)) \leq A + \alpha \|x\|^2,$$ (3.6)

where Grönwall's inequality applies anew. The assumption (3.5) is weaker than (3.3) since the former implies the latter by the Cauchy-Schwarz inequality. Indeed, as in the proof of Proposition 3.2 it is readily checked that for the bi-molecular birth-death system (2.32), we get $(x, F(x)) = k_1(a+b) - k_2(a+b)ab$ which this time readily can be bounded linearly in terms of $\|x\|^2 = a^2 + b^2$.

Unfortunately, in the case of an infinite state space and strong dependencies between the species the assumption (3.5) is also often unrealistic.

Proposition 3.3 *Neither* (2.36a) *nor* (2.36b) *admits a bound of the kind* (3.5).

Proof. As in the proof of Proposition 3.2 we look at a ray $x^{\mathrm{T}} = (a,b,c) = (N,N,3N)$ parametrized by a non-negative integer N. For (2.36a) we compute $(x, F(x)) = (x, -\mathbb{N}w(x)) = (a+b-c)(k_2c - k_1ab) = k_1N^3 - 3k_2N^2$, which clearly cannot be bounded linearly in $\|x\|^2 = 11N^2$. The same argument applies also to (2.36b). $\square$

This negative result can perhaps best be appreciated as a kind of loss of information about the dependencies between the species in the functional form of the condition (3.5). The number of A- and B-molecules is strongly correlated with the number of C-molecules such that, in fact, in (2.36a) $a+b+2c$ is a preserved quantity. By contrast, in (3.5) the growth of $\|x\|^2$ is estimated from the sum of the growth of the individual elements of *x* as if they where independent.

A way around this limitation can be found provided that we leave the general case $x \in \mathbf{R}^D$. We therefore

specify the discussion to the positive quadrant $x \in \mathbf{R}_+^D$ and assume from now on that it can be shown *a priori* that the initial data x_0 belongs to this set and that the subsequent trajectory $x(t)$ never leaves it (compare Assumption 2.1).

It then follows that $\|x\|_1 = (\mathbf{1}, x) = \mathbf{1}^{\mathrm{T}} x$, where $\mathbf{1}$ is the vector of length D containing all ones. This vector also defines a suitable "outward" vector for states $x \in \mathbf{R}_+^D$ since solutions to the ODE (2.2) cannot grow without simultaneously growing also in the direction of $\mathbf{1}$.

Again, in view of Grönwall's inequality Lemma 3.1, we tentatively require that $(\mathbf{1}, F(x)) \leq \text{constant} \times \|x\|_1$ for $\|x\|_1$ sufficiently large. Equivalently, for any x,

$$(\mathbf{1}, F(x)) \leq A + \alpha \|x\|_1, \tag{3.7}$$

implying the bounded dynamics

$$\frac{\mathrm{d}}{\mathrm{d}t} \|x\|_1 = (\mathbf{1}, F(x)) \leq A + \alpha \|x\|_1. \tag{3.8}$$
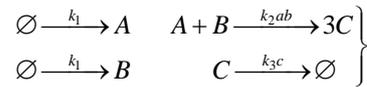
We remark in passing that the criterion (3.7) is sharp in the sense that if the reversed inequality can be shown to be true, then the growth of solutions can be estimated from below.

Example 3.1 As a point in favor of this approach we compute for the bi-molecular birth-death system (2.32), $(\mathbf{1}, F(x)) = (\mathbf{1}, -\mathbb{N}w(x)) = 2k_1 - 2k_2 ab$ which evidently falls under the assumption (3.7) with $(A, \alpha) = (2k_1, 0)$. For the reversible case (2.36a) we similarly get $(\mathbf{1}, F(x)) = -k_1 ab + k_2 c$ such that (3.7) applies with $(A, \alpha) = (0, k_2)$. Finally, and in the same fashion, the open case (2.36b) is seen to be covered by letting $(A, \alpha) = (k_4, (k_2 - k_3) \vee 0)$.

The chosen "outward" vector $\mathbf{1}$ is by no means special. Clearly, any strictly positive vector l may be used in its place since $\|x\|_1$ and $\|x\|_l := l^{\mathrm{T}} x$ are equivalent norms over $\mathbf{R}_+^D$. This is a general and useful observation as it may be used to discard parts of a system that are closed without any restrictions on the associated propensities.

Example 3.2 For the reversible system (2.36a), we have already noted that $a + b + 2c$ is a conserved quantity such that the choice $l = [1, 1, 2]^{\mathrm{T}}$ yields $\mathrm{d}/\mathrm{d}t \|x\|_l = 0$. The open case (2.36b) also benefits from this weighted norm in that we get $\mathrm{d}/\mathrm{d}t \|x\|_l \leq 2k_4$.

Example 3.3 A slightly more involved model reads as follows:

$$\left. \begin{array}{ll} \varnothing \xrightarrow{k_1} A & A + B \xrightarrow{k_2 ab} 3C \\ \varnothing \xrightarrow{k_1} B & C \xrightarrow{k_3 c} \varnothing \end{array} \right\}$$

This example has been constructed such that the quadratic reaction increases $\|x\|_1$ and hence (3.7) does not apply. However, taking $l = [3, 3, 2]^{\mathrm{T}}$ we get

$$\frac{\mathrm{d}}{\mathrm{d}t} \|x_t\|_l = 6k_1 - 2k_3 c \leq 6k_1.$$

This example hints at a general technique for obtaining suitable candidates for the weight vector l. Simply form the matrix $\mathbb{N}_2$ consisting of the columns of $\mathbb{N}$ that are affected by superlinear propensities. If a vector $l > 0$ annihilating these propensities exists, it can be found in the null-space of $-\mathbb{N}_2^{\mathrm{T}}$, readily available by linear algebra techniques. We omit the details.

3.2. Continuity

For well-posedness of the ODE (2.2) we also need continuity with respect to the initial data. We cannot ask for uniform Lipschitz continuity since $\|F(x) - F(y)\| \leq L \|x - y\|$ clearly implies (3.3) which we have already refuted. For the same reason, a uniform *one-sided Lipschitz condition* $(x - y, F(x) - F(y)) \leq \lambda \|x - y\|^2$ cannot be assumed to hold since it implies (3.5). The problem here is the global nature of the estimate and it therefore seems to be reasonable to localize this assumption. For instance, one might ask for

$$(x - y, F(x) - F(y)) \leq \lambda_R \|x - y\|^2 \quad \text{whenever } \|x\|, \|y\| \leq R, \tag{3.9}$$

presumably with some growth restrictions on λ_R. Although very general, such an analysis is likely to be less

informative when it comes to estimating actual constants in later results. We shall therefore consider the following simpler version,

$$\left(x-y, F(x)-F(y)\right) \le \left(M + \mu \|x+y\|_1\right)\|x-y\|^2, \tag{3.10}$$

where the form of λ_R has been restricted to better suit the present purposes. Trivially, the norms $\|\cdot\|_1$ and $\|\cdot\|$ are equivalent and hence the specific choice made in (3.10) is just a matter of convenience. Since the idea here is to use *a priori* bounds on x and y when deriving perturbation bounds, using $\|\cdot\|_1$ (or $\|\cdot\|_l$) is natural.

Theorem 3.4 *Suppose that the ODE* (2.2) *satisfies* (3.7) *and* (3.10) *and that initial data* $x_0 \in \mathbf{R}_+^D$ *implies a solution* $x(t) \in \mathbf{R}_+^D$. *Then for any* $t \in [0,T]$ *there is a unique such solution* $x(t)$. *Moreover, define* $C(t; x_0, y_0) \equiv M + \mu\|x_t + y_t\|_1$, *where* x_t *and* y_t *are two trajectories associated with initial data* $x_0, y_0 \in \mathbf{R}_+^D$, *respectively. Then*

$$\|x(t) - y(t)\| \le \|x_0 - y_0\| \exp\left(\int_0^t C(s; x_0, y_0)\,ds\right)$$
$$= \|x_0 - y_0\|\left[1 + \left(M + \mu\left(\|x_0 + y_0\|_1\right)\right)t + \mathcal{O}(t^2)\right]. \tag{3.11}$$

Proof. Combining (3.7) with Grönwall's inequality we get the *a priori* estimate

$$\|x(t) + y(t)\|_1 \le \|x_0 + y_0\|_1 + \left(\alpha\|x_0 + y_0\|_1 + 2A\right)\left(e^{\alpha t} - 1\right)/\alpha.$$

Hence the (bounded) solution to

$$\frac{d}{dt}\|x - y\|^2 = 2\left(x - y, F(x) - F(y)\right) \le 2C(t; x_0, y_0)\|x - y\|^2.$$

is readily found through its integrating factor. The order estimate is a consequence of the fact that

$$\int_0^t \|x_s\|_1 - \|x_0\|_1 \, ds = \mathcal{O}(t^2),$$

since the trajectory is continuous. $\square$

3.3. Bounds for Elementary Reactions

As briefly discussed by the end of Section 3.1, finding bounds on A and α in (3.7) as well as a suitable weight-vector l amounts to basic inequalities and some fairly straightforward linear algebra manipulations. In this section we therefore consider precise bounds in (3.10) for the elementary propensities (2.3). Since (3.10) is linear in F, a reasonable approach is to consider linear and quadratic propensities separate (constant propensities trivially satisfy (3.10) with $M = \mu = 0$).

