VARIANCE OF FINITE DIFFERENCE METHODS FOR REACTION NETWORKS WITH NON-LIPSCHITZ RATE FUNCTIONS*

DAVID F. ANDERSON† AND CHAOJIE YUAN‡

Abstract. Parametric sensitivity analysis is a critical component in the study of mathematical models of physical systems. Due to its simplicity, finite difference methods are used extensively for this analysis in the study of stochastically modeled reaction networks. Different coupling methods have been proposed to build finite difference estimators, with the "split coupling," also termed the "stacked coupling," yielding the lowest variance in the vast majority of cases. Analytical results related to this coupling are sparse and include an analysis of the variance of the coupled processes under the assumption of globally Lipschitz intensity functions [*SIAM J. Numer. Anal.*, 50 (2012), pp. 2237–2258]. Because of the global Lipschitz assumption utilized in [*SIAM J. Numer. Anal.*, 50 (2012), pp. 2237–2258], the main result there is only applicable to a small percentage of the models found in the literature, and it was conjectured that similar results should hold for a much wider class of models. In this paper we demonstrate this conjecture to be true by proving the variance of the coupled processes scales in the desired manner for a large class of non-Lipschitz models. We further extend the analysis to allow for time dependence in the parameters. In particular, binary systems with or without time-dependent rate parameters, a class of models that accounts for the vast majority of systems considered in the literature, satisfy the assumptions of our theory.

Key words. Monte Carlo, finite difference, parametric sensitivity, reaction network, stochastic model

AMS subject classifications. 60H35, 65C99, 92C40

DOI. 10.1137/19M1303940

1. Introduction. Stochastic models of biochemical interaction networks are now used ubiquitously in biology, especially cell biology [4, 5, 11, 14, 20, 22, 28]. The most common such model treats the system as a discrete-space, continuous-time Markov chain with the state vector giving the counts of the constituent species and with transitions for the chain modeled through the possible reactions of the system. The key system parameters of these models, the kinetic parameters, govern the rates at which the different reactions take place, and computing derivatives of system outputs with respect to these parameters, i.e., parametric sensitivity analysis, plays a critical role in many problems related to optimization and uncertainty quantification [1, 3, 7, 8, 13, 18, 25, 26]. In particular, such derivatives are often utilized in an inner loop in optimization problems, and their estimation is one of the key bottlenecks for many computational experiments in systems biology.

Suppose our state vector at time t is denoted by $X^\theta(t)$, where θ is a vector of parameters for the model. For simplicity, assume for now that θ is one-dimensional. Suppose also that f is a function and $f(X^\theta(t))$ is some output of interest. For example, if we are interested in the total molecular count, we can take $f(x) = \|x\|_1$; whereas if

†Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706 USA (anderson@math.wisc.edu).
‡Department of Mathematics, Indiana University, Bloomington, IN 47405-7000 USA (yuan13@iu.edu).

we are interested in the count of the first species we can take $f(x) = x_1$. Returning to the case of a general function f, the sensitivity of $J(\theta) = \mathbb{E}[f(X^\theta(t))]$ with respect to θ is defined as

$$J'(\theta) = \frac{\partial}{\partial\theta}\mathbb{E}[f(X^\theta(t))]. \tag{1.1}$$

By far the most common method for the computational estimation of the above derivative is finite difference combined with Monte Carlo. Specifically, we can estimate

$$J'(\theta) \approx \frac{\mathbb{E}[f(X^{\theta+\varepsilon}(t))] - \mathbb{E}[f(X^\theta(t))]}{\varepsilon} = \mathbb{E}\left[\frac{f(X^{\theta+\varepsilon}(t)) - f(X^\theta(t))}{\varepsilon}\right]$$

via

$$\mathbb{E}\left[\frac{f(X^{\theta+\varepsilon}(t)) - f(X^\theta(t))}{\varepsilon}\right] \approx D_N(\varepsilon) = \frac{1}{N\varepsilon}\sum_{i=1}^{N}\left[f(X^{\theta+\varepsilon}_{[i]}(t)) - f(X^\theta_{[i]}(t))\right],$$

where $X^\theta_{[i]}$ denotes the ith realization of X^θ. The variance of this estimator is

$$\mathrm{Var}\left(D_N(\varepsilon)\right) = \frac{1}{N\varepsilon^2}\mathrm{Var}\left(f(X^{\theta+\varepsilon}(t)) - f(X^\theta(t))\right),$$

so long as the pairs $\{(X^{\theta+\varepsilon}_{[i]}, X^\theta_{[i]})\}_{i=1}^{N}$ are independent. For a given ε, there are two ways to decrease the variance of the estimator:

  (i) increase N, or

  (ii) decrease $\mathrm{Var}(f(X^{\theta+\varepsilon}(t)) - f(X^\theta(t)))$.

Note that if X^θ and $X^{\theta+\varepsilon}$ are constructed independently, then the variance is simply [6]

$$\mathrm{Var}\left(f(X^{\theta+\varepsilon}(t))\right) + \mathrm{Var}\left(f(X^\theta(t))\right).$$

However, if the processes are *coupled* so that they are constructed on the same probability space but not independent, then

$$\mathrm{Var}\left(f(X^{\theta+\varepsilon}(t)) - f(X^\theta(t))\right) = \mathrm{Var}\left(f(X^{\theta+\varepsilon}(t))\right) + \mathrm{Var}\left(f(X^\theta(t))\right)$$
$$- 2\mathrm{Cov}(f(X^{\theta+\varepsilon}(t)), f(X^\theta(t))),$$

and it is possible to decrease the variance with little or no extra computational burden.

The two most commonly used couplings for parametric sensitivity analysis of reaction networks are the *common reaction path coupling* [24] and *coupled finite difference* (CFD)[2], also termed the *split coupling*, both of which rely on the random time change representation of Kurtz [19]. In more recent work, the *stacked coupling* was introduced [9]. This coupling is similar to the CFD coupling but utilizes thinning and is more amenable to models with time dependent rate parameters. A construction similar to the stacked coupling, and which was developed independently and simultaneously, can be found in [27].

Unfortunately, analytical results related to these couplings are sparse. In [2], it was shown that $\mathrm{Var}(f(X^{\theta+\varepsilon}(t)) - f(X^\theta(t)))$ is $O(\varepsilon)$ if the CFD coupling is used and if the intensity functions of the model are globally Lipschitz in both the state variable and the parameter θ. However, these assumptions are satisfied only if the

model stays inside a bounded set (for example, if a conservation relation exists for all species). *Therefore, these assumptions exclude most models found in the literature.* Similar analytic results related to the stability and convergence of coupled processes were given in section 4 of Engblom's 2014 paper [12].

Our main analytical result, Theorem 3.1, extends the main result from [2] to a much broader class of models, including all binary networks, which constitutes the vast majority of models utilized in the literature. Specifically, Theorem 3.1 guarantees that for any model satisfying Assumption 1, which are mild regularity conditions satisfied by nearly all models found in the literature, $\mathbb{E}[\|X^{\theta+\varepsilon}(t) - X^{\theta}(t)\|_1^p] = O(\varepsilon)$, for any $p \geq 1$. In particular, the main growth condition we assume a model satisfies is that any reaction that increases the net total number of molecules has a rate function that grows linearly. We note that this condition was also considered in [12, 16, 17, 23].

Finally, we note that finite differences is not the only method for estimating the derivatives in (1.1). In particular, it is often possible to move the derivative inside the expectation and perform a change of measure. This leads to a Girsanov transformation method [21, 24]. However, in the present setting of reaction networks this method typically has a very large variance and so is not often utilized. Further, it is sometimes possible to switch the order of differentiation and expectation in (1.1) directly, and this leads to pathwise differentiation methods [15]. However, this method of differentiation is generally not applicable in the present setting since there are commonly "interruptions" in the reactions [8]. However, see [8] for a hybrid pathwise differentiation method that is applicable to the vast majority of models in the literature.

The remainder of the paper is organized as follows. In section 2, we will provide required terminology, the formal mathematical model, and the assumptions on the model we require for our main theorem to hold. In section 3, we state and prove our main result, Theorem 3.1.

2. Mathematical model.

2.1. Notation. We use standard notation. The nonnegative integers, real numbers, and positive real numbers will be represented by $\mathbb{Z}_{\geq 0}$, $\mathbb{R}$, and $\mathbb{R}_{>0}$, respectively. For $d \geq 2$, we denote by $\mathbb{Z}_{\geq 0}^d$, $\mathbb{R}^d$, and $\mathbb{R}_{>0}^d$ the corresponding d-dimensional vector spaces. For $x \in \mathbb{Z}_{\geq 0}^d$ the factorial $x!$ is defined

$$(2.1) \qquad x! = \prod_{i=1}^{d} x_i!.$$

For $u, v \in \mathbb{R}^d$, we write $u \geq v$ if $u_i \geq v_i$ for each i, we define the inner product as $u \cdot v = \sum_{i=1}^{d} u_i v_i$, and the ℓ_1 norm of u is

$$\|u\|_1 = \sum_{i=1}^{d} |u_i|.$$

For $A \subset \mathbb{R}$ and $x \in \mathbb{R}$ we denote the indicator function $1_A(x)$ by

$$1_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

2.2. Reaction networks and stochastic mass-action kinetics. We begin with the definition of a reaction network. See [5] for any necessary background.

