SENSITIVITY ANALYSIS OF QUASI-STATIONARY-DISTRIBUTIONS (QSDS)

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 finite 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) instead of invariant probability measures. In this paper we modify the coupling based technique developed in [8] to estimate an upper bound of the 1-Wasserstein distance between two QSDs. Some numerical results for sensitivity with different population sizes are provided.

1. INTRODUCTION

A mass-action network is a system of finite many species and reactions whose rule of update satisfies the mass-action law. Mass-action network covers a large number of chemical reaction network, 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 the stoichiometric equations. The time of these random events satisfy mass-action laws. Therefore, a mass-action network can be mathematically described by a continuous-time Markov process, which is driven by finite many Poisson processes.

The randomness in updating the network is called the demographic noise in population and epidemiology models. It is well known that demographic noise leads to finite time extinction in a very 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, the noise could become the dominate term and leads 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 distribution conditioning on not hitting the absorbing set yet. As discussed in [20], when the extinction rate is low, the quasi-stationary distribution can be well approximated by the invariant probability measure of a modified process that artificially "pushes" the trajectory away from the extinction.

It has been known for decades that when the population size is large, the continuous-time Markov process converges to the mass-action ordinary differential equations (ODEs). In addition, by setting up a martingale problem, one can show that the re-scaled 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. We refer [2,10] for further details.

The goal of this paper is to study the sensitivity of QSDs against the diffusion approximation. We are interested in how QSDs of the Markov process and its diffusion approximation differs from each other. The motivation is that an exact simulation at the molecule level is usually computationally expensive even if the stochastic simulation algorithm (SSA) is implemented optimally [11,18,26]. It
is even harder to numerically compute the QSD when the number of molecule is large. On the other hand, the simulation of a diffusion process is much easier. The technique of computing the invariant probability measure or QSD of a stochastic differential equation is also well developed \cite{19,20,28}. Hence it is important to have a quantitative upper bound of the difference between the QSD of a mass-action system and that of its diffusion approximation.

The way of sensitivity analysis is developed from on the coupling-based method in \cite{8}. We need both finite time truncation error and the rate of contraction of the transition kernel of the diffusion process. The finite time error is given by the KMT algorithm in \cite{24}. With the explicit construction of coupled trajectories of the Poisson process and the diffusion process, the finite time error up to fixed time $T$ can be computed. The rate of contraction is modified from the data-driven method proposed in \cite{19}. 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 volume (i.e., molecule count).

The organization of this paper is as follows. A short preliminary about reaction networks, rates for the law of mass action, Poisson process, diffusion process and coupling times is 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 Poisson process and Wiener process are shown in the appendix.

2. Preliminary

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

\begin{equation}
\sum_{i=1}^{d} c_{ki} S_i \rightarrow \sum_{i=1}^{d} c'_{ki} S_i, \quad k = 1, \cdots, K
\end{equation}

where $c_{ki}$ and $c'_{ki}$ are non-negative integers that denote the number of molecules of species $S_i$ consumed and produced by reaction $k$, respectively. Let $V$ be the volume of the reaction system.

Let $X(t) = (x_1(t), \cdots, x_d(t)) \in \mathbb{R}^d$ be the state of the mass action system at time $t$, such that the $i$-th entry of $X(t)$ represents the concentration of species $S_i$, $i = 1, \cdots, d$. In other words the number of molecules of $S_i$ is $V x_i := N_i$. Let $\lambda_k$ be the rate at which the $k$-th reaction occurs, that is, it gives the propensity of the $k$-th reaction as a function of the concentrations of molecules of the chemical species.

2.2. Rates for the law of mass action. The law of mass action means the rate of a reaction should be proportional to the number of distinct subsets of the participating molecules. More precisely, the rate of reaction $k$ reads

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

where $\kappa_k$ is a rate constant, and $N_i$ be the number of molecule of the $i$th 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

$$
P[\text{reaction } k \text{ occurs in } [t, t + \Delta t]] = \lambda_k \Delta t + O(\Delta t^2).
$$
2.3. Poisson process. We use Poisson counting process to represent \( X(t) \), because \( X(t) \) is a continuous time discrete state Markov chain. Let \( X_i(t) \) be \( i \)-th entry of \( X(t) \), then

\[
X_i(t) = X_i(0) + \frac{1}{V} \sum_k R_k(t)(c'_k - c_k),
\]

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

\[
R_k(t) = P_k(V \int_0^t f_k(X(s))ds),
\]

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

\[
(2.2) \quad X(t) = X(0) + \sum_k l_k P_k(V \int_0^t f_k(X(s))ds)
\]

where \( P_k(t), \ k = \{1, \ldots, K\} \) 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 \( k \).

2.4. Diffusion process. When \( V \) is large, a Poisson process can be approximated by a diffusion process. The follow lemma in [16,17] gives the strong approximation theorem for Poisson processes.

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

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

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

**Remark 2.1.** By the scaling property of 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

\[
P_k\left(V \int_0^t f_k(X(s))ds\right) \approx V \int_0^t f_k(X(s))ds + \int_0^t \sqrt{V} f_k(X(s))dB(s)
\]

\[
= V \int_0^t f_k(X(s))ds + B_k\left(V \int_0^t f_k(X(s))ds\right)
\]

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 as follow theorem in [24].

**Theorem 2.2.** Let \( X(t) \) be a Poisson process represented by (2.3), let \( Y(t) \) be a diffusion process with initial condition satisfying \( X(0) = Y(0) \) and solves the following stochastic differential equation

\[
(2.4) \quad Y(t) = Y(0) + \sum_k \frac{l_k}{V} \left[ V \int_0^t f_k(Y(s))ds + B_k\left(V \int_0^t f_k(Y(s))ds\right) \right]
\]
where the $B_k(\cdot)$ are independent standard Wiener processes. As $V \to \infty$,
\begin{equation}
\sup |X(t) - Y(t)| = O\left(\frac{\log V}{V}\right).
\end{equation}

The error of diffusion approximation is proportional to $\frac{\log V}{V}$, which converges to 0 as $V \to \infty$. In macroscopic chemical reaction system $V$ 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 consider 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. Paper [24] gives the constructive procedure to generate discretized approximation. For any finite capacity $V$, the error of diffusion approximation is proportional to $\log V$. In section 3, trajectories of the two processes $X(t)$ and $Y(t)$ on the same probability space 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.

2.5. Coupling times. In this paper, we use the coupling argument to connect finite time error and the distance between QSDs. Let $\mu$ and $\nu$ be two probability measures on a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. A coupling between $\mu$ and $\nu$ is a probability measure $\gamma$ on the product space $(\mathcal{X} \times \mathcal{X}, \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X}))$ such that two marginal distribution of $\gamma$ are $\mu$ and $\nu$ respectively.