Proposition 3.5 (*linear case*) *Write a set of* R *linear propensities as* $w_r(x) = q_r^T x$, $r = 1, \cdots, R$, *each with the corresponding stoichiometric vector* $\mathbb{N}_r$. *Then* $F(x) := -\sum_r \mathbb{N}_r w_r(x)$ *satisfies*

$$\left(x - y, F(x) - F(y)\right) \le M\|x - y\|^2 \text{ with } M = \mu_2\left[-\mathbb{N}Q^T\right]$$

in terms of the Euclidean logarithmic norm $\mu_2[\cdot]$ *and the matrix* Q *containing the vectors* q_r *as columns. In particular, in the case of a single linear propensity and, if as is usually the case,* $q_j = k\delta_{jn}$ *is all-zero except for a single rate constant* k *in the nth position, then this reduces to* $M = k\left(-\mathbb{N}_{r,n} + \|\mathbb{N}_r\|\right)/2$.

Proof. The first assertion is immediate since the smallest such constant M by definition is the logarithmic norm (see e.g. [32]). To compute $\mu_2\left[-\mathbb{N}_r q_r^T\right]$ when q has the form indicated, we determine the extremal eigenvalue of $-\left(\mathbb{N}_r q_r^T + q_r \mathbb{N}_r^T\right)/2$. By the (signed) scaling invariance of the logarithmic norm we may without loss of generality take $k \equiv 1$. The spectral relation for an eigenpair (λ, z) can be written as

$$-\frac{1}{2}\mathbb{N}_{r,j} z_n = \lambda z_j, \quad j = 1, 2, \cdots, D, \ j \ne n,$$

$$-\frac{1}{2}\mathbb{N}_{r,n} z_n - \frac{1}{2}\mathbb{N}_r^T z = \lambda z_n.$$

For non-zero λ the first relation can be solved for z_j. When inserted into the second relation, using that $z_n \ne 0$ (or otherwise $z = 0$), we get a quadratic equation for λ with a single extremal root. $\square$

Example 3.4 The simple special case in Proposition 3.5 is generally sharp except for when there are linear

reactions affecting all species considered in the model. For example, in a one-dimensional state space, the single decay $A \to \varnothing$ with propensity $w_1(a) = ka$ allows the optimal value $M = -k$. In general D-dimensional space, a chain with unit rate constants of the form $A_1 \to A_2 \to \cdots \to A_D \to \varnothing$, or a closed loop in which the last transition is replaced with $A_D \to A_1$, both admit bounds $M \leq 0$ as an inspection of the Gershgorin-discs of $-\left(\mathbb{N} + \mathbb{N}^{\mathrm{T}}\right)/2$ shows.

Other than for those special examples, for the most important linear cases, **Table 1** summarizes the bounds as obtained from the special case in Proposition 3.5 (with all reaction constants normalized to unity).

Proposition 3.6 (*quadratic case*). *Write a general quadratic propensity as* $w_r(x) = x^{\mathrm{T}} S x$ *with* S *a symmetric matrix. Then* $F_r(x) = -\mathbb{N}_r w_r(x)$ *satisfies* (3.10) *with* $M = 0$ *and*

$$\mu = \|x + y\|_1^{-1} \mu_2\left[-\mathbb{N}_r (x + y)^{\mathrm{T}} S\right] \leq \|\mathbb{N}_r\| \|S\|. \text{ For the special case that } S_{ij} = k\left(\delta_{im}\delta_{jn} + \delta_{jm}\delta_{in}\right)/2 \text{ there holds}$$

$$\mu \leq k \max_{j \in \{m,n\}}\left(-\mathbb{N}_{r,j} + \|\mathbb{N}_r\|\right)/4.$$

Proof. Since S is symmetric we have $x^{\mathrm{T}} S x - y^{\mathrm{T}} S y = (x + y)^{\mathrm{T}} S(x - y)$. Hence an explicit expression for μ is obtained as follows:

$$\mu \leq \sup_{x,y \in \mathbf{Z}_+^D} \|x + y\|_1^{-1} \|x - y\|^{-2}\left(x - y, -\mathbb{N}_r (x + y)^{\mathrm{T}} S(x - y)\right)$$

$$= \sup_{w \in \mathbf{Z}_+^D} \sup_{v \in \mathbf{Z}^D} \|w\|_1^{-1} \|v\|^{-2}\left(v, -\mathbb{N}_r w^{\mathrm{T}} S v\right) \leq \sup_{\substack{u \geq 0 \\ \|u\|_1 \leq 1}} \mu_2\left[-\mathbb{N}_r u^{\mathrm{T}} S\right].$$

The indicated upper bound is derived from the fact that $\left|\mu_2[\cdot]\right| \leq \|\cdot\|$ [32]. For the useful special case, define first the vector $q = q_1 + q_2$ in terms of $q_{1,j} = k\delta_{jm}(x_n + y_n)/2$, and $q_{2,j} = k\delta_{jn}(x_m + y_m)/2$. Using the fact that the logarithmic norm is sub-additive we can reuse the calculation in the proof of Proposition 3.5,

$$\mu_2\left[-\mathbb{N}_r (x + y)^{\mathrm{T}} S\right] = \mu_2\left[-\mathbb{N}_r q^{\mathrm{T}}\right] \leq \mu_2\left[-\mathbb{N}_r q_1^{\mathrm{T}}\right] + \mu_2\left[-\mathbb{N}_r q_2^{\mathrm{T}}\right]$$

$$= (x_n + y_n)k\left(-\mathbb{N}_{r,m} + \|\mathbb{N}_r\|\right)/4 + (x_m + y_m)k\left(-\mathbb{N}_{r,n} + \|\mathbb{N}_r\|\right)/4. \qquad \square$$
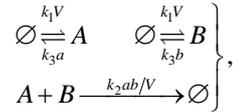
Example 3.5 The most important quadratic cases are summarized in **Table 2**. For the dimerizations in the lower half of the table there is also a linear part M in (3.10).

Example 3.6 The bi-molecular birth-death model (2.32) admits the constants $(M, \mu) = \left(0, k_2\left(\sqrt{2} - 1\right)/4\right)$ in (3.10). Similarly, the reversible cases (2.36a) and (2.36b) both obeys (3.10) with $(M, \mu) = \left(\sqrt{3} - 1\right)/2 \times \left(k_2, k_1/2\right)$. All these results are sharp except for the open case (2.36b) for which one can obtain a slightly smaller constant M by using the general formula stated in Proposition 3.5.

Example 3.7 As a highly prototypical example we consider the following natural extension of the bi-molecular birth-death model (2.32),

$$\left.\begin{array}{cc} \varnothing \underset{k_3 a}{\overset{k_1 V}{\rightleftharpoons}} A & \varnothing \underset{k_3 b}{\overset{k_1 V}{\rightleftharpoons}} B \\ A + B \xrightarrow{k_2 ab/V} \varnothing \end{array}\right\},$$

where in this example it is informative to consider the dependence on the system's size V. It is straightforward to show the bounds $(A, \alpha) \leq (2k_1 V, -k_3)$ in (3.7) and hence that the system is effectively bounded despite being of open character. This is seen from the fact that, for states $\|x\|_1 \geq 2k_1/k_3 \cdot V$, the dynamics is *dissipative* in the $\|\cdot\|_1$-norm. Furthermore, from Proposition 3.5 and 3.6 we get the sharp bounds $(M, \mu) \leq \left(-k_1 V, k_2/V \cdot \left(\sqrt{2} - 1\right)/4\right)$ in (3.10). It follows that for states $\{x, y\}$ such that $\|x + y\|_1 \lesssim 9.7 k_1/k_2 \cdot V^2$, the dynamics is *contractive* in the Euclidean norm. For density dependent propensities we expect that $\|x\| \sim V$ in any norm as V grows, and hence the region of contractivity grows in a relative sense. Intuitively one expects that these results offer an insight into the evolution of the process that is relevant also in the stochastic setting.