DEFINITION 2.1. *A reaction networks is a triple* $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ *where*
- $\mathcal{S} = \{A_1, \ldots, A_d\}$ *is a set of* species.
- $\mathcal{C}$*, the* complexes*, is a nonempty, finite set of linear combinations of the species over the nonnegative integers. We make the usual abuse of notation by corresponding* $y \in \mathcal{C}$ *with the vector in* $\mathbb{Z}_{\geq 0}^d$ *whose ith component gives the integer count of species* A_i *in complex* y*.*
- $\mathcal{R}$*, the* reactions*, is a subset of* $\mathcal{C} \times \mathcal{C}$*. We typically denote* $(y, y') \in \mathcal{R}$ *by* $y \to y'$ *and will denote by* K *the cardinality of* $\mathcal{R}$*.*
  After enumerating the reactions, we denote the k*th reaction by* $y_k \to y'_k \in \mathcal{R}$*, we denote its reaction vector by* $\zeta_k = y'_k - y_k \in \mathbb{Z}^d$*, and we will call* y_k *and* y'_k *the source and product complexes for the* k*th reaction, respectively.*

The most common stochastic model of a reaction network treats the system as a continuous-time Markov chain with state space $\mathbb{Z}_{\geq 0}^d$, where $X(t) \in \mathbb{Z}_{\geq 0}^d$ gives the counts of the different chemical species at time $t \geq 0$. Transitions for the model are determined by the reactions and their intensity functions. Specifically, for the kth reaction, we let $\lambda_k : \mathbb{Z}^d \times [0, \infty) \to \mathbb{R}_{\geq 0}$ be the corresponding intensity function. Then the transition rate from state x to x' is

$$(2.2) \qquad Q(x, x', t) = \sum_{k:\zeta_k = x' - x} \lambda_k(x, t),$$

where the sum is over those reactions with reaction vector equal to $x' - x \in \mathbb{Z}_{\geq 0}^d$. The most common choice of intensity function is given by stochastic mass-action kinetics, in which case

$$(2.3) \qquad \lambda_k(x) = \kappa_k \frac{x!}{(x - y_k)!} 1_{\{x \geq y_k\}},$$

where y_k is the source complex. Because this is the most common choice of rate function, reactions with source complex of the form $\emptyset$ are termed *zeroth order reactions* and reactions with source complex $y_k = e_i$ for some $i \in \{1, \ldots, d\}$ are termed *first order reactions.*

Note that mass-action kinetics does not depend upon time, which is why the t dependence has been dropped. When stochastic mass-action kinetics is used, the term κ_k is called a *rate constant* and is typically placed next to the reaction arrow in the reaction diagram (see Examples 1, 2, and 3 below).

The most commonly used time-dependent intensity function is only a slight modification of (2.3) and assumes that the rate parameters are functions of time,

$$(2.4) \qquad \lambda_k(x, t) = \kappa_k(t) \frac{x!}{(x - y_k)!} 1_{\{x \geq y_k\}}.$$

A few examples illustrate the model.

Example 1. Consider the following reaction network with stochastic mass-action kinetics:

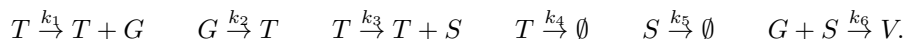$$A \xrightarrow{2} 2A \qquad 2A \xrightarrow{\theta} A.$$

Because we are assuming mass-action kinetics, the intensity functions are

$$\lambda_1(x) = 2x \quad \text{and} \quad \lambda_2(x) = \theta x(x - 1) \quad \text{for} \quad x \in \mathbb{Z}_{\geq 0}.$$

Neither intensity is a function of time, so there is no t-dependence in the notation. $\square$

Example 2 (intracellular viral kinetics). Consider the following model given in [17]. This model consists of 4 species: the viral template (T), the viral genome (G), the viral structural protein (S), and the virus (V). There are 6 reactions:
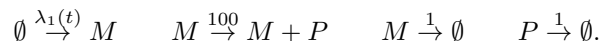
$$T \xrightarrow{k_1} T + G \qquad G \xrightarrow{k_2} T \qquad T \xrightarrow{k_3} T + S \qquad T \xrightarrow{k_4} \emptyset \qquad S \xrightarrow{k_5} \emptyset \qquad G + S \xrightarrow{k_6} V.$$

Assuming mass-action kinetics, and after ordering the species as T, G, S, V, we have

$$\lambda_1(x) = \kappa_1 x_1, \ \lambda_2(x) = \kappa_2 x_2, \ \lambda_3(x) = \kappa_3 x_1,$$
$$\lambda_4(x) = \kappa_4 x_1, \ \lambda_5(x) = \kappa_5 x_3, \ \lambda_6(x) = \kappa_6 x_2 x_3.$$

Notice that G and S can be created with the presence of only one copy of a viral template. Hence, the state space will be unbounded. $\qquad\square$

The following example incorporates circadian rhythm into our intensity function by assuming time dependence of a rate parameter.

Example 3. Consider the following standard network for gene transcription and translation:

$$\emptyset \xrightarrow{\lambda_1(t)} M \qquad M \xrightarrow{100} M + P \qquad M \xrightarrow{1} \emptyset \qquad P \xrightarrow{1} \emptyset.$$

We assume the last three reactions have mass-action kinetics (2.3) with associated rate constants given above their reaction arrow. However, we assume the intensity function for the first reaction follows (2.4). In particular,

$$\lambda_1(t) = \kappa_1(t) = 60 + 15 \sin\left(\tfrac{2\pi t}{24}\right).$$

Using such an intensity function for the first reaction allows us to model the system going through dark-light cycles that oscillate over a 24-hour time-period. $\qquad\square$

2.3. Representations for the process. We provide two representations for the process detailed in the previous section. See [4, 5] for more details on either. The first is the random time change representation of Thomas Kurtz [19], in which X is the solution to the following stochastic equation:

$$(2.5) \qquad X(t) = X(0) + \sum_{k=1}^{K} Y_k\left(\int_0^t \lambda_k(X(s), s)ds\right)\zeta_k,$$

where $\{Y_k\}_{k=1}^{K}$ are independent unit-rate Poisson processes.

The second representation arises by letting X be the solution to the following equation:

(2.6)

$$X(t) = X(0) + \sum_{k=1}^{K} \zeta_k \int_{[0,t]\times[0,\infty)} 1_{[q_{k-1}(s-), q_{k-1}(s-)+\lambda_k(X(s-),s-))}(x)N(ds \times dx),$$

where N is a unit-rate space-time Poisson point process and $q_0(s) = 0$ for all $s \geq 0$ and for $k \in \{1, \ldots, K\}$,

$$q_k(s) = \sum_{\ell=1}^{k} \lambda_\ell(X(s), s).$$

2.4. Coupled processes and assumptions on the model. We detail a number of couplings already found in the literature.

The first coupling of $(X^\theta, X^{\theta+\varepsilon})$ we detail, usually termed the *common reaction path coupling* [24], arises from using the same choice of Poisson processes in the representation (2.5). Specifically,

(2.7)
$$X^{\theta+\varepsilon}(t) = X^{\theta+\varepsilon}(0) + \sum_{k=1}^{K} Y_k \left(\int_0^t \lambda_k^{\theta+\varepsilon}(X^{\theta+\varepsilon}(s), s)ds \right) \zeta_k$$

$$X^{\theta}(t) = X^{\theta}(0) + \sum_{k=1}^{K} Y_k \left(\int_0^t \lambda_k^{\theta}(X^{\theta}(s), s)ds \right) \zeta_k,$$

where $\{Y_k\}_{k=1}^{K}$ is a sequence of independent unit-rate Poisson processes.