**Definition 2.1.** (Wasserstein distance) Let $d$ be a metric on the state space $S$. For probability measures $\mu$ and $\nu$ on $S$, the Wasserstein distance between $\mu$ and $\nu$ for $d$ is given by
\begin{equation}
\begin{align*}
&d_{\text{w}}(\mu, \nu) = \inf\{\mathbb{E}\|d(x, y)\| : \gamma \text{ is a coupling of } \mu \text{ and } \nu\} \\
&\quad = \inf\left\{\int d(x, y)\gamma(dx, dy) : \gamma \text{ is a coupling of } \mu \text{ and } \nu\right\},
\end{align*}
\end{equation}

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

Let $Z_t^{(1)}$ and $Z_t^{(2)}$ be two stochastic processes. A coupling between $Z_t^{(1)}$ and $Z_t^{(2)}$ can be defined in the same way on the space of paths. 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.

**Definition 2.2.** (Coupling time) The coupling time of a Markov coupling $(Z_t^{(1)}, Z_t^{(2)})$ is a random variable given by
\begin{equation}
\tau_c^{\text{def}} = \inf\{t \geq 0 : Z_t^{(1)} = Z_t^{(2)}\}.
\end{equation}

**Definition 2.3.** (Successful coupling) A coupling $(Z_t^{(1)}, Z_t^{(2)})$ of Markov processes $Z^{(1)}$ and $Z^{(2)}$ is said to be successful if
\begin{equation}
\mathbb{P}(\tau_c < \infty) = 1.
\end{equation}

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

**Definition 2.4.** (Reflection coupling) Let $Z_t^{(1)}$ and $Z_t^{(2)}$ be two solutions of a stochastic differential equation
\begin{equation}
dZ_t = f(Z_t)dt + \sigma(Z_t)dB_t
\end{equation}
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{equation}
\begin{align*}
dZ_t^{(1)} &= f(Z_t^{(1)})dt + \sigma(Z_t^{(1)})dB_t \\
dZ_t^{(2)} &= f(Z_t^{(2)})dt + \sigma(Z_t^{(2)})(I - 2ee^T)dB_t
\end{align*}
\end{equation}
where $B$ is a standard Wiener process, and

$$e = \frac{1}{\|\sigma^{-1}(Z_{t}^{(1)} - Z_{t}^{(2)})\|}\sigma^{-1}(Z_{t}^{(1)} - Z_{t}^{(2)})$$

is a unite vector.

We remark that the reflection coupling requires $\sigma(Z_t)$ in equation (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 equation (2.4) equals the number of reactions. Hence we often need to find an equivalent diffusion process with an invertible $\sigma$. See numerical examples for additional details.

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

Definition 2.5. (Maximal coupling) The maximal coupling looks for the maximal coupling probability for the next step of $Z_{t-1}^{(1)}$ and $Z_{t-1}^{(2)}$. Assume $Z_{t-1}^{(1)}$ and $Z_{t-1}^{(2)}$ are known and the probability density function of $Z_{t}^{(1)}$ and $Z_{t}^{(2)}$ is easy to compute. Following [13, 15], the update of $Z_{t}^{(1)}$ and $Z_{t}^{(2)}$ in Algorithm 1 maximizes the probability of coupling.

Algorithm 1 Maximal coupling

Input: $Z_{t-1}^{(1)}$ and $Z_{t-1}^{(2)}$
Output: $Z_{t}^{(1)}$ and $Z_{t}^{(2)}$, and $\tau_c$ if coupled

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

2.6. 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 equation (2.3) also has a strong approximation given by equation (2.4). As the processes $P_k(\cdot)$ and $B_k(\cdot)$ are continuous time processes, we apply the $\tau$-leaping approximation for equation (2.3) with the same step size $h$. This gives

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

(2.10)

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

$$\hat{Y}_{n+1} = \hat{Y}_n + \sum_k \frac{l_k}{V} (Vh f_k(\hat{Y}_n))$$

$$+ \sum_k \frac{l_k}{V} \left[ B_k \left( Vh \sum_{m=0}^{n} f_k(\hat{Y}_m) \right) - B_k \left( Vh \sum_{m=0}^{n-1} f_k(\hat{Y}_m) \right) \right]$$

(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 actually 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 re-scaling, one obtains a pair of discretized trajectories of $P_k(t)$ and $B_k(t)$ respectively. We refer [24] for a detailed review of the KMT algorithm.

### 3. Sensitivity of diffusion approximation

#### 3.1. Quasi-stationary distribution

Let $X = \{X(t)\}$ and $\hat{X} = \{\hat{X}_n\}$ (resp. $Y = \{Y(t)\}$ and $\hat{Y} = \{\hat{Y}_n\}$) be the stochastic process given by (3.2) (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 that how much the long time dynamics of $X$ is preserved by its diffusion approximation. This problem is more complicated than it looks because both $X$ and $Y$ 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 equation (2.3) or equation (2.4) to have finite time extinction. To see this, consider the 1D version of equation (2.4):

$$dY(t) = f(Y(t))dt + \frac{1}{\sqrt{V}}\sqrt{f(Y(t))}dB_t. \quad (3.1)$$

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

$$\frac{d}{dt}E[H(Y(t))] = f(Y(t)) \left( \frac{2}{(Y(t))^3} - \frac{1}{(Y(t))^2} \right).$$

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

$$\frac{d}{dt}E[H(Y(t))] \geq 2c(E[H(Y(t))]^2),$$

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

Therefore, to prevent finite time extinction, usually one needs constant influx of each species. That is why often we need to study the quasi-stationary distribution (QSD) instead of the invariant probability distribution. Below we introduce the QSD and its sampling method only for $X$, as the case of $Y$ is analogous.

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

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

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

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

$$P_\mu[X(t) \in C|\tau > t] = \mu(C), \ t \geq 0, \quad (3.2)$$

**Definition 3.2.** If there is a probability measure $\mu$ exists such that

$$\lim_{t \to \infty} P_x[X(t) \in C|\tau > t] = \mu(C), \ \forall x \in \mathbb{R}^d_+, \quad (3.3)$$

in which case we also say that $\mu$ is a quasi-limiting distribution (QLD).
Remark 3.1. The limiting probability measure given by equation (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 \( X \) to \( \partial \mathcal{X} \) is \( \infty \) with probability one, \( \{ \tau > t \} \) is the full probability space. As a result, QSD in equation (3.2) becomes the invariant probability measure and QLD in equation (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 rule of regeneration such that they both sample from QSDs as the time approaches to infinity. To sample from QSD, we need to regenerate a sample once it hits the absorbing set. Therefore, in addition to \( X_n \), we also need to update a temporal occupation measure

\[
\mu_n = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_k}.
\]

If the numerical scheme gives \( X_{n+1} \in \partial \mathcal{X} \), we immediately regenerate \( X_{n+1} \) from \( \mu_n \). More precisely, let the transition kernel of the numerical scheme of \( X_n \) (without resampling) be \( Q \). Then \( Q \) has an absorbing set \( \partial \mathcal{X} \) such that \( Q(\partial \mathcal{X}, \partial \mathcal{X}) = 1 \). The transition kernel of \( X_n \) is the sum of \( Q \) and the regeneration measure such that

\[
P[X_{n+1} \in A \mid X_n = x] = \hat{Q}(x, A) + \hat{Q}(x, \partial \mathcal{X}) \mu_n(A).
\]

The following convergence result follows from [3].

