SIAM/ASA J. Uncertainty Quantification
Vol. 11, No. 4, pp. 1164–1194

Sensitivity Analysis of Quasi-stationary distributions (QSDs) of Mass-action Systems*

Yao Li† and Yaping Yuan†

Abstract. This paper studies the sensitivity analysis of mass-action systems against their diffusion approximations, particularly the dependence on population sizes. As a continuous-time Markov chain, a mass-action system can be described by an equation driven by finitely many Poisson processes, which has a diffusion approximation that can be pathwisely matched. The magnitude of noise in mass-action systems is proportional to the square root of the molecule count/population, which makes a large class of mass-action systems have quasi-stationary distributions (QSDs) besides invariant probability measures. In this paper, we modify the coupling-based technique developed in [M. Dobson, Y. Li, and J. Zhai, *SIAM/ASA J. Uncertain. Quantif.*, 9 (2021), pp. 135–162] to estimate an upper bound of the 1-Wasserstein distance between two QSDs. Some numerical results of sensitivity with different population sizes are provided.

Key words. quasi-stationary distribution, law of mass action, sensitivity analysis, Monte Carlo simulation

MSC codes. 60J28, 60J60, 65C50

DOI. 10.1137/22M1535875

1. Introduction.

A mass-action network is a system of finitely many species and reactions whose update rule satisfies the mass-action law. The term "mass-action network" refers to a large number of the chemical reaction networks, epidemiology models, and population models. At the molecule level, reactions in the mass-action network are random events that modify the state of the network according to stoichiometric equations. The time of these random events satisfies mass-action laws. Therefore, a mass-action network can be mathematically described by a continuous-time Markov process, which is driven by a finite number of Poisson processes.

The randomness in updating the network is called demographic noise in population and epidemiology models. It is well known that demographic noise leads to finite time extinction in a large class of population models (see, for example, the discussion in section 3.1). This is because the magnitude of the demographic noise is proportional to the population size. As a result, when the population is small, in many mass-action systems, noise could become the dominant term and lead to finite time extinction with strictly positive probability. Therefore, the asymptotic property of the mass-action network with finite time extinction is usually described by the quasi-stationary distribution (QSD), which is the conditional limiting

*Received by the editors November 18, 2022; accepted for publication (in revised form) May 30, 2023; published electronically October 20, 2023.
https://doi.org/10.1137/22M1535875
Funding: The work of the first author was partially supported by NSF DMS-1813246 and DMS-2108628.
†Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01002 USA (yaoli@math.umass.edu, yuan@math.umass.edu).

distribution conditioning on not hitting the absorbing set yet. As discussed in [19], when the extinction rate is low, the QSD can be well approximated by the invariant probability measure of a modified process that artificially "pushes" the trajectory away from the extinction.

For decades, it has been known that when the population size is large, the continuous-time Markov process converges into the mass-action ordinary differential equations (ODEs). In addition, by setting up a martingale problem, one can show that the rescaled difference between the continuous-time Markov process and the mass-action ODE converges to a stochastic differential equation. Therefore, at any finite time, the continuous-time Markov process of a mass-action network is approximated by a stochastic differential equation. This is called the diffusion approximation of a mass-action network. For further details, we refer the reader to [2, 10]. In general, let V denote the volume of the mass-action system; the finite time error of the diffusion approximation is $O(V^{-1}\log V)$. However, the finite time error estimation does not translate to the distance between QSDs immediately. The reason is that the sensitivity of the invariant probability measure or the QSD depends on both the finite time error and the rate of contraction of the transition kernel. (See [8] or section 3 of this paper.) However, as $V \to \infty$ the noise term in the diffusion approximation vanishes. Giving a uniform bound of the spectral gap of the Fokker–Planck operator with a vanishing diffusion term is known to be very challenging. In fact, a uniform bound of the speed of contraction with respect to growing V may not even exist. For example, when the mass-action system is bi-stable, the speed of contraction can decrease exponentially quickly $\sim \exp(-cV)$ (c is a positive constant) as V increases.

The goal of this paper is to study the sensitivity of QSDs against diffusion approximation—specifically, how the QSDs of the Markov process and its diffusion approximation differ from each other. The motivation for this study is that simulating at the molecule level can be computationally expensive, even with optimal implementation of the stochastic simulation algorithm (SSA) [11, 17, 26]. Moreover, it is even more challenging to numerically compute the QSD when the number of molecules is large. In contrast, simulating a diffusion process is much easier, and the technique for computing the invariant probability measure or QSD of a stochastic differential equation is also well developed [18, 19, 29]. As stated above, because the noise term in the diffusion approximation vanishes as V increases, it is difficult to give a rigorous bound of the distance between two QSDs. Even in the simple case when a rigorous estimate is possible, the pre-factor is also unknown and difficult to estimate. Hence, it is important to have a computable quantitative upper bound of the difference between the QSD of a mass-action system and that of its diffusion approximation.

The method of sensitivity analysis is developed from the coupling-based method in [8], which requires both the finite time truncation error and the rate of contraction of the transition kernel of the diffusion process. The finite time error is modified from the Komlós–Major–Tusnády (KMT) algorithm in [24]. With the explicit construction of coupled trajectories of the Poisson process and the diffusion process, we can compute the finite time error up to fixed time T. The main modification from the original KMT algorithm is the adaption of regenerations. We need to couple the regeneration of the Poisson process and that of its diffusion approximation in order to make them stay together after a regeneration. The rate of contraction is modified from the data-driven method proposed in [18]. We design a suitable coupling scheme for the modified diffusion process that regenerates from the QSD right after

hitting the absorbing set. Because of the coupling inequality, the exponential tail of the coupling time can be used to estimate the rate of contraction. The sensitivity analysis is demonstrated by several numerical examples. Generally speaking, the distance between two processes is much larger for smaller volumes (i.e., molecule count).

The organization of this paper is as follows. Preliminaries of reaction networks, rates for the law of mass action, the Poisson process, the diffusion process, and coupling times are provided in section 2. Section 3 introduces the algorithms for computing the finite time error and the rate of contraction in two different cases. All numerical examples are demonstrated in section 4. Section 5 is the conclusion. All explicit expressions of the Poisson process and the diffusion process are shown in the appendix.

2. Preliminaries.

2.1. Stochastic mass reaction networks and Poisson process.

2.1.1. Stochastic mass reaction networks. We consider a mass-action network of K reactions involving d distinct species, $S_1, \cdots, S_d$,

$$\sum_{i=1}^{d} c_{ki} S_i \to \sum_{i=1}^{d} c'_{ki} S_i, \quad k = 1, \cdots, K, \tag{2.1}$$

where c_{ki} and c'_{ki} are nonnegative integers that represent the number of molecules of species S_i consumed and produced by reaction k, respectively. Let V be the volume of the reaction system. The state of the mass-action system at time t is denoted by $X(t) = (x_1(t), \cdots, x_d(t)) \in \mathbb{R}^d$, where the ith entry of $X(t)$ indicates the concentration of species S_i, $i = 1, \cdots, d$, i.e., the number of molecules of S_i is given by $Vx_i := N_i$. The rate at which the kth reaction occurs is denoted by λ_k, which gives the propensity of the kth reaction as a function of the concentrations of molecules of the chemical species. In other words, λ_k indicated the likelihood of the kth reaction occurring based on the current concentrations.

2.1.2. Rates for the law of mass action. The law of mass action states that the rate of a chemical reaction should be proportional to the product of the concentrations of the reactants, with each concentration raised to a power equal to the number of molecules of the corresponding reactant in the balanced chemical equation. More precisely, the rate of reaction k can be written as

$$\lambda_k = \kappa_k V \prod_{i=1}^{d} \left(\frac{N_i}{V} \right)^{c_{ki}} := V f_k(X),$$

where κ_k is a rate constant, and N_i denotes the number of molecules of the ith species in the system. Let $\Delta t \ll 1$ be a very short time period. More precisely, given all information of the system up to time t, we have

$$\mathbb{P}[\text{ reaction } k \text{ occurs in } [t, t + \Delta t)] = \lambda_k \Delta t + O(\Delta t^2).$$

2.1.3. Poisson process. We use a Poisson counting process to represent $X(t)$, since it is a continuous-time Markov chain with discrete states. Let $X_i(t)$ be the ith entry of $X(t)$; then

$$X_i(t) = X_i(0) + \frac{1}{V}\sum_k R_k(t)(c'_{ki} - c_{ki}),$$

where $R_k(t)$ is the number of times the reaction k has occurred by time t and $R_k(0) = 0$. Because the number of molecules of species changes with time, $R_k(t)$ is an inhomogeneous Poisson process that is given by

$$R_k(t) = P_k\left(V\int_0^t f_k(X(s))ds\right), \tag{2.2}$$

where $P_k(\cdot)$ is a unit-rate Poisson point process. It is well known that $P_k(\cdot)$ satisfies the following three properties: (1) $P_k(0) = 0$, (2) $P_k(\cdot)$ has independent increments, and (3) $P_k(s+t) - P_k(s)$ is a Poisson random variable with parameter t. And the whole system is given by

$$X(t) = X(0) + \sum_k \frac{l_k}{V} P_k\left(V\int_0^t f_k(X(s))ds\right), \tag{2.3}$$

where $P_k(t)$, $k = \{1, \cdots, K\}$, are independent unit-rate Poisson processes, and $l_k = c'_k - c_k \in \mathbb{R}^d$ denotes the coefficient change of molecules at reaction k.

2.2. Diffusion process. When V is large, the Poisson process can be approximated by a diffusion process. The following strong approximation theorem from [15, 16] provides a bound on the strong approximation error.

Lemma 2.1. *A unit Poisson process $P(\cdot)$ and a Wiener process $B(\cdot)$ can be constructed so that*

$$\left|\frac{P(Vt) - Vt}{\sqrt{V}} - \frac{1}{\sqrt{V}}B(Vt)\right| \le \frac{\log(Vt \vee 2)}{\sqrt{V}}\Gamma,$$

where Γ is a random variable such that $\mathbb{E}(e^{c\Gamma}) < \infty$ for some constant $c > 0$.

Remark 2.2. By the scaling property of the Wiener process, $\frac{1}{\sqrt{V}}B(Vt)$ is also a standard Wiener process.

With the lemma above and Ito's formula, we have the diffusion approximation

$$\begin{aligned} P_k\left(V\int_0^t f_k(X(s))ds\right) &\approx V\int_0^t f_k(X(s))ds + \int_0^t \sqrt{V f_k(X(s))}dB(s) \\ &= V\int_0^t f_k(X(s))ds + B_k\left(V\int_0^t f_k(X(s))ds\right). \end{aligned}$$

This gives the diffusion approximation of the mass-action system $X(t)$:

$$Y(t) = Y(0) + \sum_k \frac{l_k}{V}\left[V\int_0^t f_k(Y(s))ds + B_k\left(V\int_0^t f_k(X(s))ds\right)\right].$$

In the chemistry literature, Y is known as the Langevin approximation for the continuous-time Markov chain model. Theoretically, the distance between these two approximations is bounded according to the theorem in [24].

Theorem 2.3. *Let $X(t)$ be a Poisson process represented by (2.3), and let $Y(t)$ be a diffusion process with initial condition satisfying $X(0) = Y(0)$ and solving the following stochastic differential equation:*

$$Y(t) = Y(0) + \sum_k \frac{l_k}{V}\left[V\int_0^t f_k(Y(s))ds + B_k\left(V\int_0^t f_k(Y(s))ds\right)\right], \tag{2.4}$$

where the $B_k(\cdot)$ are independent standard Wiener processes. Let U be an open connected set that contains the deterministic trajectory

$$y(t) = y(0) + \sum_k l_k \int_0^t f_k(y(s))ds$$

for all $0 \le t \le T$. Assume all f_k are Lipschitz continuous on U. Let $\tau_V = \inf\{t \,|\, X(t) \notin U \text{ or } Y(t) \notin U\}$. For any sufficiently large V large, we have

$$\sup_{0\le t\le \tau_V \wedge T} |X(t) - Y(t)| = O\left(\frac{\log V}{V}\right) \tag{2.5}$$

for any fixed time horizon T.

Note that the probability that $\tau_V > T$ approaches 1 as $V \to \infty$. Hence, for sufficiently large V, the error of diffusion approximation is proportional to $\frac{\log V}{V}$, which converges to 0 as $V \to \infty$. In macroscopic chemical reaction systems, V is at the magnitude of Avogadro's number. Therefore, the entire diffusion term can be safely ignored. However, in many ecologic systems or cellular chemical reaction systems, the effective volume cannot be simply treated as infinity. This motivates us to investigate the sensitivity of the quasi-stationary distributions (QSDs) against the diffusion approximation. For any finite capacity V, the finite time error of the diffusion approximation can be explicitly simulated. In [24], a constructive procedure is presented for generating discretized trajectories of the two processes $X(t)$ and $Y(t)$ on the same probability space, such that they stay close to each other trajectory by trajectory with probability one. We apply the algorithm to compute the finite time error in section 3.

2.3. Coupling times. In this paper, we use the coupling argument to relate finite time error with the distance between QSDs. Let μ and ν be two probability measures on a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. A *coupling* between μ and ν is a probability measure γ on the product space $(\mathcal{X}\times\mathcal{X}, \mathcal{B}(\mathcal{X})\times\mathcal{B}(\mathcal{X}))$ such that two marginal distributions of γ are μ and ν, respectively.

Definition 2.4 (Wasserstein distance). *Let d be a metric on the state space $\mathcal{X}$. For probability measures μ and ν on $\mathcal{X}$, the Wasserstein distance between μ and ν for d is given by*

$$\begin{aligned} d_w(\mu,\nu) &= \inf\{\mathbb{E}_\gamma[d(x,y)] : \gamma \textit{ is a coupling of } \mu \textit{ and } \nu\} \\ &= \inf\left\{\int d(x,y)\gamma(dx,dy) : \gamma \textit{ is a coupling of } \mu \textit{ and } \nu\right\}. \end{aligned} \tag{2.6}$$

In this paper, without further specification, we assume that $\mathcal{X}$ is equipped with the norm $\|\cdot\|$ and the 1-Wasserstein distance is induced by $d(x,y) = \min\{1, \|x-y\|\}$.