4. Stochastic Stability

We now consider the properties of the stochastic jump SDE (2.16). For convenience we start by collecting all assumptions in Section 4.1. In the stochastic setting the requirements for existence and uniqueness are slightly stronger than in the deterministic case such that the one-sided bound (3.10) needs to be augmented with an

Table 1. Linear propensities and bounds of M in (3.10).

Reaction	Bound on M
$A \rightarrow \varnothing$	0
$A \rightarrow B$	$\left(\sqrt{2}-1\right)/2$
$A \rightarrow B+C$	$\left(\sqrt{3}-1\right)/2$

Table 2. Quadratic propensities and bounds of M, μ in (3.10).

Reaction	Bound on μ	M
$A + B \rightarrow \varnothing$	$\left(\sqrt{2}-1\right)/4$	
$A + B \rightarrow C$	$\left(\sqrt{3}-1\right)/4$	
$A + B \rightarrow A$	$1/4$	
$A + B \rightarrow A + C$	$\sqrt{2}/4$	
$A + A \rightarrow \varnothing$	0	2
$A + A \rightarrow B$	$\sqrt{5}/2-1$	$\sqrt{5}/2+1$

unsigned version, implying essentially the assumption of at most quadratically growing propensities. We demonstrate that this assumption is reasonable by constructing a model involving cubic propensities and with unbounded second moments. On the positive side we show in Section 4.2 that the assumptions are strong enough to guarantee finite moments of any order during finite time intervals.

We prove existence and uniqueness of solutions to the jump SDE (2.16) in Section 4.3. A sufficient condition for the existence of asymptotic bounds of the pth order moment is given in Section 4.4 where we also derive some stability estimates.

4.1. Working Assumptions

We state formally the set of assumptions on the jump SDE (14) as follows.

Assumption 4.9 For arguments x, $y \in \mathbf{Z}_+^D$, $F(x) := -\mathbb{N}w(x)$, and weighted norm $\|x\|_l := l^T x$ we assume that

(1) $-l^T \mathbb{N}w(x) \leq A + \alpha\|x\|_l$ ("*bounded growth*"),

(2) $\left(-l^T \mathbb{N}\right)^2 w(x)/2 \leq B + \beta_1\|x\|_l + \beta_2\|x\|_l^2$ ("*absolutely bounded growth*"),

(3) $\left(\mathbf{1}^T \mathbb{N}^2\right)|w(x)-w(y)| \leq L\left(1+\|x+y\|_1\right)\|x-y\|$,

(4) $\left(x-y, F(x)-F(y)\right) \leq \left(M + \mu\|x+y\|_1\right)\|x-y\|^2$.

The parameters $\{A, B, \beta_1, \beta_2, L\}$ are assumed to be positive (with β_2 possibly zero) but we allow also negative values of $\{\alpha, M, \mu\}$. The vector l is normalized such that $\min_i l_i = 1$; hence the bound $\|x\|_1 \leq \|x\|_l$ is sharp.

After the original draft of the current paper was finished, the author became aware of two other papers discussing very similar conditions [33] [34]. In particular, Assumption 4.1 (1), (2) are also found in ([33], Condition (1). In fact, these very conditions can be shown to be exactly what is needed to apply the earlier and quite general theory found in ([35], Theorem 7.1).

In Assumption 4.9 (2) the case $\beta_2 = 0$ will merit special attention. For well-posedness it turns out that we will need to require a higher regularity of the initial data when $\beta_2 > 0$ (see Theorem 4.7) and the condition for ergodicity becomes more restrictive (see Theorem 4.9). In practice, $\beta_2 = 0$ implies that opposing quadratic reactions of the type

$$2X \xrightarrow{\ x(x-1)\ } 3X \atop 2X \xrightarrow{\ x(x-1)\ } X \Bigg\},$$

(4.1)

are impossible. Similarly, when $\beta_1 = 0$ reactions of the type

$$X \xrightarrow{\ x\ } 2X$$

(4.2)

are excluded.

Note that (2) and (4) are redundant in the sense that they are both implied by (3). However, as we saw in Section 3.3, in (4) it is often possible to find sharper constants M and μ by considering this bound in isolation. Also, although (3) is stronger than (4), it is in particular valid for quadratic propensities as can be seen from the representation used in the proof of Proposition 3.6,

$$\left| w_r(x) - w_r(y) \right| = \left| (x+y)^{\mathrm{T}} S(x-y) \right| \leq \|S\| \|x+y\|_1 \|x-y\|.$$

(4.3)

The Danger with Cubic Propensities

Assumption 4.1 (2) specifies the discussion to propensities with at most quadratic growth, at least when measured in the direction of the weight vector l. To show that this is natural we now demonstrate that additional care should be taken when considering *cubic* propensities.

Example 4.2 *Consider the model*

$$3X \xrightarrow{\ x(x-1)(x-2)/2\ } X \atop 3X \xrightarrow{\ x(x-1)(x-2)\ } 4X \Bigg\},$$
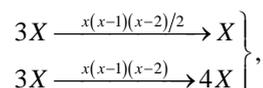
such that the stoichiometric vector is given by $\mathbb{N} = [-2,1]$, *and hence that the drift* $-\mathbb{N}w(x) = 0$.

Proposition 4.1 *For the model in Example* 4.2, *if* $X_0 \geq 3$, *then the second moment explodes in finite time.*

Proof. Assume that both the second and the third moment are bounded for $t \in [0,T)$ with $T > 0$. From (2.7) we get the governing equation

$$\frac{\mathrm{d}}{\mathrm{d}t} EX_t^2 = E\left[3X_t(X_t - 1)(X_t - 2) \right],$$

such that the growth of the second moment remains bounded only provided that the third moment remains finite. It is convenient to look at the cumulative third order moment. From (2.7),

$$\frac{\mathrm{d}}{\mathrm{d}t} EC_3(X_t) := \frac{\mathrm{d}}{\mathrm{d}t} EX_t(X_t - 1)(X_t - 2) = E\left[9C_3(X_t)(X_t - 2/3) \right].$$

By the arithmetic-geometric mean inequality, $x - 2/3 \geq x - 1 \geq \left[x(x-1)(x-2) \right]^{1/3}$, such that by Jensen's inequality,

$$\frac{\mathrm{d}}{\mathrm{d}t} EC_3(X_t) \geq 9E\left[C_3(X_t)^{4/3} \right] \geq 9\left[EC_3(X_t) \right]^{4/3}.$$

We put $u^3 = EC_3(X_t)$ and get the differential inequality

$$\frac{\mathrm{d}(1/u)}{\mathrm{d}t} \leq -3,$$

which can be integrated and rearranged to produce the bound

$$E\left[X_t(X_t - 1)(X_t - 2) \right] \geq \frac{X_0(X_0 - 1)(X_0 - 2)}{1 - 3t X_0(X_0 - 1)(X_0 - 2)}.$$

(4.4)

Hence the third, and consequently also the second moment explode for some finite t whenever $X_0 \geq 3$. $\square$

Interestingly, we note that if $X_0 = 3$, then the probability that the cubic decay transition occurs first is $1/3$, and if this happens the state of the system will be stuck with a single molecule indefinitely.

4.2. Moment Bounds

In this section we consider general moment bounds derived from (2.20) using localization. To get some guidance, let us first assume that the differential form of Dynkin's formula (2.7) is valid. Since any trajectory $(X_t)_{t \geq 0}$ by the basic Assumption 2.1 will belong to $\mathbf{Z}_+^D$, we may use that $\|X_t\|_l = (l, X_t)$. Hence from (2.7) with $f(x) = (l, x)$ we get that

$$\frac{\mathrm{d}}{\mathrm{d}t} E \|X_t\|_l = (l, F(X_t)) \leq A + \alpha \|X_t\|_l, \tag{4.5}$$

by Assumption 4.1 (1). Clearly, the differential form of Grönwall's inequality in Lemma 3.1 applies here. A correct version of this argument unfortunately looses the sign of α.