**Proposition 3.1** (Theorem 2.5 in [3]). Let \( \hat{\mu} \) be the QSD of the numerical process \( X_n \). Under suitable assumptions about \( X_n \), the occupation measure \( \mu_n \) converges to the QSD \( \hat{\mu} \) as \( n \to \infty \).

To study the sensitivity of diffusion approximation, we also need a theoretical process \( \bar{X} = \{ \bar{X}_n \} \) that directly regenerate from the QSD \( \hat{\mu} \) once exit to the boundary. Recall that \( \bar{Q} \) is the transition kernel of \( \bar{X}_n \) (without resampling). The transition kernel of \( \bar{X} \) is

\[
\bar{P}(x, \cdot) = \hat{Q}(x, \cdot) + \hat{Q}(x, \partial \mathcal{X}) \hat{\mu}(\cdot)
\]

for all \( x \in \mathbb{R}_+^d \). Note that \( \bar{X}_n \) is not a Markov process (but \( \bar{X}_n, \mu_n \) is a Markov process). But \( \bar{X} \) is a homogeneous Markov process with an invariant probability measure \( \bar{\mu} \). The case of \( Y(t) \) is analogous. We denote the numerical process that resample from a temporal occupation measure by \( \bar{Y} = \{ \bar{Y}_n \} \), and the Markov process that directly resample from QSD by \( \bar{Y} = \{ \bar{Y}_n \} \).

### 3.2. Decomposition of error term.

Let \( P_X \) and \( \bar{P}_X \) be the transition kernels of \( X(t) \) and \( \bar{X}_n \) respectively. Let \( P_Y \) and \( \bar{P}_Y \) be that of \( Y(t) \) and \( \bar{Y}_n \) respectively. Denote the QSDs of \( X(t) \), \( \bar{X}_n \), \( Y(t) \) and \( \bar{Y}_n \) by \( \pi_X \), \( \pi\bar{X} \), \( \pi_Y \), and \( \pi\bar{Y} \), respectively. The quantity that we are interested in is \( d_w(\pi_X, \pi_Y) \).

Let \( T \) be a fixed constant. Motivated by [14], the following decomposition follows easily by the triangle inequality and the invariance.

\[
d_w(\pi_X, \pi_Y) \leq d_w(\pi_X, \pi\bar{X}) + d_w(\pi\bar{X}, \pi\bar{Y}) + d_w(\pi\bar{Y}, \pi_Y)
\]

The sensitivity of invariant probability against time discretization has been addressed in [8]. When the time step size of the time discretization is small enough, the invariant probability measure \( \pi_Y \) is close to the numerical invariant probability measure \( \pi\bar{Y} \). The case of QSD is analogous. Hence the third term \( d_w(\pi\bar{Y}, \pi_Y) \) is proportional to step size \( h \). The estimation of the first term \( d_w(\pi_X, \pi\bar{X}) \) can be obtained by some linear algebraic calculation.
Theorem 3.1. Let $X(t)$ be a continuous time Markov chain with finite state space and $\hat{X}$ be its tau-leaping approximation with step size $h$. Suppose that $\pi$ and $\hat{\pi}$ be the true QSD and the numerical approximation of QSD respectively. Let $h$ be the time step size of numerical process. If the generating matrix of $X(t)$ is irreducible, then

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

for $0 < h \ll 1$.

Proof. This proof follows the standard argument of 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 $\pi$ is true QSD and $\hat{\pi}$ is the numerical approximation of QSD, we have

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

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

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

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 $\pi$ is reduced to the derivative of $A(t)$.

Since $\pi$ is normalized to 1 in $l_1$ norm, it follows from [6] Section 3 that

$$\pi'(0) = S^2 A'(0) \pi(0),$$

where $S = \lambda I - A(0)$, and $S^2$ is the group inverse of $S$. (We refer [6] for further discussion of the group inverse and derivative of 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 $S^2 = O(h^{-1})$. 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. \hfill \Box

Therefore, we have that $d_w(\pi_X, \hat{\pi}_X) = O(h)$ and $d_w(\pi_Y, \hat{\pi}_Y) = O(h)$, which make the second error term be the key part. The second error term is the difference between numerical Poisson process of a mass-action system and its corresponding numerical diffusion process.

Proposition 3.2. Let $T > 0$ be a fixed constant. We can decompose $d_w(\pi_X, \hat{\pi}_X)$ via the following inequality:

$$d_w(\pi_X, \hat{\pi}_X) \leq d_w(\pi_X \hat{P}^T_X, \hat{\pi}_X \hat{P}^T_Y) + d_w(\pi_X \hat{P}^T_X, \hat{\pi}_Y \hat{P}^T_Y)$$

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

There are two different ways to think about the distance $d_w(\pi_X, \hat{\pi}_X)$. One method is considering $\pi_X$ and $\hat{\pi}_X$ as conditional distributions on set $\mathbb{R}_+^d / \partial X$, i.e. $\pi_X(A) = \{X \in A | t < \tau_X\}$ and $\hat{\pi}_X(A) = \{\hat{Y} \in A | t < \tau_Y\}$, where $\tau_X$ and $\tau_Y$ are the killing time for processes $\hat{X}$ and $\hat{Y}$, respectively. The other way is to use the $\tilde{X}$ and $\tilde{Y}$ that regenerate from QSDs. No conditioning is needed as $\tilde{\mu}_X$ and $\tilde{\mu}_Y$ are now the invariant probability measures of $\tilde{X}$ and $\tilde{Y}$ respectively. There are some fundamental difficulty when computing the conditional finite time error because it is hard to couple $\tilde{X}_n$ and $\tilde{Y}_n$ when one regenerates while the other does not. Hence we choose to use $\tilde{X}$ and $\tilde{Y}$ instead.
3.3. Finite time error. We consider the modified processes $\tilde{X}$ and $\tilde{Y}$, which are regenerated from the corresponding QSDs when they hit the boundary. Let $\tilde{\pi}_X$ and $\tilde{\pi}_Y$ be the invariant measures of $\tilde{X}$ and $\tilde{Y}$. Let $\tilde{\Gamma}(dx,dy) = \tilde{\pi}_X^2(\tilde{P}_X^T \times \tilde{P}_Y^T)$, where $\tilde{\pi}_X^2$ is the coupled measure of $\tilde{\pi}_X$ 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 $\tilde{\pi}_X^2(\{(x,x) | x \in A\}) = \tilde{\pi}_X(A)$, and $\tilde{P}_X^T \times \tilde{P}_Y^T$ is any coupled process such that two marginal processes are $\tilde{X}$ and $\tilde{Y}$ respectively. The following proposition follows easily.