The second coupling of $(X^\theta, X^{\theta+\varepsilon})$ we detail, termed the *stacked coupling* [9], utilizes a space-time Poisson point process as its randomness, as in (2.6). Specifically,

(2.8)
$$X^{\theta+\varepsilon}(t) = X^{\theta+\varepsilon}(0) + \sum_{k=1}^{K} \zeta_k \int_{[0,t]\times[0,\infty)} 1_{[q_{k-1}^{\theta,\varepsilon}(s-), q_{k-1}^{\theta,\varepsilon}(s-) + \lambda_k^{\theta+\varepsilon}(X^{\theta+\varepsilon}(s-), s-))}(x) N(ds \times dx)$$

$$X^{\theta}(t) = X^{\theta}(0) + \sum_{k=1}^{K} \zeta_k \int_{[0,t]\times[0,\infty)} 1_{[q_{k-1}^{\theta,\varepsilon}(s-), q_{k-1}^{\theta,\varepsilon}(s-) + \lambda_k^{\theta}(X^{\theta}(s-), s-))}(x) N(ds \times dx),$$

where

(2.9)
$$\overline{\lambda}_\ell^{\theta,\varepsilon}(s) = \max \left\{ \lambda_\ell^{\theta+\varepsilon}\left(X^{\theta+\varepsilon}(s), s\right), \lambda_\ell^{\theta}\left(X^{\theta}(s), s\right) \right\}$$

$$q_k^{\theta,\varepsilon}(s) = \sum_{\ell=1}^{k} \overline{\lambda}_\ell^{\theta,\varepsilon}(s) \qquad \text{with} \qquad q_0^{\theta,\varepsilon}(s) = 0.$$

Remark 2.2. One key advantage of the stacked coupling is that whenever both processes jump at the same time, they must jump via the same reaction channel. That is, when the kth reaction channel is chosen to fire for X^θ, then $X^{\theta+\varepsilon}$ can either update through the kth reaction channel or stay put and vice versa.

For future reference, we define

(2.10)
$$N_{q_K}^{\theta,\varepsilon}(t) = \int_{[0,t]\times[0,\infty)} 1_{[0,q_K^{\theta,\varepsilon}(s-))}(x) N(ds \times dx)$$

to be the number of points of the point process N in our region of interest up to time t.

If the intensity functions are independent of time, then the stacked coupling (2.8) is equivalent to the CFD, also termed the split coupling [2],

(2.11)

$$X^{\theta+\varepsilon}(t) = X^{\theta+\varepsilon}(0) + \sum_{k=1}^{K} \zeta_k Y_k^1 \left(\int_0^t \lambda_k^{\theta+\varepsilon}(X^{\theta+\varepsilon}(s),s) \wedge \lambda_k^{\theta}(X^{\theta}(s),s) ds \right)$$

$$+ \sum_{k=1}^{K} \zeta_k Y_k^2 \left(\int_0^t \left(\lambda_k^{\theta+\varepsilon}(X^{\theta+\varepsilon}(s),s) - \lambda_k^{\theta+\varepsilon}(X^{\theta+\varepsilon}(s),s) \wedge \lambda_k^{\theta}(X^{\theta}(s),s) \right) ds \right)$$

$$X^{\theta}(t) = X^{\theta}(0) + \sum_{k=1}^{K} \zeta_k Y_k^1 \left(\int_0^t \lambda_k^{\theta+\varepsilon}(X^{\theta+\varepsilon}(s),s) \wedge \lambda_k^{\theta}(X^{\theta}(s),s) ds \right)$$

$$+ \sum_{k=1}^{K} \zeta_k Y_k^3 \left(\int_0^t \left(\lambda_k^{\theta}(X^{\theta}(s),s) - \lambda_k^{\theta+\varepsilon}(X^{\theta+\varepsilon}(s),s) \wedge \lambda_k^{\theta}(X^{\theta}(s),s) \right) ds \right),$$

where $\{Y_k^1, Y_k^2, Y_k^3\}_{k=1}^{K}$ are independent unit-rate Poisson processes.

As noted in the introduction, analytical results related to these couplings are sparse. However, the following is proven in [2].

THEOREM 2.3. *Let $(\mathcal{S},\mathcal{C},\mathcal{R})$ be a reaction network with d species and K reactions. Let $\{X^\theta\}$ be a family of stochastic models whose intensity functions $\lambda_k^\theta(x,t)$ are parametrized by $\theta \in \mathbb{R}^m_{>0}$. Suppose $(X^{\theta+\varepsilon}, X^\theta)$ are coupled as in (2.11) and there exist constants K_1, K_2 such that for all $x,y \in \mathbb{Z}^d_{\geq 0}$*

(2.12) $$\sum_{k=1}^{K} \left| \lambda_k^\theta(x) - \lambda_k^\theta(y) \right| \leq K_1 |x-y|,$$

and for all $x \in \mathbb{Z}^d_{\geq 0}$ and all $\varepsilon < 1$

(2.13) $$\sum_{k=1}^{K} \left| \lambda_k^{\theta+\varepsilon}(x) - \lambda_k^\theta(x) \right| \leq K_2 \varepsilon.$$

Then, for any $T > 0$ and any function $f : \mathbb{R}^d \to \mathbb{R}$ that is C^1 (bounded first derivative), there is a $C_{T,f} > 0$ for which

$$\mathbb{E}\left[\sup_{t \leq T} \left(f(X^{\theta+\varepsilon}(t)) - f(X^\theta(t)) \right)^2 \right] \leq C_{T,f} \varepsilon.$$

Notice that (2.12) implies the intensity functions are globally Lipschitz in the state variable, and (2.13) implies the intensity functions are globally Lipschitz in the parameter; these assumptions are highly restrictive, and most models in the literature are excluded.

Example 4. Consider Example 1, and recall the reaction intensity for the second reaction is

$$\lambda_2(x,t) = \theta x(x-1).$$

Notice that the function is only locally Lipschitz in x and θ (since the state space is not bounded); thus Theorem 2.3 is not applicable. □

Example 5. Consider the intracellular viral kinetics model introduced in Example 2. Notice that the last reaction is binary; hence $\lambda_6(x) = \kappa_6 x_2 x_3$ is only locally Lipschitz in x. Moreover, the viral genome (G) and the viral structural protein (S) can be created by the first and third reaction without consuming any other species. Hence x_2 and x_3 can not be bounded, and $\lambda_2, \lambda_5, \lambda_6$ are only locally Lipschitz in terms of rate constants. Thus, Theorem 2.3 is not applicable. □

We will extend Theorem 2.3 to a wider class of models. However, care must be taken not to try to extend too broadly. For example, models such as $2A \to 3A$ are explosive, and their expectation and variance are not finite. With this example in mind, we denote by $\mathcal{P} = \{k \in \{1, \ldots, K\} : \zeta_k \cdot \vec{1} > 0\}$, the indices of the reactions that have a positive net effect on the count of the total population, and give our main modeling assumption below.

Assumption 1. A family of rate functions, $\{\lambda_k^\theta\}$, parameterized by $\theta \in \mathbb{R}_{>0}^\gamma$ satisfies this assumption if there is a $\overline{C} > 0$ and a compact set $\Theta \subset \mathbb{R}_{>0}^\gamma$ for which the following conditions hold:

(1) (**Linear growth for** $\mathcal{P}$) For any $k \in \mathcal{P}$

$$\sup_{\theta \in \Theta} \sup_{t \in \mathbb{R}_{>0}} \lambda_k^\theta(x, t) \leq \overline{C}(1 + \|x\|_1).$$

(2) (**Polynomial growth for** $\mathcal{R}$) For any $k \in \{1, 2, \ldots, K\}$, there is an integer $p \geq 1$ for which

$$\sup_{\theta \in \Theta} \sup_{t \in \mathbb{R}_{>0}} \lambda_k^\theta(x, t) \leq \overline{C}(1 + \|x\|_1^p).$$

(3) For $\|\varepsilon\|_1 > 0$ small enough,

$$\sup_{\theta \in \Theta} \sup_{x \in \mathbb{Z}_{\geq 0}^d} \frac{\sum_{k=1}^K |\lambda_k^{\theta+\varepsilon}(x, s) - \lambda_k^\theta(x, s)|}{\sum_{k=1}^K \max\{\lambda_k^{\theta+\varepsilon}(x, s), \lambda_k^\theta(x, s)\}} \leq \overline{C}\|\varepsilon\|_1.$$

Remark 2.4. We will sometimes work with sets of rate functions with a particular choice of rate constants, and these will be denoted by $\{\lambda_k\}$. In this case, the set Θ is taken to be a single point, and the criterion $\sup_{\theta \in \Theta}$ can be dropped from each of the three conditions.

We note that condition (1) in Assumption 1 was also utilized in [12, 16, 17, 23].

The next lemma shows that stochastic mass-action kinetics (2.3) satisfies conditions (2) and (3) of Assumption 1. Hence, any model with this choice of kinetics satisfies Assumption 1 if only first or zeroth order reactions generate a net gain in total molecule counts. In particular, binary systems, a class of models that accounts for the vast majority of systems considered in the literature, satisfy this assumption, since a reaction with a binary source complex will either maintain or decrease total molecule counts when it occurs. For an example of a model that does *not* satisfy condition (1) when mass-action kinetics is utilized, simply consider $2A \rightleftarrows 3A$. In particular, the reaction $2A \to 3A$ violates the assumption.