Let $Z_t^{(1)}$ and $Z_t^{(2)}$ be two stochastic processes on the same state space, where $t \in \mathbb{T}$, $\mathbb{T} = \mathbb{R}_+$ or $\mathbb{T} = \mathbb{Z}_+$. A coupling between $Z_t^{(1)}$ and $Z_t^{(2)}$ can be defined in the same way on the space of paths. The first time when $Z_t^{(1)}$ meets $Z_t^{(2)}$ is called the coupling time.

Definition 2.5 (coupling time). *The coupling time τ_c of a coupling $(Z_t^{(1)}, Z_t^{(2)})$ is a random variable in $\mathbb{T}$ such that*

$$\tau_c \overset{\text{def}}{=} \inf\left\{t \geq 0 \,|\, Z_{t+s}^{(1)} = Z_{t+s}^{(2)} \text{ for all } s \geq 0\right\}. \tag{2.7}$$

Throughout this paper, we assume $Z_{t+s}^{(1)} = Z_{t+s}^{(2)}$ for all $s > 0$ if $Z_t^{(1)} = Z_t^{(2)}$. In other words, $Z_t^{(1)}$ and $Z_t^{(2)}$ stay together after their first meet. Note that all coupled Markov processes satisfy this assumption.

Definition 2.6 (successful coupling). *A coupling $(Z_t^{(1)}, Z_t^{(2)})$ of processes $Z^{(1)}$ and $Z^{(2)}$ is said to be successful if*

$$\mathbb{P}(\tau_c < \infty) = 1. \tag{2.8}$$

We use the following reflection coupling to couple two diffusion processes when they are far away from each other.

Definition 2.7 (reflection coupling). *Let $Z_t^{(1)}$ and $Z_t^{(2)}$ be two solutions of a stochastic differential equation*

$$\mathrm{d}Z_t = f(Z_t)\mathrm{d}t + \sigma(Z_t)\mathrm{d}B_t$$

when starting from different initial distributions. A reflection coupling of $Z_t^{(1)}$ and $Z_t^{(2)}$ is made by reflecting the noise term about the orthogonal hyperplane at the midpoint between $Z_t^{(1)}$ and $Z_t^{(2)}$:

$$\begin{aligned} \mathrm{d}Z_t^{(1)} &= f\left(Z_t^{(1)}\right)\mathrm{d}t + \sigma\left(Z_t^{(1)}\right)\mathrm{d}B_t, \\ \mathrm{d}Z_t^{(2)} &= f\left(Z_t^{(2)}\right)\mathrm{d}t + \sigma\left(Z_t^{(2)}\right)\left(I - 2\mathbf{e}\mathbf{e}^T\right)\mathrm{d}B_t, \end{aligned} \tag{2.9}$$

where B is a standard Wiener process, and

$$\mathbf{e} = \frac{1}{\|\sigma^{-1}\left(Z_t^{(1)} - Z_t^{(2)}\right)\|}\sigma^{-1}\left(Z_t^{(1)} - Z_t^{(2)}\right)$$

is a unit vector.

We remark that the reflection coupling requires $\sigma(Z_t)$ in (2.9) to be an invertible matrix. This is often not satisfied in the diffusion approximation (2.4) because the number of Wiener processes in (2.4) may not match the number of relations, making $\sigma(Z_t)$ not invertible. Hence, it is necessary to find an equivalent diffusion process with an invertible σ. See numerical examples for additional details.

The following maximal coupling is used to couple two processes that are close to each other.

Definition 2.8 (maximal coupling). *A maximal coupling between two probability distributions μ and ν on the state space $\mathcal{X}$ is a pair of random variables $(Z^{(1)}, Z^{(2)})$ that maximizes $\mathbb{P}[Z^{(1)} = Z^{(2)}]$, subject to the constraint that the law of $(Z^{(1)}, Z^{(2)})$ is a coupling of μ and ν.*

In the context of couplings of Markov processes, the maximal coupling looks for the maximal coupling probability for the next step of $Z_t^{(1)}$ and $Z_t^{(2)}$. Assume $Z_{t-1}^{(1)}$ and $Z_{t-1}^{(2)}$ are known

Algorithm 1 Maximal coupling

Input: $Z_{t-1}^{(1)}$ and $Z_{t-1}^{(2)}$
Output: $Z_t^{(1)}$ and $Z_t^{(2)}$, and τ_c if coupled
  Compute probability density functions $p^{(1)}(z)$ and $p^{(2)}(z)$
  Sample $Z_t^{(1)}$ and calculate $r = \mathcal{U}p^{(1)}(Z_t^{(1)})$, where $\mathcal{U}$ is uniformly distributed on [0,1]
  if $r < p^{(2)}(Z_t^{(1)})$ **then**
    $Z_t^{(2)} = Z_t^{(1)}, \tau_c = t$
  else
    Sample $Z_t^{(2)}$ and calculate $r' = \mathcal{V}p^{(2)}(Z_t^{(2)})$, where $\mathcal{V}$ is uniformly distributed on [0,1]
    while $r' < p^{(1)}(Z_t^{(2)})$ **do**
      Resample $Z_t^{(2)}$ and $\mathcal{V}$. Recalculate $r' = \mathcal{V}p^{(2)}(Z_t^{(2)})$
    end while
    τ_c is still undetermined
  end if

and the probability density function of $Z_t^{(1)}$ and $Z_t^{(2)}$ is easy to compute. Following [13, 14], the update of $Z_t^{(1)}$ and $Z_t^{(2)}$ in Algorithm 1 maximizes the probability of coupling.

2.4. Paired trajectories of Poisson process and of the diffusion process. Recall that according to Lemma 2.1 a unit-rate Poisson process has a strong diffusion approximation. Hence (2.3) also has a strong approximation given by (2.4). As the processes $P_k(\cdot)$ and $B_k(\cdot)$ are continuous time processes, we apply the τ-leaping approximation for (2.3) with the same step size h. This gives

$$\hat{X}_{n+1} = \hat{X}_n + \sum_k \frac{l_k}{V}\left[P_k\left(Vh\sum_{m=0}^{n} f_k(\hat{X}_m)\right) - P_k\left(Vh\sum_{m=0}^{n-1} f_k(\hat{X}_m)\right)\right] \tag{2.10}$$

with $\hat{X}_0 = X_0$. Similarly, the discretized approximation of (2.4) using the Euler–Maruyama method reads

$$\begin{aligned}\hat{Y}_{n+1} = \hat{Y}_n &+ \sum_k \frac{l_k}{V}(Vhf_k(\hat{Y}_n)) \\ &+ \sum_k \frac{l_k}{V}\left[B_k\left(Vh\sum_{m=0}^{n} f_k(\hat{Y}_m)\right) - B_k\left(Vh\sum_{m=0}^{n-1} f_k(\hat{Y}_m)\right)\right]\end{aligned} \tag{2.11}$$

with initial condition $\hat{Y}_0 = Y_0$.

The paired trajectories of $P_k(t)$ and $B_k(t)$ can be numerically generated by applying the KMT algorithm. The KMT algorithm generates a sequence of standard Poisson random variables $\{P_n\}$ and a sequence of standard normal random variables $\{W_n\}$, such that $\sum_{n=1}^{N} P_n$ is approximated by $N + \sum_{n=1}^{N} W_n$ for each finite N. Then after a rescaling, one obtains a pair of discretized trajectories of $P_k(t)$ and $B_k(t)$, respectively. We refer the reader to [24] for a detailed review of the KMT algorithm.

3. Sensitivity of QSD for diffusion approximations.

3.1. Quasi-stationary distribution. Let $\mathbf{X} = \{X(t)\}$ and $\hat{\mathbf{X}} = \{\hat{X}_n\}$ (resp., $\mathbf{Y} = \{Y(t)\}$ and $\hat{\mathbf{Y}} = \{\hat{Y}_n\}$) be the stochastic process given by (2.3) (resp., (2.4)) and a numerical approximation with step size h, respectively. Needless to say, a diffusion process is much easier to study than a Poisson process with jumps. One natural question here is how much the long-time dynamics of $\mathbf{X}$ are preserved by its diffusion approximation. This problem is more complicated than it looks because both $\mathbf{X}$ and $\mathbf{Y}$ have natural domain $\mathbb{R}^d_+$. When the number of molecules of one species reaches 0, the process exits from its domain due to extinction. It is common for (2.3) or (2.4) to have finite time extinction. To see this, consider the 1D version of (2.4):

$$\mathrm{d}Y(t) = f(Y(t))\mathrm{d}t + \frac{1}{\sqrt{V}}\sqrt{f(Y(t))}\mathrm{d}B_t\,. \tag{3.1}$$

Let $H(x) = x^{-1}$ be a test function. Applying Ito's formula and then taking the expectation, we have

$$\begin{aligned}\mathbb{E}[\mathrm{d}H(Y(t))] &= \mathbb{E}\left[-\frac{1}{Y(t)^2}f(Y(t))\mathrm{d}t + \frac{1}{V}\sqrt{f(Y(t))}\mathrm{d}B_t + \frac{f(Y(t))}{2V}\frac{2}{Y(t)^3}\mathrm{d}t\right]\\ &= f(Y(t))\left(\frac{1}{VY(t)^3} - \frac{1}{Y(t)^2}\right)\mathrm{d}t\,.\end{aligned}$$

If $f(Y(t)) = cY(t)$ for a constant c, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[H(Y(t))] \geq \frac{c}{V}\left(\mathbb{E}[H(Y(t))]\right)^2,$$

which blows up to ∞ in finite time. Hence $Y(t)$ has a strictly positive extinction probability in finite time. The calculation above fits the setting of many mass-action systems.

To prevent finite-time extinction, a constant influx of each species is usually needed, such as the artificial influx proposed by [28]. This is why it is often necessary to study the QSD instead of the invariant probability distribution. Below, we introduce the QSD and its sampling method only for $\mathbf{X}$, as the case for $\mathbf{Y}$ is analogous.

Let $\partial\mathcal{X} = \mathbb{R}^d \setminus \mathbb{R}^d_+$ be the absorbing set of $\mathbf{X}$. The QSD is an invariant probability measure conditioning on $\mathbf{X}$ that has not hit the absorbing set yet. We further define

$$\tau_X = \inf\{t > 0 : X(t) \in \partial\mathcal{X}\}$$

as the first passage time to $\partial\mathcal{X}$.

Definition 3.1. *A probability measure μ on $\mathbb{R}^d_+$ is called a quasi-stationary distribution (QSD) of the Markov process $\mathbf{X}$ with an absorbing set $\partial\mathcal{X}$ if for every measurable set $C \subset \mathbb{R}^d_+$*

$$\mathbb{P}_\mu[X(t) \in C | \tau_X > t] = \mu(C), \quad t \geq 0. \tag{3.2}$$

Definition 3.2. *A probability measure μ is called a quasi-limiting distribution (QLD) if*

$$\lim_{t\to\infty} \mathbb{P}_x[X(t) \in C | \tau_X > t] = \mu(C) \quad \forall x \in \mathbb{R}^d_+. \tag{3.3}$$

Remark 3.3. The limiting probability measure given by (3.3), or the QLD, is also called the Yaglom limit. A QLD must be a QSD. Under some mild assumptions about ergodicity, a QSD is also a QLD [5].

If the first passage time of $\boldsymbol{X}$ to $\partial\mathcal{X}$ is ∞ with probability one, $\{\tau_X > t\}$ is the full probability space. As a result, the QSD in (3.2) becomes the invariant probability measure, and the QLD in (3.3) becomes the limiting probability measure (which is also invariant). Therefore, when the mass-action system admits an invariant probability measure instead of a QSD, all our arguments and algorithms still apply.

When we define the numerical processes (2.10) and (2.11), we need to specify the regeneration rule such that they both sample from QSDs as the time approaches infinity. To sample from QSD, we need to regenerate a sample once it hits the absorbing set. Therefore, in addition to $\hat{X}_n$, we also need to update a temporal occupation measure

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{\hat{X}_k}.$$

If the numerical scheme yields $\hat{X}_{n+1} \in \partial\mathcal{X}$, we immediately regenerate $\hat{X}_{n+1}$ from μ_n. Specifically, we use $\hat{Q}$ to represent the transition kernel of the numerical scheme of $\hat{X}_n$ (without resampling), which has an absorbing set $\partial\mathcal{X}$ with $\hat{Q}(\partial\mathcal{X}, \partial\mathcal{X}) = 1$. The transition kernel of $\hat{X}_n$ is the sum of $\hat{Q}$ and the regeneration measure such that

$$\mathbb{P}[\hat{X}_{n+1} \in A \,|\, \hat{X}_n = x] = \hat{Q}(x, A) + \hat{Q}(x, \partial\mathcal{X})\mu_n(A).$$

The following convergence result follows from [3].

Proposition 3.4 (Theorem 2.5 in [3]). *Let $\hat{\pi}_X$ be the QSD of the numerical process $\hat{X}_n$. Under suitable assumptions about $\hat{X}_n$, the occupation measure μ_n converges to the QSD $\hat{\pi}_X$ as $n \to \infty$.*

To study the sensitivity of diffusion approximation, we also need a theoretical process $\tilde{\mathbf{X}} = \{\tilde{X}_n\}$ that regenerates directly from the QSD $\hat{\pi}_X$ once it exits to the boundary. Recall that $\hat{Q}$ is the transition kernel of $\hat{X}_n$ (without resampling). The transition kernel of $\tilde{\mathbf{X}}$ is

$$\tilde{P}(x, \cdot) = \hat{Q}(x, \cdot) + \hat{Q}(x, \partial\mathcal{X})\hat{\pi}_X(\cdot)$$

for all $x \in \mathbb{R}^d_+$. Note that $\hat{X}_n$ is not a Markov process (but $(\hat{X}_n, \mu_n)$ is a Markov process). But $\tilde{\mathbf{X}}$ is a homogeneous Markov process with an invariant probability measure $\hat{\pi}_X$. The case of $Y(t)$ is analogous. We denote the numerical process that resamples from a temporal occupation measure by $\hat{\mathbf{Y}} = \{\hat{Y}_n\}$, and the Markov process that directly resamples from QSD by $\tilde{\mathbf{Y}} = \{\tilde{Y}_n\}$. The QSD of $\hat{\mathbf{Y}}$, as well as the invariant probability measure of $\tilde{\mathbf{Y}}$, is denoted by $\hat{\pi}_Y$.