**Proposition 3.3.** Let $(\tilde{X}_n, \tilde{Y}_n)$ be a coupling of $\tilde{X}$ and $\tilde{Y}$ with transition kernel $\tilde{P}_X^T \times \tilde{P}_Y^T$, then

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

**Proof.** By the definition of Wasserstein distance

$$d_w(\tilde{\pi}_X \tilde{P}_X^T, \tilde{\pi}_Y \tilde{P}_Y^T) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} d(x,y) \tilde{\pi}_X^2(\tilde{P}_X^T \times \tilde{P}_Y^T)(dx,dy)$$

$$= \int_{\mathbb{R}^d} \mathbb{E}_{(x,y)} d(\tilde{X}_T, \tilde{Y}_T) \tilde{\pi}_X(dx) = \mathbb{E}_{\tilde{\pi}_X}[d(\tilde{X}_T, \tilde{Y}_T)].$$

The key of 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 matching Wiener processes and Poisson processes. Here it remains to define how $\tilde{X}_n$ and $\tilde{Y}_n$ couple when they regenerate from QSDs. Since we do not have QSD in priori, we will use $\tilde{X}_n$ and $\tilde{Y}_n$ to approximate $\tilde{X}_n$ and $\tilde{Y}_n$. In other words, we regenerate samples from the temporal occupation measure. To minimize error during sample regeneration, we define the following coupled process $(\tilde{X}_n, \tilde{\mu}_X^n)$ and $(\tilde{Y}_n, \tilde{\mu}_Y^n)$, such that $\tilde{X}_n$ and $\tilde{Y}_n$ follows equations (2.10) and (2.11) respectively by using paired processes $B_k(t)$ and $P_k(t)$ for each $k$, and $\tilde{\mu}_X^n$, $\tilde{\mu}_Y^n$ 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 $N_Y$ 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 increase $N_X$ by one and choose the $N_X$-th element of $S$, $Z_{NX}$ to regenerate $\tilde{X}_{n+1}$, by letting $\tilde{X}_{n+1} = \tilde{X}_{\lfloor Z_{NX} n \rfloor}$. Then it is easy to see that $(\tilde{X}_n, \tilde{\mu}_X^n)$ and $(\tilde{Y}_n, \tilde{\mu}_Y^n)$ is a Markov coupling and the marginal processes $(\tilde{X}_n, \tilde{Y}_n)$ is a coupling of equations (2.10) and (2.11).

Details of computation are shown in Algorithm 2. When $N$ is large, initial values $\tilde{X}_1^1, \cdots, \tilde{X}_1^M$ in Algorithm 2 are from a trajectory of the time-$T$ skeleton of $\tilde{X}_T$. Hence $\tilde{X}_1^1, \tilde{X}_2^1, \cdots, \tilde{X}_1^M$ are approximately sampled from $\tilde{\pi}_X$. The error term $d(\tilde{X}_T, \tilde{Y}_T)$ is evolved from the initial value pair $\tilde{X}_1^m = \tilde{Y}_1^m = \tilde{X}_T^{m-1}$ is recorded. Therefore,

$$\frac{1}{M} \sum_{m=1}^M d(\tilde{X}_T, \tilde{Y}_T)$$

is an estimator of

$$\mathbb{E}_{\tilde{\pi}_X}[d(\tilde{X}_T, \tilde{Y}_T)],$$

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

3.4. Coupling inequality and contraction rate. Similar to the coupling inequality of the total variation norm, the distance $d$ 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 $\tau_c$ be the coupling time. The following Lemma follows easily.

**Proposition 3.4.** For a Markov coupling $(Z_t^{(1)}, Z_t^{(2)})$, we have

$$d_w(\text{law}(Z_t^{(1)}), \text{law}(Z_t^{(2)})) \leq \mathbb{P}(Z_T^{(1)} \neq Z_T^{(2)}) = \mathbb{P}(\tau_c > T).$$
Proposition 3.5. Assume that

\[d_w(\pi_X, \hat{\pi}_X) \text{ and } d_w(\pi_Y, \hat{\pi}_Y) \text{ are in order } O(h), \text{ then the error} \]

\[d_w(\pi_X, \pi_Y) \leq \frac{d_w(\hat{\pi}_X \tilde{P}^T_X, \hat{\pi}_X \tilde{P}^T_Y)}{1 - \alpha} + O(h),\]

where \(\alpha < 1\) is the contraction rate of the transition kernel \(\tilde{P}^T_X\) and \(d_w(\hat{\pi}_X \tilde{P}^T_X, \hat{\pi}_X \tilde{P}^T_Y)\) is the finite time error.

Proof. By the triangle inequality,

\[d_w(\pi_X, \pi_Y) \leq d_w(\pi_X, \hat{\pi}_X) + d_w(\hat{\pi}_X, \hat{\pi}_Y) + d_w(\hat{\pi}_Y, \pi_Y).\]

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

\[d_w(\hat{\pi}_X, \hat{\pi}_Y) \leq d_w(\hat{\pi}_X \tilde{P}^T_X, \hat{\pi}_X \tilde{P}^T_Y) + d_w(\hat{\pi}_X \tilde{P}^T_Y, \hat{\pi}_Y \tilde{P}^T_Y).\]
If the transition kernel $\hat{P}_Y^T$ has enough contraction such that 
\[ d_u(\hat{\pi}_X \hat{P}_Y^T, \hat{\pi}_Y \hat{P}_Y^T) \leq \alpha d_u(\hat{\pi}_X, \hat{\pi}_Y) \]
for some $\alpha < 1$, then we have 
\[ d_u(\hat{\pi}_X, \hat{\pi}_Y) \leq \frac{d_u(\hat{\pi}_X \hat{P}_Y^T, \hat{\pi}_Y \hat{P}_Y^T)}{1 - \alpha}. \]
(3.8)
Therefore,
\[ d_u(\pi_X, \pi_Y) \leq \frac{d_u(\hat{\pi}_X \hat{P}_Y^T, \hat{\pi}_Y \hat{P}_Y^T)}{1 - \alpha} + O(h), \]
\[ \square \]

Therefore, in order to estimate $d_u(\pi_X, \pi_Y)$, we need to look for suitable numerical estimators of the finite time error and the speed of contraction of $\hat{P}_Y^T$. The finite time error can be easily estimated by Algorithm 2. And the speed of contraction $\alpha$ comes from the geometric ergodicity of the Markov process $\hat{Y}$ is approximated by that of $\hat{Y}$ because of the convergence result in Proposition 3.1. If our numerical estimation gives 
\[ d_u(\hat{\pi}_X \hat{P}_Y^T, \hat{\pi}_Y \hat{P}_Y^T) \approx d_u(\hat{\pi}_X \hat{P}_Y^T, \hat{\pi}_Y \hat{P}_Y^T) \leqCe^{-\gamma T}, \]
then we set $\alpha = e^{-\gamma T}$. Similar as in [8], we use the following coupling method to estimate the contraction rate $\alpha$. 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 $\{ (x, y) \in \mathbb{R}^{2d} | y = x \}$ be the coupling time. Then by Proposition 3.4 
\[ d_u(\hat{\pi}_X \hat{P}_Y^T, \hat{\pi}_Y \hat{P}_Y^T) \leq \mathbb{P}(\tau_c > T). \]