LEMMA 2.5. *Let* $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ *be a reaction network with d species and K reactions. Assume the intensity functions are given by stochastic mass-action kinetics* (2.3). *Let* $\Theta \subset \mathbb{R}_{>0}^K$ *be a compact set, where we correspond each* $\theta = (\kappa_1, \kappa_2, \ldots, \kappa_K) \in \Theta$ *with a choice of mass-action rate constants. Then, conditions (2) and (3) of Assumption 1 are satisfied.*

Proof. Condition (2) of Assumption 1 follows easily since we have

$$\lambda_k(x) = \kappa_k \frac{x!}{(x - y_k)!} 1_{\{x \geq y_k\}} \leq \kappa_k \prod_{i=1}^d [x_i(x_i - 1) \ldots (x_i - y_{ki} + 1)]$$

$$\leq \kappa_k \prod_{i=1}^d x_i^{y_{ki}} \leq \kappa_k \|x\|_1^{\|y_k\|_1},$$

where we take $0^0 = 1$, and since $\theta \in \Theta$, which is compact.

To verify condition (3) of Assumption 1, notice that

$$|\lambda_k^{\theta+\varepsilon}(x,s) - \lambda_k^\theta(x,s)| = |\varepsilon_k| \frac{x!}{(x-y_k)!} 1_{\{x \geq y_k\}} \quad \text{and}$$

$$\max\{\lambda_k^{\theta+\varepsilon}(x,s), \lambda_k^\theta(x,s)\} = \max\{\kappa_k + \varepsilon_k, \kappa_k\} \frac{x!}{(x-y_k)!} 1_{\{x \geq y_k\}},$$

where we choose ε_k small enough so that $\kappa_k + \varepsilon_k > 0$. Hence

$$\max_{x \in \mathbb{Z}_{\geq 0}^d} \frac{\sum_{k=1}^K |\lambda_k^{\theta+\varepsilon}(x,s) - \lambda_k^\theta(x,s)|}{\sum_{k=1}^K \max\{\lambda_k^{\theta+\varepsilon}(x,s), \lambda_k^\theta(x,s)\}}$$

$$= \max_{x \in \mathbb{Z}_{\geq 0}^d} \frac{\sum_{k=1}^K |\varepsilon_k| \frac{x!}{(x-y_k)!} 1_{\{x \geq y_k\}}}{\sum_{k=1}^K \max\{\kappa_k + \varepsilon_k, \kappa_k\} \frac{x!}{(x-y_k)!} 1_{\{x \geq y_k\}}}$$

$$\leq \max_{x \in \mathbb{Z}_{\geq 0}^d} \max_k \left\{ \frac{|\varepsilon_k|}{\max\{\kappa_k + \varepsilon_k, \kappa_k\}} \right\} = \max_k \left\{ \frac{|\varepsilon_k|}{\max\{\kappa_k + \varepsilon_k, \kappa_k\}} \right\}.$$

For the last inequality, we used the fact that

$$a_k \leq c b_k \qquad \Rightarrow \qquad \frac{\sum_{k=1}^K a_k}{\sum_{k=1}^K b_k} \leq c \frac{\sum_{k=1}^K b_k}{\sum_{k=1}^K b_k} = c.$$

Hence, again by the compactness of $\Theta \in \mathbb{R}_{>0}^K$, condition (3) of Assumption 1 holds so long as we choose $\|\varepsilon\|_1$ small enough, . $\square$

3. Main results. In this section, we prove our main result, stated below.

THEOREM 3.1. *Let $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network with d species and K reactions. Let $\{X^\theta\}$ be a family of stochastic models whose associated intensity functions $\{\lambda_k^\theta\}$ are parametrized by $\theta \in \Theta \subset \mathbb{R}_{>0}^\gamma$, for some $\gamma \in \mathbb{Z}_{>0}$, where Θ is compact. Suppose that Assumption 1 holds and that $(X^\theta, X^{\theta+\varepsilon})$ are coupled via the stacked coupling (2.8) with $X^\theta(0) = X^{\theta+\varepsilon}(0) = x_0 \in \mathbb{Z}_{\geq 0}^d$. Then, for any θ in the interior of Θ and any $r \geq 1$, there is a $C_{r,t} > 0$ and $\bar{\varepsilon} \in \mathbb{R}_{>0}$ so that*

$$(3.1) \qquad \mathbb{E}\left[\|X^{\theta+\varepsilon}(t) - X^\theta(t)\|_1^r\right] \leq C_{r,t} \|\varepsilon\|_1,$$

when $\|\varepsilon\|_1 \leq \bar{\varepsilon}$.

The condition that θ be in the interior of Θ is not vital. In particular, if θ were on the boundary of Θ, then a larger $\widetilde{\Theta} \supset \Theta$ could be chosen. Note also that the choice of $\bar{\varepsilon}$ can be used to guarantee that $\theta + \varepsilon \in \Theta$.

Note that an immediate corollary of Theorem 3.1, acquired by taking $r = 2$, is the following.

COROLLARY 3.2. *Let $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network with d species and K reactions. Let $\{X^\theta\}$ be a family of stochastic models whose associated intensity functions $\{\lambda_k^\theta\}$ are parametrized by $\theta \in \mathbb{R}_{>0}^\gamma$, for some $\gamma \in \mathbb{Z}_{>0}$. Suppose that Assumption 1 holds and that $(X^\theta, X^{\theta+\varepsilon})$ are coupled via the stacked coupling (2.8) with $X^\theta(0) = X^{\theta+\varepsilon}(0) = x_0 \in \mathbb{Z}_{\geq 0}^d$. Finally, let $f : \mathbb{R}^d \to \mathbb{R}$ be Lipschitz. Then there is a $C_{T,f} > 0$ and a $\bar{\varepsilon} \in \mathbb{R}_{>0}$ so that*

$$Var\left(f(X^{\theta+\varepsilon}(t)) - f(X^\theta(t))\right) \leq C_{T,f} \|\varepsilon\|_1$$

when $\|\varepsilon\|_1 \leq \bar{\varepsilon}$.

The rest of the section is organized as follows. In section 3.1, we will discuss the key component of the proof, which establishes an upper bound for the rate of growth of the process. In section 3.2, we will then introduce some technical lemmas. The pieces will then be put together in section 3.3.

3.1. Growth bound of the stochastic process. Let τ_m be the first time the process leaves the L^1 ball with radius m. We are interested in the decay rate of $P(\tau_m \leq t)$, as $m \to \infty$, and will show in Lemma 3.3 that it decays exponentially so long as the model satisfies condition (1) of Assumption 1. Moreover, by Lemma 3.4, the bound is sharp. We note that Lemma 3.3 is similar to a result in [17], where, under the same assumptions, the decay was shown to be polynomial in m.

LEMMA 3.3. *Let $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network with d species and K reactions. Let $\{X^\theta\}$ be a family of stochastic models whose associated intensity functions $\{\lambda_k^\theta\}$ are parametrized by $\theta \in \Theta \subset \mathbb{R}_{>0}^\gamma$, for some $\gamma \in \mathbb{Z}_{>0}$, where Θ is compact. Assume $\{\lambda_k^\theta\}$ satisfies condition (1) of Assumption 1. Define*

$$\tau_m^\theta = \inf\left\{t \geq 0 : \|X^\theta(t)\|_1 = \sum_{i=1}^d X_i^\theta(t) \geq m\right\}.$$

Then, for any initial condition $x_0 \in \mathbb{Z}_{\geq 0}^d$, any $\theta \in \Theta$, and any $t \geq 0$, there exist constants $C > 0$ and $\delta \in (0, 1)$, both of which are independent of θ, such that

$$(3.2) \qquad P(\tau_m^\theta \leq t) \leq C\delta^m$$

for all $m \in \mathbb{Z}_{>0}$ large enough.

In the proof, we will show that one possible choice of C and δ is

$$(3.3) \qquad C = (1 - e^{-ct\overline{C}K})^{-1 - \frac{\|x_0\|_1}{\ell}} \text{ and } \delta = \left(1 - e^{-ct\overline{C}K}\right)^{\frac{1}{\ell}},$$

where $\overline{C}$ comes from Assumption 1, $c = \max\{\|x_0\|_1 + 1, \ell\}$, and $\ell = \max_k\{\zeta_k \cdot \vec{1}\}$. By Remark 2.4, we may apply Lemma 3.3 when only a particular model (with a particular choice of rate constants) is being considered. In this case, we take Θ to be a single point.

We prove Lemma 3.3 at the end of the section. The proof will proceed by comparing the model of interest with a particular linear (i.e., first order) model. We therefore begin with a sequence of lemmas related to linear models. The first concerns a pure birth process with a particular choice of rate constants.