3.2. Decomposition of error terms. Let P_X and $\hat{P}_X$ be the transition kernels of $X(t)$ and $\hat{X}_n$, respectively. Let P_Y and $\hat{P}_Y$ be that of $Y(t)$ and $\hat{Y}_n$, respectively. Denote the QSDs of $X(t)$, $\hat{X}_n$, $Y(t)$, and $\hat{Y}_n$ by π_X, $\hat{\pi}_X$, π_Y, and $\hat{\pi}_Y$, respectively. The quantity that we are interested in is $d_w(\pi_X, \pi_Y)$. But only the distance between numerical QSDs, i.e., $d_w(\hat{\pi}_X, \hat{\pi}_Y)$, is computable.

The following decomposition follows easily by the triangle inequality.

$$d_w(\pi_X, \pi_Y) \le d_w(\pi_X, \hat{\pi}_X) + d_w(\hat{\pi}_X, \hat{\pi}_Y) + d_w(\hat{\pi}_Y, \pi_Y)\,. \tag{3.4}$$

The sensitivity of invariant probability against time discretization has been addressed by [21] decades ago. A computable upper bound is given by [8]. In general, assuming some ergodic conditions, when the at time step size of h the discretization is sufficiently small, the invariant probability measure of a stochastic differential equation, denoted by ν, is approximated by the numerical invariant probability measure, denoted by ν^h, with an error $O(h)$. The case of QSD is analogous. The third term $d_w(\pi_Y, \hat{\pi}_Y)$ is proportional to step size h under suitable ergodicity and regularity conditions.

For a mass-action system on finite state space, the estimation of the first term $d_w(\pi_X, \hat{\pi}_X)$ can be obtained by some linear algebraic calculations.

Theorem 3.5. *Let $X(t)$ be a continuous-time Markov chain on finite state space and $\hat{X}$ be its tau-leaping approximation with step size h. Suppose that π and $\hat{\pi}$ are the true QSD and the numerical approximation of the QSD, respectively. If the generating matrix of $X(t)$ is irreducible, then*

$$\|\pi - \hat{\pi}\| = O(h)$$

for $0 < h \ll 1$.

Proof. This proof follows the standard argument of the eigenvector perturbation result. The case of stationary distribution is proved in [22]. Here we follow the argument in [6] to prove a similar result for QSDs. Let Q be the generating matrix of $X(t)$. Because π is true QSD and $\hat{\pi}$ is the numerical approximation of QSD, we have

$$\pi^T e^{hQ} = \lambda \pi^T,\ \hat{\pi}^T (I + hQ) = \hat{\lambda} \hat{\pi}^T,$$

where λ and $\hat{\lambda}$ are simple eigenvalues. Define a function

$$A(t) \stackrel{\text{def}}{=} I + hQ + tR(h),$$

where $R(h)$ is an $O(1)$ matrix given by the Taylor expansion $e^{hQ} = I + hQ + h^2 R(h)$. Then we have $A(0) = I + hQ$ and $A(h^2) = e^{hQ}$. Note that $A(0)$ is irreducible for all sufficiently small h because Q is also irreducible. Let $\pi(t)$ be the first eigenvector of $A(t)$ normalized to 1 in l_1 norm. Then the sensitivity of π is reduced to the derivative of $A(t)$.

Since π is normalized to 1 in l_1 norm, it follows from [6, section 3] that

$$\pi'(0) = S^\sharp A'(0) \pi(0)\,,$$

where $S = \lambda I - A(0)$, and $S^\sharp$ is the group inverse of S. (We refer the reader to [6] for further discussion of the group inverse and derivative of the Perron vector.)

When h is small, we have $1 - \lambda = O(h)$. Hence $S = I - O(h) - I - hQ$ is an $O(h)$ small matrix. This means $S^\sharp = O(h^{-1})$. In addition $A'(0) = R = O(1)$ by definition. Hence $\pi'(0) = O(h^{-1})$. Since $\hat{\pi} = \pi(h^2)$, we have

$$\|\pi - \hat{\pi}\| = O(h^{-1}) \times O(h^2) = O(h).$$

This completes the proof. ∎

Since the focus of this paper is on the sensitivity of QSDs against the diffusion approximation, throughout this paper, we assume that $d_w(\pi_X, \hat{\pi}_X) = O(h)$ and $d_w(\pi_Y, \hat{\pi}_Y) = O(h)$. The time step h is assumed to be small such that the sensitivity against the time discretization does not play a leading role. In other words, the sensitivity of QSDs against the diffusion approximation is assumed to be well approximated by the distance between the QSD of the τ-leaping process of a mass-action system and the QSD of the Euler–Maruyama process of its diffusion approximation.

Let $T > 0$ be a fixed constant. Note that QSDs $\hat{\pi}_X$ and $\hat{\pi}_Y$ are also the invariant probability measures of regenerating processes $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, respectively. Therefore, we can decompose $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ via the following inequality:

$$d_w(\hat{\pi}_X, \hat{\pi}_Y) \le d_w\left(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T\right) + d_w\left(\hat{\pi}_X \tilde{P}_Y^T, \hat{\pi}_Y \tilde{P}_Y^T\right). \tag{3.5}$$

The term $d_w(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T)$ is the finite time error and the term $d_w(\hat{\pi}_X \tilde{P}_Y^T, \hat{\pi}_Y \tilde{P}_Y^T)$ can be bounded by coupling methods.

There are two different ways to think about the distance $d_w(\hat{\pi}_X, \hat{\pi}_Y)$. One method is considering $\hat{\pi}_X$ and $\hat{\pi}_Y$ as conditional distributions on set $\mathbb{R}_+^d/\partial\mathcal{X}$, i.e., $\hat{\pi}_X(A) = \mathbb{P}[\{\hat{X} \in A | t < \tau_X\}]$ and $\hat{\pi}_Y(A) = \mathbb{P}\{\hat{Y} \in A | t < \tau_Y\}$, where τ_X and τ_Y are the killing times for processes $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$, respectively. The other way is to use the $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ that regenerates from QSDs. No conditioning is needed as $\hat{\mu}_X$ and $\hat{\mu}_Y$ are now the invariant probability measures of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, respectively. There are some fundamental difficulties when computing the conditional finite time error because it is hard to couple $\hat{X}_n$ and $\hat{Y}_n$ when one regenerates while the other does not. Hence we choose to use $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ instead.

3.3. Finite time error. We consider the coupled modified processes $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, which are regenerated from the corresponding QSDs when they hit the boundary. Let $\hat{\pi}_X$ and $\hat{\pi}_Y$ be the invariant measures of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$. Let $\tilde{P}_X^T \circ \tilde{P}_Y^T$ denote the transition kernel of the coupled process in the product space $\mathcal{X} \times \mathcal{X}$. Denote $\hat{\pi}_X^2$ by the coupled measure of $\hat{\pi}_X$ on the "diagonal" of $\mathbb{R}^d \times \mathbb{R}^d$ that is supported by the hyperplane $\{(x, y) \in \mathbb{R}^{2d} | y = x\}$ such that $\hat{\pi}_X^2(\{(x, x) | x \in A\}) = \hat{\pi}_X(A)$. Then the pushforward measure $\hat{\pi}_X^2(\tilde{P}_X^T \circ \tilde{P}_Y^T)$ is still a coupling of $\hat{\pi}_X \tilde{P}_X^T$ and $\hat{\pi}_X \tilde{P}_Y^T$. Hence the following proposition follows easily.

Proposition 3.6. *Let $(\tilde{X}_n, \tilde{Y}_n)$ be a coupling of $\tilde{X}_n$ and $\tilde{Y}_n$ with transition kernel $\tilde{P}_X^T \circ \tilde{P}_Y^T$; then*

$$d_w\left(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T\right) \le \mathbb{E}_{\hat{\pi}_X}\left[d\left(\tilde{X}_T, \tilde{Y}_T\right)\right].$$

Proof. Since $\hat{\pi}_X^2(\tilde{P}_X^T \circ \tilde{P}_Y^T)$ is a coupling of $\hat{\pi}_X \tilde{P}_X^T$ and $\hat{\pi}_X \tilde{P}_Y^T$, by the definition of Wasserstein distance

$$\begin{aligned} d_w\left(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T\right) &\le \int_{\mathbb{R}^d \times \mathbb{R}^d} d(x, y) \hat{\pi}_X^2\left(\tilde{P}_X^T \circ \tilde{P}_Y^T\right)(dx, dy) \\ &= \int_{\mathbb{R}^d} \mathbb{E}_{(x,x)} d\left(\tilde{X}_T, \tilde{Y}_T\right) \hat{\pi}_X(dx) = \mathbb{E}_{\hat{\pi}_X}\left[d\left(\tilde{X}_T, \tilde{Y}_T\right)\right]. \end{aligned}$$

■

The key to estimating the finite time error effectively is to create a good coupled process $(\tilde{X}_n, \tilde{Y}_n)$. That is why we need to use the KMT algorithm to generate a pair of matched

Wiener processes and Poisson processes. Here it remains to define how $\tilde{X}_n$ and $\tilde{Y}_n$ couple when they regenerate from the QSDs. Since we do not have QSD a priori, we will use $\hat{X}_n$ and $\hat{Y}_n$ to approximate $\tilde{X}_n$ and $\tilde{Y}_n$. In other words, we regenerate samples from the temporal occupation measure. If the regenerations of $\tilde{X}_n$ and $\tilde{Y}_n$ are not coordinated, we will lose control of the error after one regeneration. Because even if the distance between $\tilde{X}n$ and $\tilde{Y}n$ is small, when they both hit the boundary and regenerate either simultaneously or a few steps apart, the regenerated processes can be far away from each other if they are regenerated independently. If the process regenerates frequently, the finite time error will quickly get out of control.

To minimize error during sample regeneration, we must couple the sample regenerations of $\tilde{X}_n$ and $\tilde{Y}_n$ as well. Here we define the coupled processes $(\hat{X}_n, \mu_n^X)$ and $(\hat{Y}_n, \mu_n^Y)$, such that $\hat{X}_n$ and $\hat{Y}_n$ follow (2.10) and (2.11), respectively, by using paired processes $B_k(t)$ and $P_k(t)$ for each k, and μ_n^X, μ_n^Y are two occupation measures. $S = (Z_1, \cdots, Z_N)$ (N is large enough) is a finite sequence of uniform random variables on $(0,1)$. Let N_X and N_Y be the total number of regenerations up to time n. In other words when $\tilde{X}_{n+1}$ enters $\partial\mathcal{X}$ at step n and needs regeneration, we increment N_X by one and choose the N_Xth element of S, i.e., Z_{N_X}, to regenerate $\tilde{X}_{n+1}$ by letting $\tilde{X}_{n+1} = \tilde{X}_{\lfloor Z_{N_X} n \rfloor}$. Then it is easy to see that $(\hat{X}_n, \mu_n^X)$ and $(\hat{Y}_n, \mu_n^Y)$ is a Markov coupling and the marginal processes $(\hat{X}_n, \hat{Y}_n)$ are a coupling of (2.10) and (2.11). Nontechnically, this coupled regeneration algorithm samples the kth regeneration of $\tilde{X}_n$ and $\tilde{Y}_n$ from the same relative position of the occupation measure. This is one of the key algorithms developed in this paper. It can "glue" two coupled numerical trajectories together even after several regenerations.

Details of the computation are shown in Algorithm 2. When N is large, initial values $\hat{X}_1^1, \cdots, \hat{X}_1^M$ in Algorithm 2 are from a trajectory of the time-T skeleton of $\hat{X}^T$. Hence $\hat{X}_1^1, \hat{X}_1^2, \cdots, \hat{X}_1^M$ are approximately sampled from $\hat{\pi}_X$. The error term $d(\hat{X}_T^m, \hat{Y}_T^m)$ evolved from the initial value pair $\hat{X}_1^m = \hat{Y}_1^m = \hat{X}_T^{m-1}$ is recorded. Therefore,

$$\frac{1}{M}\sum_{m=1}^{M} d\left(\hat{X}_T^m, \hat{Y}_T^m\right) \tag{3.6}$$

is an estimator of

$$\mathbb{E}_{\hat{\pi}_X}\left[d\left(\tilde{X}_T, \tilde{Y}_T\right)\right], \tag{3.7}$$

which is an upper bound of $d_w(\hat{\pi}_X^T \tilde{P}_X^T, \hat{\pi}_X^T \tilde{P}_Y^T)$.

3.4. Coupling inequality and contraction rate. Similar to the coupling inequality of the total variation norm, the distance d_w we use in this paper also satisfies the coupling inequality. Let $(Z_t^{(1)}, Z_t^{(2)})$ be a coupling of two stochastic processes and let τ_c be the coupling time. The following lemma follows easily.

