

# ON THE STABILITY OF STOCHASTIC JUMP KINETICS

STEFAN ENGBLOM\*

**Abstract.** Motivated by the lack of a suitable framework for analyzing popular stochastic models of Systems Biology, we devise conditions for existence and uniqueness of solutions to certain jump stochastic differential equations (jump 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 perturbation results can be derived in this setting such that these can readily be compared 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.

**Key words.** nonlinear stability, perturbation, continuous-time Markov chain, jump process, uncertainty, rate equation

**AMS subject classifications.** 60J27, 92C42, 60J28, 92C45

**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 [19, 26, 27] 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 [12, 13], 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. Either 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 [25, 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 [2, Chap. 6], [28, Chap. 3–5] also treats the evolution of general jump-diffusion SDEs in *continuous* state spaces.

---

\*Division of Scientific Computing, Department of Information Technology, Uppsala University, SE-751 05 Uppsala, Sweden, email: [stefane@it.uu.se](mailto:stefane@it.uu.se)

A study of the flow properties of jump SDEs is found in [23], where the setting is scalar and the state continuous. In [14] 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 [21, 22]. Numerical aspects in a similar setting are discussed in [11].

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 [17, 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, 20], time-parallel simulation techniques [7], and parameter perturbations [1, *manuscript*].

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 [18] 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 [16], that with SDEs, superlinearly growing coefficients may in fact cause the forward Euler method to diverge.

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. This is a challenge in biological systems where, for evolutionary reasons, one often finds that the operating point remains close to critical points in phase-space where the efficiency is the highest. It goes without saying that, for such dynamical systems, an analysis *by analogy* might be highly misleading.

We also like to argue that such 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.

The paper is organized as follows: we devote Section 2 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 flow 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 for the constants occurring among our assumptions are also devised. In Section 4 we put our assumptions together and prove existence and uniqueness results, as well as some basic perturbation estimates. 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 [5], Genetics [10] and Sociodynamics [8], 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 interest-

ing 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  $X(t) \in \mathbf{Z}_+^D = \{0, 1, 2, \dots\}^D$  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*,  $w_r : \mathbf{Z}_+^D \rightarrow \mathbf{R}_+$ . This is the transition probability per unit of time for moving from the state  $x$  to  $x - \mathbb{N}_r$ ;

$$\mathbf{P} [\text{Reaction } r \text{ occurs in } [0, dt] | X(t) = x] = w_r(x) dt + o(dt). \quad (2.1)$$

We specify such a reaction by the convenient notation



where  $\mathbb{N}_r \in \mathbf{Z}^D$  is the transition step and is the  $r$ th 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)–(2.2) as a stochastic version of the time-homogeneous ordinary differential equation

$$x'(t) = - \sum_{r=1}^R \mathbb{N}_r w_r(x) = -\mathbb{N}w(x) =: F(x), \quad (2.3)$$

where  $w(x) \equiv [w_1(x), \dots, w_R(x)]^T$  is the column vector of reaction propensities.

The physical premises leading to a description in the form of discrete transition laws (2.2) 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  $x = [a, b]^T$ ,



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 [9, 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 [4, 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, for all finite arguments  $x$ , we assume that  $w_r(x)$  is finite.*

To state the chemical master equation (CME), let for brevity  $p(x, t) = \mathbf{P}(X(t) = x | X(0) = x_0)$  be the probability that a certain number  $x$  of molecules is present at time  $t$  conditioned upon an initial state  $X(0) = x_0$ . The CME is then given by [17, Chap. V]

$$\frac{\partial p(x, t)}{\partial t} = \sum_{r=1}^R w_r(x + \mathbb{N}_r) p(x + \mathbb{N}_r, t) - w_r(x) p(x, t) =: \mathbb{M}^T p(x, t). \quad (2.8)$$

The convention of the transpose of the operator to the right of (2.8) is the standard mathematical formulation of *Kolmogorov's forward differential system* [4, 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*  $\mathbb{M}^T$  in the sense that  $(\mathbb{M}^T p, q) = (p, \mathbb{M}q)$  in the Euclidean inner product over the state space. An explicit representation is

$$\mathbb{M}q = \sum_{r=1}^R w_r(x) [q(x - \mathbb{N}_r) - q(x)]. \quad (2.9)$$

Under assumptions to be prescribed later on it follows that the dynamics of the expected value of some time-independent unknown function  $f$ , conditioned upon the initial state  $x_0$ , can be written

$$\begin{aligned} \frac{d}{dt} E^{x_0} [f(X_t)] &= \sum_{x \in \mathbf{Z}_+^D} \frac{\partial p(x, t)}{\partial t} f(x) = (\mathbb{M}^T p, f) = \\ &= (p, \mathbb{M}f) = \sum_{r=1}^R E^{x_0} [w_r(X_t) (f(X_t - \mathbb{N}_r) - f(X_t))]. \end{aligned} \quad (2.10)$$

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

**2.2. The sample path representation.** In the context of analyzing models in stochastic chemical kinetics, the path-wise jump SDE representation seems to have been first put to use in [24, *unpublished manuscript*], and it was later further detailed in [20]. 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 (2.2) 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  $X(t)$ ,

$$X_t = X_0 - \sum_{r=1}^R \mathbb{N}_r \pi_r(t). \quad (2.11)$$

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

$$\mathbf{P}[\pi_r(t + dt) - \pi_r(t) = 1 | \mathcal{F}_t] = w_r(X_{t-}) dt + o(dt), \quad (2.12)$$

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 a *random time change representation* [9,

Chap. 6.2], we can produce the counting process from a standard unit-rate Poisson process  $\Pi_r$ ,

$$\pi_r(t) = \Pi_r \left( \int_0^t w_r(X_{s-}) ds \right). \quad (2.13)$$

The *marked* counting measure [3, Chap. VIII]  $\mu_r(dt \times dz; \omega)$  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  $m_r(dt \times dz)$  of  $\mu_r(dt \times dz)$  is the Lebesgue measure scaled by the corresponding propensity,  $m_r(dt \times dz) = w_r(X_{t-}) dt \times dz$ . Using this formalism, (2.11) and (2.13) can be written in the jump SDE form

$$dX_t = - \int_I \mathbb{N} \boldsymbol{\mu}(dt \times dz), \quad (2.14)$$

where  $\boldsymbol{\mu} = [\mu_1, \dots, \mu_R]^T$ . 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 depends nonlinearly on the state [3, Chap. II.3].

Using that the point processes are independent and therefore have no common jump times [4, Chap. 8.1.3], we can obtain a more transparent notation in terms of a *scalar* counting measure. Define for this purpose and for any state  $x$  the cumulative intensities

$$W_r(x) = \sum_{s=1}^r w_s(x), \quad (2.15)$$

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  $\hat{w}_r : \mathbf{Z}_+^D \times I \rightarrow \{0, 1\}$  defined according to

$$\hat{w}_r(x; z) = \begin{cases} 1 & \text{if } W_{r-1}(x) < zW(x) \leq W_r(x), \\ 0 & \text{otherwise.} \end{cases} \quad (2.16)$$

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

$$\hat{F}(x; z) = -\mathbb{N} \hat{w}(x; z), \quad (2.17)$$

such that

$$F(x) = \int_I \hat{F}(x; z) W(x) dz, \quad (2.18)$$

where  $F(x)$  is defined in (2.3).

The jump SDE (2.14) can now be written in terms of a scalar counting random measure  $\mu$  through a state-dependent *thinning procedure* [6, Chap. 7.5],

$$dX_t = - \int_I \mathbb{N} \hat{w}(X_{t-}; z) \mu(dt \times dz). \quad (2.19)$$

Eq. (2.19) expresses exponentially distributed reaction times that arrive according to a point process of intensity  $m(dt \times dz) = W(X_{t-}) dt \times dz$  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 (2.16).

We will frequently decompose (2.19) into its “drift” and “jump” parts,

$$dX_t = -\mathbb{N}w(X_t) dt - \int_I \mathbb{N}\hat{w}(X_{t-}; z)(\mu - m)(dt \times dz). \quad (2.20)$$

The second term in (2.20) is driven by the compensated measure  $(\mu - m)$  and is a local martingale provided in essence that the path is absolutely integrable (see [3, Chap. VIII.1, Corollary C4] for details).

**2.2.1. Equality of measures and a comparative thinning.** When considering stability properties we will need to compare different trajectories with respect to *the same counting measure*. It is not completely clear in either (2.19) or (2.14) how this is most naturally achieved. Since equality of measure is easy to understand for a unit-rate Poisson process, the viewpoint of *local time* expressed in (2.13) provides a satisfactory answer; two processes  $X_t$  and  $Y_t$  share the same measure if and only if they are evolved using identical Poisson processes  $\Pi_r$ ,  $r = 1, \dots, R$  in (2.11) and (2.13).