LEMMA 3.4. *Suppose X satisfies*

$$X(t) = 1 + Y\left(\kappa \int_0^t X(s)ds\right),$$

where Y is a unit-rate Poisson process, and $\kappa > 0$. For any integer $M \geq 2$, define $\tau_M = \inf\{t \geq 0 : X(t) \geq M\}$. Then

$$P(\tau_M \leq t) = (1 - e^{-\kappa t})^{M-1}.$$

Proof. We first prove the results when $\kappa = 1$. Suppose the unit-rate Poisson process Y has holding times $e_1, e_2, \ldots$, where e_i are unit exponentials. Then

$$\tau_M = \frac{e_1}{1} + \frac{e_2}{2} + \frac{e_3}{3} + \cdots + \frac{e_{M-1}}{M-1} = \sum_{i=1}^{M-1} \frac{e_i}{i}.$$

Notice that τ_M is the sum of exponential random variables with distinct parameters $\rho_i = i$. Thus τ_M is hypoexponentially distributed with density (see [10]):

$$f_{\tau_M}(t) = \sum_{i=1}^{M-1} \rho_i e^{-\rho_i t} \left(\prod_{j \neq i}^{M-1} \frac{\rho_j}{\rho_j - \rho_i} \right).$$

Thus,

$$
\begin{aligned}
P(\tau_M \leq t) &= \int_0^t f_{\tau_M}(s) ds = \int_0^t \sum_{i=1}^{M-1} \rho_i e^{-\rho_i s} \left(\prod_{j \neq i}^{M-1} \frac{\rho_j}{\rho_j - \rho_i} \right) ds \\
&= \sum_{i=1}^{M-1} \left[\prod_{j \neq i}^{M-1} \frac{\rho_j}{\rho_j - \rho_i} \right] \int_0^t \rho_i e^{-\rho_i s} ds = \sum_{i=1}^{M-1} \left(\prod_{j \neq i}^{M-1} \frac{\rho_j}{\rho_j - \rho_i} \right) (1 - e^{-\rho_i t}) \\
&= \sum_{i=1}^{M-1} \left(\prod_{j \neq i}^{M-1} \frac{\rho_j}{\rho_j - \rho_i} \right) - \sum_{i=1}^{M-1} \left(\prod_{j \neq i}^{M-1} \frac{\rho_j}{\rho_j - \rho_i} \right) e^{-\rho_i t}.
\end{aligned}
$$

To simplify we note that

$$
\begin{aligned}
\prod_{j \neq i}^{M-1} \frac{\rho_j}{\rho_j - \rho_i} &= \prod_{j \neq i}^{M-1} \frac{j}{j - i} = \frac{1}{1-i} \frac{2}{2-i} \cdots \frac{i-1}{-1} \frac{i+1}{1} \frac{i+2}{2} \cdots \frac{M-1}{M-1-i} \\
&= (-1)^{i-1} \frac{1}{i-1} \frac{2}{i-2} \cdots \frac{i-1}{1} \frac{i+1}{1} \frac{i+2}{2} \cdots \frac{M-1}{M-1-i} \\
&= (-1)^{i-1} \frac{(M-1)!}{(M-1-i)! i!} = -(-1)^i \binom{M-1}{i}.
\end{aligned}
$$

Thus,

(3.4)
$$
\begin{aligned}
P(\tau_M \leq t) &= - \sum_{i=1}^{M-1} (-1)^i \binom{M-1}{i} + \sum_{i=1}^{M-1} (-1)^i \binom{M-1}{i} e^{-it} \\
&= - \sum_{i=0}^{M-1} (-1)^i \binom{M-1}{i} + \sum_{i=0}^{M-1} (-1)^i \binom{M-1}{i} e^{-it} \\
&= (1 - e^{-t})^{M-1},
\end{aligned}
$$

where the third equality follows since the $i = 0$ terms cancel out, and the fourth equality follows by applying the binomial theorem twice.

The result is therefore shown when $\kappa = 1$. When $\kappa \neq 1$, we have

$$\tau_M = \frac{e_1}{\kappa} + \frac{e_2}{2\kappa} + \frac{e_3}{3\kappa} + \cdots + \frac{e_{M-1}}{(M-1)\kappa} = \frac{1}{\kappa} \sum_{i=1}^{M-1} \frac{e_i}{i},$$

and, simply by scaling,

$$P(\tau_M \leq t) = P \left(\sum_{i=1}^{M-1} \frac{e_i}{i} \leq \kappa t \right) = (1 - e^{-\kappa t})^{M-1},$$

where we used (3.4). $\qquad \Box$

In the next lemma we will consider models with a general positive initial condition and with linear intensity functions but larger jumps sizes (denoted by ℓ). In particular, the model below can be thought of as arising from the reaction network $A \xrightarrow{\kappa} (\ell+1)A$.

LEMMA 3.5. *Suppose X satisfies*

$$(3.5) \qquad X(t) = x_0 + \ell Y \left(\kappa \int_0^t X(s)ds \right),$$

where Y is a unit-rate Poisson process, and $\kappa > 0$. For any integer $M > x_0$, define $\tau_M = \inf\{t \geq 0 : X(t) \geq M\}$. Then for $M \geq \ell + x_0$,

$$P(\tau_M \leq t) \leq C\delta^M$$

for some constant $C = C(t, \ell, \kappa, x_0) > 0$ and $\delta = \delta(t, \ell, \kappa, x_0) \in (0, 1)$.

Proof. Let Y be the Poisson process in (3.5), and define Z to be the solution to

$$Z(t) = 1 + Y \left(\kappa \int_0^t Z(s)ds \right).$$

Note that Z is using the same Poisson process as X. However, the two processes are different because Z has an initial condition of $Z(0) = 1$ and only jumps by size 1. In particular, the process Z satisfies the assumptions of Lemma 3.4. By Lemma 3.4, if we denote $\mu_m = \inf\{t \geq 0 : Z(t) \geq m\}$, we have

$$(3.6) \qquad P(\mu_m \leq t) = (1 - e^{-\kappa t})^{m-1}.$$

Denote the unit-rate exponential holding times of the Poisson process Y by $e_1, e_2, \ldots$. Then, for $m \geq 2$,

$$\tau_{(m-1)\ell+x_0} = \frac{1}{\kappa}\left(\frac{e_1}{x_0} + \frac{e_2}{x_0+\ell} + \cdots + \frac{e_{m-1}}{x_0+(m-2)\ell} \right) = \frac{1}{\kappa}\sum_{i=1}^{m-1} \frac{e_i}{x_0+(i-1)\ell}$$

$$\mu_m = \frac{1}{\kappa}\left(\frac{e_1}{1} + \frac{e_2}{2} + \frac{e_3}{3} + \cdots + \frac{e_{m-1}}{m-1} \right) = \frac{1}{\kappa}\sum_{i=1}^{m-1} \frac{e_i}{i}.$$

Hence if we let $c = \max\{x_0, \ell\}$,

$$\tau_{(m-1)\ell+x_0} = \frac{1}{\kappa}\sum_{i=1}^{m-1} \frac{e_i}{x_0+(i-1)\ell} \geq \frac{1}{\kappa}\sum_{i=1}^{m-1} \frac{e_i}{c+(i-1)c} = \frac{1}{c}\mu_m.$$

Hence, by (3.6)

$$P(\tau_{(m-1)\ell+x_0} \leq t) \leq P(\mu_m \leq ct) = (1 - e^{-ct\kappa})^{m-1}.$$

Thus, for $M \geq \ell + x_0$, we may take $m = \lfloor \frac{M-x_0}{\ell} \rfloor + 1$, which must be greater than or equal to 2, and conclude

$$P(\tau_M \leq t) \leq P(\tau_{(m-1)\ell+x_0} \leq t) \leq (1 - e^{-ct\kappa})^{m-1} \leq (1 - e^{-ct\kappa})^{\frac{M-x_0}{\ell}-1} = C\delta^M,$$

where $\delta = (1 - e^{-ct\kappa})^{\frac{1}{\ell}}$ and $C = (1 - e^{-ct\kappa})^{-1-\frac{x_0}{\ell}}$. $\square$

LEMMA 3.6. *Let $x_0 \geq 0$, and suppose X satisfies*

$$(3.7) \qquad X(t) = x_0 + \ell Y \left(\int_0^t \kappa(1 + X(s))ds \right),$$

where Y is a unit-rate Poisson process, and $\kappa > 0$. For any integer $M > x_0$, define $\tau_M = \inf\{t \geq 0 : X(t) \geq M\}$. Then for $M \geq \ell + x_0$,

$$(3.8) \qquad P(\tau_M \leq t) \leq C\delta^M$$

for some constant $C = C(t, \ell, \kappa, x_0) > 0$ and $\delta = \delta(t, \ell, \kappa, x_0) \in (0, 1)$.