Proposition 3.7. *For a Markov coupling $(Z_t^{(1)}, Z_t^{(2)})$, we have*

$$d_w\left(\text{law}\left(Z_T^{(1)}\right), \text{law}\left(Z_T^{(2)}\right)\right) \le \mathbb{P}\left(Z_T^{(1)} \ne Z_T^{(2)}\right) = \mathbb{P}(\tau_c > T).$$

Algorithm 2 Estimate finite time error

Input: Initial value $\hat{X}_0$
Output: An estimator of $d_w(\hat{\pi}_X^T \tilde{P}_X^T, \hat{\pi}_Y^T \tilde{P}_Y^T)$
  Set initial value $\hat{X}_1^1 = \hat{Y}_1^1$
  Generate a sequence of uniformly distributed random variables S
  for $m = 1$ to M **do**
    Using the KMT algorithm to generate paired trajectories $\{P_k\}$ and $\{B_k\}$
    If $m \neq 1$, reset initial value $\hat{X}_1^m = \hat{Y}_1^m = \hat{X}_T^{m-1}$
    Let $N_X = N_Y = 0$
    for $n = 1$ to T **do**
      Update $\hat{X}_{n+1}^m$ and $\hat{Y}_{n+1}^m$ using equations (2.10) and (2.11) respectively
      if $\hat{X}_{n+1}^m \in \partial\mathcal{X}$ **then**
        $N_X = N_X + 1$
        Let $\hat{X}_{n+1}^m = \hat{X}_{\lfloor Z_{N_X} n \rfloor}^m$
      end if
      if $\hat{Y}_{n+1}^m \in \partial\mathcal{X}$ **then**
        $N_Y = N_Y + 1$
        Let $\hat{Y}_{n+1}^m = \hat{Y}_{\lfloor Z_{N_Y} n \rfloor}^m$
      end if
    end for
    Let $d(\hat{X}_T^m, \hat{Y}_T^m) = \min(1, \|\hat{X}_T^m - \hat{Y}_T^m\|)$
  end for
  return $\frac{1}{M}\sum_{m=1}^M d(\hat{X}_T^m, \hat{Y}_T^m)$

Proof. By the definition of the Wasserstein distance,

$$\begin{aligned} d_w\left(\text{law}\left(Z_T^{(1)}\right), \text{law}\left(Z_T^{(2)}\right)\right) &\leq \int d(\xi,\eta)\mathbb{P}\left(\left(Z_T^{(1)}, Z_T^{(2)}\right) \in (d\xi, d\eta)\right) \\ &= \int_{\xi\neq\eta} d(\xi,\eta)\mathbb{P}\left(\left(Z_T^{(1)}, Z_T^{(2)}\right) \in (d\xi, d\eta)\right) \\ &\leq \int_{\xi\neq\eta} \mathbb{P}\left(\left(Z_T^{(1)}, Z_T^{(2)}\right) \in (d\xi, d\eta)\right) \\ &= \mathbb{P}\left(Z_T^{(1)} \neq Z_T^{(2)}\right). \end{aligned}$$

■

Proposition 3.8. *Assume that $d_w(\pi_X, \hat{\pi}_X)$ and $d_w(\pi_Y, \hat{\pi}_Y)$ are in order $O(h)$; then the error*

$$d_w(\pi_X, \pi_Y) \leq \frac{d_w\left(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T\right)}{1-\alpha} + O(h),$$

where $\alpha < 1$ is the contraction rate of the transition kernel $\tilde{P}_Y^T$ in the space of probability measures equipped with metric d_w, and $d_w(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T)$ is the finite time error.

Proof. By the triangle inequality,

$$d_w(\pi_X, \pi_Y) \le d_w(\pi_X, \hat{\pi}_X) + d_w(\hat{\pi}_X, \hat{\pi}_Y) + d_w(\hat{\pi}_Y, \pi_Y).$$

Because both $d_w(\pi_X, \hat{\pi}_X)$ and $d_w(\pi_Y, \hat{\pi}_Y)$ are $O(h)$, we only need to estimate the second term $d_w(\hat{\pi}_X, \hat{\pi}_Y)$. By the triangle inequality again, we have

$$d_w(\hat{\pi}_X, \hat{\pi}_Y) \le d_w\left(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T\right) + d_w\left(\hat{\pi}_X \tilde{P}_Y^T, \hat{\pi}_Y \tilde{P}_Y^T\right).$$

If the transition kernel $\tilde{P}_Y^T$ has enough contraction such that

$$d_w\left(\hat{\pi}_X \tilde{P}_Y^T, \hat{\pi}_Y \tilde{P}_Y^T\right) \le \alpha d_w(\hat{\pi}_X, \hat{\pi}_Y)$$

for some $\alpha < 1$, then we have

$$d_w(\hat{\pi}_X, \hat{\pi}_Y) \le \frac{d_w\left(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T\right)}{1-\alpha}. \tag{3.8}$$

Therefore,

$$d_w(\pi_X, \pi_Y) \le \frac{d_w\left(\hat{\pi}_X \tilde{P}_X^T, \hat{\pi}_X \tilde{P}_Y^T\right)}{1-\alpha} + O(h). \qquad \blacksquare$$

Therefore, to estimate $d_w(\pi_X, \pi_Y)$, we need to find suitable numerical estimators for the finite time error and the speed of contraction of $\tilde{P}_Y^T$. The finite time error can be easily estimated by Algorithm 2. The speed of contraction α comes from the geometric ergodicity of the Markov process $\tilde{\mathbf{Y}}$, which is approximated by that of $\hat{\mathbf{Y}}$ due to the convergence result in Proposition 3.4. If our numerical estimation yields

$$d_w\left(\hat{\pi}_X \tilde{P}_Y^T, \hat{\pi}_Y \tilde{P}_Y^T\right) \approx d_w\left(\hat{\pi}_X \hat{P}_Y^T, \hat{\pi}_Y \hat{P}_Y^T\right) \le Ce^{-\gamma T},$$

we can then set $\alpha = e^{-\gamma T}$. Motivated by [8], we use the following coupling method to estimate the contraction rate α. Let $\hat{Z} = (\hat{Y}^{(1)}, \hat{Y}^{(2)})$ be a Markov process in $\mathbb{R}^{2d}$ such that $\hat{Y}^{(1)}$ and $\hat{Y}^{(2)}$ are two copies of $\hat{Y}$. Let the first passage time to the "diagonal" hyperplane $\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{2d} | \mathbf{y} = \mathbf{x}\}$ be the coupling time. Then by Proposition 3.7

$$d_w\left(\hat{\pi}_X \hat{P}_Y^T, \hat{\pi}_Y \hat{P}_Y^T\right) \le \mathbb{P}(\tau_c > T).$$

As discussed in [18], a hybrid coupling scheme is necessary to ensure that two numerical trajectories couple. This is achieved by implementing coupling methods such as reflection coupling or synchronous coupling in the first phase, under the condition that the trajectories have not become extinct yet. Next, when the distance between coupled trajectories is small enough, we compare the probability density function for the next step and couple these two numerical trajectories with the maximal possible probability, using maximal coupling. After repeating this process many times, we obtain multiple samples of τ_c, denoted by $\boldsymbol{\tau}_c$. To estimate the contraction rate α, we use the exponential tail of $\mathbb{P}[\tau_c > t]$ and look for a constant $\gamma > 0$ such that

Algorithm 3 Estimation of contraction rate α

Input: Initial values $x, y \in \mathcal{X} \setminus \partial\mathcal{X}$

Output: An estimation of contraction rate α

```
Choose threshold d > 0
for m = 1 to M do
  τ_c^m = 0, t = 0, (Ŷ_0^(1), Ŷ_0^(2)) = (x, y)
  Flag = 0
  while Flag = 0 do
    if Ŷ_t^(1) and Ŷ_t^(2) ∈ X/∂X then
      if |Ŷ_t^(1) − Ŷ_t^(2)| > d then
        Compute (Ŷ_{t+1}^(1), Ŷ_{t+1}^(2)) using reflection coupling or independent coupling t ← t + 1
      else
        Compute (Ŷ_{t+1}^(1), Ŷ_{t+1}^(2)) using maximal coupling
        if coupled successfully then
          Flag = 1
          τ_c^m = t
        else
          t ← t + 1
        end if
      end if
    end if
  end while
end for
Use τ_c^1, ..., τ_c^M to compute P(τ_c > t | t < min(τ_{Y^(1)}, τ_{Y^(2)}))
Fit the tail of log P(τ_c > t | t < min(τ_{Y^(1)}, τ_{Y^(2)})) versus t by linear regression. Compute the slope γ.
```

$$-\gamma = \lim_{t\to\infty} \frac{1}{t} \log(\mathbb{P}[\tau_c > t])$$

if the limit exists. See Algorithm 3 for the details of the implementation of coupling. Note that we cannot simply compute the contraction rate starting from $t = 0$ because only the tail of coupling time can be considered exponentially distributed. In addition, $\hat{Y}$ is a good approximation of $\tilde{Y}$ only if t is large. Our approach is to check the exponential tail in a log-linear plot. After having $\boldsymbol{\tau}_c$, it is easy to choose a sequence of times $t_0, t_1, \cdots, t_n$ and calculate $n_i = |\{\tau_c^m > t_i | 0 \le m \le M\}|$ for each $i = 0, \cdots, n$. Then $p_i = n_i/M$ is an estimator of $\mathbb{P}_{\hat{\pi}_Y}[\tau_c > t_i]$. Now let p_i^u (resp., p_i^l) be the upper (resp., lower) bound of the confidence interval of p_i such that

$$p_i^u = \tilde{p} + z\sqrt{\frac{\tilde{p}}{\tilde{n}_i}(1-\tilde{p})} \quad \left(\text{resp., } p_i^l = \tilde{p} - z\sqrt{\frac{\tilde{p}}{\tilde{n}_i}(1-\tilde{p})}\right),$$

where $z = 1.96$, $\tilde{n}_i = n_i + z^2$, and $\tilde{p} = \frac{1}{\tilde{n}}(n_i + \frac{z^2}{2})$ [1]. If $p_i^l \le e^{-\gamma t_i} \le p_i^u$ for each $0 \le i \le n$, we say that the exponential tail starts at $t = t_{i_0}$. We accept the exponential tail with rate $e^{-\gamma T}$ if the confidence interval $p_{i_0}^u - p_{i_0}^l$ is sufficient small. Otherwise, we need to run Algorithm 3 for a longer time to eliminate the initial bias in τ_c.

4. Numerical examples.

4.1. SIR model.

Consider an epidemic model in which the entire population is divided into three distinct classes: S (susceptible), I (infected), and R (recovered). After nondimensionalization, the ODE version of a SIR model is given by

$$\begin{aligned} \frac{dS}{dt} &= (\alpha - \beta SI - \mu S), \\ \frac{dI}{dt} &= (\beta SI - (\mu + \rho + \gamma)I), \\ \frac{dR}{dt} &= (\gamma I - \mu R), \end{aligned} \tag{4.1}$$

where α is the birth rate, μ is the disease-free death rate, ρ is the excess death rate for the infected class, γ is the recovery rate for the infected population, and β is the effective contact rate between the susceptible class and infected class [7]. Note that R completely depends on S and I. So we just consider the evolutions of S and I.

Now we let V be the total population and consider the corresponding stochastic mass-action network. There are four reactions involved in this network. The stochastic mass-action network can be defined by a Poisson process $X_n = (S_n, I_n)$.

$$\begin{aligned} &\emptyset \overset{\alpha}{\Rightarrow} S, \ S + I \overset{\beta}{\Rightarrow} 2I, \\ &S \overset{\mu}{\Rightarrow} \emptyset, \ I \overset{\mu+\rho+\gamma}{\Rightarrow} \emptyset. \end{aligned} \tag{4.2}$$

Applying the numerical representation in (2.10), we have the approximate rate functions of Poisson process $\hat{X}_n$:

$$\begin{aligned} q_{1,n} &= \sum_{m=0}^{n-1} Vh\alpha, \ q_{2,n} = \sum_{m=0}^{n-1} Vh\beta S_m I_m, \\ q_{3,n} &= \sum_{m=0}^{n-1} Vh\mu S_m, \ q_{4,n} = \sum_{m=0}^{n-1} Vh(\mu + \rho + \gamma) I_m. \end{aligned}$$

Let $P_i, i = 1, 2, 3, 4$, be independent unit rate Poisson processes. Then $\hat{X}_n$ is driven by the discrete approximation of $\{P_i\}_{i=1}^4$. The rule of update of the numerical approximation $\hat{X}_n$ follows

$$\hat{X}_{n+1} = \begin{pmatrix} S_{n+1} \\ I_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{f}_1(P_1, \cdots, P_4, q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{f}_2(P_1, \cdots, P_4, q_{1,n}, \cdots, q_{4,n}) \end{pmatrix}, \tag{4.3}$$

where $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ come from discrete approximation in (2.10). To improve the readability of the present paper, we move detailed expressions of $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ to the appendix.

As described in section 2.1, each Poisson process $P_i, i = 1,2,3,4$, is pathwisely approximated by a Wiener process $B_i, i = 1,2,3,4$. Further, the discrete approximation $\hat{X}_n$ is pathwisely approximated by an Euler–Maruyama scheme $\hat{Y}_n$ that reads

$$\hat{Y}_{n+1} = \begin{pmatrix} S_{n+1} \\ I_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g}_1(q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{g}_2(q_{1,n}, \cdots, q_{4,n}) \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{\sigma}_1(B_1, \cdots, B_4, q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{\sigma}_2(B_1, \cdots, B_4, q_{1,n}, \cdots, q_{4,n}) \end{pmatrix}, \tag{4.4}$$

where functions $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{\sigma}_1$, and $\boldsymbol{\sigma}_2$ follow the expression in (2.11). We refer the reader to the appendix for the detailed form of these functions.

By the stationary increments property of the standard Wiener process, we know that every finite difference of B_i is normally distributed. In addition Wiener processes $B_i, i = 1,2,3,4$, are independent. Therefore, (4.4) can be simplified to

$$\hat{Y}_{n+1} = \begin{pmatrix} S_{n+1} \\ I_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g}_1(q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{g}_2(q_{1,n}, \cdots, q_{4,n}) \end{pmatrix} + \frac{1}{V} M \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{pmatrix}, \tag{4.5}$$

where $W_i, i = 1, \cdots, 4$, are independent standard normal random variables, and M is a matrix that depends only on S_n and I_n. We refer readers to the appendix for the full expression of M.