To obtain a *comparative* version of (2.19) we will have to make the thinning dependent on both trajectories. This is achieved by replacing the cumulative intensities in (2.15) with

$$W_r^{(2)}(x, y) = \sum_{s=1}^r \max\{w_s(x), w_s(y)\}, \quad (2.21)$$

and use the new total intensity  $W^{(2)}(x, y) \equiv W_R^{(2)}(x, y)$  as the intensity of the counting measure  $\mu$ ;  $m(dt \times dz) = W^{(2)}(X_{t-}, Y_{t-}) dt \times dz$ . We also modify (2.16) accordingly,

$$\hat{w}_r(x, y; z) = \begin{cases} 1 & \text{if } W_{r-1}^{(2)}(x, y) < zW^{(2)}(x, y) \leq W_{r-1}^{(2)}(x, y) + w_r(x), \\ 0 & \text{otherwise.} \end{cases} \quad (2.22)$$

Conveniently, by the symmetries of the arguments of  $W_r^{(2)}$ , the correct indicator function for the path  $Y_t$  is just  $\hat{w}(y, x; z)$ .

**2.2.2. 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  $P \rightarrow \infty$  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 [28, Chap. 2.7] and [25, Chap. II.7]), we shall get around with the following simple version [2, Chap. 4.4.2]. By the properties of the semi-martingale pure jump process we have for  $\hat{t} = t \wedge \tau_P$

$$f(X_{\hat{t}}) - f(X_0) = \sum_{0 < s \leq \hat{t}} f(X_s) - f(X_{s-}) = \int_0^{\hat{t}} \int_I f(X_s) - f(X_{s-}) \mu(ds \times dz), \quad (2.23)$$

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

$$df(X_t) = \int_I f(X_{t-} - \mathbb{N}\hat{w}(X_{t-}; z)) - f(X_{t-}) \mu(dt \times dz). \quad (2.24)$$

Alternatively, decomposing (2.23) into drift- and jump parts and taking expectation values we get, since the compensated measure is a local martingale,

$$\begin{aligned}
Ef(X_{\hat{t}}) - Ef(X_0) &= E \int_0^{\hat{t}} \int_I f(X_{s-} - \mathbb{N}\hat{w}(X_{s-}; z)) - f(X_{s-}) m(ds \times dz) \\
&= E \int_0^{\hat{t}} \int_I [f(X_{s-} - \mathbb{N}\hat{w}(X_{s-}; z)) - f(X_{s-})] W(X_{s-}) ds \times dz \\
&= E \int_0^{\hat{t}} \sum_{r=1}^R [(f(X_s - \mathbb{N}_r) - f(X_s))w_r(X_s)] ds. \tag{2.25}
\end{aligned}$$

This is Dynkin's formula [4, Chap. 9.2.2] for the stopped process and we note that (2.10) is just a differential version.

**2.2.3. The validity of the master equation.** With this much formalism developed, we may conveniently quote the following result:

**THEOREM 2.2** ([4, Chap. 8.3.2, Theorem 3.3]). *Under Assumption 2.1, and if additionally, for  $t \in [0, T]$  it holds that*

$$EW(X_t) < \infty, \tag{2.26}$$

then (2.8) is valid for  $t \in [0, T]$ .

Since the governing equation (2.10) for the expected value of  $f(X_t)$  is a direct consequence of (2.8), we can similarly conclude the following:

**COROLLARY 2.3.** *Under the assumptions of Theorem 2.2, and if, moreover, in an arbitrary norm  $\|\cdot\|$ ,*

$$E\|f(X_t)\| < \infty, \quad t \in [0, T], \tag{2.27}$$

then (2.10) 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  $x_0$ .

Although from now on our main target of interest will be the path itself, in Section 4.4 we shall find some use of (2.10) to obtain realistic moment estimates for models of dissipative character.

**2.3. Concrete examples.** Consider the bi-molecular birth-death system,



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 (2.24) with  $f(x) = x_1 - x_2$  yields

$$dU_t = df(X_t) = \int_I -[-1, 1, 0] \hat{w}(X_{t-}; z) \mu(dt \times dz), \quad (2.29)$$

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

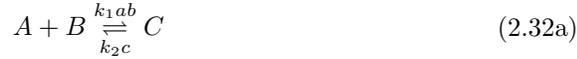
$$\emptyset \xrightleftharpoons[k_1]{k_1} U, \quad (2.30)$$

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

$$U_t = U_0 + \Pi_1(k_1 t) - \Pi_2(k_1 t) \sim \mathcal{N}(U_0, 2k_1 t), \quad \text{as } t \rightarrow \infty, \quad (2.31)$$

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  $t \rightarrow \infty$ .

**2.3.1. Reversible versions.** From time to time below we shall be concerned with the following *closed* version of (2.28), consisting of a single reversible reaction,



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



and will prove to be a useful example in the stochastic setting since formally, all states in  $\mathbf{Z}_+^3$  are reachable. For (2.32a) we have

$$\mathbb{N} \equiv \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad w(x) \equiv [k_1 a b, k_2 c]^T, \quad (2.33)$$

while (2.32b) is represented by

$$\mathbb{N} \equiv \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ -1 & 1 & 1 & -1 \end{bmatrix}, \quad w(x) \equiv [k_1 a b, k_2 c, k_3 c, k_4]^T. \quad (2.34)$$

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

**3. Deterministic stability.** In this section we shall be concerned with the deterministic drift part of the dynamics (2.20). We are interested in techniques for judging the stability of the time-homogeneous ODE (2.3), the so-called reaction rate equations implied by the rates (2.2). Stability and continuity with respect to initial data are considered in Sections 3.1 and 3.2, and techniques for explicitly obtaining all our postulated constants are discussed in Section 3.3.

Initially we will consider states  $x \in \mathbf{R}^D$ , but we will soon find it productive 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 even 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*

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

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

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

*Proof.* The constant  $A$  can be disregarded by using the substitution  $v(t) = u(t) + A/\alpha$ . For the differential version one integrates the given inequality using  $\exp(-\alpha t)$  as an integrating factor. For the integral version, define

$$(v(t) - v(0)e^{\alpha t} \leq) w(t) := v(0) + \int_0^t \alpha v(s) ds - v(0)e^{\alpha t}.$$

Under the assumed integral inequality we readily get  $w'(t) \leq \alpha w(t)$  such that  $w(t) \leq 0$  for  $t \geq 0$ .  $\square$

The most immediate way of comparing the growth of solutions to the ODE (2.3) to those of a linear ODE is to require that the norm of the driving function is bounded in terms of its argument;

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

since then by the triangle inequality,

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

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 (2.28) does not satisfy (3.3).*

*Proof.* We compute  $\|F(x)\| = \|\mathbb{N}w(x)\| = \sqrt{2}|k_1 - k_2ab|$  for a state  $x = [a, b]^T$ . Hence for  $a = b = N = 0, 1, \dots$  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 (2.28) 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 composant 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  $x$ ,

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

from which one deduces

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

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 (2.28), 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 (2.32a) nor (2.32b) admits a bound of the kind (3.5).*

*Proof.* As in the proof of Proposition 3.2 we look at a ray  $x^T = (a, b, c) = (N, N, 3N)$  parametrized by a non-negative integer  $N$ . For (2.32a) 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 (2.32b).  $\square$

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 (2.32a)  $a+b+2c$  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 were 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  $\|x_t\|_1 = (\mathbf{1}, x_t) = \mathbf{1}^T x_t$ , 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 (2.3) 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  $x$ ,

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

implying the bounded dynamics

$$\frac{d}{dt} \|x\|_1 = (\mathbf{1}, F(x)) \leq A + \alpha \|x\|_1. \quad (3.8)$$

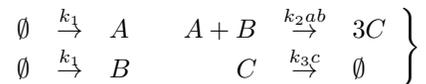
We remark in passing that the criterion (3.7) is sharp in the sense that if the reversed inequality can be shown to be true, then the growth of solutions can be estimated from below.

EXAMPLE 3.4. *As a point in favor of this approach we compute for the bi-molecular birth-death system (2.28),  $(\mathbf{1}, F(x)) = (\mathbf{1}, -\mathbb{N}w(x)) = 2k_1 - 2k_2ab$  which evidently falls under the assumption (3.7) with  $(A, \alpha) = (2k_1, 0)$ . For the reversible case (2.32a) we similarly get  $(\mathbf{1}, F(x)) = -k_1ab + k_2c$  such that (3.7) applies with  $(A, \alpha) = (0, k_2)$ . Finally, and in the same fashion, the open case (2.32b) 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  $\|x\|_{(l)} := l^T x$  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 3.5. For the reversible system (2.32a), we have already noted that  $a+b+2c$  is a conserved quantity such that the choice  $l = [1, 1, 2]^T$  yields  $d/dt\|x\|_{(l)} = 0$ . The open case (2.32b) also benefits from this weighted norm in that we get  $d/dt\|x\|_{(l)} \leq 2k_4$ .