Proof. Define $Z(t) = X(t) + 1$; then we can rewrite (3.7) to get the stochastic equation for $Z(t)$, which is

$$(3.9) \qquad Z(t) = (1 + x_0) + \ell Y \left(\int_0^t \kappa Z(s)ds \right).$$

For any integer k, define $\mu_k = \inf\{t \geq 0 : Z(t) \geq k\}$. Then, by Lemma 3.5,

$$(3.10) \qquad P(\mu_k \leq t) \leq C\delta^k,$$

so long as $k \geq \ell + x_0 + 1$. By the definition of $Z(t)$, $\tau_M = \mu_{M+1}$ and thus, so long as $M + 1 \geq \ell + x_0 + 1$,

$$P(\tau_M \leq t) = P(\mu_{M+1} \leq t) \leq C\delta^{M+1} = \tilde{C}\delta^M,$$

where $\tilde{C} = C\delta = (1 - e^{-ct\kappa})^{-1-\frac{x_0}{\ell}}$ and $\delta = (1 - e^{-ct\kappa})^{\frac{1}{\ell}}$ for $c = \max\{x_0 + 1, \ell\}$. ☐

Remark 3.7. Suppose $\{Y_k\}_{k=1}^K$ are independent unit-rate Poisson processes, and suppose that the process $\overline{X}$ satisfies the representation

$$\overline{X}(t) = x_0 + \sum_{k=1}^K \ell Y_k \left(\int_0^t \kappa_k(1 + \overline{X}(s))ds \right).$$

Then the process has the same distribution as the process X satisfying the representation

$$X(t) = x_0 + \ell Y \left(\int_0^t \kappa(1 + X(s))ds \right),$$

where Y is a unit-rate Poisson process, and $\kappa = \sum_{k=1}^K \kappa_k$. By Lemma 3.6, X satisfies (3.2), implying that $\overline{X}$ does as well.

We now turn to the proof of Lemma 3.3.

Proof of Lemma 3.3. Denote $\mathcal{P} = \{k \in \{1, \ldots, K\} : \zeta_k \cdot \vec{1} > 0\}$, and let $\ell = \max_{k \in \mathcal{P}}\{\zeta_k \cdot \vec{1}\}$. We choose a specific $\theta \in \Theta$ and suppose that X^θ and Z are defined (coupled) via

$$X^\theta(t) = x_0 + \sum_{k=1}^K Y_k \left(\int_0^t \lambda_k^\theta(X^\theta(s), s)ds \right) \zeta_k,$$

$$(3.11) \qquad Z(t) = \|x_0\|_1 + \ell \sum_{k \in \mathcal{P}} Y_k \left(\int_0^t \overline{C}(1 + Z(s))ds \right),$$

where $\{Y_k\}$ are independent unit-rate Poisson processes, and $\overline{C}$ is defined as in Assumption 1. Define $\mu_m = \inf\{t \geq 0 : Z(t) \geq m\}$. Then, so long as $m \geq \|x_0\|_1 + \ell$, Lemma 3.6 (and Remark 3.7) implies

$$(3.12) \qquad P(\mu_m \leq t) \leq C\delta^m,$$

where $c = \max\{\|x_0\|_1 + 1, \ell\}$ and

$$(3.13) \qquad C = (1 - e^{-ct\overline{C}K})^{-1-\frac{\|x_0\|_1}{\ell}}, \qquad \delta = \left(1 - e^{-ct\overline{C}K}\right)^{\frac{1}{\ell}}.$$

Turning to X^θ, we take the 1-norm and find

$$\|X^\theta(t)\|_1 = X^\theta(t) \cdot \vec{1} = x_0 \cdot \vec{1} + \sum_{k=1}^{K} Y_k\left(\int_0^t \lambda_k^\theta(X^\theta(s), s)ds\right)(\zeta_k \cdot \vec{1})$$

$$(3.14) \qquad \leq \|x_0\|_1 + \ell \sum_{k \in \mathcal{P}} Y_k\left(\overline{C}\int_0^t (1 + \|X^\theta(s)\|_1)ds\right).$$

Note that if we can show $\|X^\theta(t)\|_1 \leq Z(t)$ for all $t \geq 0$, then $\tau_m^\theta \geq \mu_m$, and (3.2) will be implied by (3.12). Hence the rest of the proof will consist of showing $\|X^\theta(t)\|_1 \leq Z(t)$ for all $t \geq 0$. We will proceed by induction, and the arguments will rely crucially on the coupling in (3.11).

Denote α_n as the nth jump time for X^θ due to reactions in $\mathcal{P}$. That is,

$$(3.15) \quad \alpha_n = \inf\{t > \alpha_{n-1} : X^\theta(t) = X^\theta(t-) + \zeta_k \text{ for some } k \in \mathcal{P}\}, \text{ with } \alpha_0 = 0.$$

1. Consider $n = 1$. Any reaction before time α_1 will not increase $\|X^\theta(t)\|_1$, since $\zeta_k \cdot \vec{1} \leq 0$ for any $k \notin \mathcal{P}$. On the other hand, $Z(t)$ is monotonically increasing, and hence $\|X^\theta(t)\|_1 \leq Z(t)$ for any $t < \alpha_1$. Consequently,

$$\|X^\theta(\alpha_1)\|_1 = \|x_0\|_1 + \sum_{k=1}^{K} Y_k\left(\int_0^{\alpha_1} \lambda_k^\theta(X^\theta(s), s)ds\right)(\zeta_k \cdot \vec{1})$$

$$(3.16) \qquad \leq \|x_0\|_1 + \ell \sum_{k \in \mathcal{P}} Y_k\left(\overline{C}\int_0^{\alpha_1} (1 + Z(s))ds\right) = Z(\alpha_1).$$

2. Now let $n \geq 2$, and suppose that $\|X^\theta(t)\|_1 \leq Z(t)$ for all $t \leq \alpha_{n-1}$. Then by the same reasoning as above, for any $t < \alpha_n$, $\|X^\theta(t)\|_1 \leq Z(t)$, and similarly to (3.16), $\|X^\theta(\alpha_n)\|_1 \leq Z(\alpha_n)$.

Since $Z(t)$ is nonexplosive and $\|X^\theta(t)\|_1 \leq Z(t)$, we have that $\alpha_n \to \infty$, as $n \to \infty$, and the proof for the particular value of θ chosen is now complete. Noting that the bound in (3.16) is uniform in θ completes the proof. $\square$

3.2. More lemmas. In this section, we will prove some technical lemmas which will be used in section 3.3 to prove the main result, Theorem 3.1. The next two lemmas are provided for completeness.

LEMMA 3.8. *Suppose $k \geq 1$ and $l \in \mathbb{Z}_{>0}$ are fixed and $\delta \in (0, 1)$. Then*

$$\sum_{n=1}^{\infty} n^k \delta^{(n^{1/\ell})} < \infty.$$

LEMMA 3.9. *Suppose $x \in [0,1]$ and $n \geq 2$. Then*

$$1 - (1-x)^n \leq nx.$$

To prove Theorem 3.1, we will need some qualitative properties of $N_{q_K}^{\theta,\varepsilon}(t)$ introduced in (2.10) for the coupled processes. The final lemma concerns the moments of $N_{q_K}^{\theta,\varepsilon}(t)$.

LEMMA 3.10. *Let $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network with d species and K reactions. Let $\{X^\theta\}$ be a family of stochastic models whose associated intensity functions $\{\lambda_k^\theta\}$ are parametrized by $\theta \in \Theta \subset \mathbb{R}_{>0}^\gamma$, for some $\gamma \in \mathbb{Z}_{>0}$, where Θ is compact. Assume $\{\lambda_k^\theta\}$ satisfies conditions (1) and (2) of Assumption 1. Fix $\theta \in \Theta$ and $\theta + \varepsilon \in \Theta$, and suppose $(X^{\theta+\varepsilon}, X^\theta)$ are coupled via the stacked coupling (2.8), with $N_{q_K}^{\theta,\varepsilon}(t)$ defined as in (2.10). Then for any $r \geq 1$, $\mathbb{E}[(N_{q_K}^{\theta,\varepsilon}(t))^r] < \infty$. Consequently, $N_{q_K}^{\theta,\varepsilon}(t) < \infty$ almost surely.*

Note that there is no requirement for ε to be small. Instead, we just need both parameters to be contained within Θ.