Proposition 4.2 *If Assumption* 4.1 (1) *is true, then*

$$E \|X_t\|_l \leq \|X_0\|_l \exp(\alpha_+ t) + A(\exp(\alpha_+ t) - 1)/\alpha_+,$$

where $\alpha_+ = \alpha \vee 0$.

Here and below we shall make use of the stopping time $\tau_P = \inf_{t \geq 0} \{\|X_t\|_l > P\}$ and define $\hat{t} = t \wedge \tau_P$.

Proof. From (2.20) with $f(x) = (l, x)$ we get that

$$E \|X_{\hat{t}}\|_l = \|X_0\|_l + E \int_0^{\hat{t}} (l, F(X_s)) \mathrm{d}s \leq \|X_0\|_l + E \int_0^t A + \alpha_+ \|X_{\hat{s}}\|_l \, \mathrm{d}s, \tag{4.6}$$

By the integral form of Grönwall's inequality in Lemma 3.1 we deduce in terms of $Y_t := X_{t \wedge \tau_P}$ that

$$E \|Y_t\|_l \leq \|X_0\|_l \exp(\alpha_+ t) + A(\exp(\alpha_+ t) - 1)/\alpha_+ \tag{4.7}$$

such that the same bound holds for X_t by letting $P \to \infty$. $\square$

We attempt a similar treatment for obtaining bounds in mean square. Assuming tactically that (2.7) is valid, writing $\|x\|^2 = x^{\mathrm{T}} x$ we get after some work that

$$\frac{\mathrm{d}}{\mathrm{d}t} E \|X_t\|^2 = E \left[\mathbf{1}^{\mathrm{T}} \mathbb{N}^2 w(X_t) - 2 X_t^{\mathrm{T}} \mathbb{N} w(X_t) \right] \tag{4.8}$$

where $\mathbb{N}_{ij}^2 \equiv (\mathbb{N}_{ij})^2$. We expect from Grönwall's inequality that $E \|X_t\|^2$ grows at most exponentially with αt whenever

$$\mathbf{1}^{\mathrm{T}} \mathbb{N}^2 w(x) - 2 x^{\mathrm{T}} \mathbb{N} w(x) \leq A + \alpha \|x\|^2. \tag{4.9}$$

However, this tentative condition is often violated in practice since the second term $-x^{\mathrm{T}} \mathbb{N} w(x) = (x, F(x))$, and since we already know from Proposition 3.3 that this quantity does not admit bounds in terms of $\|x\|^2$ even for very simple problems.

More realistic conditions arise when seeking to bound $\|X_t\|_l^2$ instead.

Proposition 4.3 *If for some constants* γ *and* C,

$$\left(l^{\mathrm{T}} \mathbb{N} \right)^2 w(x) - 2 l^{\mathrm{T}} \mathbb{N} w(x) \|x\|_l \leq C + \gamma \|x\|_l^2, \tag{4.10}$$

$\left(l^{\mathrm{T}} \mathbb{N} \right)^2$ *understood elementwise*), *then* $E \|X_t\|_l^2 \leq \|X_0\|_l^2 \exp(\gamma_+ t) + C(\exp(\gamma_+ t) - 1)/\gamma_+$.

The proof of Proposition 4.3 follows the same pattern as for Proposition 4.2, but using this time $f(x) = \|x\|_l^2 = (l, x)^2$ in (2.20). The condition (4.10) is typically more realistic than (4.9) since we recognize the term $-l^{\mathrm{T}} \mathbb{N} w(x) \|x\|_l = (l, F(x)) \|x\|_l$, which under the evidently reasonable Assumption 4.1 (1) is $\leq (A + \alpha \|x\|_l) \|x\|_l$. It follows that if $\left(l^{\mathrm{T}} \mathbb{N} \right)^2 w(x)$ grows at most quadratically with $\|x\|_l$, then this assumption is sufficient to yield bounds in mean square. Stated formally,

Proposition 4.4 *Under Assumption* 4.1 (1) *and* (2) *the condition* (4.10) *of Proposition 4.3 is true with* $\gamma = 2\beta_2 + 2\alpha + 2$ *and* $C = 2B + \beta_1^2 + A^2$.

Proof. This is straightforward: we get by the assumptions and Hölder's inequality,

$$\left(\boldsymbol{l}^{\mathrm{T}}\mathbb{N}\right)^2 w(x) - 2\boldsymbol{l}^{\mathrm{T}}\mathbb{N}w(x)\|x\|_l \le 2B + 2\beta_1\|x\|_l + 2\beta_2\|x\|_l^2 + 2\left(A + \alpha\|x\|_l\right)\|x\|_l,$$

where an application of Young's inequality yields the indicated bounds. $\square$

As a strong point in favor of our running assumptions we now demonstrate that the above reasoning can be generalized: these two conditions implies finite time stability in any order moment. We note that in a recent manuscript [34], related conditions for the same results are proposed.

Theorem 4.5 (*Moment estimate*). *Under Assumption* 4.1 (1) *and* (2), *for any integer* $p \ge 1$,

$$E\|X_t\|_l^p \le \left(\|X_0\|_l^p + 1\right)\exp(Ct) - 1, \tag{4.11}$$

where $C > 0$ is a constant depending on the assumptions and on p.

The proof of Theorem 4.5 and some later results will simplify using the following bound.

Lemma 4.6 *Let* $H(x) \equiv (x+y)^p - x^p$ *with* $x \in \mathbf{R}_+$ *and* $y \in \mathbf{R}$. *Then for integer* $p \ge 1$ *we have the bounds*

$$H(x) \le pyx^{p-1} + 2^{p-4}p(p-1)y^2\left[x^{p-2} + |y|^{p-2}\right], \tag{4.12}$$

$$|H(x)| \le p|y|2^{p-2}\left[x^{p-1} + |y|^{p-1}\right]. \tag{4.13}$$

Proof. Both results follow from Taylor expansions;

$$H(x) = pyx^{p-1} + \frac{p(p-1)}{2}y^2\left[x + \theta_1 y\right]^{p-2},$$

$$|H(x)| = p|y|\left|x + \theta_2 y\right|^{p-1},$$

respectively, where $\theta_{1,2} \in [0,1]$. Using the triangle inequality and the elementary inequality $(a+b)^p \le 2^{p-1}\left(a^p + b^p\right)$ the lemma is proved. $\square$

Proof of Theorem 4.5. Using (2.20) with $f(X_t) \equiv \left[\boldsymbol{l}^{\mathrm{T}}X_t\right]^p$ we get

$$E\|X_{\hat{t}}\|_l^p = \|X_0\|_l^p + E\int_0^{\hat{t}}\underbrace{\sum_{r=1}^{R}w_r(X_s)\left[\left[\boldsymbol{l}^{\mathrm{T}}\left(X_s - \mathbb{N}_r\right)\right]^p - \left[\boldsymbol{l}^{\mathrm{T}}X_s\right]^p\right]}_{=:G(X_s)}\mathrm{d}s. \tag{4.14}$$

Using Lemma 4.6 (4.12) and Assumption 4.1 (1) and (2) we obtain

$$G(x) \le p\left(A + \alpha\|x\|_l\right)\|x\|_l^{p-1} + 2^{p-3}p(p-1)\left(B + \beta_1\|x\|_l + \beta_2\|x\|_l^2\right)\left(\|x\|_l^{p-2} + \Delta^{p-2}\right),$$

where $\Delta := \left\|\boldsymbol{l}^{\mathrm{T}}\mathbb{N}\right\|_\infty$. Expanding and using Young's inequality with exponents $\{p/(p-1), p/(p-2)\}$ and conjugate exponents $\{p, p/2\}$,

$$E\|X_{\hat{t}}\|_l^p \le \|X_0\|_l^p + \int_0^t C\left(1 + E\|X_{\hat{s}}\|_l^p\right)\mathrm{d}s,$$

for some constant C which thus depends on the assumptions. Applying Grönwall's inequality and letting $P \to \infty$ we obtain the stated result.