EXAMPLE 3.6. A slightly more involved model reads as follows:



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

$$\frac{d}{dt}\|x_t\|_{(l)} = 6k_1 - 2k_3c \leq 6k_1.$$

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  $-\mathbb{N}_2^T$ , readily available by linear algebra techniques. We omit the details.

**3.2. Continuity.** For well-posedness of the ODE (2.3) 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

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

presumably with some growth restrictions on  $\lambda_R$ . Although theoretically precise, such an analysis is likely to be somewhat inconvenient when it comes to estimating actual constants in perturbation results. We shall therefore consider the following version,

$$(x - y, F(x) - F(y)) \leq (M + \mu\|x + y\|_1)\|x - y\|^2, \quad (3.10)$$

where the form of  $\lambda_R$  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 3.7. Suppose that the ODE (2.3) satisfies (3.7) and (3.10) and that initial data  $x_0 \in \mathbf{R}_+^D$  implies a trajectory  $x(t) \in \mathbf{R}_+^D$ . Then for any  $t \in [0, T]$  there is a unique 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

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

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

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

Hence the (bounded) solution to

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

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  $\alpha$  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 (2.4)–(2.7). 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 3.8 (linear case).** *Write a set of  $R$  linear propensities as  $w_r(x) = q_r^T x$ ,  $r = 1, \dots, R$ , each with the corresponding stoichiometric vector  $\mathbb{N}_r$ . Then  $F(x) := -\sum_r \mathbb{N}_r w_r(x)$  satisfies  $(x - y, F(x) - F(y)) \leq M\|x - y\|^2$  with  $M = \mu_2[-\mathbb{N}Q^T]$  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  $n$ th position, then this reduces to  $M = k(-\mathbb{N}_{r,n} + \|\mathbb{N}_r\|)/2$ .*

*Proof.* The first assertion is immediate since the smallest such constant  $M$  by definition is the logarithmic norm (see e.g. [29]). To compute  $\mu_2[-\mathbb{N}_r q_r^T]$  when  $q$  has the form indicated, we determine the extremal eigenvalue of  $-(\mathbb{N}_r q_r^T + q_r \mathbb{N}_r^T)/2$ . 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  $(\lambda, z)$  can be written as

$$\begin{aligned} -\frac{1}{2}\mathbb{N}_{r,j}z_n &= \lambda z_j, \quad j = 1, 2, \dots, D, \quad j \neq n, \\ -\frac{1}{2}\mathbb{N}_{r,n}z_n - \frac{1}{2}\mathbb{N}_r^T z &= \lambda z_n. \end{aligned}$$

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

**EXAMPLE 3.9.** *The simple special case in Proposition 3.8 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  $A \rightarrow \emptyset$  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  $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_D \rightarrow \emptyset$ , or a closed loop in which the last transition is replaced with  $A_D \rightarrow A_1$ , both admit bounds  $M \leq 0$  as an inspection of the Gershgorin-discs of  $-(\mathbb{N} + \mathbb{N}^T)/2$  shows.*

Other than for those special examples, for the most important linear cases, the following table summarizes the bounds as obtained from the special case in Proposition 3.8 (with all reaction constants normalized to unity):

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

PROPOSITION 3.10 (quadratic case). Write a general quadratic propensity as  $w_r(x) = x^T S x$  with  $S$  a symmetric matrix. Then  $F_r(x) = -\mathbb{N}_r w_r(x)$  satisfies (3.10) with  $M = 0$  and  $\mu = \|x + y\|_1^{-1} \mu_2 [-\mathbb{N}_r(x + y)^T S] \leq \|\mathbb{N}_r\| \|S\|$ . For the special case that  $S_{ij} = k(\delta_{im}\delta_{jn} + \delta_{jm}\delta_{in})/2$  with  $m \neq n$  there holds  $\mu \leq k \max_{j \in \{m,n\}} (-\mathbb{N}_{r,j} + \|\mathbb{N}_r\|)/4$ .

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

$$\begin{aligned} \mu &\leq \sup_{x,y \in \mathbf{Z}_+^D} \|x + y\|_1^{-1} \|x - y\|^{-2} (x - y, -\mathbb{N}_r(x + y)^T S (x - y)) \\ &= \sup_{w \in \mathbf{Z}_+^D} \sup_{v \in \mathbf{Z}^D} \|w\|_1^{-1} \|v\|^{-2} (v, -\mathbb{N}_r w^T S v) \leq \sup_{\substack{u \geq 0 \\ \|u\|_1 \leq 1}} \mu_2 [-\mathbb{N}_r u^T S]. \end{aligned}$$

The indicated upper bound is derived from the fact that  $|\mu_2[\cdot]| \leq \|\cdot\|$  [29]. 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 3.8,

$$\begin{aligned} \mu_2 [-\mathbb{N}_r(x + y)^T S] &= \mu_2 [-\mathbb{N}_r q^T] \leq \mu_2 [-\mathbb{N}_r q_1^T] + \mu_2 [-\mathbb{N}_r q_2^T] \\ &= (x_n + y_n)k(-\mathbb{N}_{r,m} + \|\mathbb{N}_r\|)/4 + (x_m + y_m)k(-\mathbb{N}_{r,n} + \|\mathbb{N}_r\|)/4. \end{aligned}$$

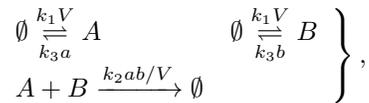
□

EXAMPLE 3.11. The most important quadratic cases are summarized in the table below. For the dimerizations in the lower half of the table there is also a linear part  $M$  in (3.10).

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

EXAMPLE 3.12. The bi-molecular birth-death model (2.28) admits the constants  $(M, \mu) = (0, k_2(\sqrt{2} - 1)/4)$  in (3.10). Similarly, the reversible cases (2.32a) and (2.32b) both obeys (3.10) with  $(M, \mu) = (\sqrt{3} - 1)/2 \times (k_2, k_1/2)$ . All these results are sharp except for the open case (2.32b) for which one can obtain a slightly smaller constant  $M$  by using the general formula stated in Proposition 3.8.

EXAMPLE 3.13. As a highly prototypical example we consider the following natural extension of the bi-molecular birth-death model (2.28),



where in this example it is informative to consider the dependence on the system's size  $V$ . It is straightforward to show the bounds  $(A, \alpha) \leq (2k_1V, -k_3)$  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 3.8 and 3.10 we get the sharp bounds  $(M, \mu) \leq (-k_1V, k_2/V \cdot (\sqrt{2} - 1)/4)$  in (3.10). It follows that for states  $\{x, y\}$  such that  $\|x + y\|_1 \lesssim 9.7k_1/k_2 \cdot V^2$ , 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 (2.19). 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 unsigned version, implying 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 (2.19) in Section 4.3. A sufficient condition for the existence of asymptotical bounds of the  $p$ th order moment is given in Section 4.4 where we also derive some basic perturbation estimates.

**4.1. Working assumptions.** We state formally the set of assumptions on the jump SDE (2.19) as follows:

ASSUMPTION 4.1. For arguments  $x, y \in \mathbf{Z}_+^D$ ,

- (i)  $(\mathbf{1}, F(x)) \leq A + \alpha\|x\|_1$ ,
- (ii)  $\|w(x) - w(y)\|_1 \leq (L + l\|x + y\|_1)\|x - y\|$ ,
- (iii)  $\|w(x)\|_1 \leq \Gamma + \gamma\|x\|_1^2$ ,
- (iv)  $(x - y, F(x) - F(y)) \leq (M + \mu\|x + y\|_1)\|x - y\|^2$ .

We assume the parameters  $\{A, L, l, \Gamma, \gamma\}$  to be positive but allow also negative values of  $\{\alpha, M, \mu\}$ .

The only unique assumptions here are (i) and (ii) since the latter clearly implies (iii)–(iv). However, as we saw in Section 3.3, in (iv) it is often possible to find sharper constants  $M$  and  $\mu$  by considering this bound in isolation. We note also that although (ii) is stronger than (iv), it is in particular valid for quadratic propensities as can be seen from the representation used in the proof of Proposition 3.10,

$$|w_r(x) - w_r(y)| = |(x + y)^T S(x - y)| \leq \|S\| \|x + y\|_1 \|x - y\|. \quad (4.1)$$

**4.1.1. The danger with cubic propensities.** Assumption 4.1 (iii) specifies the discussion to propensities with at most quadratic growth. To show that this is natural we now demonstrate that care should be taken when considering cubic propensities.