Proof. We adopt the notation from Lemma 3.3, where

$$\tau_m^\theta = \inf \left\{ t \geq 0 : \|X^\theta(t)\|_1 = \sum_{i=1}^d X_i^\theta(t) \geq m \right\},$$

and let $\tau_m = \min\{\tau_m^\theta, \tau_m^{\theta+\varepsilon}\}$. Then by Lemma 3.3, for m large enough,

$$P(\tau_m \leq t) \leq P(\tau_m^\theta \leq t) + P(\tau_m^{\theta+\varepsilon} \leq t) \leq 2C\delta^m.$$

The remainder of the proof focuses on bounding the tail probability $P(N_{q_K}^{\theta,\varepsilon}(t) \geq n)$, as $n \to \infty$. Define

$$n^* = n^{\frac{1}{p+1}},$$

where p is the order of the polynomial in condition (2) of Assumption 1. With this choice of n^*, by Lemma 3.3,

$$(3.17) \qquad P(N_{q_K}^{\theta,\varepsilon}(t) \geq n, \tau_{n^*} \leq t) \leq P(\tau_{n^*} \leq t) \leq 2C\delta^{n^{\frac{1}{p+1}}},$$

so long as n is large enough.

If $\tau_{n^*} > t$, then $\|X(s)\|_1 \leq n^*$ for all $s \in [0,t]$, and so by condition (2) of Assumption 1,

$$N_{q_K}^{\theta,\varepsilon}(t) \leq \int_{[0,t]\times[0,\infty)} 1_{\left[0, \sum_{k=1}^K \overline{C}(1+n^{\frac{p}{p+1}})\right)}(x) N(ds \times dx).$$

Denote the process on the right-hand side of the above equation by $Z_n(t)$, and note that it is a Poisson process with rate $\overline{C}K(1+n^{\frac{p}{p+1}})$. Since, $Z_n(t)$ is a Poisson random variable, we know

$$P(Z_n(t) \geq a) \leq e^{-a \log \frac{a}{\mathbb{E}[Z_n(t)]} + a - \mathbb{E}[Z_n(t)]}$$

so long as $a > \mathbb{E}[Z_n(t)]$. Therefore, for n large enough,

$$
\begin{aligned}
P\left(N_{q_K}^{\theta,\varepsilon}(t) \geq n, \tau_{n^*} > t\right) &\leq P\left(Z_n(t) \geq n, \tau_{n^*} > t\right) \\
&\leq P\left(Z_n(t) \geq n\right)
\end{aligned}
$$

$$
(3.18) \qquad\qquad \leq \exp\left(-n \log \frac{n}{t\overline{C}K(1+n^{\frac{p}{p+1}})} + n - t\overline{C}K(1+n^{\frac{p}{p+1}})\right).
$$

Note that the leading order on the exponent is $-n \log n$.

Collecting (3.17) and (3.18), we have

$$
\begin{aligned}
P(N_{q_K}^{\theta,\varepsilon}(t) \geq n) &= P(N_{q_K}^{\theta,\varepsilon}(t) \geq n, \tau_{n*} \leq t) + P(N_{q_K}^{\theta,\varepsilon}(t) \geq n, \tau_{n*} > t) \\
&\leq \exp\left(-n \log\left(\frac{n}{t\overline{C}K(1+n^{\frac{p}{p+1}})}\right) + n - t\overline{C}K(1+n^{\frac{p}{p+1}})\right) + 2C\delta^{n^{\frac{1}{p+1}}}.
\end{aligned}
$$

We therefore have

$$
\begin{aligned}
\mathbb{E}\left[\left(N_{q_K}^{\theta,\varepsilon}(t)\right)^r\right] &= \sum_{n=1}^{\infty} n^r P(N_{q_K}^{\theta,\varepsilon}(t) = n) \leq \sum_{n=1}^{\infty} n^r P(N_{q_K}^{\theta,\varepsilon}(t) \geq n) \\
&\leq \sum_{n=1}^{\infty} n^r \left[\exp\left(-n\log\left(\frac{n}{t\overline{C}K(1+n^{\frac{p}{p+1}})}\right) + n - t\overline{C}K(1+n^{\frac{p}{p+1}})\right) + 2C\delta^{n^{\frac{1}{p+1}}}\right] \\
&\leq \sum_{n=1}^{\infty} n^r \exp\left(-n\log\left(\frac{n}{t\overline{C}K(1+n^{\frac{p}{p+1}})}\right) + n - t\overline{C}K(1+n^{\frac{p}{p+1}})\right) + 2C\sum_{n=1}^{\infty} n^r \delta^{n^{\frac{1}{p+1}}}.
\end{aligned}
$$

The first series converges by the root test, and the second series converges by Lemma 3.8. □

3.3. Proof of the main result. We restate the main theorem for completeness.

THEOREM 3.1. *Let* $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ *be a reaction network with d species and K reactions. Let $\{X^\theta\}$ be a family of stochastic models whose associated intensity functions $\{\lambda_k^\theta\}$ are parametrized by $\theta \in \Theta \subset \mathbb{R}_{>0}^\gamma$, for some $\gamma \in \mathbb{Z}_{>0}$, where Θ is compact. Suppose that Assumption 1 holds and that $(X^\theta, X^{\theta+\varepsilon})$ are coupled via the stacked coupling (2.8) with $X^\theta(0) = X^{\theta+\varepsilon}(0) = x_0 \in \mathbb{Z}_{\geq 0}^d$. Then, for any θ in the interior of Θ and any $r \geq 1$, there is a $C_{r,t} > 0$ and $\overline{\varepsilon} \in \mathbb{R}_{>0}$ so that*

$$
\mathbb{E}\left[\|X^{\theta+\varepsilon}(t) - X^\theta(t)\|_1^r\right] \leq C_{r,t}\|\varepsilon\|_1,
$$

when $\|\varepsilon\|_1 \leq \overline{\varepsilon}$.

Proof. Denote $\mathcal{P} = \{k \in \{1, \ldots, K\} : \zeta_k \cdot \vec{1} > 0\}$, and let $\hat{\ell} = \max_{k \in \{1, \ldots, K\}}\{|\zeta_k| \cdot \vec{1}\}$.

Fix $t > 0$ and recall the counting process $N_{q_K}^{\theta,\varepsilon}(t)$ from (2.10). Denote by μ_i the time of the ith jump of $N_{q_K}^{\theta,\varepsilon}$. Next, let

$$
\beta_{\theta,\varepsilon} = \min_{i=1,2,\ldots,N_{q_K}^{\theta,\varepsilon}(t)}\{X^\theta(\mu_i) \neq X^{\theta+\varepsilon}(\mu_i)\},
$$

where the minimum of the empty set is taken to be infinity. Hence, $\beta_{\theta,\varepsilon}$ is the number of steps it took for X^θ and $X^{\theta+\varepsilon}$ to decouple or is infinity if they have not decoupled by time t.

Since $N_{q_K}^{\theta,\varepsilon}(t) < \infty$ almost surely by Lemma 3.10, the expectation can be calculated as

$$\mathbb{E}\left[\|X^{\theta+\varepsilon}(t) - X^{\theta}(t)\|_1^r\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[\|X^{\theta+\varepsilon}(t) - X^{\theta}(t)\|_1^r 1_{\{N_{q_K}^{\theta,\varepsilon}(t)=n\}}\right]$$

$$(3.19) \qquad = \sum_{n=1}^{\infty} \mathbb{E}\left[\|X^{\theta+\varepsilon}(t) - X^{\theta}(t)\|_1^r 1_{\{N_{q_K}^{\theta,\varepsilon}(t)=n\}} 1_{\{\beta_{\theta,\varepsilon} \leq n\}}\right]$$

$$\leq \sum_{n=1}^{\infty} (\hat{\ell}n)^r P(N_{q_K}^{\theta,\varepsilon}(t) = n, \beta_{\theta,\varepsilon} \leq n)$$

$$(3.20) \qquad = \sum_{n=1}^{\infty} \hat{\ell}^r n^r P(N_{q_K}^{\theta,\varepsilon}(t) = n) P(\beta_{\theta,\varepsilon} \leq n | N_{q_K}^{\theta,\varepsilon}(t) = n).$$

The last inequality is a consequence of Remark 2.2, since $X^{\theta+\varepsilon}(0) = X^{\theta}(0)$ and each jump recorded by either processes will increase their difference $\|X^{\theta+\varepsilon} - X^{\theta}\|_1$ by at most $\hat{\ell}$. Since we are indicating on the fact that there was precisely n jumps total, the total change in the 1-norm is at most $\hat{\ell}n$.