4.3. Existence and Uniqueness

We shall now prove that the jump SDE (2.16) under Assumption 4.1 has a uniquely defined and locally bounded solution. To this end and following ([8], Section 3.1.2), we introduce the following spaces of path-wise locally bounded processes:

$$S_{\mathcal{F}}^{p,\mathrm{loc}}\left(\mathbf{Z}_+^D\right) = \Bigg\{ X(t,\omega) : X \text{ is } \mathcal{F}_t\text{-adapted and } \mathbf{Z}_+^D\text{-valued such that}$$

$$E\sup_{t\in[0,T]}\left\|X(t,\omega)\right\|_1^p < \infty \text{ for } \forall T < \infty\Bigg\}. \tag{4.15}$$

Theorem 4.7 (Existence). *Let X_t be a solution to* (2.16) *under Assumption* 4.1 (1) *and* (2) *with* $\beta_2 = 0$. *Then if* $\|X_0\|_l^p < \infty$, $\{X_t\}_{t\geq 0} \in S_{\mathcal{F}}^{p,\text{loc}}\left(\mathbf{Z}_+^D\right)$. *If* $\beta_2 > 0$ *then the conclusion remains under the additional requirement that* $\|X_0\|_l^{p+1} < \infty$.

Proof. Below we let C denote a positive constant which may be different on each occasion used. As before we use the stopping time $\tau_P = \inf_{t\geq 0}\left\{\|X_t\|_l > P\right\}$ and put $\hat{t} = t \wedge \tau_P$. We get from Itô's formula (with G defined in (4.14))

$$\|X_{\hat{t}}\|_l^p = \|X_0\|_l^p + \int_0^{\hat{t}} G(X_s)\,\mathrm{d}s + \int_0^{\hat{t}}\int_I \sum_{r=1}^R \hat{w}_r(X_{s-};z)\left[\left[\boldsymbol{l}^{\mathrm{T}}(X_{s-} - \mathbb{N}_r)\right]^p - \left[\boldsymbol{l}^{\mathrm{T}}X_{s-}\right]^p\right](\mu - m)(\mathrm{d}s \times \mathrm{d}z).$$

Since the propensities are bounded for bounded arguments (Assumption 2.1), using the stopping time we find that the jump part is absolutely integrable and hence a local martingale $M_{\hat{t}}$. We estimate its quadratic variation under Assumption 4.1 (2),

$$E[M]_{\hat{t}}^{1/2} \leq E\left[\int_0^{\hat{t}}\int_I \sum_{r=1}^R \hat{w}_r(X_s;z)\left|\left[\boldsymbol{l}^{\mathrm{T}}(X_s - \mathbb{N}_r)\right]^p - \left[\boldsymbol{l}^{\mathrm{T}}X_s\right]^p\right|\mu(\mathrm{d}s \times \mathrm{d}z)\right]$$

$$= E\left[\int_0^{\hat{t}}\int_I \sum_{r=1}^R \hat{w}_r(X_s;z)W(X_s)\left|\left[\boldsymbol{l}^{\mathrm{T}}(X_s - \mathbb{N}_r)\right]^p - \left[\boldsymbol{l}^{\mathrm{T}}X_s\right]^p\right|\mathrm{d}s \times \mathrm{d}z\right]$$

$$= E\left[\int_0^{\hat{t}} \sum_{r=1}^R w_r(X_s)\left|\left[\boldsymbol{l}^{\mathrm{T}}(X_s - \mathbb{N}_r)\right]^p - \left[\boldsymbol{l}^{\mathrm{T}}X_s\right]^p\right|\mathrm{d}s\right]$$

$$\leq E\left[\int_0^{\hat{t}} \sum_{r=1}^R p\left|\boldsymbol{l}^{\mathrm{T}}\mathbb{N}_r\right|w_r(X_s)2^{p-2}\left[\|X_s\|_l^{p-1} + \left|\boldsymbol{l}^{\mathrm{T}}\mathbb{N}_r\right|^{p-1}\right]\mathrm{d}s\right] \tag{4.16}$$

$$\leq E\left[\int_0^{\hat{t}} C\left(B + \beta_1\|X_s\|_l + \beta_2\|X_s\|_l^2\right)\left[\|X_s\|_l^{p-1} + \Delta^{p-1}\right]\mathrm{d}s\right] \tag{4.17}$$

$$\leq E\int_0^{\hat{t}} C\left(1 + \|X_s\|_l^p + \beta_2\|X_s\|_l^{p+1}\right)\mathrm{d}s \tag{4.18}$$

where $\Delta = \|\boldsymbol{l}^{\mathrm{T}}\mathbb{N}\|_\infty$. In (4.16) Lemma 4.6 (4.13) was applied and Assumption 4.1 (2) entered in (4.17). Assume first that $\beta_2 = 0$. Then for the drift part we have already constructed a suitable bound in Theorem 4.5 such that

$$\|X_{\hat{t}}\|_l^p \leq \|X_0\|_l^p + \int_0^{\hat{t}} C\left(1 + \|X_s\|_l^p\right)\mathrm{d}s + |M_{\hat{t}}|.$$

Taking supremum and expectation values we get from Burkholder's inequality ([6], Chap. IV.4) that

$$E\sup_{s\in[0,\hat{t}]}\|X_s\|_l^p \leq \|X_0\|_l^p + \int_0^{\hat{t}} C\left(1 + E\sup_{s'\in[0,s]}\|X_{s'}\|_l^p\right)\mathrm{d}s.$$

Writing $\|X\|_l^p(t) \equiv \sup_{s\in[0,t]}\|X_s\|_l^p$ we conclude that

$$E\|X\|_l^p(t \wedge \tau_P) \leq \|X_0\|_l^p + \int_0^t C\left(1 + E\|X\|_l^p(s \wedge \tau_P)\right)\mathrm{d}s.$$

By Grönwall's inequality we have thus shown that $E\|X\|_l^p(t \wedge \tau_P)$ can be bounded in terms of the initial data and time t. The result now follows by letting $P \to \infty$ and using Fatou's lemma.

Next assume that $\beta_2 > 0$. Then we have to rely more directly on Theorem 4.5 in (4.18),

$$E[M]_{\hat{t}}^{1/2} \leq \int_0^{\hat{t}} C\left(1 + E\|X_s\|_l^{p+1}\right)\mathrm{d}s \leq \left(\mathrm{e}^{C\hat{t}} - 1\right)\left(\|X_0\|_l^{p+1} + 1\right),$$

where, although there is now a dependency on $\|X_0\|^{p+1}$, the rest of the argument carries through.

For the case that the initial data X_0 is non-deterministic we see that the general quadratic case Assumption 4.1 (2) with $\beta_2 > 0$ requires a one order higher moment of the initial data in order for a solution in $S_{\mathcal{F}}^{p,\text{loc}}\left(\mathbf{Z}_+^D\right)$ to exist.

Theorem 4.8 (Uniqueness) *Let Assumption* 4.1 (1)-(4) *hold true. Then any two paths X_t and Y_t coupled*

according to the description in Section 2.2.2 *with* $X_0 = Y_0$ *are equal.*

We shall be using the observation that, for $x \in \mathbf{Z}_+^D$, we have that $\|x\|_1 \le \|x\|^2$ (referred below to as the "integer inequality").

Proof. Under the same stopping time as before we get from Itô's formula using the coupling described in Section 2.2.2 that

$$E\left\|X_{\hat{t}} - Y_{\hat{t}}\right\|^2 = E\int_0^{\hat{t}} 2\left(X_s - Y_s, F(X_s) - F(Y_s)\right) + \left(\mathbf{1}^T \mathbb{N}^2\right)\left|w(X_s) - w(Y_s)\right| ds$$

$$\le E\int_0^{\hat{t}} 2\left(M + \mu\|X_s + Y_s\|_1\right)\|X_s - Y_s\|^2 + \left\|\mathbf{1}^T \mathbb{N}^2\right\|_\infty L\left(1 + \|X_s + Y_s\|_1\right)\|X_s - Y_s\| ds.$$

From the integer inequality we find that there is a constant $C \ge 0$ depending on P such that

$$E\left\|(X - Y)(t \wedge \tau_P)\right\|^2 \le \int_0^{t \wedge \tau_P} CE\|X_s - Y_s\|^2 \, ds \le \int_0^t CE\left\|(X - Y)(s \wedge \tau_P)\right\|^2 \, ds.$$

Using that $E\|X_0 - Y_0\|^2 = 0$ and Grönwall's inequality we conclude that the only solution is the zero solution. Letting $P \to \infty$ and using Fatou's lemma the statement is therefore proved. $\square$

In a certain sense the previous result is trivial; from the Poisson representation (2.8) we see that up to the first explosion, a path is uniquely determined from an initial state and a series of Poisson distributed events. However, and as we shall see below, the above proof is prototypical for more involved situations. An example would be when devising hybrid approximations in continuous state space. Indeed, in the above proof, note that if the integer inequality did not hold we would naturally have to rely on the Cauchy-Schwartz inequality instead. With $u(t)^2 = E\|X_t - Y_t\|^2$ this leads to bounds of the typical kind

$$u(t)^2 \le \int_0^t C\left(u(s) + u(s)^2\right) ds, \tag{4.20}$$

for which $u(t) = \exp(Ct/2) - 1$ is an admissible solution. This observation shows that the integer inequality as used in the proof is crucial; without it the integral inequality (4.19) admits growing solutions.