EXAMPLE 4.2. Consider the model

$$\left. \begin{array}{l} 3X \xrightarrow{x(x-1)(x-2)/2} X \\ 3X \xrightarrow{x(x-1)(x-2)} 4X \end{array} \right\},$$

such that the stoichiometric vector is given by  $\mathbb{N} = [-2, 1]$ , and hence that the drift  $-\mathbb{N}w(x) = 0$ .

**PROPOSITION 4.3.** *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 (2.10) we get the governing equation

$$\frac{d}{dt} E X_t^2 = E [3X_t(X_t - 1)(X_t - 2)],$$

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 (2.10),

$$\frac{d}{dt} E C_3(X_t) := \frac{d}{dt} E X_t(X_t - 1)(X_t - 2) = E [9C_3(X_t)(X_t - 2/3)].$$

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

$$\frac{d}{dt} E C_3(X_t) \geq 9E [C_3(X_t)^{4/3}] \geq 9[E C_3(X_t)]^{4/3}.$$

We put  $u^3 = E C_3(X_t)$  and get the differential inequality

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

which can be integrated and rearranged to produce the bound

$$E X_t(X_t - 1)(X_t - 2) \geq \frac{X_0(X_0 - 1)(X_0 - 2)}{1 - 3t X_0(X_0 - 1)(X_0 - 2)}. \quad (4.2)$$

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 (2.25) using localization. To get some guidance, let us first assume that the differential form of Dynkin's formula (2.10) 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  $\|X_t\|_1 = (\mathbf{1}, X_t)$ . Hence from (2.10) with  $f(x) = (\mathbf{1}, x)$  we get that

$$\frac{d}{dt} E \|X_t\|_1 = (\mathbf{1}, F(X_t)) \leq A + \alpha \|X_t\|_1, \quad (4.3)$$

by Assumption 4.1 (i). Clearly, the differential form of Grönwall's inequality in Lemma 3.1 applies here. A correct version of this argument unfortunately loses the sign of  $\alpha$ .

**PROPOSITION 4.4.** *If Assumption 4.1 (i) is true, then*

$$E \|X_t\|_1 \leq \|X_0\|_1 \exp(\alpha_+ t) + A(\exp(\alpha_+ t) - 1)/\alpha_+,$$

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

Here and below we shall make use of the stopping time  $\tau_P = \inf_{t \geq 0} \{\|X_t\|_1 > P\}$  and define  $\hat{t} = t \wedge \tau_P$ .

*Proof.* From (2.25) with  $f(x) = (\mathbf{1}, x)$  we get that

$$E\|X_{\hat{t}}\|_1 = \|X_0\|_1 + E \int_0^{\hat{t}} (\mathbf{1}, F(X_s)) ds \leq \|X_0\|_1 + E \int_0^t A + \alpha_+ \|X_{\hat{s}}\|_1 ds, \quad (4.4)$$

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

$$E\|Y_t\|_1 \leq \|X_0\|_1 \exp(\alpha_+ t) + A(\exp(\alpha_+ t) - 1)/\alpha_+ \quad (4.5)$$

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

We attempt a similar treatment for obtaining bounds in mean square. Assuming tactically that (2.10) is valid, writing  $\|x\|_2^2 = x^T x$  we get after some work that

$$\frac{d}{dt} E\|X_t\|_2^2 = E[\mathbf{1}^T \mathbb{N}^2 w(X_t) - 2X_t^T \mathbb{N} w(X_t)] \quad (4.6)$$

where  $\mathbb{N}_{ij}^2 \equiv (\mathbb{N}_{ij})^2$ . We expect from Grönwall's inequality that  $E\|X_t\|_2^2$  grows at most exponentially with  $\alpha t$  whenever

$$\mathbf{1}^T \mathbb{N}^2 w(x) - 2x^T \mathbb{N} w(x) \leq A + \alpha \|x\|^2. \quad (4.7)$$

However, this tentative condition is often violated in practice since the second term  $-x^T \mathbb{N} w(x) = (x, F(x))$ , 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.

Surprisingly, more realistic conditions arise when looking for bounds with respect to  $\|X_t\|_1^2$  instead.

PROPOSITION 4.5. *If for some constants  $\beta$  and  $B$ ,*

$$(\mathbf{1}^T \mathbb{N})^2 w(x) - 2\mathbf{1}^T \mathbb{N} w(x) \|x\|_1 \leq B + \beta \|x\|_1^2, \quad (4.8)$$

(with  $(\mathbf{1}^T \mathbb{N})^2$  understood elementwise), then  $E\|X_t\|_1^2 \leq \|X_0\|_1^2 \exp(\beta_+ t) + B(\exp(\beta_+ t) - 1)/\beta_+$ .

The proof of Proposition 4.5 follows the same pattern as for Proposition 4.4, but using this time  $f(x) = \|x\|_1^2 = (\mathbf{1}, x)^2$  in (2.25). The condition (4.8) is typically more realistic than (4.7) since we recognize the term  $-\mathbf{1}^T \mathbb{N} w(x) \|x\|_1 = (\mathbf{1}, F(x)) \|x\|_1$ , which under the evidently reasonable Assumption 4.1 (i) is  $\leq (A + \alpha \|x\|_1) \|x\|_1$ . It follows that if  $(\mathbf{1}^T \mathbb{N})^2 w(x)$  grows at most quadratically with  $\|x\|_1$ , then this assumption is sufficient to yield bounds in mean square. Stated formally,

PROPOSITION 4.6. *Under Assumption 4.1 (i) and (iii) the condition (4.8) of Proposition 4.5 is true with  $\beta = \|(\mathbf{1}^T \mathbb{N})^2\|_\infty \gamma + \varepsilon + 2\alpha$  and  $B = \|(\mathbf{1}^T \mathbb{N})^2\|_\infty \Gamma + \varepsilon^{-1} A^2$  for any  $\varepsilon > 0$ .*

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

$$(\mathbf{1}^T \mathbb{N})^2 w(x) - 2\mathbf{1}^T \mathbb{N} w(x) \|x\|_1 \leq \|(\mathbf{1}^T \mathbb{N})^2\|_\infty \|w(x)\|_1 + 2(A + \alpha \|x\|_1) \|x\|_1,$$

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: the condition sufficient for 1st order stability paired with quadratic propensities implies finite time stability in any order moment.

**THEOREM 4.7 (Moment estimate).** *Under Assumption 4.1 (i) and (iii), and for any integer  $p > 2$ ,*

$$E\|X_t\|_1^p \leq \|X_0\|_1^p \exp(\beta_+ t) + B(\exp(\beta_+ t) - 1)/\beta_+, \quad (4.9)$$

with  $\beta_+ = \beta \vee 0$ , and bounds  $\beta \leq (p - 1 + p\alpha) + (\Gamma + \gamma)(2^p - 2 - p)\|\mathbf{1}^T \mathbb{N}\|_\infty^p$  and  $B \leq A^p + \Gamma\|\mathbf{1}^T \mathbb{N}\|_\infty^p$ .

*Proof.* Using (2.25) with  $f(X_t) \equiv [\mathbf{1}^T X_t]^p$  we get

$$E\|X_t\|_1^p = \|X_0\|_1^p + E \int_0^t \sum_{r=1}^R w_r(X_s) \left[ [\mathbf{1}^T (X_s - \mathbb{N}_r)]^p - [\mathbf{1}^T X_s]^p \right] =: E \int_0^t G(X_s) ds.$$

Expanding and using Young's inequality with exponent  $p/(p-1)$  and conjugate exponent  $p$ , respectively, we find that

$$\begin{aligned} G(x) &= \sum_{r=1}^R w_r(x) \sum_{j=1}^p \binom{p}{j} (\mathbf{1}^T x)^{p-j} (-\mathbf{1}^T \mathbb{N}_r)^j \\ &\leq A^p + (p-1+p\alpha)\|x\|_1^p + \sum_{j=2}^p \binom{p}{j} \|x\|_1^{p-j} \|\mathbf{1}^T \mathbb{N}\|_\infty^j \|w(x)\|_1 \\ &\leq A^p + (p-1+p\alpha)\|x\|_1^p + \left(1 + (2^p - 2 - p)\|x\|_1^{p-2}\right) \|\mathbf{1}^T \mathbb{N}\|_\infty^p (\Gamma + \gamma\|x\|_1^2) \\ &\leq B + \beta\|x\|_1^p. \end{aligned}$$

We thus get

$$E\|X_t\|_1^p \leq \|X_0\|_1^p + \int_0^t B + \beta_+ E\|X_s\|_1^p ds,$$

where Grönwall's integral inequality applies. Letting  $P \rightarrow \infty$  we obtain the result.  $\square$

As demonstrated in Examples 3.5 and 3.6, the vector  $\mathbf{1}$  is not always the best choice for a system's "outward" direction. In such cases we note that the entire statement of Theorem 4.7, including the proof, holds verbatim if a suitably normalized vector  $l > 0$  is used in place of  $\mathbf{1}$  and the norm  $\|\cdot\|_1$  is replaced with the equivalent norm  $\|\cdot\|_{(l)}$ .