As discussed in [19], we need a hybrid coupling scheme to make sure that two numerical trajectories couple. Under the condition that two trajectories coupled before extinction time, some coupling methods such as reflection coupling or synchronous coupling are implemented in the first phase to bring two trajectories together. Then we compare the probability density function for the next step and couple these two numerical trajectories with the maximal possible probability (called maximal coupling). After doing this for many times, we have many samples of $\tau_c$ denote by $\tau_c$. We use the exponential tail of $\mathbb{P}(\tau_c > t)$ to estimate the contraction rate $\alpha$. We look for a constant $\gamma > 0$ such that 
\[ -\gamma = \lim_{t \to \infty} \frac{1}{t} \log(\mathbb{P}(\tau_c > t)) \]
if the limit exists. See Algorithm 3 for the details of implementation of coupling. Note that we cannot simply compute the contraction rate start from $t = 0$ because only the tail of coupling time can be considered as exponential distributed. In addition $\hat{Y}$ is a good approximation of $\hat{Y}$ only if $t$ is large. Our approach is to check the exponential tail in a log-linear plot. After having $\tau_c$, it is easy to choose a sequence of times $t_0, t_1, \ldots, t_n$ and calculate $n_i = |\{ \tau_c^m > t_i | 0 \leq m \leq M \}|$ for each $i = 0, \ldots, n$. Then $p_i = n_i/M$ is an estimator of $\mathbb{P}(\tau_c > t_i)$. Now let $p^u_i$ (resp. $p^l_i$) be the upper (resp. lower) bound of the confidence interval of $p_i$ such that 
\[ p^u_i = \tilde{p} + z \sqrt{\frac{\tilde{p} (1 - \tilde{p})}{n_i}} \] (resp. $p^l_i = \tilde{p} - z \sqrt{\frac{\tilde{p} (1 - \tilde{p})}{n_i}}$). If $p^l_i \leq e^{-\gamma t_i} \leq p^u_i$ for each $0 \leq i \leq n$, 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 $p^u_{i_0} - p^l_{i_0}$ is sufficient small. Otherwise we need to run Algorithm 3 for longer time to eliminate the initial bias in $\tau_c$. 

SENSITIVITY ANALYSIS OF QUASI-STATIONARY-DISTRIBUTIONS (QSDS) 11
Algorithm 3 Estimation of contraction rate $\alpha$

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

**Output:** An estimation of contraction rate $\alpha$

Choose threshold $d > 0$

for $m = 1$ to $M$ do

\[ \tau^m_c = 0, t = 0, (\hat{Y}^{(1)}_0, \hat{Y}^{(2)}_0) = (x, y) \]

Flag = 0

while Flag=0 do

if $\hat{Y}^{(1)}_t$ and $\hat{Y}^{(2)}_t \in \mathcal{X}/\partial \mathcal{X}$ then

if $|\hat{Y}^{(1)}_t - \hat{Y}^{(2)}_t| > d$ then

Compute $(\hat{Y}^{(1)}_{t+1}, \hat{Y}^{(2)}_{t+1})$ using reflection coupling or independent coupling

$t \leftarrow t + 1$

else

Compute $(\hat{Y}^{(1)}_{t+1}, \hat{Y}^{(2)}_{t+1})$ using maximal coupling

if coupled successfully then

Flag=1

$\tau^m_c = t$

else

$t \leftarrow t + 1$

end if

end if

end if

end while

end for

Use $\tau^1_c, \cdots, \tau^M_c$ to compute $\mathbb{P}(\tau_c > t | t < \min(\tau_{Y^{(1)}}, \tau_{Y^{(2)}}))$

Fit the tail of $\log \mathbb{P}(\tau_c > t | t < \min(\tau_{Y^{(1)}}, \tau_{Y^{(2)}}))$ versus $t$ by linear regression. Compute the slope $\gamma$.

4. Numerical Examples

4.1. SIR model. Consider an epidemic model in which the whole population is divided into three distinct classes $S$ (susceptible), $I$ (infected) and $R$ (recovered), respectively. After non-dimensionalization, the ODE version of an SIR model reads

\[
\begin{align*}
\frac{dS}{dt} &= (\alpha - \beta SI - \mu S) \\
\frac{dI}{dt} &= (\beta SI - (\mu + \rho + \gamma)I) \\
\frac{dR}{dt} &= (\gamma I - \mu R)
\end{align*}
\]

where $\alpha$ is the birth rate, $\mu$ is the disease-free death rate, $\rho$ is the excess death rate for the infected class, $\gamma$ is the recover rate for the infected population, and $\beta$ is the effective contact rate between the susceptible class and infected class [7]. Note $R$ 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 are involved in this network. The stochastic mass action network
can be defined by a Poisson process $X_n = (S_n, I_n)$.

\begin{align}
\emptyset & \overset{\alpha}{\Rightarrow} S, \ S + I \overset{\beta}{\Rightarrow} 2I \\
S & \overset{\mu}{\Rightarrow} \emptyset, \ I \overset{\mu + \rho + \gamma}{\Rightarrow} \emptyset
\end{align}

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

\begin{align}
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{align}

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

\begin{align}
\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} f_1(P_1, \cdots, P_4, q_{1,n}, \cdots, q_{4,n}) \\ f_2(P_1, \cdots, P_4, q_{1,n}, \cdots, q_{4,n}) \end{pmatrix},
\end{align}

where $f_1$ and $f_2$ comes from discrete approximation in equation (2.10). To improve the readability of the present paper, we move detailed expressions of $f_1$ and $f_2$ to the appendix.

As described in Section 2.1, each Poisson processes $P_i, i = 1, 2, 3, 4$ is path-wisely 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 $\hat{Y}_n$ reads

\begin{align}
\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} g_1(q_{1,n}, \cdots, q_{4,n}) \\ g_2(q_{1,n}, \cdots, q_{4,n}) \end{pmatrix} + \frac{1}{V} \begin{pmatrix} \sigma_1(B_1, \cdots, B_4, q_{1,n}, \cdots, q_{4,n}) \\ \sigma_2(B_1, \cdots, B_4, q_{1,n}, \cdots, q_{4,n}) \end{pmatrix},
\end{align}

where functions $g_1, g_2, \sigma_1,$ and $\sigma_2$ follows the expression in equation (2.11). We refer the appendix for the detailed form of these functions.

By the stationary increments property of 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, equation (4.4) can be simplified to:

\begin{align}
\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} g_1(q_{1,n}, \cdots, q_{4,n}) \\ 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{align}

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 $\alpha$ 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 re-written as

\begin{align}
\hat{Y}_{n+1} &= \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} g_1(q_{1,n}, \cdots, q_{4,n}) \\ g_2(q_{1,n}, \cdots, q_{4,n}) \end{pmatrix} \\
&+ \frac{1}{\sqrt{\text{tr}(N) + 2\sqrt{\det(N)}}} (N + \det(N)I) d \begin{pmatrix} W_1 \\ W_2 \end{pmatrix},
\end{align}
where $\text{tr}(N)$ is the trace of $N$ and $\text{det}(N)$ is the determinant of $N$, and $Id$ is the identity matrix. It is easy to see that the diffusion process $\hat{Y}$ in equations (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 $\partial X$ be the union of 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 behaviour 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. Left bottom of Figure 1 shows the QSD of diffusion process $\hat{Y}$. The QSD of the Poisson process is shown on right top of Figure 1. The difference of 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 Poisson process and the diffusion process moves far away from the absorbing set $\partial X$.