In order to estimate the distance between two QSDs, we need to find the contraction rate α for diffusion process $\hat{Y}$ above. However, the diffusion matrix M in $\hat{Y}$ is not square, which makes a reflection coupling difficult. Here we define an equivalent diffusion process that is driven by a 2D Wiener process but has the same law as $\hat{Y}$. In our simulation, we compute the 2 by 2 covariance matrix $N = MM^T$ and set the square root of N to be the new diffusion matrix. Then $\hat{Y}$ can be rewritten as

$$\begin{aligned} \hat{Y}_{n+1} = & \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g}_1(q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{g}_2(q_{1,n}, \cdots, q_{4,n}) \end{pmatrix} \\ & + \frac{1}{\sqrt{\mathrm{tr}(N) + 2\sqrt{\det(N)}}} (N + \det(N) Id) \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}, \end{aligned} \tag{4.6}$$

where $\mathrm{tr}(N)$ is the trace of N and $\det(N)$ is the determinant of N, and Id is the identity matrix. It is easy to see that the diffusion process $\hat{Y}$ in (4.5) and (4.6) are equivalent. Hence we do not change its notation here. The modification of $\hat{Y}$ allows us to run Algorithm 3 to compute the coupling time distribution.

It remains to compute the finite time error. Let $\partial\mathcal{X}$ be the union of the x-axis and y-axis. The model parameters are set as $\alpha = 7, \beta = 3, \mu = 1, \rho = 1, \gamma = 2$. Processes $\hat{X}$ and $\hat{Y}$ admit QSDs $\hat{\pi}_X$ and $\hat{\pi}_Y$, respectively. Long trajectories $P(i\Delta)$ and $B(i\Delta)$ for $i = \{1, \cdots, 2^{20}\}$ and $\Delta = 0.01$ are constructed when we consider the trajectory-by-trajectory behavior of two processes. The time step size is $h = 0.001$ and the fixed time is set as $T = 0.5$.

The result for $V = 1000$ is demonstrated in Figure 1. The bottom left of Figure 1 shows the QSD of diffusion process $\hat{Y}$. The QSD of the Poisson process is shown on the top right of

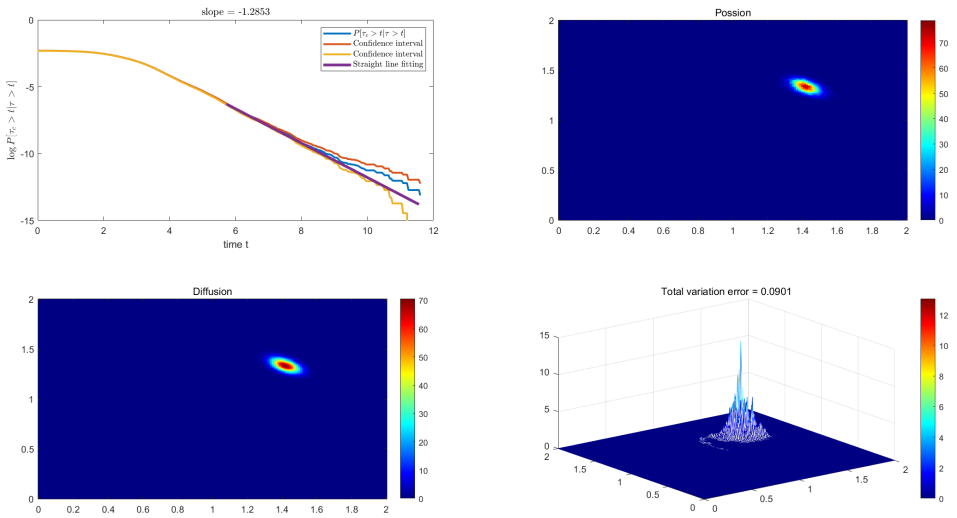

Figure 1. *(Case $V = 1000$) Upper panel: (Left) $\mathbb{P}(\tau_c > t|\tau < t)$ versus t. (Right) QSD of Poisson process. Lower panel: (Left) QSD of diffusion process. (Right) Total variation of two QSDs.*

Figure 1. The difference between these two QSDs is shown at the bottom of Figure 1. We can see that the total variation distance between two QSDs is 0.0901, which is considered to be small. This is reasonable because, with high probability, the trajectories of both the Poisson process and the diffusion process move far away from the absorbing set $\partial\mathcal{X}$.

The total variation distance between two QSDs is consistent with the prediction developed in this paper. We first use Algorithm 3 to compute the distribution of the coupling time, which is shown at the top left of Figure 1. Then we use Algorithm 2 to compute the finite time error. The finite time error is 0.0026 for $V = 1000$. As a result, the upper bound given in (3.8) is 0.0054 for $V = 1000$, which is smaller than the empirical total variation error 0.0901 in this case.

Then we carry out similar computations for $V = 10$ on a coarse mesh. The result is shown in Figure 2. To compare with the case for $V = 1000$ on the same mesh, we rescaled the probability density function obtained from the Monte Carlo simulation. The probability density in one bin in the coarse mesh is evenly distributed into many bins in the refined mesh. The difference between the two QSDs is shown at the bottom of Figure 2. It is not hard to see that the total variation distance becomes significantly larger when the volume gets smaller. Similarly to above, we use Algorithm 3 to compute the distribution of the coupling time distribution (Figure 2, top left) and use Algorithm 2 to compute the finite time error. The finite time error is 0.1748 for $V = 10$. As a result, the upper bound given in (3.8) is 0.3639 for $V = 10$. This is consistent with the numerical finding shown in Figure 2, bottom right.

As we consider the effect of the capacity volume, the finite time error and the contraction rate for different volumes are compared in Table 1. The last column $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ is computed using (3.8). Being consistent with Theorem 2.3, the 1-Wasserstein distance between two QSDs is smaller as V gets larger.

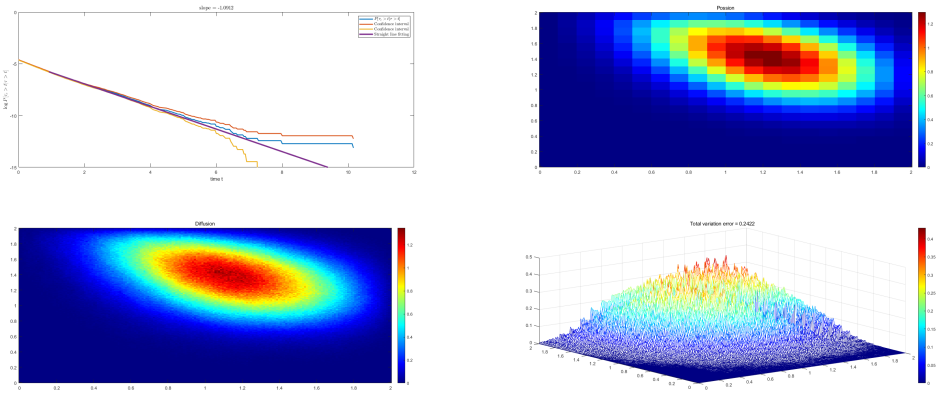

Figure 2. *(Case $V = 10$) Upper panel: (Left) $\mathbb{P}(\tau_c > t|\tau < t)$ versus t. (Right) QSD of Poisson process. Lower panel: (Left) QSD of diffusion process. (Right) Total variation of two QSDs.*

Table 1
SIR model. Numerical results for different volumes.

Volume V	Finite time error	Contraction rate γ	$d_w(\hat{\pi}_X, \hat{\pi}_Y)$
1000	0.0026	1.2853	0.0054
400	0.0079	1.2418	0.0170
100	0.0279	1.1613	0.0634
10	0.1748	1.0912	0.3639

4.2. Oregonator system. In this example, we consider a well-known example of a chemical oscillator called the Belousov–Zhabotinsky (BZ) reaction model or "Oregonator" [4, 9, 12]. The ODE version of an Oregonator system is given by

$$\begin{aligned}
\frac{dS_1}{dt} &= S_1S_2 - C_2S_1S_2 + C_3S_1 - 2C_4S_1^2,\\
\frac{dS_2}{dt} &= -C_1S_2 - C_2S_1S_2 + C_5hS_3,\\
\frac{dS_3}{dt} &= 2C_3S_1 - C_5S_3.
\end{aligned}$$

We refer to Figure 3, top left for sample trajectories of the Oregonator on $\mathbb{R}^3_+$. The parameter values are chosen as $C_1 = 2560, C_2 = 800000, C_3 = 16000, C_4 = 2000, C_5 = 9000, \delta = 0.4$.

Let V be the volume. Six reactions in this process are shown as follows.

$$S_2 \overset{C_1}{\Rightarrow} S_1,\ S_1 + S_2 \overset{C_2}{\Rightarrow} \emptyset,\ S_1 \overset{C_3}{\Rightarrow} 2S_1 + 2S_3,$$
$$2S_1 \overset{C_4}{\Rightarrow} \emptyset,\ S_3 \overset{C_5\delta}{\Rightarrow} S_2,\ S_3 \overset{C_5(1-\delta)}{\Rightarrow} \emptyset.$$

Applying the numerical representation in (2.10), we have the approximate rate functions of Poisson process $\hat{X}_n = (S_{1,n}, S_{2,n}, S_{3,n})$:

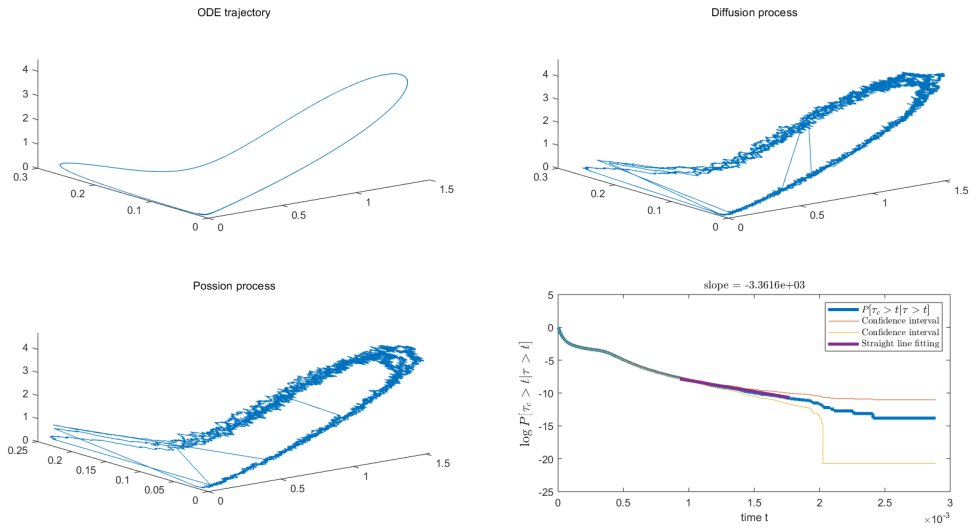

Figure 3. *(Case $V = 1000$) Upper panel: (Left) ODE trajectories. (Right) Trajectories of Poisson process. Lower panel: (Left) Trajectories of diffusion process. (Right) $\mathbb{P}(\tau_c > t|\tau < t)$ vs.t.*

$$q_{1,n} = \sum_{m=0}^{n-1} VhC_1S_{2,m},\ q_{2,n} = \sum_{m=0}^{n-1} VhC_2S_{1,m}S_{2,m},\ q_{3,n} = \sum_{m=0}^{n-1} VhC_3S_{1,m},$$
$$q_{4,n} = \sum_{m=0}^{n-1} VhC_4S_{1,m}^2,\ q_{5,n} = \sum_{m=0}^{n-1} VhC_5\delta S_{3,m},\ q_{6,n} = \sum_{m=0}^{n-1} VhC_5(1-\delta)S_{3,m}.$$

We remark that term $S_{1,m}$ is the numerical value of species S_1 at time step m, and cases of other terms are analogous. Hence the Poisson process $\hat{X}$ of the Oregonator model can be written as

$$\hat{X}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \end{pmatrix} = \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{f_1}(P_1, \cdots, P_6, q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{f_2}(P_1, \cdots, P_6, q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{f_3}(P_1, \cdots, P_6, q_{1,n}, \cdots, q_{6,n}) \end{pmatrix},$$

where $P_i, i = \{1, \cdots, 6\}$, are independent unit rate Poisson processes. $\boldsymbol{f_1}$, $\boldsymbol{f_2}$, and $\boldsymbol{f_3}$ come from discrete approximation in (2.10). To improve the readability of the present paper, we move detailed expressions of $\boldsymbol{f_1}$, $\boldsymbol{f_2}$, and $\boldsymbol{f_3}$ to the appendix.

The diffusion approximation $\hat{Y}$ can be written as

(4.7)
$$\hat{Y}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \end{pmatrix} = \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g_1}(q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{g_2}(q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{g_3}(q_{1,n}, \cdots, q_{6,n}) \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{\sigma_1}(B_1, \cdots, B_6, q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{\sigma_2}(B_1, \cdots, B_6, q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{\sigma_3}(B_1, \cdots, B_6, q_{1,n}, \cdots, q_{6,n}) \end{pmatrix},$$

where $B_i, i = \{1, \cdots, 6\}$, are independent standard Wiener processes, and functions $\boldsymbol{g_1}$, $\boldsymbol{g_2}$, $\boldsymbol{g_3}$, $\boldsymbol{\sigma_1}$, $\boldsymbol{\sigma_2}$, and $\boldsymbol{\sigma_3}$ follow the expression in (2.11). We refer the reader to the appendix for the detailed form of these functions.

By the stationary increments property and independence of Wiener processes $B_i, i = \{1, \cdots, 6\}$, equation (4.7) can be simplified to

$$\hat{Y}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \end{pmatrix} = \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g_1}(q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{g_2}(q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{g_3}(q_{1,n}, \cdots, q_{6,n}) \end{pmatrix} + \frac{1}{V} M \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \\ W_6 \end{pmatrix}, \tag{4.8}$$

where $W_i, i = 1, \cdots, 6$, are independent standard normal random variables, and M is a matrix that depends only on S_n and I_n. We refer readers to the appendix for the full expression of M.