We turn to the conditional probability above and begin by noting that

$$(3.21) \qquad P(\beta_{\theta,\varepsilon} > n | N_{q_K}^{\theta,\varepsilon}(t) = n) = \prod_{i=1}^{n} P(\beta_{\theta,\varepsilon} > i | N_{q_K}^{\theta,\varepsilon}(t) = n, \beta_{\theta,\varepsilon} > i - 1).$$

Next, for each $i \in \{1, \ldots, n\}$, condition (3) of Assumption 1 gives

$$P(\beta_{\theta,\varepsilon} = i | N_{q_K}^{\theta,\varepsilon}(t) = n, \beta_{\theta,\varepsilon} > i - 1)$$
$$= \frac{\sum_{k=1}^{K} |\lambda_k^{\theta+\varepsilon}(X^{\theta}(\mu_{i-1}), \mu_i) - \lambda_k^{\theta}(X^{\theta}(\mu_{i-1}), \mu_i)|}{\sum_{k=1}^{K} \max\{\lambda_k^{\theta+\varepsilon}(X^{\theta}(\mu_{i-1}), \mu_i), \lambda_k^{\theta}(X^{\theta}(\mu_{i-1}), \mu_i)\}}$$
$$\leq \overline{C}\|\varepsilon\|_1.$$

Hence,

$$P\left(\beta_{\theta,\varepsilon} > i | N_{q_K}^{\theta,\varepsilon}(t) = n, \beta_{\theta,\varepsilon} > i - 1\right) \geq 1 - \overline{C}\|\varepsilon\|_1,$$

and (3.21) yields

$$P(\beta_{\theta,\varepsilon} > n | N_{q_K}^{\theta,\varepsilon}(t) = n) \geq (1 - \overline{C}\|\varepsilon\|_1)^n.$$

Hence, choosing ε so that $\overline{C}\|\varepsilon\|_1 < 1$,

$$P(\beta_{\theta,\varepsilon} \leq n | N_{q_K}^{\theta,\varepsilon}(t) = n) \leq 1 - (1 - \overline{C}\|\varepsilon\|_1)^n \leq \overline{C}n\|\varepsilon\|_1,$$

where the last inequality comes from Lemma 3.9. Plugging the above back into (3.20) yields

$$\mathbb{E}\left[\|X^{\theta+\varepsilon}(t) - X^{\theta}(t)\|_1^r\right] \leq \overline{C}\|\varepsilon\|_1 \sum_{n=1}^{\infty} \hat{\ell}^r n^{r+1} P(N_{q_K}^{\theta,\varepsilon}(t) = n)$$
$$= C\hat{\ell}^r\|\varepsilon\|_1 \mathbb{E}\left[\left(N_{q_K}^{\theta,\varepsilon}(t)\right)^{r+1}\right],$$

which is finite by Lemma 3.10.   □

The following corollary gives easy to check structural conditions on a mass-action network which guarantee the bound (3.1) holds.

COROLLARY 3.11. *Let $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a reaction network with d species and K reactions. Let $\{X^\theta\}$ be a family of stochastic models whose associated intensity functions $\{\lambda_k^\theta\}$ are given by stochastic mass-action kinetics (2.3). Suppose that $\theta = (\kappa_1, \ldots, \kappa_k) \in \Theta \subset \mathbb{R}_{>0}^K$, where Θ is compact. Suppose that only zeroth and first order reactions produce a net gain in molecules (i.e., those with $\zeta_k \cdot 1 > 0$) and that $(X^\theta, X^{\theta+\varepsilon})$ are coupled via the stacked coupling (2.8) with $X^\theta(0) = X^{\theta+\varepsilon}(0) = x_0 \in \mathbb{Z}_{\geq 0}^d$. Then, for any θ in the interior of Θ and any $r \geq 1$, there is a $C_{r,t} > 0$ and $\overline{\varepsilon} \in \mathbb{R}_{>0}$ so that*

$$\mathbb{E}\left[\|X^{\theta+\varepsilon}(t) - X^\theta(t)\|_1^r\right] \leq C_{r,t}\|\varepsilon\|_1,$$

when $\|\varepsilon\|_1 \leq \overline{\varepsilon}$.

Proof. The proof is immediate by Lemma 2.5 and Theorem 3.1. □

REFERENCES

[1] D. F. ANDERSON, *A modified next reaction method for simulating chemical systems with time dependent propensities and delays*, J. Chem. Phys., 127 (2007), 214107.

[2] D. F. ANDERSON, *An efficient finite difference method for parameter sensitivities of continuous time Markov chains*, SIAM J. Numer. Anal., 50 (2012), pp. 2237–2258.

[3] D. F. ANDERSON AND M. KOYAMA, *An asymptotic relationship between coupling methods for stochastically modeled population processes*, IMA J. Numer. Anal., 35 (2015), pp. 1757–1778.

[4] D. F. ANDERSON AND T. G. KURTZ, *Continuous Time Markov Chain Models for Chemical Reaction Networks*, Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology. Springer, New York, 2011.

[5] D. F. ANDERSON AND T. G. KURTZ, *Stochastic Analysis of Biochemical Systems*, Springer, New York, 2015.

[6] D. F. ANDERSON, T. SEPPÄLÄINEN, AND B. VALKÓ, *Introduction to Probability*, Cambridge University Press, Cambridge, UK, 2017.

[7] D. F. ANDERSON AND E. S. WOLF, *A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks*, J. Chem. Phys., 137 (2012), 224112.

[8] D. F. ANDERSON AND E. S. WOLF, *Hybrid pathwise sensitivity methods for discrete stochastic models of chemical reaction systems*, J. Chem. Phys., 142 (2015), 034103.

[9] D. F. ANDERSON AND C. YUAN, *Low variance couplings for stochastic models of intracellular processes with time-dependent rate functions*, Bull. Math. Biol., (2018), pp. 1–29.

[10] M. BIBINGER, *Notes on the Sum and Maximum of Independent Exponentially Distributed Random Variables with Different Scale Parameters*, preprint, arXiv:1307.3945, 2013.

[11] M. B. ELOWITZ, A. J. LEVINE, E. D. SIGGIA, AND P. S. SWAIN, *Stochastic gene expression in a single cell*, Science, 297 (2002), pp. 1183–1186.

[12] S. ENGBLOM, *On the stability of stochastic jump kinetics*, App. Math., 5 (2014), pp. 3217–3239.

[13] X.-J. FENG, S. HOOSHANGI, D. CHEN, G. LI, R. WEISS, AND H. RABITZ, *Optimizing genetic circuits by global sensitivity analysis*, Biophy. J., 87 (2004), pp. 2195–2202.

[14] 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.

[15] P. GLASSERMAN AND Y. C. HO, *Gradient Estimation via Perturbation Analysis*, Vol. 116, Springer Science & Business Media, 1991.

[16] A. GUPTA, C. BRIAT, AND M. KHAMMASH, *A scalable computational framework for establishing long-term behavior of stochastic reaction networks*, PLoS Comput. Biol., 10 (2014), e1003669.

[17] A. GUPTA AND M. KHAMMASH, *Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks*, SIAM J. Sci. Comput., 35 (2013), pp. A2598–A2620.

[18] M. KOMOROWSKI, M. J. COSTA, D. A. RAND, AND M. P. STUMPF, *Sensitivity, robustness, and identifiability in stochastic chemical kinetics models*, Proc. Nat. Acad. Sci., 108 (2011), pp. 8645–8650.

[19] T. G. KURTZ, *Representations of Markov processes as multiparameter time changes*, Ann. Prob., 8 (1980), pp. 682–715.

[20] J. PAULSSON, *Summing up the noise in gene networks*, Nature, 427 (2004), p. 415.

[21] S. PLYASUNOV AND A. P. ARKIN, *Efficient stochastic sensitivity analysis of discrete event systems*, J. Comp. Phys., 221 (2007), pp. 724–738.

[22] A. RAJ, C. S. PESKIN, D. TRANCHINA, D. Y. VARGAS, AND S. TYAGI, *Stochastic MRNA synthesis in mammalian cells*, PLoS Biol., 4 (2006), e309.

[23] M. RATHINAM, *Moment growth bounds on continuous time Markov processes on non-negative integer lattices*, Quart. Appl. Math., 73 (2015), pp. 347–364.

[24] M. RATHINAM, P. W. SHEPPARD, AND M. KHAMMASH, *Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks*, J. Chem. Phys., 132 (2010), 034103.

[25] R. SRIVASTAVA, D. F. ANDERSON, AND J. B. RAWLINGS, *Comparison of finite difference based methods to obtain sensitivities of stochastic chemical kinetic models*, J. Chem. Phys., 138 (2013), 074110.

[26] J. STELLING, E. D. GILLES, AND F. J. DOYLE, *Robustness properties of circadian clock architectures*, Proc. Nat. Acad. Sci., 101 (2004), pp. 13210–13215.

[27] V. H. THANH, R. ZUNINO, AND C. PRIAMI, *Efficient finite difference method for computing sensitivities of biochemical reactions*, Proc. Roy. Soc. A, 474 (2018), 20180303.

[28] D. J. WILKINSON, *Stochastic Modelling for Systems Biology*, CRC Press, Boca Raton, FL, 2018.