**4.3. Existence and uniqueness.** We shall now prove that the jump SDE (2.19) under Assumption 4.1 has a uniquely defined and locally bounded solution. To this end and following [28, Sect. 3.1.2], we introduce the following space of path-wise locally bounded processes:

$$S_{\mathcal{F}}^{2,\text{loc}}(\mathbf{Z}_+^D) = \left\{ \begin{array}{l} X(t, \omega) : X \text{ is } \mathcal{F}_t\text{-adapted and } \mathbf{Z}_+^D\text{-valued such that} \\ E \sup_{t \in [0, T]} \|X(t, \omega)\|_1^2 < \infty \text{ for } \forall T < \infty \end{array} \right\}. \quad (4.10)$$

**THEOREM 4.8 (Existence).** *Let  $X_t$  be a solution to (2.19) under Assumption 4.1 (i) and (iii). Then if  $\|X_0\|^2 < \infty$ ,  $\{X_t\}_{t \geq 0} \in S_{\mathcal{F}}^{2,\text{loc}}(\mathbf{Z}_+^D)$ .*

*Proof.* As before we use the stopping time  $\tau_P = \inf_{t \geq 0} \{\|X_t\|_1 > P\}$  and put  $\hat{t} = t \wedge \tau_P$ . We get from Itô's formula

$$\begin{aligned} \|X_{\hat{t}}\|_1^2 &= \|X_0\|_1^2 + \int_0^{\hat{t}} 2(\mathbf{1}, F(X_s))\|X_s\|_1 + (-\mathbf{1}^T \mathbb{N})^2 w(X_s) ds + \\ &\quad \int_0^{\hat{t}} \int_I 2(\mathbf{1}, \hat{F}(X_{s-}; z))\|X_{s-}\|_1 + (\mathbf{1}, \hat{F}(X_{s-}; z))^2 (\mu - m)(ds \times dz). \end{aligned}$$

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,

$$\begin{aligned} E[M_{\hat{t}}]^{1/2} &= E \left[ \int_0^{\hat{t}} \int_I \left[ 2(\mathbf{1}, \hat{F}(X_{s-}; z))\|X_{s-}\|_1 + (\mathbf{1}, \hat{F}(X_{s-}; z))^2 \right]^2 \mu(ds \times dz) \right]^{1/2} \\ &= E \left[ \int_0^{\hat{t}} \int_I \left[ 2(-\mathbf{1}^T \mathbb{N})\|X_{s-}\|_1 + (-\mathbf{1}^T \mathbb{N})^2 \right]^2 \hat{w}(X_{s-}; z) W(X_{s-}) dz ds \right]^{1/2} \\ &\leq E \int_0^{\hat{t}} |2(-\mathbf{1}^T \mathbb{N})\|X_s\|_1 + (-\mathbf{1}^T \mathbb{N})^2| w(X_s)^{1/2} ds \\ &\leq E \int_0^{\hat{t}} C_1 + C_2 \|X_s\|_1^2 ds \end{aligned}$$

for constants  $C_1, C_2 \geq 0$  depending on the network topology  $\mathbb{N}$  and on the constants  $\Gamma$  and  $\gamma$  in Assumption 4.1 (iii). For the drift part we have already constructed a suitable bound in Proposition 4.6 such that

$$\|X_{\hat{t}}\|_1^2 \leq \|X_0\|_1^2 + \int_0^{\hat{t}} B + \beta \|X_s\|_1^2 ds + |M_{\hat{t}}|.$$

Taking supremum and expectation values we get from Burkholder's inequality [25, Chap. IV.4] with some constants  $C_3, C_4 \geq 0$  that

$$E \sup_{s \in [0, \hat{t}]} \|X_s\|_1^2 \leq \|X_0\|_1^2 + \int_0^{\hat{t}} C_3 + C_4 E \sup_{s' \in [0, s]} \|X_{s'}\|_1^2 ds.$$

Writing  $\|X\|_t^2 \equiv \sup_{s \in [0, t]} \|X_s\|_1^2$  we conclude that

$$E \|X\|_{t \wedge \tau_P}^2 \leq \|X_0\|_1^2 + \int_0^t C_3 + C_4 E \|X\|_{s \wedge \tau_P}^2 ds.$$

By Grönwall's integral inequality we have thus shown that  $E \|X\|_{t \wedge \tau_P}^2$  can be bounded in terms of the initial data and time  $t$ . The result now follows by letting  $P \rightarrow \infty$  and using Fatou's lemma.  $\square$

We note in passing that Theorem 4.8 can readily be generalized to solutions in the space  $S_{\mathcal{F}}^{p, \text{loc}}(\mathbf{Z}_+^D)$  for any  $p \geq 1$ .

**THEOREM 4.9 (Uniqueness).** *Let Assumption 4.1 (i)–(iv) hold true. Then the solution to (2.19) is path-wise unique.*

We shall be using the observation that, for  $x \in \mathbf{Z}_+^D$ , we have that  $\|x\|_1 \leq \|x\|_2^2$  (referred below to as the “integer inequality”).

*Proof.* Under the same stopping time as before we get from Itô's formula after discarding the martingale part

$$E\|X_{\hat{t}} - Y_{\hat{t}}\|^2 = E \int_0^{\hat{t}} 2(X_s - Y_s, F(X_s) - F(Y_s)) + (\mathbf{1}^T \mathbb{N}^2) |w(X_s) - w(Y_s)| ds \quad (4.11)$$

$$\leq E \int_0^{\hat{t}} 2(M + \mu \|X_s + Y_s\|_1) \|X_s - Y_s\|^2 + \|(\mathbf{1}^T \mathbb{N}^2)\|_{\infty} (L + l \|X_s + Y_s\|_1) \|X_s - Y_s\| ds. \quad (4.12)$$

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

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

Using that  $X_0 = Y_0$  and Grönwall's integral inequality we conclude that the only solution is the zero solution. Letting  $P \rightarrow \infty$  and using Fatou's lemma the statement is therefore proved.  $\square$

**4.4. Basic perturbation estimates.** We next aim at deriving some stability estimates with respect to perturbations in initial data and of the coefficients. To this effect we shall need a slightly more general Grönwall-type estimate than before.

LEMMA 4.10. *Let  $u$  and  $\alpha$  be non-negative functions in some positive interval including zero. Suppose that  $d/dt u(t)^2 \leq \alpha(t)u + \beta u^2$  for  $t \geq 0$ . Then*

$$u(t) \leq e^{\beta/2t} \left( u(0) + \frac{1}{2} \int_0^t \alpha(s) e^{-\beta/2s} ds \right). \quad (4.13)$$

*The same conclusion holds with  $\beta \geq 0$  and under the weaker integral condition*

$$u(t)^2 \leq u(0)^2 + \int_0^t \alpha(s)u(s) + \beta u(s)^2 ds. \quad (4.14)$$

*Proof.* Performing the differentiation and dividing through by  $u$  we get the implied inequality  $u' \leq \alpha(t)/2 + \beta/2 u$ . This inequality has an integrating factor  $\exp(-\beta/2t)$  and is therefore readily solved. To prove the second part, define

$$w(t) := \left( u(0)^2 + \int_0^t \alpha(s)u(s) + \beta u(s)^2 ds \right)^{1/2} - e^{\beta/2t} \left( u(0) + \frac{1}{2} \int_0^t \alpha(s) e^{-\beta/2s} ds \right).$$

We get under the assumptions that  $w'(t) \leq \beta/2 w(t)$  such that  $w(t) \leq 0$ .  $\square$

Note the special case of the lemma  $\alpha(t) = 1$ ,  $\beta = 0$ , and  $u(0) = 0$ , implying the (sharp) bound  $u \leq t/2$ . This observation shows that the integer inequality as used in the proof of Theorem 4.9 is crucial for path-wise uniqueness; without it the integral inequality (4.12) admits growing solutions.

Although Theorem 4.7 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  $S_{\mathcal{F}}^{2,\text{loc}}(\mathbf{Z}_+^D)$  and the assumption of at most quadratic growth. Under these conditions the differential form (2.10) may be used (cf. Corollary 2.3) such that the differential Grönwall inequality applies.

**THEOREM 4.11.** *Under Assumption 4.1 (i) and (iii), suppose that for an integer  $p \geq 1$ ,*

$$2\alpha + \gamma \|(\mathbf{1}^T \mathbb{N})^2\|_{\infty} (p-1) \leq -\kappa_p < 0. \quad (4.15)$$

Then  $E\|X_t\|_1^p$  is asymptotically bounded.