4.4. Stability

Although Theorem 4.5 shows that any moments are bounded in finite time, a relevant question from the modeling point of view is whether the first few moments remain bounded indefinitely. We give a result to this effect which relies on the existence of solutions in $S_{\mathcal{F}}^{p,\text{loc}}\left(\mathbf{Z}_+^D\right)$ which implies that the differential form (2.7) of Dynkin's formula may be used (cf. Corollary 2.2) such that in turn the differential Grönwall inequality applies. We mention anew that a very similar result has recently appeared in ([33], Theorem 2).

Theorem 4.9 (Ergodicity) *Under Assumption* 4.1 (1), (2), *suppose that*

$$\alpha + \beta_2(p-1) \le -\kappa_p < 0. \tag{4.21}$$

Then for integer $p \ge 1$, $E\|X_t\|_l^p$ *remains bounded as* $t \to \infty$.

Proof. The case $\beta_2 > 0$ is slightly more complicated to obtain so we shall concentrate on this. We omit the case $p = 1$ since it follows from (4.5) under the present assumptions. The idea of the proof is to asymptotically bound $E\left(C + \|X_t\|_l\right)^p$ with a certain positive constant $C = C(p)$ to be decided upon below. By (2.7) we get with $Z_t = C + \|X_t\|_l$,

$$\frac{dEZ_t^p}{dt} = E\sum_{r=1}^R w_r(X_t)\left[\left[Z_t - \mathbf{l}^T \mathbb{N}_r\right]^p - Z_t^p\right]$$

$$= E\sum_{r=1}^R w_r(X_t)\left[-\mathbf{l}^T \mathbb{N}_r pZ_t^{p-1} + \frac{\left(-\mathbf{l}^T \mathbb{N}_r\right)^2}{2} p(p-1)\left(Z_t - \theta_r \mathbf{l}^T \mathbb{N}_r\right)^{p-2}\right], \tag{4.22}$$

by Taylor's formula for some $\theta_r \in [0,1]$. Using the assumptions we get the bound

$$\frac{dEZ_t^p}{dt} \le pE\left[\left(A + \alpha\|X_t\|_l\right)Z_t^{p-1} + (p-1)\left(B + \beta_1\|X_t\|_l + \beta_2\|X_t\|_l^2\right)\left(Z_t + \|\mathbf{l}^T \mathbb{N}\|_\infty\right)^{p-2}\right]. \tag{4.23}$$

For the first term in (4.23) we get from the scaled Young's inequality with exponent $p/(p-1)$ and conjugate exponent p that

$$\left(A+\alpha\|X_t\|_l\right)Z_t^{p-1}=\alpha Z_t^p+(A-\alpha C)Z_t^{p-1}\le\alpha Z_t^p+\frac{\varepsilon(p-1)}{p}Z_t^p+\frac{\varepsilon^{1-p}}{p}|A-\alpha C|^p,$$

for some $\varepsilon>0$. As for the second term in (4.23) we first estimate for $\beta_2>0$

$$B+\beta_1\|X_t\|_l+\beta_2\|X_t\|_l^2\le\beta_2\left(\|X_t\|_l+B/\beta_1\right)\left(\|X_t\|_l+\beta_1/\beta_2\right).$$

Next by the arithmetic-geometric mean inequality we get

$$\left(\|X_t\|_l+B/\beta_1\right)\left(\|X_t\|_l+\beta_1/\beta_2\right)\left(Z_t+\|\boldsymbol{l}^\mathrm{T}\mathbb{N}\|_\infty\right)^{p-2}$$

$$\le\left(\frac{B}{p\beta_1}+\frac{\beta_1}{p\beta_2}+\frac{p-2}{p}\left(\|\boldsymbol{l}^\mathrm{T}\mathbb{N}\|_\infty+C\right)+\|X_t\|_l\right)^p=Z_t^p,$$

provided that we choose C as the solution to the equation

$$\frac{B}{p\beta_1}+\frac{\beta_1}{p\beta_2}+\frac{p-2}{p}\left(\|\boldsymbol{l}^\mathrm{T}\mathbb{N}\|_\infty+C\right)=C. \tag{4.24}$$

Taken together we thus have

$$\frac{\mathrm{d}EZ_t^p}{\mathrm{d}t}\le p\left(-\kappa_p+\frac{\epsilon(p-1)}{p}\right)EZ_t^p+\epsilon^{1-p}|A-\alpha C|^p.$$

Since $\kappa_p>0$ we may pick a small enough ϵ such that the bracketed expression remains negative. By Grönwall's inequality this then proves the result with $\beta_2>0$. To prove the case $\beta_2=0$ the same idea of proof applies and results in

$$\frac{\mathrm{d}EZ_t^p}{\mathrm{d}t}\le p\left(-\kappa_p+\frac{\epsilon(p-1)}{p}+\frac{\epsilon(p-1)^2}{p}\beta_1\right)EZ_t^p+\epsilon^{1-p}\left(|A-\alpha C|^p+\beta_1(p-1)\right),$$

for a certain new constant C satisfying an equation similar to (4.24). $\square$

We next aim at deriving some stability estimates with respect to perturbations in the *reaction coefficients*. An early account of this was given by Kurtz in [31], see also [18] for a recent discussion in a bounded setting. Given the linear dependence on the coefficients k_r, $r=1,\cdots,4$ in the elementary reactions (2.3) a suitable model seems to be that a perturbation $k_r\to k_r+\delta_r$ in a propensity $w_r(x)$ spreads linearly in a relative sense,

$$|w_r(x,k_r)-w_r(x,k_r+\delta_r)|\le\text{constant}\times\delta_r w_r(x,k_r). \tag{4.25}$$

We make this formal by requiring that

$$\left(\mathbf{1}^\mathrm{T}\mathbb{N}^2\right)|w(x)-w_\delta(x)|\le\delta\|w(x)\|_1\le\delta C\left(1+\|x\|_1^2\right), \tag{4.26}$$

where δ is a suitable measure of the total perturbation vector and where the perturbed propensity vector function is given by

$$w_\delta(x)\equiv\left[w_1\left(x,k_1+\delta_1\right),\cdots,w_R\left(x,k_R+\delta_R\right)\right]^\mathrm{T}. \tag{4.27}$$

The existence of an absolute constant C in (4.26) follows from Assumption 4.1 (3). We further conveniently assume that the entire statement of Assumption 4.1 carries over to the perturbed system, and for convenience we shall also assume that all constants are the same. By the triangle inequality and Assumption 4.1 (3) we obtain from (4.26) the bound

$$\left(\mathbf{1}^\mathrm{T}\mathbb{N}^2\right)|w(x)-w_\delta(y)|\le L\left(1+\|x+y\|_1\right)\|x-y\|+\delta C\left(1+\|y\|_1^2\right)\le C\left(\delta+\|x-y\|^2\right), \tag{4.28}$$

with C some constant and where the simplification in (4.28) assumes an *a priori* bound (e.g. stopping time)

$\|x + y\|_1 \leq P$ and additionally requires the integer inequality.

The starting point for the analysis will be Itô's formula under the coupling described in Section 2.2.2. The techniques used below generalize well to pth order moment estimates, but for ease of exposition we let $p = 2$.