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 in Figure 1 Top Left. 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 equation (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 re-scaled 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 two QSDs are shown at the bottom of Figure 2. It is not hard to see the total variation distance becomes significantly larger when the volume gets smaller. Same as 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 3. The last column $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ is computed using (3.8). Being consistent with Theorem 2.2, the 1-Wasserstein distance between two QSDs is smaller as $V$ getting larger.

| Volume $V$ | Finite Time Error | Contraction Rate $\gamma$ | $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ |
|------------|-------------------|---------------------------|----------------------------------|
| 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                           |

Table 1. SIR model. Numerical results for different volumes

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

\begin{align*}
\frac{dS_1}{dt} &= S_1 S_2 - C_2 S_1 S_2 + C_3 S_1 - 2C_4 S_1^2 \\
\frac{dS_2}{dt} &= -C_1 S_2 - C_2 S_1 S_2 + C_5 h S_3 \\
\frac{dS_3}{dt} &= 2C_6 S_1 - C_5 S_3.
\end{align*}
We refer Figure 3 Top Left for a sample trajectories of the Oregonator on \( \mathbb{R}^2_+ \). 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 following.

\[
S_2 \xrightarrow{C_1} S_1, \quad S_1 + S_2 \xrightarrow{C_2} \emptyset, \quad S_1 \xrightarrow{C_3} 2S_1 + 2S_3 \\
2S_1 \xrightarrow{C_4} \emptyset, \quad S_3 \xrightarrow{C_5} S_2, \quad S_3 \xrightarrow{C_5(1-\delta)} \emptyset
\]

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

\[
q_{1,n} = \sum_{m=0}^{n-1} V h C_1 S_{2,m}, \quad q_{2,n} = \sum_{m=0}^{n-1} V h C_2 S_{1,m} S_{2,m}, \quad q_{3,n} = \sum_{m=0}^{n-1} V h C_3 S_{1,m} \\
q_{4,n} = \sum_{m=0}^{n-1} V h C_4 S_{1,m}^2, \quad q_{5,n} = \sum_{m=0}^{n-1} V h C_5 \delta S_{3,m}, \quad q_{6,n} = \sum_{m=0}^{n-1} V h C_5(1-\delta) S_{3,m}
\]

We remark terms \( S_{1,m} \) 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

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

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

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

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

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

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

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

where \( W_i, i = 1, \ldots, 6 \) are independent standard normal random variables, and \( M \) is a matrix that depends only on \( S_0 \) and \( I_0 \). We refer readers to the appendix for the full expression of \( M \).

Let \( \partial \Omega \) be union of 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 \( i = \{1, \ldots, 2^{29}\} \) and \( \Delta = 0.001 \) are constructed when we consider the trajectory-by-trajectory behaviour of two processes. The time step size is \( h = 10^{-8} \) and the fixed time is set as \( T = 2 \times 10^{-4} \) when \( V = 1000 \), \( T = 4 \times 10^{-5} \) when \( V = 400 \), \( T = 1 \times 10^{-5} \) when \( V = 100 \), and \( T = 2 \times 10^{-6} \) when \( V = 10 \). Note that large rate coefficients \( C_i \) make the numerical results easily to 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 shows the solution of the ordinary differential equation. For any initial point, the trajectory eventually converges to the limit cycle. In terms of thermodynamics, the oscillation is induced through dissipation of energy and is often called a self-sustained oscillator [23]. The trajectories of Poisson process and the diffusion process up to fixed time \( T \) are shown on the Top Right and Bottom Left. It looks that the trajectories are close and this is reasonable because with high probability, the trajectories of both Poisson process and the diffusion process moves far away from the absorbing set \( \partial X \). 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 the different situations for volume \( V = 1000 \) and \( V = 10 \), we plot the trajectories for both processes for each species. Trajectories for \( V = 1000 \) is shown in the upper row of Figure 4 and lower row shows the case for \( V = 10 \). It is not hard to see the Poisson process is quite close to the diffusion process when \( V = 1000 \). But when the volume is too small, not much Poisson jumps can be observe 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 \( V = 1000 \). Same as above, we compute the contraction rate \( \gamma \) of the coupling time distribution to be \( 2.0927 \times 10^5 \). This is due to the large magnitude of noise in the diffusion approximation. As a result, 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, 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 upper bound of \( d_w(\hat{\pi}_X, \hat{\pi}_Y) \) is quite larger when \( V = 10 \). This is consistent
Figure 4. \((V = 1000 \text{ vs. } V = 10)\) 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 \(V = 10\).

| Volume \(V\) | Cut-off Time \(T\) | Finite Time Error | Contraction Rate \(\gamma\) | \(d_w(\tilde{\pi}_X, \tilde{\pi}_Y)\) |
|-------------|------------------|------------------|--------------------------|-----------------------------|
| 1000        | \(2 \times 10^{-4}\) | 0.0057           | 3.3616 \times 10^3      | 0.0116          |
| 400         | \(4 \times 10^{-5}\) | 0.0088           | 2.0599 \times 10^4      | 0.0157          |
| 100         | \(1 \times 10^{-5}\) | 0.0099           | 6.0150 \times 10^4      | 0.0195          |
| 10          | \(2 \times 10^{-6}\) | 0.0563           | 2.0927 \times 10^5      | 0.1646          |

Table 2. Oregonator model: Numerical results for different volumes

with Theorem 2.2, the supreme distance between two processes will be smaller as \(V\) is getting larger.

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 \cite{21} and independently by Lotka to describe a chemical reaction \cite{27}, the general Lotka-Volterra model is widely used in ecology, biology, chemistry, physics, etc. \cite{25}. In this example we consider here a chaotic system in which 4 species with whole population \(V\) compete for a finite set of resources. The ODE version of the system reads

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

\[
\begin{bmatrix}
1 & 1.09 & 1.52 & 0 \\
0.72 & 1 & 0.44 & 1.36 \\
1.53 & 0 & 1 & 0.47 \\
1.27 & 2 & 3 & 1
\end{bmatrix}
\]

For \( i = 1, \ldots, 4 \), all reactions in this system are shown as follows.

\[
S_i \xrightarrow{r_i} 2S_i, \quad S_1 + S_i \xrightarrow{a_{1i}r_i} S_1, \quad S_2 + S_i \xrightarrow{a_{2i}r_i} S_2, \quad S_3 + S_i \xrightarrow{a_{3i}r_i} S_3, \quad S_4 + S_i \xrightarrow{a_{4i}r_i} S_4
\]

The corresponding rate functions are

\[
\begin{align*}
q_{i,1}^n &= \sum_{m=0}^{n-1} V h r_i S_i m \\
q_{i,2}^n &= \sum_{m=0}^{n-1} V h r_i a_{i1} S_{i1} m S_i m \\
q_{i,3}^n &= \sum_{m=0}^{n-1} V h r_i a_{i2} S_{i2} m S_i m \\
q_{i,4}^n &= \sum_{m=0}^{n-1} V h r_i a_{i3} S_{i3} m S_i m \\
q_{i,5}^n &= \sum_{m=0}^{n-1} V h r_i a_{i4} S_{i4} m S_i m
\end{align*}
\]