*Proof.* We omit the case  $p = 1$  since it follows from (4.3) under the present assumptions. The idea of the proof is to asymptotically bound  $E(C + \|X_t\|_1)^p$  with a certain positive constant  $C = C(p)$  to be decided upon below. By (2.10) we get with  $Z_t = C + \|X_t\|_1$ ,

$$\begin{aligned} \frac{dEZ_t^p}{dt} &= E \sum_{r=1}^R w_r(X_t) \left[ [Z_t - \mathbf{1}^T \mathbb{N}_r]^p - Z_t^p \right] \\ &= E \sum_{r=1}^R w_r(X_t) \left[ -\mathbf{1}^T \mathbb{N}_r p Z_t^{p-1} + \frac{(-\mathbf{1}^T \mathbb{N}_r)^2}{2} p(p-1) (Z_t - \theta_r \mathbf{1}^T \mathbb{N}_r)^{p-2} \right], \end{aligned}$$

where Taylor's formula was used and where  $\theta_r \in [0, 1]$  are constants. Using the assumptions we get the bound

$$\begin{aligned} \frac{dEZ_t^p}{dt} &\leq \frac{p}{2} E \left[ 2(A + \alpha \|X_t\|_1) Z_t^{p-1} \right. \\ &\quad \left. + \|(\mathbf{1}^T \mathbb{N})^2\|_{\infty} (p-1) (\Gamma + \gamma \|X_t\|_1^2) (Z_t + \|\mathbf{1}^T \mathbb{N}\|_{\infty})^{p-2} \right]. \quad (4.16) \end{aligned}$$

For the first term in (4.16) we get from the scaled Young's inequality with exponent  $p/(p-1)$  and conjugate exponent  $p$  that

$$\begin{aligned} 2(A + \alpha \|X_t\|_1) Z_t^{p-1} &= 2\alpha Z_t^p + 2(A - \alpha C) Z_t^{p-1} \\ &\leq 2\alpha Z_t^p + \frac{2\epsilon(p-1)}{p} Z_t^p + \frac{2\epsilon^{1-p}}{p} |A - \alpha C|^p, \end{aligned}$$

for some  $\epsilon > 0$ . As for the second term in (4.16) we first estimate

$$(\Gamma + \gamma \|X_t\|_1^2) \leq \gamma \left( (\Gamma/\gamma)^{1/2} + \|X_t\|_1 \right)^2.$$

Next by the arithmetic-geometric mean inequality we get

$$\begin{aligned} &\left( (\Gamma/\gamma)^{1/2} + \|X_t\|_1 \right)^2 (Z_t + \|\mathbf{1}^T \mathbb{N}\|_{\infty})^{p-2} \\ &\leq \left( \frac{2}{p} \left( \frac{\Gamma}{\gamma} \right)^{1/2} + \frac{p-2}{p} (\|\mathbf{1}^T \mathbb{N}\|_{\infty} + C) + \|X_t\|_1 \right)^p = Z_t^p, \end{aligned}$$

provided that we choose  $C$  as the solution to the equation

$$\frac{2}{p} \left( \frac{\Gamma}{\gamma} \right)^{1/2} + \frac{p-2}{p} (\|\mathbf{1}^T \mathbb{N}\|_{\infty} + C) = C.$$

Taken together we thus have

$$\frac{dEZ_t^p}{dt} \leq \left( -\kappa_p + \frac{2\epsilon(p-1)}{p} \right) EZ_t^p + \text{constant},$$

where the constant depends on  $\epsilon$ . Since  $\kappa_p > 0$  we may pick a small enough  $\epsilon$  such that the bracketed expression remains negative. By Grönwall's inequality this then proves the result.  $\square$

For technical reasons, two more *a priori* moment estimates are required as complements to Theorems 4.7 and 4.11.

LEMMA 4.12. *Let Assumption 4.1 (i) and (iii) hold true. For  $p \geq 1$ , put*

$$M_1^p(X_0) := E[\|X_t\|_1 - \|X_0\|_1]^p. \quad (4.17)$$

and

$$M_2^p(X_0) := E\|X_t - X_0\|^p. \quad (4.18)$$

Then both  $M_1^p$  and  $M_2^p$  are  $\mathcal{O}(t)$ .

*Proof.* From Theorem 4.7 ( $p > 2$ ) and Propositions 4.4 and 4.5 ( $p = 1$  and 2, respectively), we have the estimate

$$E\|X_t\|_1^p \leq \|X_0\|_1^p + (\beta_p \|X_0\|_1^p + B_p) \frac{e^{\beta_p t} - 1}{\beta_p}, \quad (4.19)$$

with explicit expressions for the constants  $\{\beta_p, B_p\}$  and with the possibility that  $\beta_p < 0$  whenever the condition in Theorem 4.11 applies. We note a technical point here: *this estimate is one-sided in  $\|X_t\|_1^p - \|X_0\|_1^p$  as it relies on the signed Assumption 4.1 (i)*. To get a two-sided version we proceed by integrating (2.10),

$$\begin{aligned} M_1^p(X_0) &= E \int_0^t \sum_{r=1}^R w_r(X_s) \left[ [\|X_s - \mathbb{N}_r\|_1 - \|X_0\|_1]^p - [\|X_s\|_1 - \|X_0\|_1]^p \right] ds \\ &\leq \int_0^t C_p + c_p E\|X_s\|^{p+1} ds = \mathcal{O}(t), \end{aligned}$$

by the mean-value theorem, Assumption 4.1 (iii), and various simple estimates which we omit for brevity. Using the equivalence of norms, a very similar proof is readily obtained for  $M_2^p$ .  $\square$

THEOREM 4.13 (*Initial data perturbation*). *Let Assumption 4.1 (i)–(iv) hold true and let two solutions of (2.19)  $X_t$  and  $Y_t$  be given, both generated by the same counting measure but with different initial data. Put  $\Sigma_0 := \|X_0 + Y_0\|_1$ . Then*

$$\begin{aligned} (E\|X_t - Y_t\|^2)^{1/2} &\leq \exp\left((M + \mu\Sigma_0)t\right) \times \left( \|X_0 - Y_0\| + \frac{[X_0 \neq Y_0]}{2} R \right) \\ &= \|X_0 - Y_0\| \left[ 1 + (M + \mu\Sigma_0)t + \mathcal{O}(t^2) \right] \\ &\quad + \frac{[X_0 \neq Y_0]}{2} \left[ (L' + l'\Sigma_0)t + \mathcal{O}(t^{3/2}) \right], \end{aligned} \quad (4.20)$$

in terms of

$$R = \int_0^t \left[ L' + l'\Sigma_0 + \mathcal{O}(s^{1/2}) \right] e^{-(M + \mu\Sigma_0)s} ds, \quad (4.21)$$

and where  $L' \equiv \|\mathbf{1}^T \mathbb{N}^2\|_\infty L$  and similarly for  $l'$ , and where the indicator function  $[f] \equiv 1$  iff  $f$  is true and 0 otherwise.

*Proof.* Our starting point is the following differential version of (4.12),

$$\begin{aligned} \frac{d}{dt} E \|X_t - Y_t\|^2 &\leq E \left[ 2(M + \mu \Sigma_0) \|X_t - Y_t\|^2 \right. \\ &\quad \left. + [(L' + l' \Sigma_0) + \mu R_1 + l' R_2] \|X_t - Y_t\| \right], \end{aligned} \quad (4.22)$$

in terms of

$$\begin{aligned} R_1 &= [\|X_t + Y_t\|_1 - \|X_0 + Y_0\|_1] \|X_t - Y_t\|, \\ R_2 &= \|X_t + Y_t\|_1 - \|X_0 + Y_0\|_1. \end{aligned}$$