Hence under the model for coefficient perturbations (4.26)-(4.27) we have that

$$
\begin{aligned}
\|X_t - Y_t\|^2 &= \int_0^t 2\left(X_s - Y_s, F(X_s) - F_\delta(Y_s)\right) + \left(\mathbf{1}^{\mathrm{T}}\mathbb{N}^2\right)\left|w(X_s) - w_\delta(Y_s)\right| \mathrm{d}s \\
&+ \int_0^t \int_I 2\left(X_{s-} - Y_{s-}, -\mathbb{N}\left(\hat{w}^\delta(X_{s-}, Y_{s-}; z) - \hat{w}^\delta(Y_{s-}, X_{s-}; z)\right)\right) \\
&+ \left\|-\mathbb{N}\left(\hat{w}^\delta(X_{s-}, Y_{s-}; z) - \hat{w}^\delta(Y_{s-}, X_{s-}; z)\right)\right\|^2 (\mu_\delta - m_\delta)(\mathrm{d}s \times \mathrm{d}z).
\end{aligned}
\tag{4.29}
$$

Theorem 4.10 (Continuity). *Let two trajectories X_t and Y_t be given, with the same initial data and coupled according to the discussion in Section* 2.2.2. *Let the propensities for Y_t be perturbed by δ as indicated in* (4.26), (4.27). *Then*

$$
\lim_{\delta \to 0+} E\|X_t - Y_t\|^2 = 0.
\tag{4.30}
$$

Proof. We use the stopping time $\tau_P = \inf_{t \geq 0}\left\{\|X_t + Y_t\|_1 > P\right\}$ and put $\hat{t} = t \wedge \tau_P$. From (4.29) we get

$$
\begin{aligned}
E\|X_{\hat{t}} - Y_{\hat{t}}\|^2 &= \int_0^{\hat{t}} E\left[2\left(X_s - Y_s, F(X_s) - F_\delta(Y_s)\right) + \left(\mathbf{1}^{\mathrm{T}}\mathbb{N}^2\right)\left|w(X_s) - w_\delta(Y_s)\right|\right]\mathrm{d}s \\
&\leq \int_0^{\hat{t}} E\left[2\left(M + \mu\|X_s + Y_s\|_1\right)\|X_s - Y_s\|^2 + 2\delta\left(\mathbf{1}^{\mathrm{T}}\mathbb{N}^2\right)^{1/2}\|w(Y_s) - w_\delta(Y_s)\|\|X_s - Y_s\| + C\left(\delta + \|X_s - Y_s\|^2\right)\right]\mathrm{d}s,
\end{aligned}
\tag{4.31}
$$

where (4.28) was used. Simplifying further for a bounded δ we get

$$
\begin{aligned}
E\left\|(X - Y)(t \wedge \tau_P)\right\|^2 &\leq \int_0^{t \wedge \tau_P} C\left(\delta + E\|X_s - Y_s\|^2\right)\mathrm{d}s \\
&\leq \int_0^t C\left(\delta + E\left\|(X - Y)(s \wedge \tau_P)\right\|^2\right)\mathrm{d}s.
\end{aligned}
$$

Using that $X_0 = Y_0$ and Grönwall's inequality we conclude that

$$
E\|X_{\hat{t}} - Y_{\hat{t}}\|^2 \leq \delta\left(\exp(Ct) - 1\right).
$$

To get rid of the stopping time we write in terms of indicator functions,

$$
\begin{aligned}
E\|X_t - Y_t\|^2 &= E\left[\|X_t - Y_t\|^2[t < \tau_P] + \|X_t - Y_t\|^2[t \geq \tau_P]\right] \\
&\leq E\|X_{\hat{t}} - Y_{\hat{t}}\|^2 + \left(E\|X_t - Y_t\|^4\right)^{1/2}\left(\mathbf{P}[t \geq \tau_P]\right)^{1/2}.
\end{aligned}
\tag{4.32}
$$

Using $\mathbf{P}[t \geq \tau_P] = \mathbf{P}\left[\sup_{s \in [0,t]}\|X_s + Y_s\|_1 > P\right]$ we get from Markov's inequality and the previous estimate

$$
E\|X_t - Y_t\|^2 \leq \delta\left(\exp(Ct) - 1\right) + \left(E\|X_t - Y_t\|^4\right)^{1/2} P^{-1/2}\left(E\sup_{s \in [0,t]}\|X_s + Y_s\|_1\right)^{1/2}.
\tag{4.33}
$$

Relying on the existence result in Theorem 4.7 we find that for any given $\varepsilon > 0$ we can select P (and hence also C) such that the right term is $< \varepsilon/2$. We can next find $\delta_0 > 0$ such that for all $\delta \leq \delta_0$, also the left term is $< \varepsilon/2$. Hence for all $\delta \leq \delta_0$, $E\|X_t - Y_t\|^2 < \varepsilon$ as claimed. $\square$

As a by-product of the proof we see that if the process is *bounded*, then for P large enough the probability in (4.32) is zero.

Corollary 4.11 (Perturbation estimate, bounded version) *If in Theorem* 4.10, *the processes X_t and Y_t are bounded, then for a constant $C > 0$,*

$$
E\|X_t - Y_t\|^2 \leq \delta\left(\exp(Ct) - 1\right).
\tag{4.34}
$$

The constant C in (4.34) can be bounded explicitly by inspection of (4.28) and (4.31).

For an *unbounded* system it is apparently much more difficult to obtain explicit estimates. However, by controlling also the martingale part we can strengthen Theorem 4.10 in another direction.

Theorem 4.12 (Continuity/sup) *Under the same assumptions as Theorem* 4.10 *we have that*

$$\lim_{\delta \to 0+} E \sup_{s \in [0,t]} \|X_s - Y_s\|^2 = 0. \tag{4.35}$$

Proof. The quadratic variation of the martingale part in (4.29) can be bounded as

$$
\begin{aligned}
E[M]_{\hat{t}}^{1/2} &\le E\Bigg[\int_0^{\hat{t}} \int_I \Big| 2\Big(X_{s-} - Y_{s-}, -\mathbb{N}\big(\hat{w}^\delta\left(X_{s-}, Y_{s-}; z\right) - \hat{w}^\delta\left(Y_{s-}, X_{s-}; z\right)\big)\Big) \\
&\quad + \left\| \mathbb{N}\big(\hat{w}^\delta\left(X_{s-}, Y_{s-}; z\right) - \hat{w}^\delta\left(Y_{s-}, X_{s-}; z\right)\big) \right\|^2 \Big| \mu_\delta\left(\mathrm{d}s \times \mathrm{d}z\right) \Bigg] \\
&\le E\Bigg[\int_0^{\hat{t}} \int_I \Big(2\|X_s - Y_s\| \left\| \mathbb{N}\big(\hat{w}^\delta\left(X_s, Y_s; z\right) - \hat{w}^\delta\left(Y_s, X_s; z\right)\big) \right\| \\
&\quad + \left\| \mathbb{N}\big(\hat{w}^\delta\left(X_s, Y_s; z\right) - \hat{w}^\delta\left(Y_s, X_s; z\right)\big) \right\|^2 \Big) W^\delta\left(X_s, Y_s\right) \mathrm{d}s \times \mathrm{d}z \Bigg] \\
&= E\Bigg[\int_0^{\hat{t}} \Big(2\|X_s - Y_s\| \big(\mathbf{1}^{\mathrm{T}} \mathbb{N}^2 \big)^{1/2} + \big(\mathbf{1}^{\mathrm{T}} \mathbb{N}^2 \big) \Big) \big| w(X_s) - w_\delta\left(Y_s\right) \big| \mathrm{d}s \Bigg] \\
&\le \int_0^{\hat{t}} C\Big(\delta + E\|X_s - Y_s\|^2 \Big) \mathrm{d}s,
\end{aligned}
$$

after using (4.28) and the integer inequality anew. For the drift part we may use the corresponding bound developed in the proof of Theorem 4.10. After taking supremum and expectation values of (4.29) and using Burkholder's inequality we therefore arrive at

$$E\|X - Y\|^2 \left(t \wedge \tau_P\right) \le \int_0^{\hat{t}} C\Big(\delta + E\|X - Y\|^2(s) \Big) \mathrm{d}s \le \int_0^t C\Big(\delta + E\|X - Y\|^2\left(s \wedge \tau_P\right) \Big) \mathrm{d}s \le \delta\big(\exp(Ct) - 1\big)$$

by Grönwall's inequality and using the notation $\|X\|(t) \equiv \sup_{s \in [0,t]} \|X_t\|$. We now rely on the same strategy as in the proof of Theorem 4.10 to similarly arrive at

$$E\|X - Y\|^2(t) \le \delta\big(\exp(Ct) - 1\big) + \Big(E\|X - Y\|^4(t) \Big)^{1/2} P^{-1/2} \Big(E\|X + Y\|_1(t) \Big)^{1/2},$$

and the conclusion follows as before.