Let $\partial\mathcal{X}$ be the union of the x-axis, y-axis, and z-axis. Processes $\hat{X}$ and $\hat{Y}$ admit QSDs $\hat{\pi}_X$ and $\hat{\pi}_Y$, respectively. Long trajectories $P(i\Delta)$ and $B(i\Delta)$ for $i = \{1, \cdots, 2^{29}\}$ and $\Delta = 0.001$ are constructed when we consider the trajectory-by-trajectory behavior of two processes. The time step size is $h = 10^{-8}$ and the fixed time is set as $T = 0.0002$. Large rate coefficients C_i make the numerical results easily go beyond the length of long trajectory $B(i\Delta)$, so we pick small time step size h and the fixed finite time T.

Figure 3 top left shows the solution of the ODE. For any initial point, the trajectory eventually converges to the limit cycle. In terms of thermodynamics, the oscillation is induced through the dissipation of energy and is often called a self-sustained oscillator [23]. The trajectories of the Poisson process and the diffusion process up to fixed time $T = 0.0002$ are shown at the top right and bottom left. It looks like the trajectories are close, and this is reasonable because, with high probability, the trajectories of both the Poisson process and the diffusion process move far away from the absorbing set $\partial\mathcal{X}$. There are only a few regeneration events (the lines crossing the limit cycle). We compute the distribution of the coupling time. The coupling time distribution and its exponential tail are shown in Figure 3, top left. Then we use Algorithm 2 to compute the finite time error. The finite time error is 0.0057 for $V = 1000$. As a result, the upper bound given in (3.8) is 0.0116 for $V = 1000$. For $V = 10$, the finite time error is 0.4531 and the upper bound given in (3.8) is 0.4531.

To compare different scenarios for volume $V = 1000$ and $V = 10$, we plot the trajectories of each species for both processes. The trajectories for $V = 1000$ are shown in the upper row of Figure 4, and the lower row shows the case for $V = 10$. It is evident that the Poisson process is quite close to the diffusion process when $V = 1000$. However, when the volume is too small, not many Poisson jumps can be observed in the Poisson process, while significant noise can be seen in the diffusion approximation. As a result, the finite time error for $V = 10$ is 0.0563, which is around ten times larger than that for $V = 1000$. Similarly, we compute the contraction rate γ of the coupling time distribution to be 2.0927×10^5. This is due to the large magnitude of noise in the diffusion approximation. Therefore, the upper bound given in (3.8) is 0.4531 for $V = 10$. We conclude that the diffusion approximation does not approximate the QSD well when the volume is not large enough.

As we consider the effect of the capacity volume, we compare the finite time error and the contraction rate for different volumes in Table 2. The last column, $d_w(\hat{\pi}_X, \hat{\pi}_Y)$, is computed via (3.8). It is clear that the upper bound of $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ is significantly larger when $V = 10$.

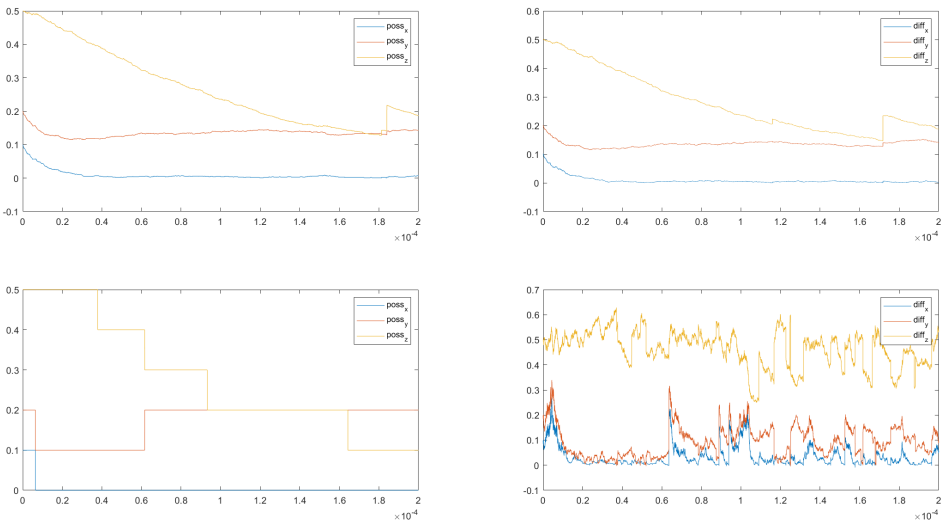

Figure 4. *($V = 1000$ versus $V = 10$) Upper panel: (Left) Trajectories of Poisson process for $V = 1000$. (Right) Trajectories of diffusion process for $V = 1000$. Lower panel: (Left) Trajectories of Poisson process for $V = 10$. (Right) Trajectories of diffusion process for $V = 10$.*

Table 2
Oregonator model: Numerical results for different volumes.

Volume V	Finite time error	Contraction rate γ	$d_w(\hat{\pi}_X, \hat{\pi}_Y)$
1000	0.0057	$3.3616*10^3$	0.0116
400	0.0088	$2.0599*10^4$	0.0157
100	0.0099	$6.0150*10^4$	0.0195
10	0.0563	$2.0927*10^5$	0.1646

This is consistent with Theorem 2.3, which states that the supreme distance between two processes will be smaller as V increases.

4.3. 4D Lotka–Volterra competitive dynamics. Originally derived by Volterra in 1926 to describe the interaction between a predator species and a prey species [20] and independently by Lotka to describe a chemical reaction [27], the general Lotka–Volterra model is widely used in ecology, biology, chemistry, physics, etc. [25]. In this example, we consider here a chaotic system in which 4 species with a whole population V compete for a finite set of resources. The ODE version of the system is as follows:

$$\frac{dS_i}{dt} = r_i S_i \left(1 - \sum_{j=1}^{4} a_{ij} S_j\right), \quad i = 1, 2, 3, 4.$$

Here r_i represents the growth rate of species i and a_{ij} represents the extent to which species j competes for resources used by species i. The parameter values are

$$r=(r_i)_{i=1}^4=\begin{pmatrix}1\\0.72\\1.53\\1.27\end{pmatrix},\ A=(a_{ij})_{i,j=1}^4=\begin{pmatrix}1&1.09&1.52&0\\0&1&0.44&1.36\\2.33&0&1&0.47\\1.21&0.51&0.35&1\end{pmatrix}.$$

For $i=1,\cdots,4$, all reactions in this system are shown as follows:

$$S_i\overset{r_i}{\Rightarrow}2S_i,\ S_1+S_i\overset{a_{i1}r_i}{\Rightarrow}S_1,\ S_2+S_i\overset{a_{i2}r_i}{\Rightarrow}S_2,\ S_3+S_i\overset{a_{i3}r_i}{\Rightarrow}S_3,\ S_4+S_i\overset{a_{i4}r_i}{\Rightarrow}S_4.$$

The corresponding rate functions are

$$\begin{aligned}q_{i,1}^n&=\sum_{m=0}^{n-1}Vhr_iS_{i,m},\\q_{i,2}^n&=\sum_{m=0}^{n-1}Vhr_ia_{i1}S_{1,m}S_{i,m},\\q_{i,3}^n&=\sum_{m=0}^{n-1}Vhr_ia_{i2}S_{2,m}S_{i,m},\\q_{i,4}^n&=\sum_{m=0}^{n-1}Vhr_ia_{i3}S_{3,m}S_{i,m},\\q_{i,5}^n&=\sum_{m=0}^{n-1}Vhr_ia_{i4}S_{4,m}S_{i,m}.\end{aligned}$$

As three zeros appear in coefficient matrix A, this system actually includes 17 reactions. Therefore, the Poisson process $\hat{X}_n=(S_{1,n},S_{2,n},S_{3,n},S_{4,n})$ can be written as

$$\hat{X}_{n+1}=\begin{pmatrix}S_{1,n+1}\\S_{2,n+1}\\S_{3,n+1}\\S_{4,n+1}\end{pmatrix}=\begin{pmatrix}S_{1,n}\\S_{2,n}\\S_{3,n}\\S_{4,n}\end{pmatrix}+\frac{1}{V}\begin{pmatrix}\boldsymbol{f_1}(P_1,\cdots,P_{17},q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{f_2}(P_1,\cdots,P_{17},q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{f_3}(P_1,\cdots,P_{17},q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{f_4}(P_1,\cdots,P_{17},q_{i,1}^n,\cdots,q_{i,5}^n)\end{pmatrix},$$

where $i=1,\cdots,4$, P_j, $j=\{1,\cdots,17\}$, are independent unit rate Poisson processes, and $\boldsymbol{f_1}$, $\boldsymbol{f_2}$, $\boldsymbol{f_3}$, and $\boldsymbol{f_4}$ come from discrete approximation in (2.10). To improve the readability of the present paper, we move detailed expressions of $\boldsymbol{f_1}$–$\boldsymbol{f_4}$ to the appendix.

The diffusion approximation $\hat{Y}$ can be written as

(4.9)

$$\hat{Y}_{n+1}=\begin{pmatrix}S_{1,n+1}\\S_{2,n+1}\\S_{3,n+1}\\S_{4,n+1}\end{pmatrix}=\begin{pmatrix}S_{1,n}\\S_{2,n}\\S_{3,n}\\S_{4,n}\end{pmatrix}+\frac{1}{V}\begin{pmatrix}\boldsymbol{g_1}(q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{g_2}(q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{g_3}(q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{g_4}(q_{i,1}^n,\cdots,q_{i,5}^n)\end{pmatrix}+\frac{1}{V}\begin{pmatrix}\boldsymbol{\sigma_1}(B_1,\cdots,B_{17},q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{\sigma_2}(B_1,\cdots,B_{17},q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{\sigma_3}(B_1,\cdots,B_{17},q_{i,1}^n,\cdots,q_{i,5}^n)\\\boldsymbol{\sigma_4}(B_1,\cdots,B_{17},q_{i,1}^n,\cdots,q_{i,5}^n)\end{pmatrix},$$

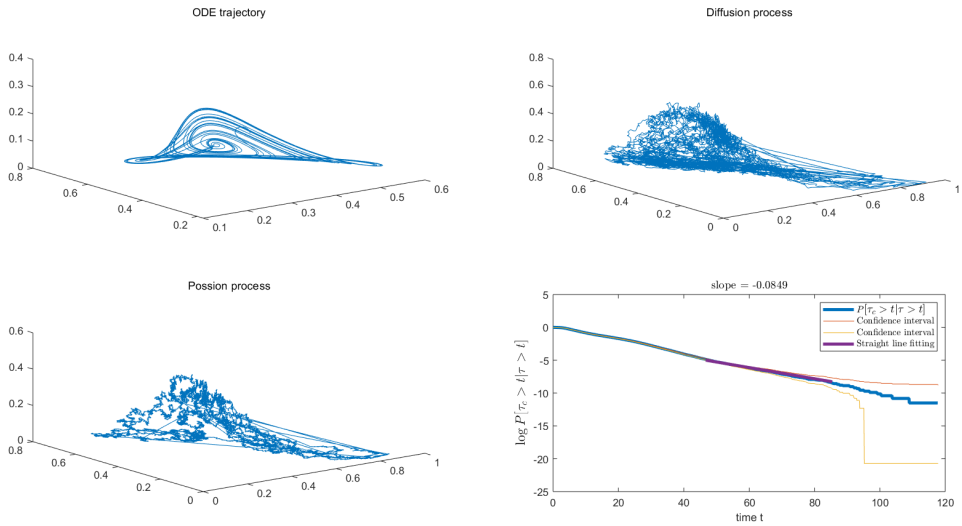

Figure 5. *(Case $V = 1000$) Upper panel: (Left) ODE trajectories. (Right) Poisson process. Lower panel: (Left) Diffusion process. (Right) $\mathbb{P}(\tau_c > t|\tau < t)$ versus t.*

where $i = 1, \cdots, 4$, B_j, $j = \{1, \cdots, 17\}$, are independent standard Wiener process, and functions $\boldsymbol{g_1}$, $\boldsymbol{g_2}$, $\boldsymbol{g_3}$, $\boldsymbol{g_4}$, $\boldsymbol{\sigma_1}$, $\boldsymbol{\sigma_2}$, $\boldsymbol{\sigma_3}$, and $\boldsymbol{\sigma_4}$ follow the expression in (2.11). We refer the reader to the appendix for the detailed form of these functions.

By the stationary increments property and independence of Wiener processes $B_i, i = \{1, \cdots, 6\}$, equation (4.9) can be simplified to

$$\hat{Y}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \\ S_{4,n+1} \end{pmatrix} = \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \\ S_{4,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g}_1(q_1, \cdots, q_4) \\ \boldsymbol{g}_2(q_1, \cdots, q_4) \end{pmatrix} + \frac{1}{V} M \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_{16} \\ W_{17} \end{pmatrix}, \tag{4.10}$$

where $W_i, i = 1, \cdots, 17$, are independent standard normal random variables, and M is a matrix that depends only on S_n and I_n. We refer readers to the appendix for the full expression of M.

Let $\partial\mathcal{X}$ be the union of 4 axes. Processes $\hat{X}$ and $\hat{Y}$ admit QSDs $\hat{\pi}_X$ and $\hat{\pi}_Y$, respectively. Long trajectories $P(i\Delta)$ and $B(i\Delta)$ for $i = \{1, \cdots, 2^{22}\}$ and $\Delta = 0.01$ are constructed when we consider the trajectory-by-trajectory behavior of two processes. The time step size is $h = 0.001$ and the fixed time is set as $T = 1$.