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

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

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

The diffusion approximation \( \dot{Y}_n \) can be written as

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

where \( i = 1, \ldots, 4 \), \( B_j \), \( j = \{\ldots, 17\} \) are independent standard Wiener process, functions \( g_1, g_2, g_3, g_4, \sigma_1, \sigma_2, \sigma_3 \) and \( \sigma_4 \) follows the expression in equation (2.11). We refer 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:

\[
\dot{Y}_{n+1} = \begin{pmatrix} S_{1,n+1} \\ S_{2,n+1} \\ S_{3,n+1} \\ S_{4,n+1} \end{pmatrix} = \begin{pmatrix} S_{1,n} \\ S_{2,n} \\ S_{3,n} \\ S_{4,n} \end{pmatrix} + \frac{1}{V} \begin{pmatrix} g_1(q_1, \cdots, q_4) \\ g_2(q_1, \cdots, q_4) \end{pmatrix} + \frac{1}{V} M \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_{17} \end{pmatrix}
\]

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 $\partial X$ be 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 behaviour 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 ordinary differential equation projected onto $x_1 x_2 x_3$ space. The trajectories of Poisson process and the diffusion process are shown on the Top Right and Bottom Left. It looks that the trajectories are close and this is reasonable because with high probability, the trajectories of both Poisson process and the diffusion process moves far away from the absorbing set $\partial X$. We compute the distribution of the coupling time. The coupling time distribution and its exponential tail are shown in Figure 5 Top Left, that gives the contraction rate $\gamma = 0.0849$. Then we apply Algorithm 1 to compute the finite time error. The finite time error 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 situations 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 not hard to see the trajectory-by-trajectory behavior between Poisson process and diffusion process is quite remarkable when $V = 1000$. However, more regeneration happens in Poisson process when $V = 10$. So it’s not surprised us that the finite time error for $V = 10$ is 0.1286, that around 40 times larger than the case for $V = 1000$. Trajectories of the Poisson process have high probability moving along the boundary in this case. Same as above, we compute the contraction rate $\gamma$ of the coupling time distribution to be 1.7905. As a result, 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 upper bound of $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ is quite larger when $V = 10$. This is consistent with Theorem 2.2, the supreme distance between two processes will be smaller as $V$ is getting larger.

| volume $V$ | finite time error | contraction rate $\gamma$ | $d_w(\hat{\pi}_X, \hat{\pi}_Y)$ |
|-----------|-----------------|-----------------|------------------|
| 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           |

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

5. Conclusion

In this paper we develop a coupling-based approach to quantitatively estimate the distance between the QSD of a stochastic mass-action process and that of its diffusion approximation. The dependence of QSDs in terms of the volume of the mass-action system is studied. To address the
Figure 5. (Case $V = 1000$) Upper panel: (Left) ODE trajectories. (Right) Poisson process. Lower panel: (Left) Diffusion process. (Right) $P(\tau_c > t | \tau < t)$ vs. $t$.

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

The challenge of QSDs, we use the idea of regeneration from QSDs after exiting to construct a process with stationary distribution. This is the main change from our previous work \cite{8,19}. Both the coupling algorithm and the path-wise matching of a stochastic mass-action system and its diffusion approximation need to be adapted to the regeneration from QSDs. We compare the finite time error and the rate of contraction for different population size $V$. All numerical results shows that the
distance between two QSDs is smaller for larger population. In general, the effect of demographic noise must be seriously addressed when the population is small.

The study of path-wise 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? A 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 random events occur at continuous time. However, there may be some way of building a "discrete reflection" and coupling two tau-leaping trajectories, i.e., two trajectories of equation (2.10) effectively. This will be addressed in our future work.

APPENDIX A. EXPRESSIONS OF MASS-ACTION SYSTEMS AND THEIR DIFFUSION APPROXIMATIONS

To improve the readability, we put the explicit formulas of the Poisson approximation and the diffusion approximation for each model in this section.

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

$$
\hat{X}_{n+1} = \left( \frac{S_{n+1}}{I_{n+1}} \right) \\
= \left( \frac{S_n}{I_n} \right) + \frac{1}{\sqrt{V}} \left( \begin{array}{c} [P_{1}(q_{1,n+1}) - P_{1}(q_{1,n})] - [P_{2}(q_{2,n+1}) - P_{2}(q_{2,n})] - [P_{4}(q_{4,n+1}) - P_{4}(q_{4,n})] \end{array} \right) \\
:= \left( \frac{S_n}{I_n} \right) + \frac{1}{\sqrt{V}} \left( \begin{array}{c} [P_{1}(q_{1,n+1}) - P_{1}(q_{1,n})] - [P_{2}(q_{2,n+1}) - P_{2}(q_{2,n})] - [P_{4}(q_{4,n+1}) - P_{4}(q_{4,n})] \end{array} \right)
$$

where $P_i$, $i = \{1, 2, 3, 4\}$ are independent unit rate Poisson processes.
The rule of update of the numerical approximation $\tilde{Y}_n$ follows

$$
\tilde{Y}_{n+1} = \begin{pmatrix} \tilde{S}_{n+1} \\ \tilde{I}_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} [q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] - [q_{3,n+1} - q_{3,n}] \\ [q_{2,n+1} - q_{2,n}] - [q_{4,n+1} - q_{4,n}] \\ [B_1(q_{1,n+1}) - B_1(q_{1,n})] - [B_2(q_{2,n+1}) - B_2(q_{2,n})] - [B_3(q_{3,n+1}) - B_3(q_{3,n})] \\ [B_2(q_{2,n+1}) - B_2(q_{2,n})] - [B_4(q_{4,n+1}) - B_4(q_{4,n})] \end{pmatrix},
$$

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

As two classes $S_i$ and $I_n$ and four reactions are considered in this SIR model, the corresponding diffusion matrix $M$ should be a $2 \times 4$ matrix. Specifically, the diffusion matrix $M$ reads as

$$
\tilde{Y}_{n+1} = \begin{pmatrix} \tilde{S}_{n+1} \\ \tilde{I}_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} [q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] - [q_{3,n+1} - q_{3,n}] \\ [q_{2,n+1} - q_{2,n}] - [q_{4,n+1} - q_{4,n}] \end{pmatrix},
$$

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 process $P_i$ and $B_i$ in the approximations. The rule of update of the numerical approximation $\tilde{X}_n$ follows

$$
\tilde{X}_{n+1} = \begin{pmatrix} \tilde{S}_{n+1} \\ \tilde{Q}_{n+1} \\ \tilde{I}_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ Q_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} [q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] - [q_{3,n+1} - q_{3,n}] - 2[q_{4,n+1} - q_{4,n}] \\ [q_{2,n+1} - q_{2,n}] - [q_{4,n+1} - q_{4,n}] \end{pmatrix},
$$

where $P_i, i = \{1, \ldots, 6\}$ are independent unite rate Poisson processes. The diffusion approximation $\tilde{Y}$ can be written as

$$
\tilde{Y}_{n+1} = \begin{pmatrix} \tilde{S}_{n+1} \\ \tilde{Q}_{n+1} \\ \tilde{I}_{n+1} \end{pmatrix} = \begin{pmatrix} S_n \\ Q_n \\ I_n \end{pmatrix} + \frac{1}{V} \begin{pmatrix} [q_{1,n+1} - q_{1,n}] - [q_{2,n+1} - q_{2,n}] - [q_{3,n+1} - q_{3,n}] - 2[q_{4,n+1} - q_{4,n}] \\ [q_{2,n+1} - q_{2,n}] - [q_{4,n+1} - q_{4,n}] \end{pmatrix},
$$

where $B_i, i = \{1, \ldots, 6\}$ 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 $3 \times 3$ matrix. Specifically, the diffusion matrix $M$ is defined as follows.