Using the properties of the 1-norm over  $\mathbf{Z}_+^D$  we readily get

$$ER_2^2 = E (\|X_t\|_1 - \|X_0\|_1 + \|Y_t\|_1 - \|Y_0\|_1)^2 \leq \underbrace{2M_1^2(X_0) + 2M_1^2(Y_0)}_{=: R_3} = \mathcal{O}(t),$$

by Lemma 4.12. The term  $R_1$  needs slightly more work. Subtracting and using the previous bound we get

$$\begin{aligned} ER_1^2 &\leq R_3 \|X_0 - Y_0\|^2 + \left( E (2[\|X_t\|_1 - \|X_0\|_1]^2 + 2[\|Y_t\|_1 - \|Y_0\|_1]^2) \right)^{1/2} \\ &\quad \times (E[\|X_t - Y_t\|^2 - \|X_0 - Y_0\|^2])^{1/2} \end{aligned} \quad (4.23)$$

by the Cauchy-Schwartz inequality. For brevity, put  $\bar{X}_t := X_t - X_0$  and  $\bar{Y}_t := Y_t - Y_0$ . By the properties of the inner product we have that

$$\begin{aligned} \|\|X_t - Y_t\|^2 - \|X_0 - Y_0\|^2\| &= \|\|\bar{X}_t - \bar{Y}_t\|^2 + 2(\bar{X}_t - \bar{Y}_t, X_0 - Y_0)\| \\ &\leq \|\bar{X}_t - \bar{Y}_t\|^2 + 2\|\bar{X}_t - \bar{Y}_t\| \|X_0 - Y_0\| \\ &\leq 2(1 + \|X_0 - Y_0\|) \|\bar{X}_t\|^2 + 2(1 + \|X_0 - Y_0\|) \|\bar{Y}_t\|^2, \end{aligned}$$

using the integer inequality. Inserting this into (4.23) and simplifying a bit we get

$$\begin{aligned} ER_1^2 &\leq R_3 \|X_0 - Y_0\|^2 + 8 (E[\|X_t\|_1 - \|X_0\|_1]^4 + E[\|Y_t\|_1 - \|Y_0\|_1]^4)^{1/2} \\ &\quad \times (1 + \|X_0 - Y_0\|) (E\|\bar{X}_t\|^4 + E\|\bar{Y}_t\|^4)^{1/2} \\ &= R_3 \|X_0 - Y_0\|^2 + 8(1 + \|X_0 - Y_0\|) [M_1^4(X_0) + M_1^4(Y_0)]^{1/2} \\ &\quad \times [M_2^4(X_0) + M_2^4(Y_0)]^{1/2} \leq (1 + \|X_0 - Y_0\|^2) \mathcal{O}(t), \end{aligned}$$

where Lemma 4.12 was used. Starting from (4.22), using the Cauchy-Schwartz inequality several times and combining the bounds for  $R_1$  and  $R_2$  we infer that

$$\begin{aligned} \frac{d}{dt} E \|X_t - Y_t\|^2 &\leq 2(M + \mu \Sigma_0) E \|X_t - Y_t\|^2 \\ &\quad + [X_0 \neq Y_0] \left[ L' + l' \Sigma_0 + (\mu + l') \mathcal{O}(t^{1/2}) \right] (E \|X_t - Y_t\|^2)^{1/2}. \end{aligned}$$

For the lower order terms we have used the observation that by path-wise uniqueness, we may introduce the indicator function according to  $E \|X_t - Y_t\| = [X_0 \neq Y_0] E \|X_t - Y_t\|$

$Y_t\| \leq [X_0 \neq Y_0](E\|X_t - Y_t\|^2)^{1/2}$ . Without this variant of the integer inequality there would be a growing deviation between the two trajectories even with identical initial data. The theorem now follows as an application of Lemma 4.10 with  $u(t)^2 = E\|X_t - Y_t\|^2$ .  $\square$

We next consider perturbations in the *reaction coefficients*. Given the linear dependence on the coefficients  $k_r$ ,  $r = 1 \dots 4$  in the elementary reactions (2.4)–(2.7) a suitable model seems to be that a perturbation  $k_r \rightarrow k_r + \delta_r$  in a propensity  $w_r(x)$  spreads linearly in a relative sense,

$$|w_r(x, k_r) - w_r(x, k_r + \delta_r)| \leq \text{constant} \times \delta_r w_r(x, k_r). \quad (4.24)$$

We make this formal by simply requiring that

$$\|w(x) - w_\delta(x)\|_1 \leq \delta \|w(x)\|_1, \quad (4.25)$$

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

$$w_\delta(x) \equiv [w_1(x, k_1 + \delta_1), \dots, w_R(x, k_R + \delta_R)]^T. \quad (4.26)$$

We 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 (ii) and (iii) we have the bound

$$\begin{aligned} \|w(x) - w_\delta(y)\|_1 &\leq \|w(x) - w(y)\|_1 + \|w(y) - w_\delta(y)\|_1 \\ &\leq (L + l\|x + y\|_1)\|x - y\| + \delta(\Gamma + \gamma\|y\|_1^2). \end{aligned} \quad (4.27)$$

Also, induced from (4.25), it will be convenient to define  $\delta_F$  to be the smallest constant such that

$$\|F(x) - F_\delta(x)\| = \|\mathbb{N}[w(x) - w_\delta(x)]\| \leq \delta_F \|w(x)\|_1. \quad (4.28)$$

We shall need a slight extension of Lemma 4.10 as follows. Consider the differential inequality  $d/dt u(t)^2 \leq \alpha(t)u + \beta u^2 + \gamma(t)$  with  $\alpha(t)$  and  $\gamma(t)$  non-negative functions, but again  $\beta$  a signed constant. Using the integrating factor we obtain

$$\left(e^{-\beta/2t}u(t)\right)^2 \leq u(0)^2 + \int_0^t e^{-\beta s}\gamma(s) ds + \int_0^t e^{-\beta/2s}\alpha(s) \left(e^{-\beta/2s}u(s)\right) ds, \quad (4.29)$$

written such that the integral version of Lemma 4.10 applies. The final result is

$$u(t) \leq e^{\beta/2t} \left[ \left( u(0)^2 + \int_0^t e^{-\beta s}\gamma(s) ds \right)^{1/2} + \frac{1}{2} \int_0^t e^{-\beta/2s}\alpha(s) ds \right]. \quad (4.30)$$

Since we wish to keep the analysis sharp by allowing for negative values of  $\beta$  we also introduce a convenient notation to efficiently keep track of estimates involving exponentials. We define  $\exp_+(\alpha t) = \exp(0 \vee \alpha t)$  as a majorant to  $\exp(\alpha t)$ . For  $t \geq 0$  and  $f$  positive and non-decreasing we get the convenient rule

$$\int_0^t f(s) \exp_+(\alpha s) ds \leq f(t) t \exp_+(\alpha t). \quad (4.31)$$

This notation behaves intuitively in most cases, a notable exception being when mixing it with ordinary exponentials, e.g.,  $\exp(\alpha t) \exp_+(-\alpha t) = \exp_+(\alpha t)$ . Another example is that (4.30) takes the compact form

$$u(t) \leq \exp_+(\beta/2t) \left[ \gamma(t)^{1/2} t^{1/2} + \frac{\alpha(t)}{2} t \right], \quad (4.32)$$

provided that  $\gamma$  and  $\alpha$  are non-decreasing and that  $u(0) = 0$ . This form is used in the proof below.

**THEOREM 4.14** (*Coefficient perturbation*). *Let two solutions of (2.19)  $X_t$  and  $Y_t$  be given, both generated by the same counting measure and with the same initial data, but with the propensities for  $Y_t$  perturbed by  $\delta$  as indicated in (4.25)–(4.26). Put  $W_0 := \Gamma + \gamma \|X_0\|_1^2$ . Then, firstly, there is the “small perturbation”, or “short time” estimate*

$$\begin{aligned} (E\|X_t - Y_t\|^2)^{1/2} &\leq \exp_+ \left( (M + L'/2 + (2\mu + l') \|X_0\|_1) t \right) \\ &\quad \times \left[ (\delta' W_0)^{1/2} t^{1/2} + \delta_F W_0 t + \mathcal{O}(t^{3/2}) \right]. \end{aligned} \quad (4.33)$$

When  $t$  is large a generally sharper estimate is

$$\begin{aligned} (E\|X_t - Y_t\|^2)^{1/2} &\leq \exp_+ \left( (M + 2\mu \|X_0\|_1) t \right) \\ &\quad \times \left[ (\delta' W_0)^{1/2} t^{1/2} + (\delta_F W_0 + L'/2 + l' \|X_0\|_1) t + \mathcal{O}(t^{3/2}) \right], \end{aligned} \quad (4.34)$$

where  $L'$  and  $l'$  are defined in Theorem 4.13, and similarly,  $\delta' = \|\mathbf{1}^T \mathbb{N}^2\|_\infty \delta$ .