Figure 5 top left shows the solution of the ODE projected onto $x_1x_2x_3$ space. The trajectories of the Poisson process and the diffusion process are shown at the top right and bottom left, respectively. It appears that the trajectories are close, which is reasonable because, with high probability, the trajectories of both the Poisson process and the diffusion process move far away from the absorbing set $\partial\mathcal{X}$. We then compute the distribution of the coupling time. The coupling time distribution and its exponential tail are shown in Figure 5 top left, which

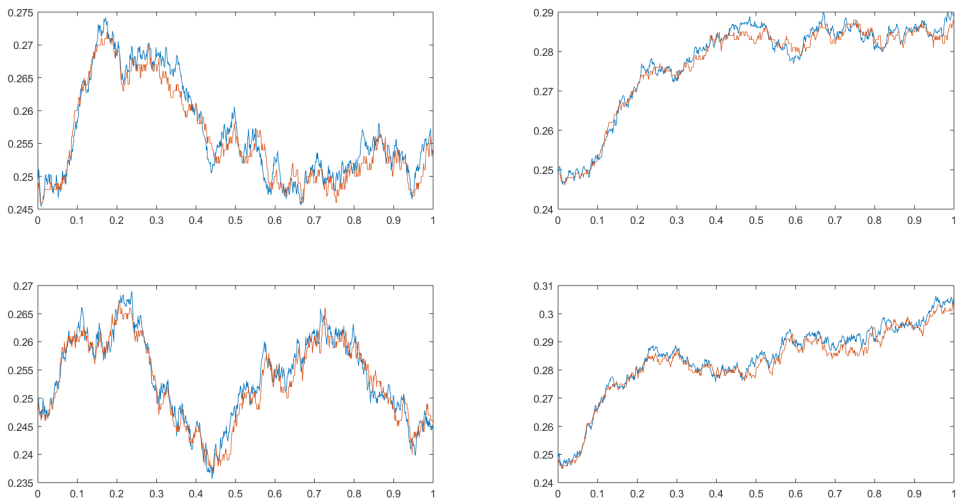

Figure 6. *(Case $V = 1000$) Poisson trajectories and diffusion trajectories for 4 species.*

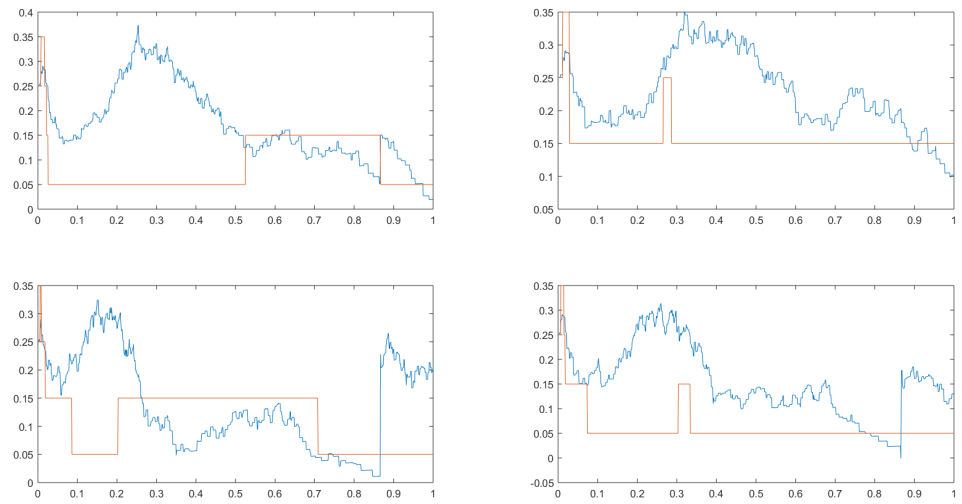

Figure 7. *(Case $V = 10$) Poisson trajectories and diffusion trajectories for 4 species.*

gives the contraction rate $\gamma = 0.0849$. Next, we apply Algorithm 1 to compute the finite time error, and the result is 0.0030 for $V = 1000$. As a result, the upper bound given in (3.8) is 0.0375 for $V = 1000$.

To compare the different scenarios for volume $V = 1000$ and $V = 10$, we plot trajectories of each species for $V = 1000$ in Figure 6, and the case for $V = 10$ is shown in Figure 7. It is apparent that the behavior of the trajectories between the Poisson process and diffusion process is notably different when $V = 1000$ as compared to the case when $V = 10$. Therefore,

Table 3
4D Lotka–Volterra model: Numerical results for different volumes.

Volume V	Finite time error	Contraction rate γ	$d_w(\hat{\pi}_X, \hat{\pi}_Y)$
1000	0.0030	0.0849	0.0375
400	0.0110	0.1831	0.0659
100	0.0502	0.3110	0.1878
10	0.1286	1.7905	0.1543

it is not surprising that the finite time error for $V = 10$ is 0.1286, which is approximately 40 times greater than the case for $V = 1000$. In this case, the trajectories of the Poisson process have a high probability of moving along the boundary. As mentioned before, we have computed the contraction rate γ of the coupling time distribution to be 1.7905. Consequently, the upper bound given in (3.8) is 0.1543 for $V = 10$.

As we consider the effect of the capacity volume, the finite time error and the contraction rate for different volumes are compared in Table 3. The last column $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ is computed via (3.8). It is not hard to see that the upper bound of $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ is quite larger when $V = 10$. This is consistent with Theorem 2.3; the supreme distance between two processes will be smaller as V gets larger.

5. Conclusion. In this paper, we propose a coupling-based approach to estimate the distance between the QSD of a stochastic mass-action process and its diffusion approximation. The study investigates how the QSDs depend on the volume of the mass-action system. To tackle the QSDs, the paper adopts the idea of regeneration from QSDs after exiting to construct a process with the stationary distribution. This approach differs from the previous work by the authors [8, 18]. The coupling algorithm and the pathwise matching of a stochastic mass-action system and its diffusion approximation need to be adapted to the regeneration from QSDs. The paper compares the finite time error and the rate of contraction for different population volumes V, and the results show that the distance between two QSDs is smaller for a larger population. Overall, the study emphasizes the need to address demographic noise when the population is small.

The study of pathwise approximation of stochastic mass-action systems by diffusion processes and the coupling of diffusion processes motivates a very interesting question. All our existing work relies on the reflection coupling of diffusion processes, which is known to be highly effective. Then how can one effectively couple two continuous-time Markov processes on a lattice? Successful coupling of two trajectories of a mass-action system will extend our framework of sensitivity analysis to many more applications. We believe it is very difficult to couple the exact stochastic mass-action system because of the difficulty of letting jumps of coupled processes occur simultaneously. However, there may be some way of building a "discrete reflection" and coupling two tau-leaping trajectories, i.e., two trajectories of (2.10), effectively. This will be addressed in our future work.

Appendix A. Expressions of mass-action systems and their diffusion approximations. In this section, we provide the explicit formulas of the Poisson approximation and the diffusion approximation for each model to improve readability.

A.1. SIR model. There are four reactions involved in the SIR system, so there are 4 pairs of Poisson process P_i and Wiener process B_i that appear in the evolution of each class. The rule of updating the numerical approximation $\hat{X}_n$ is as follows:

$$\begin{aligned}\hat{X}_{n+1} &= \begin{pmatrix} S_{n+1} \\ I_{n+1} \end{pmatrix} \\ &= \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} [P_1(q_{1,n+1}) - P_1(q_{1,n})] - [P_2(q_{2,n+1}) - P_2(q_{2,n}] - [P_3(q_{3,n+1} - P_3(q_{3,n})] \\ [P_2(q_{2,n+1}) - P_2(q_{2,n})] - [P_4(q_{4,n+1}) - P_4(q_{4,n})] \end{pmatrix} \\ &:= \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{f}_1(P_1, \cdots, P_4, q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{f}_2(P_1, \cdots, P_4, q_{1,n}, \cdots, q_{4,n}) \end{pmatrix},\end{aligned}$$

where $P_i, i = \{1, 2, 3, 4\}$, are independent unit rate Poisson processes.

And the rule of updating the numerical approximation $\hat{Y}_n$ is as follows:

$$\begin{aligned}\hat{Y}_{n+1} &= \begin{pmatrix} S_{n+1} \\ I_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} [q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] - [q_{3,n+1} - q_{3,n}] \\ [q_{2,n+1} - q_{2,n}] - [q_{4,n+1} - q_{4,n}] \end{pmatrix} \\ &\quad + \frac{1}{V} \begin{pmatrix} [B_1(q_{1,n+1}) - B_1(q_{1,n})] - [B_2(q_{2,n+1}) - B_2(q_{2,n})] - [B_3(q_{3,n+1} - B_3(q_{3,n}] \\ [B_2(q_{2,n+1}) - B_2(q_{2,n})] - [B_4(q_{4,n+1}) - B_4(q_{4,n})] \end{pmatrix} \\ &:= \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g}_1(q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{g}_2(q_{1,n}, \cdots, q_{4,n}) \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{\sigma}_1(B_1, \cdots, B_4, q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{\sigma}_2(B_1, \cdots, B_4, q_{1,n}, \cdots, q_{4,n}) \end{pmatrix},\end{aligned}$$

where $B_i,\ i = \{1, 2, 3, 4\}$, are independent standard Wiener processes.

As two classes S_n and I_n and four reactions are considered in this SIR model, the corresponding diffusion matrix M should be a 2×4 matrix. Specifically, the diffusion matrix M reads as follows:

$$\begin{aligned}\hat{Y}_{n+1} &= \begin{pmatrix} S_{n+1} \\ I_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} [q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] - [q_{3,n+1} - q_{3,n}] \\ [q_{2,n+1} - q_{2,n}] - [q_{4,n+1} - q_{4,n}] \end{pmatrix} \\ &\quad + \frac{1}{V} \begin{pmatrix} \sqrt{q_{1,n+1} - q_{1,n}} & -\sqrt{q_{2,n+1} - q_{2,n}} & -\sqrt{q_{3,n+1} - q_{3,n}} & 0 \\ 0 & \sqrt{q_{2,n+1} - q_{2,n}} & 0 & -\sqrt{q_{4,n+1} - q_{4,n}} \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{pmatrix} \\ &= \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} Vh\alpha - Vh\beta S_n I_n - Vh\mu S_n \\ Vh\beta S_n I_n - Vh(\mu + \rho + \gamma) I_n \end{pmatrix} \\ &\quad + \frac{1}{V} \begin{pmatrix} \sqrt{Vh\alpha} & -\sqrt{Vh\beta S_n I_n} & -\sqrt{Vh\mu S_n} & 0 \\ 0 & \sqrt{Vh\beta S_n I_n} & 0 & -\sqrt{Vh(\mu + \rho + \gamma) I_n} \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{pmatrix} \\ &:= \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g}_1(q_{1,n}, \cdots, q_{4,n}) \\ \boldsymbol{g}_2(q_{1,n}, \cdots, q_{4,n}) \end{pmatrix} + \frac{1}{V} M \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{pmatrix},\end{aligned}$$

where $W_i,\ i = \{1, 2, 3, 4\}$, are independent standard normal distributed random variables.

A.2. Oregonator model. For the Oregonator model, there are six reactions involved. So we have 6 pairs of Poisson processes P_i and B_i in the approximations. The rule of updating the numerical approximation $\hat{X}_n$ follows:

$$\hat{X}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \end{pmatrix} = \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \end{pmatrix} := \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{f_1}(P_1, \cdots, P_6, q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{f_2}(P_1, \cdots, P_6, q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{f_3}(P_1, \cdots, P_6, q_{1,n}, \cdots, q_{6,n}) \end{pmatrix},$$

where

$$\begin{aligned} \boldsymbol{f_1} =& P_1(q_{1,n+1}) - P_1(q_{1,n})] - [P_2(q_{2,n+1}) - P_2(q_{2,n})] + [P_3(q_{3,n+1}) - P_3(q_{3,n})] \\ & - 2[P_4(q_{4,n+1}) - P_4(q_{4,n})] \\ \boldsymbol{f_2} =& -[P_1(q_{1,n+1}) - P_1(q_{1,n})] - [P_2(q_{2,n+1}) - P_2(q_{2,n})] + [P_5(q_{5,n+1}) - P_5(q_{5,n})] \\ \boldsymbol{f_3} =& 2[P_3(q_{3,n+1}) - P_3(q_{3,n})] - [P_5(q_{5,n+1}) - P_5(q_{5,n})] - [P_6(q_{6,n+1}) - P_6(q_{6,n})], \end{aligned}$$

where $P_i, i = \{1, \cdots, 6\}$, are independent unit rate Poisson processes.

The diffusion approximation $\hat{Y}$ can be written as

$$\begin{aligned} \hat{Y}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \end{pmatrix} :=& \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g_1}(q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{g_2}(q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{g_3}(q_{1,n}, \cdots, q_{6,n}) \end{pmatrix} \\ & + \frac{1}{V} \begin{pmatrix} \boldsymbol{\sigma_1}(B_1, \cdots, B_6, q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{\sigma_2}(B_1, \cdots, B_6, q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{\sigma_3}(B_1, \cdots, B_6, q_{1,n}, \cdots, q_{6,n}) \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{g_1} =& [q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] + [q_{3,n+1} - q_{3,n}] - 2[q_{4,n+1} - q_{4,n}] \\ \boldsymbol{g_2} =& -[q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] + [q_{5,n+1} - q_{5,n}] \\ \boldsymbol{g_3} =& 2[q_{3,n+1} - q_{3,n}] - [q_{5,n+1} - q_{5,n}] - [q_{6,n+1} - q_{6,n}] \\ \boldsymbol{\sigma_1} =& [B_1(q_{1,n+1}) - B_1(q_{1,n})] - [B_2(q_{2,n+1}) - B_2(q_{2,n})] + [B_3(q_{3,n+1}) - B_3(q_{3,n})] \\ & - 2[B_4(q_{4,n+1}) - B_4(q_{4,n})] \\ \boldsymbol{\sigma_2} =& -[B_1(q_{1,n+1}) - B_1(q_{1,n})] - [B_2(q_{2,n+1}) - B_2(q_{2,n})] + [B_5(q_{5,n+1}) - B_5(q_{5,n})] \\ \boldsymbol{\sigma_3} =& 2[B_3(q_{3,n+1}) - B_3(q_{3,n})] - [B_5(q_{5,n+1}) - B_5(q_{5,n})] - [B_6(q_{6,n+1}) - B_6(q_{6,n})], \end{aligned}$$

where $B_i, \{i = 1, \cdots, 6\}$, are independent Wiener processes.