\[
\hat{Y}_{n+1} = \begin{pmatrix}
\frac{S_{1,n+1}}{S_{2,n+1}} \\
\frac{S_{2,n+1}}{S_{3,n+1}} \\
\frac{S_{3,n+1}}{S_{1,n+1}} 
\end{pmatrix} = \begin{pmatrix}
\frac{S_{1,n}}{S_{2,n}} + \frac{1}{V} \left( \left[ q_{1,n+1} - q_{1,n} \right] - \left[ q_{2,n+1} - q_{2,n} \right] + \left[ q_{3,n+1} - q_{3,n} \right] - 2\left[ q_{4,n+1} - q_{4,n} \right] \right) \\
\frac{S_{2,n}}{S_{3,n}} + \frac{1}{V} \left( \sqrt{q_{1,n+1} - q_{1,n}} - \sqrt{q_{2,n+1} - q_{2,n}} \right) \left[ \sqrt{q_{3,n+1} - q_{3,n}} \right] - \sqrt{q_{5,n+1} - q_{5,n}} - \sqrt{q_{6,n+1} - q_{6,n}} \right) \\
\frac{S_{3,n}}{S_{1,n}} + \frac{1}{V} \left( \left[ q_{1,n+1} - q_{1,n} \right] - \left[ q_{2,n+1} - q_{2,n} \right] + \left[ q_{3,n+1} - q_{3,n} \right] - 2\left[ q_{4,n+1} - q_{4,n} \right] \right) \\
\end{pmatrix} 
\]

where $W_i, \{i = 1, \ldots, 6\}$ 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 update of the numerical approximation $\hat{X}_n$ follows

\[
\hat{X}_{n+1} = \begin{pmatrix}
\frac{S_{1,n+1}}{S_{2,n+1}} \\
\frac{S_{2,n+1}}{S_{3,n+1}} \\
\frac{S_{3,n+1}}{S_{1,n+1}} 
\end{pmatrix} = \begin{pmatrix}
\frac{S_{1,n}}{S_{2,n}} + \frac{1}{V} \left( f_1(P_1, \ldots, P_7, q_{1,n+1}^{q_1}, \ldots, q_{4,n+1}^{q_4}) + \frac{1}{V} \left( f_2(P_1, \ldots, P_7, q_{1,n+1}^{q_1}, \ldots, q_{4,n+1}^{q_4}) + \frac{1}{V} \left( f_3(P_1, \ldots, P_7, q_{1,n+1}^{q_1}, \ldots, q_{4,n+1}^{q_4}) + \frac{1}{V} \left( f_4(P_1, \ldots, P_7, q_{1,n+1}^{q_1}, \ldots, q_{4,n+1}^{q_4}) \right) \right) \right) \right) \\
\end{pmatrix} 
\]

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

\[
\hat{Y}_{n+1} = \begin{pmatrix}
\frac{S_{1,n+1}}{S_{2,n+1}} \\
\frac{S_{2,n+1}}{S_{3,n+1}} \\
\frac{S_{3,n+1}}{S_{1,n+1}} 
\end{pmatrix} = \begin{pmatrix}
\frac{S_{1,n}}{S_{2,n}} + \frac{1}{V} \left( f_1(P_1, \ldots, P_7, q_{1,n+1}^{q_1}, \ldots, q_{4,n+1}^{q_4}) + \frac{1}{V} \left( f_2(P_1, \ldots, P_7, q_{1,n+1}^{q_1}, \ldots, q_{4,n+1}^{q_4}) + \frac{1}{V} \left( f_3(P_1, \ldots, P_7, q_{1,n+1}^{q_1}, \ldots, q_{4,n+1}^{q_4}) + \frac{1}{V} \left( f_4(P_1, \ldots, P_7, q_{1,n+1}^{q_1}, \ldots, q_{4,n+1}^{q_4}) \right) \right) \right) \right) \\
\end{pmatrix} 
\]

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

[1] Alan Agresti and Brent A Coull. Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126, 1998.

[2] David F Anderson and Thomas G Kurtz. Continuous time markov chain models for chemical reaction networks. In *Design and analysis of biomolecular circuits*, pages 3–42. Springer, 2011.

[3] Michel Benaim. Stochastic persistence. arXiv preprint arXiv:1806.08450, 2018.

[4] Morten Brøns and Keilma Bar-Eli. Canard explosion and excitation in a model of the belousov-zhabotinskii reaction. *The Journal of Physical Chemistry*, 95(22):8706–8713, 1991.

[5] John N Darroch and Eugene Seneta. On quasi-stationary distributions in absorbing discrete-time finite markov chains. *Journal of Applied Probability*, 2(1):88–100, 1965.

[6] Emeric Deutsch and Michael Neumann. On the first and second order derivatives of the perron vector. *Linear algebra and its applications*, 71:57–76, 1985.

[7] Nguyen Thanh Dieu, Dang Hai Nguyen, Nguyen Huu Du, and G Yin. Classification of asymptotic behavior in a stochastic sir model. *SIAM Journal on Applied Dynamical Systems*, 15(2):1062–1084, 2016.

[8] Matthew Dobson, Yao Li, and Jinyu Zhai. Using coupling methods to estimate sample quality of stochastic differential equations. *SIAM/ASA Journal on Uncertainty Quantification*, 9(1):135–162, 2021.

[9] Irving R Epstein and John A Pojman. *An introduction to nonlinear chemical dynamics: oscillations, waves, patterns, and chaos.* Oxford University Press, 1998.

[10] Michael A Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, 104(9):1876–1889, 2000.

[11] Daniel T Gillespie et al. Stochastic simulation of chemical kinetics. *Annual review of physical chemistry*, 58(1):35–55, 2007.

[12] P Gray and SK Scott. Chemical oscillations and instabilities; nonlinear chemical kinetics. *Science Progress in the Twentieth Century (1919-1933)*, 21(82):341–343, 1926.

[13] Carl D Meyer. Sensitivity of the stationary distribution of a markov chain. *SIAM Journal on Matrix Analysis and Applications*, 15(3):715–728, 1994.

[14] Jun Miyazaki. *Pattern Formations and Oscillatory Phenomena: 2. Belousov–Zhabotinsky Reaction*. Elsevier Inc. Chapters, 2013.

[15] James D Murray. *Mathematical biology II: spatial models and biomedical applications*, volume 3. Springer New York, 2001.

[16] Vito Volterra. *Variazioni e fluttuazioni del numero d’individui in specie animali conviventi.* Società anonima tipografica” Leonardo da Vinci”, 1926.
Yao Li: Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA, 01002, USA
Email address: yaoli@math.umass.edu

Yaping Yuan: Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA, 01002, USA
Email address: yuan@math.umass.edu