5. Conclusions

We have proposed a theoretical framework consisting of *a priori* assumptions and estimates for problems in stochastic chemical kinetics. The assumptions are strong enough to guarantee well-posedness for a large and physically relevant class of problems. Long time estimates and limit results for perturbations in rate constants have been studied to exemplify the theory. The assumptions are *constructive* in the sense that explicit techniques for obtaining all postulated constants have either been worked out in detail or at least indicated. We have seen that the case $\beta_2 = 0$ in Assumption 4.1 (2) is particularly promising from the analysis point of view in that the conditions for existence in Theorem 4.7 and the ergodicity in Theorem 4.9 both can be formulated naturally.

In the course of motivating our setup we have seen that most problems do not admit global Lipschitz constants and that one-sided versions do not provide a better alternative. Another conclusion worth highlighting is that it pays off to consider jump SDEs in a fully discrete setting in that there are potential complications in proving uniqueness in continuous state space. A practical implication is that care should be exercised when forming continuous approximations to these types of jump SDEs.

For future work we intend to re-visit certain classical results from the perspective of the framework developed herein; for example, thermodynamic limit results, time discretization strategies, and quasi-steady state approximations—all of which have a practical impact in a range of applications.

Acknowledgements

References

[1] Lestas, I., Vinnicombe, G. and Paulsson, J. (2010) Fundamental Limits on the Suppression of Molecular Fluctuations. *Nature*, **467**, 174-178. http://dx.doi.org/10.1038/nature09333

[2] Raj, A. and van Oudenaarden, A. (2008) Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, **135**, 216-226. http://dx.doi.org/10.1016/j.cell.2008.09.050

[3] Samoilov, M.S. and Arkin, A.P. (2006) Deviant Effects in Molecular Reaction Pathways. *Nature Biotechnology*, **24**, 1235-1240. http://dx.doi.org/10.1038/nbt1253

[4] Gillespie, D.T. (1976) A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *Journal of Computational Physics*, **22**, 403-434. http://dx.doi.org/10.1016/0021-9991(76)90041-3

[5] Gillespie, D.T. (1992) A Rigorous Derivation of the Chemical Master Equation. *Physica A*, **188**, 404-425. http://dx.doi.org/10.1016/0378-4371(92)90283-V

[6] Protter, P.E. (2005) Stochastic Integration and Differential Equations. Number 21 in Stochastic Modelling and Applied Probability. 2nd Edition, Version 2.1, Springer, Berlin.

[7] Applebaum, D. (2004) Lévy Processes and Stochastic Calculus, Volume 93 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

[8] Situ, R. (2005) Theory of Stochastic Differential Equations with Jumps and Applications. Mathematical and Analytical Techniques with Applications to Engineering. Springer, New York.

[9] Øksendal, B. and Zhang, T. (2007) The Itô-Ventzell Formula and Forward Stochastic Differential Equations Driven by Poisson Random Measures. *Osaka Journal of Mathematics*, **44**, 207-230.

[10] Hausenblas, E. (2007) SPDEs Driven by Poisson Random Measure with Non Lipschitz Coefficients: Existence Results. *Probability Theory and Related Fields*, **137**, 161-200.

[11] Marinelli, C., Prévôt, C. and Röckner, M. (2010) Regular Dependence on Initial Data for Stochastic Evolution Equations with Multiplicative Poisson Noise. *Journal of Functional Analysis*, **258**, 616-649. http://dx.doi.org/10.1016/j.jfa.2009.04.015

[12] Marinelli, C. and Röckner, M. (2010) Well-Posedness and Asymptotic Behavior for Stochastic Reaction-Diffusion Equations with Multiplicative Poisson Noise. *Electronic Journal of Probability*, **15**, 1529-1555. http://dx.doi.org/10.1214/EJP.v15-818

[13] Filipović, D., Tappe, S. and Teichmann, J. (2010) Jump-Diffusions in Hilbert Spaces: Existence, Stability and Numerics. *Stochastics*, **82**, 475-520. http://dx.doi.org/10.1080/17442501003624407

[14] van Kampen, N.G. (2004) Stochastic Processes in Physics and Chemistry. 2nd Edition, Elsevier, Amsterdam.

[15] Higham, D.J. and Kloeden, P.E. (2005) Numerical Methods for Nonlinear Stochastic Differential Equations with Jumps. *Numerische Mathematik*, **101**, 101-119. http://dx.doi.org/10.1007/s00211-005-0611-8

[16] Li, T.J. (2007) Analysis of Explicit Tau-Leaping Schemes for Simulating Chemically Reacting Systems. *Multiscale Modeling & Simulation*, **6**, 417-436. http://dx.doi.org/10.1137/06066792X

[17] Engblom, S. (2009) Parallel in Time Simulation of Multiscale Stochastic Chemical Kinetics. *Multiscale Modeling & Simulation*, **8**, 46-68. http://dx.doi.org/10.1137/080733723

[18] Anderson, D.F. (2012) An Efficient Finite Difference Method for Parameter Sensitivities of Continuous Time Markov Chains. *SIAM Journal on Numerical Analysis*, **50**, 2237-2258. http://dx.doi.org/10.1137/110849079

[19] Kawamura, A. (2009) Lipschitz Continuous Ordinary Differential Equations Are Polynomial-Space Complete. 24*th Annual IEEE Conference on Computational Complexity*, Paris, 15-18 July 2009, 149-160.

[20] Hutzenthaler, M., Jentzen, A. and Kloeden, P.E. (2011) Strong and Weak Divergence in Finite Time of Euler's Method for Stochastic Differential Equations with Non-Globally Lipschitz Continuous Coefficients. *Proceedings of the Royal Society A*, **467**, 1563-1576.

[21] Chen, W.Y. and Bokka, S. (2005) Stochastic Modeling of Nonlinear Epidemiology. *Journal of Theoretical Biology*, **234**, 455-470. http://dx.doi.org/10.1016/j.jtbi.2004.11.033

[22] Ewens, W.J. (2004) Mathematical Population Genetics I. Theoretical Introduction, Volume 27 of Interdisciplinary Applied Mathematics. 2nd Edition, Springer, New York.

[23] Escudero, C., Buceta, J., de la Rubia, F.J. and Lindenberg, K. (2004) Extinction in Population Dynamics. *Physical Review E*, **69**, 021908.

[24] Ethier, S.N. and Kurtz, T.G. (1986) Markov Processes: Characterization and Convergence. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York. http://dx.doi.org/10.1002/9780470316658

[25] Brémaud, P. (1999) Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues. Number 31 in Texts in Applied Mathematics. Springer, New York. http://dx.doi.org/10.1007/978-1-4757-3124-8

[26] Plyasunov, S. (2005) On Hybrid Simulation Schemes for Stochastic Reaction Dynamics. http://arxiv.org/abs/math/0504477

[27] Brémaud, P. (1981) Point Processes and Queues: Martingale Dynamics. Springer Series in Statistics. Springer, New York.

[28] Daley, D.J. and Vere-Jones, D. (2003) An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods. 2nd Edition, Springer, New York.

[29] Kurtz, T.G. (1978) Strong Approximation Theorems for Density Dependent Markov Chains. *Stochastic Processes and Their Applications*, **6**, 223-240. http://dx.doi.org/10.1016/0304-4149(78)90020-0

[30] Rathinam, M., Sheppard, P.W. and Khammash, M. (2010) Efficient Computation of Parameter Sensitivities of Discrete Stochastic Chemical Reaction Networks. *Journal of Chemical Physics*, **132**, 034103. http://dx.doi.org/10.1063/1.3280166

[31] Kurtz, T.G. (1982) Representation and Approximation of Counting Processes. In: Fleming, W.H. and Gorostiza, L.G., Eds., *Advances in Filtering and Optimal Stochastic Control*, Vol. 42, *Lecture Notes in Control and Information Sciences*, Springer, Berlin, 177-191.

[32] Ström, T. (1975) On Logarithmic Norms. *SIAM Journal on Numerical Analysis*, **12**, 741-753. http://dx.doi.org/10.1137/0712055

[33] Briat, C., Gupta, A. and Khammash, M. (2014) A Scalable Computational Framework for Establishing Long-Term Behavior of Stochastic Reaction Networks. *PLoS Computational Biology*, **10**, e1003669.

[34] Rathinam, M. (2014) Moment Growth Bounds on Continuous Time Markov Processes on Non-Negative Integer Lattices. To appear in *Quart. Appl. Math*. http://arxiv.org/abs/1304.5169

[35] Meyn, S.P. and Tweedie, R.L. (1993) Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes. *Advances in Applied Probability*, **25**, 518-548. http://dx.doi.org/10.2307/1427522

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.