As we focus on three classes $S_{1,n}, S_{2,n}, S_{3,n}$ and six reactions, we can confirm that the diffusion matrix M is a 3×6 matrix. Specifically, the diffusion matrix M is defined as follows:

$$\hat{Y}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \end{pmatrix} := \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g_1}(q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{g_2}(q_{1,n}, \cdots, q_{6,n}) \\ \boldsymbol{g_3}(q_{1,n}, \cdots, q_{6,n}) \end{pmatrix} + \frac{1}{V} M \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \\ W_6 \end{pmatrix},$$

where the 3×6 matrix M has expression

$$M = \begin{bmatrix} \sqrt{q_{1,n+1}-q_{1,n}} & -\sqrt{q_{2,n+1}-q_{2,n}} & \sqrt{q_{3,n+1}-q_{3,n}} & -\sqrt{2[q_{4,n+1}-q_{4,n}]} & 0 & 0 \\ -\sqrt{q_{1,n+1}-q_{1,n}} & -\sqrt{q_{2,n+1}-q_{2,n}} & 0 & 0 & \sqrt{q_{5,n+1}-q_{5,n}} & 0 \\ 0 & 0 & \sqrt{2[q_{3,n+1}-q_{3,n}]} & 0 & -\sqrt{q_{5,n+1}-q_{5,n}} & -\sqrt{q_{6,n+1}-q_{6,n}} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{VhC_1S_{2,n}} & -\sqrt{VhC_2S_{1,n}S_{2,n}} & \sqrt{VhC_3S_{1,n}} & 2\sqrt{VhC_4S_{1,n}^2} & 0 & 0 \\ -\sqrt{VhC_1S_{2,n}} & -\sqrt{VhC_2S_{1,n}S_{2,n}} & 0 & 0 & \sqrt{VhC_5\delta S_{3,n}} & 0 \\ 0 & 0 & 2\sqrt{VhC_3S_{1,n}} & 0 & -\sqrt{VhC_5\delta S_{3,n}} & -\sqrt{VhC_5(1-\delta)S_{3,n}} \end{bmatrix},$$

where $W_i, \{i = 1, \cdots, 6\}$, are independent standard normal distributed random variables.

A.3. 4D Lotka–Volterra model. For the 4D Lotka–Volterra system, there are 17 reactions involved, so we have 17 pairs of Poisson process P_i and Wiener process B_i. The rule of updating the numerical approximation $\hat{X}_n$ follows:

$$\hat{X}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \\ S_{4,n+1} \end{pmatrix} := \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \\ S_{4,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{f_1}(P_1, \cdots, P_{17}, q_{i,1}^n, \cdots, q_{i,5}^n) \\ \boldsymbol{f_2}(P_1, \cdots, P_{17}, q_{i,1}^n, \cdots, q_{i,5}^n) \\ \boldsymbol{f_3}(P_1, \cdots, P_{17}, q_{i,1}^n, \cdots, q_{i,5}^n) \\ \boldsymbol{f_4}(P_1, \cdots, P_{17}, q_{i,1}^n, \cdots, q_{i,5}^n), \end{pmatrix},$$

where

$$\begin{aligned} \boldsymbol{f_1} =& [P_1(q_{1,n+1}) - P_1(q_{1,n})] - [P_2(q_{2,n+1}) - P_2(q_{2,n})] - [P_3(q_{1,n+1}) - P_3(q_{1,n})] \\ & - [P_4(q_{4,n+1}) - P_4(q_{4,n})] \\ \boldsymbol{f_2} =& [P_5(q_{5,n+1}) - P_5(q_{5,n})] - [P_6(q_{6,n+1}) - P_6(q_{6,n})] - [P_7(q_{7,n+1}) - P_7(q_{7,n})] \\ & - [P_8(q_{8,n+1}) - P_8(q_{8,n})] \\ \boldsymbol{f_3} =& [P_9(q_{9,n+1}) - P_9(q_{9,n})] - [P_{10}(q_{10,n+1}) - P_{10}(q_{10,n})] \\ & - [P_{11}(q_{11,n+1}) - P_{11}(q_{11,n})] - [P_{12}(q_{12,n+1}) - P_{12}(q_{12,n})] \\ \boldsymbol{f_4} =& [P_{13}(q_{13,n+1}) - P_{13}(q_{13,n})] - [P_{14}(q_{14,n+1}) - P_{14}(q_{14,n})] \\ & - [P_{15}(q_{15,n+1}) - P_{15}(q_{15,n})] - [P_{16}(q_{16,n+1}) - P_{16}(q_{16,n})] \\ & - [P_{17}(q_{17,n+1}) - P_{17}(q_{17,n})], \end{aligned}$$

where $i = 1, \cdots, 4$ and $P_j, \{j = 1, \cdots, 17\}$ are independent unit rate Poisson processes. The diffusion approximation $\hat{Y}$ can be written as

$$\hat{Y}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \\ S_{4,n+1} \end{pmatrix} := \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \\ S_{4,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{g_1}(q^n_{i,1}, \cdots, q^n_{i,5}) \\ \boldsymbol{g_2}(q^n_{i,1}, \cdots, q^n_{i,5}) \\ \boldsymbol{g_3}(q^n_{i,1}, \cdots, q^n_{i,5}) \\ \boldsymbol{g_4}(q^n_{i,1}, \cdots, q^n_{i,5}) \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \boldsymbol{\sigma_1}(B_1, \cdots, B_{17}, q^n_{i,1}, \cdots, q^n_{i,5}) \\ \boldsymbol{\sigma_2}(B_1, \cdots, B_{17}, q^n_{i,1}, \cdots, q^n_{i,5}) \\ \boldsymbol{\sigma_3}(B_1, \cdots, B_{17}, q^n_{i,1}, \cdots, q^n_{i,5}) \\ \boldsymbol{\sigma_4}(B_1, \cdots, B_{17}, q^n_{i,1}, \cdots, q^n_{i,5}) \end{pmatrix},$$

where

$$\begin{aligned}
\boldsymbol{g_1} =& [q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] - [q_{3,n+1} - q_{3,n}] - [q_{4,n+1} - q_{4,n}] \\
\boldsymbol{g_2} =& [q_{5,n+1} - q_{5,n}] - [q_{6,n+1} - q_{6,n}] - [q_{7,n+1} - q_{7,n}] - [q_{8,n+1} - q_{8,n}] \\
\boldsymbol{g_3} =& [q_{9,n+1} - q_{9,n}] - [q_{10,n+1} - q_{10,n}] - [q_{11,n+1} - q_{11,n}] - [q_{12,n+1} - q_{12,n}] \\
\boldsymbol{g_4} =& [q_{13,n+1} - q_{13,n}] - [q_{14,n+1} - q_{14,n}] - [q_{15,n+1} - q_{15,n}] - [q_{16,n+1} - q_{16,n}] \\
& - [q_{17,n+1} - q_{17,n}] \\
\boldsymbol{\sigma_1} =& [B_1(q_{1,n+1}) - B_1(q_{1,n})] - [B_2(q_{2,n+1}) - B_2(q_{2,n})] - [B_3(q_{3,n+1}) - B_3(q_{3,n})] \\
& - [B_4(q_{4,n+1}) - B_4(q_{4,n})] \\
\boldsymbol{\sigma_2} =& [B_5(q_{5,n+1}) - B_5(q_{5,n})] - [B_6(q_{6,n+1}) - B_6(q_{6,n})] - [B_7(q_{7,n+1}) - B_7(q_{7,n})] \\
& - [B_8(q_{8,n+1}) - B_8(q_{8,n})] \\
\boldsymbol{\sigma_3} =& [B_9(q_{9,n+1}) - B_9(q_{9,n})] - [B_{10}(q_{10,n+1}) - B_{10}(q_{10,n})] \\
& - [B_{11}(q_{11,n+1}) - B_{11}(q_{11,n})] - [B_{12}(q_{12,n+1}) - B_{12}(q_{12,n})] \\
\boldsymbol{\sigma_4} =& [B_{13}(q_{13,n+1}) - B_{13}(q_{13,n})] - [B_{14}(q_{14,n+1}) - B_{14}(q_{14,n})] \\
& - [B_{15}(q_{15,n+1}) - B_{15}(q_{15,n})] - [B_{16}(q_{16,n+1}) - B_{16}(q_{16,n})] \\
& - [B_{17}(q_{17,n+1}) - B_{17}(q_{17,n})],
\end{aligned}$$

where $i = 1, \cdots, 4$, $B_j, \{j = 1, \cdots, 17\}$, are independent Wiener processes.

REFERENCES

[1] A. Agresti and B. A. Coull, *Approximate is better than "exact" for interval estimation of binomial proportions*, Amer. Statist., 52 (1998), pp. 119–126.

[2] D. F. Anderson and T. G. Kurtz, *Continuous time Markov chain models for chemical reaction networks*, in Design and Analysis of Biomolecular Circuits, Springer, 2011, pp. 3–42.

[3] M. Benaim, *Stochastic Persistence*, preprint, arXiv:1806.08450, 2018.

[4] M. Brøns and K. Bar-Eli, *Canard explosion and excitation in a model of the Belousov-Zhabotinskii reaction*, J. Phys. Chem., 95 (1991), pp. 8706–8713.

[5] J. N. Darroch and E. Seneta, *On quasi-stationary distributions in absorbing discrete-time finite Markov chains*, J. Appl. Probab., 2 (1965), pp. 88–100.

[6] E. Deutsch and M. Neumann, *On the first and second order derivatives of the Perron vector*, Linear Algebra Appl., 71 (1985), pp. 57–76.

[7] N. T. Dieu, D. H. Nguyen, N. H. Du, and G. Yin, *Classification of asymptotic behavior in a stochastic SIR model*, SIAM J. Appl. Dyn. Syst., 15 (2016), pp. 1062–1084, https://doi.org/10.1137/15M1043315.

[8] M. Dobson, Y. Li, and J. Zhai, *Using coupling methods to estimate sample quality of stochastic differential equations*, SIAM/ASA J. Uncertain. Quantif., 9 (2021), pp. 135–162, https://doi.org/10.1137/20M1312009.
[9] I. R. Epstein and J. A. Pojman, *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos*, Oxford University Press, 1998.
[10] M. A. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. A, 104 (2000), pp. 1876–1889.
[11] D. T. Gillespie, et al., *Stochastic simulation of chemical kinetics*, Annu. Rev. Phys. Chem., 58 (2007), pp. 35–55.
[12] P. Gray and S. Scott, *Chemical Oscillations and Instabilities: Nonlinear Chemical Kinetics*, Clarendon, Oxford, 1990.
[13] P. E. Jacob, J. O'Leary, and Y. F. Atchadé, *Unbiased Markov chain Monte Carlo methods with couplings*, J. R. Stat. Soc. Ser. B Stat. Methodol., 82 (2020), pp. 543–600.
[14] V. E. Johnson, *A coupling-regeneration scheme for diagnosing convergence in Markov chain Monte Carlo algorithms*, J. Amer. Statist. Assoc., 93 (1998), pp. 238–248.
[15] J. Komlós, P. Major, and G. Tusnády, *An approximation of partial sums of independent RV's, and the sample DF.* I, Zeitschrift für Wahrscheinlichkeitstheorie und Verw. Gebiete, 32 (1975), pp. 111–131.
[16] J. Komlós, P. Major, and G. Tusnády, *An approximation of partial sums of independent RV's, and the sample DF.* II, Zeitschrift für Wahrscheinlichkeitstheorie und Verw. Gebiete, 34 (1976), pp. 33–58.
[17] Y. Li and L. Hu, *A fast exact simulation method for a class of Markov jump processes*, J. Chem. Phys., 143 (2015), 184105.
[18] Y. Li and S. Wang, *Numerical computations of geometric ergodicity for stochastic dynamics*, Nonlinearity, 33 (2020), 6935.
[19] Y. Li and Y. Yuan, *Data-driven computational methods for quasi-stationary distribution and sensitivity analysis*, J. Dynam. Differential Equations, (2022), pp. 1–29.
[20] A. J. Lotka, *Elements of physical biology*, Science Progress in the Twentieth Century (1919–1933), 21 (1926), pp. 341–343.
[21] J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise*, Stoch. Process. Appl., 101 (2002), pp. 185–232.
[22] C. D. Meyer, *Sensitivity of the stationary distribution of a Markov chain*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 715–728, https://doi.org/10.1137/S0895479892228900.
[23] J. Miyazaki, *Pattern Formations and Oscillatory Phenomena:* 2. *Belousov-Zhabotinsky Reaction*, Elsevier, 2013.
[24] P. Mozgunov, M. Beccuti, A. Horvath, T. Jaki, R. Sirovich, and E. Bibbona, *A review of the deterministic and diffusion approximations for stochastic chemical reaction networks*, React. Kinet. Mech. Catal., 123 (2018), pp. 289–312.
[25] J. D. Murray, *Mathematical Biology* II*: Spatial Models and Biomedical Applications*, 3rd ed., Springer, New York, 2003.
[26] A. Slepoy, A. P. Thompson, and S. J. Plimpton, *A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks*, J. Chem. Phys., 128 (2008), 05B618.
[27] V. Volterra, *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*, Memoria della Reale Accademia Nazionale dei Lincei, 2 (1926), pp. 31–113.
[28] C. Xu, *Threshold dynamics of a stochastic Keizer's model with stochastic incidence*, J. Math. Chem., 55 (2017), pp. 1034–1045.
[29] J. Zhai, M. Dobson, and Y. Li, *A deep learning method for solving Fokker-Planck equations*, in Mathematical and Scientific Machine Learning, PMLR, 2022, pp. 568–597.