*Proof.* We start from a differential version of (4.11), but taking the perturbation of the coefficients into account,

$$\begin{aligned} \frac{d}{dt} E\|X_t - Y_t\|^2 &= E \left[ 2(X_t - Y_t, F(X_t) - F_\delta(Y_t)) + (\mathbf{1}^T \mathbb{N}^2) |w(X_t) - w_\delta(Y_t)| \right] \\ &\leq E \left[ 2(M + \mu \|X_t + Y_t\|_1) \|X_t - Y_t\|^2 + 2\delta_F \|X_t - Y_t\| (\Gamma + \gamma \|Y_t\|_1^2) \right. \\ &\quad \left. + (L' + l' \|X_t + Y_t\|_1) \|X_t - Y_t\| + \delta' (\Gamma + \gamma \|Y_t\|_1^2) \right], \end{aligned}$$

after using the Cauchy-Schwartz inequality and the bounds (4.27) and (4.28), respectively. In the proof of Theorem 4.13 we already have constructed bounds for two of the terms in the above expression. For the other two terms we use firstly from Propositions 4.5 and 4.6 that

$$E(\Gamma + \gamma \|Y_t\|_1^2) \leq W_0 + \mathcal{O}(t).$$

Secondly, writing

$$E(\|Y_t\|_1^2 - \|X_0\|_1^2)^2 = E(\|Y_t\|_1^4 - \|X_0\|_1^4) - 2\|X_0\|_1 E(\|Y_t\|_1^2 - \|X_0\|_1^2) = \mathcal{O}(t),$$

by Lemma 4.12. Using this and the Cauchy-Schwartz inequality we see that

$$E \left[ \|X_t - Y_t\| (\Gamma + \gamma \|Y_t\|_1^2) \right] \leq \left[ W_0 + \mathcal{O}(t^{1/2}) \right] (E\|X_t - Y_t\|^2)^{1/2}.$$

The final result is

$$\begin{aligned} \frac{d}{dt} E \|X_t - Y_t\|^2 &\leq \delta' W_0 + \delta' \mathcal{O}(t) + 2(M + 2\mu \|X_0\|_1) E \|X_t - Y_t\|^2 \\ &+ \left[ L' + 2l' \|X_0\|_1 + 2\delta_F W_0 + (\mu + l' + \delta_F) \mathcal{O}(t^{1/2}) \right] (E \|X_t - Y_t\|^2)^{1/2}. \end{aligned}$$

The asymptotically sharper estimate (4.34) now follows as an application of the extended Grönwall's inequality as expressed in (4.32). The short time estimate (4.33) is a result of using the integer inequality

$$(L' + 2l' \|X_0\|_1) E \|X_s - Y_s\| \leq (L' + 2l' \|X_0\|_1) E \|X_s - Y_s\|^2,$$

rather than the Cauchy-Schwartz inequality, and then applying (4.32). This has the effect of transferring all terms which do not go to zero with the perturbation  $\delta$  into the exponent and is therefore sharper when  $\delta$  and/or  $t$  are small.  $\square$

**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. Explicit bounds for the effect of perturbations in initial data and in rate constants have been obtained. The assumptions are also *constructive* in the sense that explicit techniques for obtaining all postulated constants have either been worked out in detail or at least indicated.

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, under the relevant assumptions, there are potential complications in proving path-wise uniqueness when the state space is continuous. A practical implication is that some care should be exercised when forming continuous approximations to jump SDEs of the types considered here.

It should be clear that we have not considered all possible perturbation results. For instance, bounds of the form  $E \sup_{s \leq t} \|\cdot\|$  were restricted to mere existence in the 2nd order moment. Also, ergodicity has not been discussed at all although we were able to show under certain conditions that the systems admit asymptotically bounded moments.

For future work we like 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.

**Acknowledgment.** The author likes to express his sincere gratitude to Takis Konstantopoulos for several fruitful and clarifying discussions. Early inputs on this work were also gratefully obtained from Henrik Hult, Ingemar Kaj, and Per Lötstedt.

This work was supported by the Swedish Research Council within the UPMARC Linnaeus center of Excellence.

#### References.

- [1] D. F. ANDERSON, *An efficient finite difference method for parameter sensitivities of continuous time Markov chains*, 2011. Available at <http://arxiv.org/abs/1109.2890>.
- [2] D. APPLEBAUM, *Lévy Processes and Stochastic Calculus*, vol. 93 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2004.

- [3] P. BRÉMAUD, *Point Processes and Queues: Martingale Dynamics*, Springer Series in Statistics, Springer, New York, 1981.
- [4] ———, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, no. 31 in Texts in Applied Mathematics, Springer, New York, 1999.
- [5] W-Y. CHEN AND S. BOKKA, *Stochastic modeling of nonlinear epidemiology*, J. Theoret. Biol., 234 (2005), pp. 455–470. doi:10.1016/j.jtbi.2004.11.033.
- [6] D. J. DALEY AND D. VERE-JONES, *An Introduction to the Theory of Point Processes*, vol. I: Elementary Theory and Methods, Springer, New York, 2nd ed., 2003.
- [7] S. ENGBLOM, *Parallel in time simulation of multiscale stochastic chemical kinetics*, Multiscale Model. Simul., 8 (2009), pp. 46–68. doi:10.1137/080733723.
- [8] C. ESCUDERO, J. BUCETA, F. J. DE LA RUBIA, AND K. LINDENBERG, *Extinction in population dynamics*, Phys. Rev. E, 69 (2004). doi:10.1103/PhysRevE.69.021908.
- [9] S. N. ETHIER AND T. G. KURTZ, *Markov Processes: Characterization and Convergence*, Wiley series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1986.
- [10] W. J. EWENS, *Mathematical Population Genetics I. Theoretical Introduction*, vol. 27 of Interdisciplinary Applied Mathematics, Springer, New York, 2nd ed., 2004.
- [11] D. FILIPOVIĆ, S. TAPPE, AND J. TEICHMANN, *Jump-diffusions in Hilbert spaces: Existence, stability and numerics*, Stochastics, 82 (2010), pp. 475–520. doi:10.1080/17442501003624407.
- [12] D. T. GILLESPIE, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), pp. 403–434. doi:10.1016/0021-9991(76)90041-3.
- [13] ———, *A rigorous derivation of the chemical master equation*, Phys. A, 188 (1992), pp. 404–425. doi:10.1016/0378-4371(92)90283-V.
- [14] E. HAUSENBLAS, *SPDEs driven by Poisson random measure with non Lipschitz coefficients: existence results*, Probab. Theory Related Fields, 137 (2007), pp. 161–200. doi:10.1007/s00440-006-0501-8.
- [15] D. J. HIGHAM AND P. E. KLOEDEN, *Numerical methods for nonlinear stochastic differential equations with jumps*, Numer. Math., 101 (2005), pp. 101–119. doi:10.1007/s00211-005-0611-8.
- [16] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Strong and weak divergence in finite time of Euler’s method for stochastic differential equations with non-globally Lipschitz continuous coefficients*, Proc. R. Soc., 467 (2011), pp. 1563–1576. doi:10.1098/rspa.2010.0348.
- [17] N. G. VAN KAMPEN, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 2nd ed., 2004.
- [18] A. KAWAMURA, *Lipschitz continuous ordinary differential equations are polynomial-space complete*, in 24th Annual IEEE Conference on Computational Complexity, IEEE, 2009, pp. 149–160. doi:10.1109/CCC.2009.34.
- [19] I. LESTAS, G. VINNICOMBE, AND J. PAULSSON, *Fundamental limits on the suppression of molecular fluctuations*, Nature, 467 (2010), pp. 174–178. doi:10.1038/nature09333.
- [20] T. LI, *Analysis of explicit tau-leaping schemes for simulating chemically reacting systems*, Multiscale Model. Simul., 6 (2007), pp. 417–436. doi:10.1137/06066792X.

- [21] C. MARINELLI, C. PRÉVÔT, AND M. RÖCKNER, *Regular dependence on initial data for stochastic evolution equations with multiplicative Poisson noise*, J. Funct. Anal., 258 (2010), pp. 616–649. doi:10.1016/j.jfa.2009.04.015.
- [22] C. MARINELLI AND M. RÖCKNER, *Well-posedness and asymptotic behavior for stochastic reaction-diffusion equations with multiplicative Poisson noise*, Electron. J. Prob., 15 (2010), pp. 1529–1555.
- [23] B. ØKSENDAL AND T. ZHANG, *The Itô-Ventzell formula and forward stochastic differential equations driven by Poisson random measures*, Osaka J. Math., 44 (2007), pp. 207–230.
- [24] S. PLYASUNOV, *On hybrid simulation schemes for stochastic reaction dynamics*, 2005. Available at <http://arxiv.org/abs/math/0504477>.
- [25] P. E. PROTTER, *Stochastic Integration and Differential Equations*, no. 21 in Stochastic Modelling and Applied Probability, Springer, Berlin, 2nd ed., 2005. Version 2.1.
- [26] A. RAJ AND A. VAN OUDENAARDEN, *Nature, nurture, or chance: Stochastic gene expression and its consequences*, Cell, 135 (2008), pp. 216–226. doi:10.1016/j.cell.2008.09.050.
- [27] M. S. SAMOILOV AND A. P. ARKIN, *Deviant effects in molecular reaction pathways*, Nature Biotech., 24 (2006), pp. 1235–1240. doi:10.1038/nbt1253.
- [28] R. SITU, *Theory of Stochastic Differential Equations with Jumps and Applications*, Mathematical and Analytical Techniques with Applications to Engineering, Springer, New York, 2005.
- [29] T. STRÖM, *On logarithmic norms*, SIAM J. Numer. Anal., 12 (1975), pp. 741–753.