Commun. Comput. Phys.
doi: 10.4208/cicp.290313.051213s

Vol. **15**, No. 4, pp. 1207-1236
April 2014

Strong Convergence and Speed up of Nested Stochastic Simulation Algorithm

Can Huang and Di Liu*

Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA.

Received 29 March 2013; Accepted (in revised version) 5 December 2013

Available online 21 January 2014

Abstract. In this paper, we revisit the Nested Stochastic Simulation Algorithm (NSSA) for stochastic chemical reacting networks by first proving its strong convergence. We then study a speed up of the algorithm by using the explicit Tau-Leaping method as the Inner solver to approximate invariant measures of fast processes, for which strong error estimates can also be obtained. Numerical experiments are presented to demonstrate the validity of our analysis.

AMS subject classifications: 65C30, 60H35

Key words: Stochastic simulation algorithm, biochemical reacting network, strong convergence.

1 Introduction

The stochastic simulation algorithm (SSA) by Gillespie [12, 13], which can be formulated as an equation of jump diffusion processes [23], provides the world a benchmark for the numerical simulation of intra-cellular biochemical networks, such as gene expression and regulation. The algorithm is exact in the sense that the probability distribution of particular realizations being simulated is the same as the solution of the chemical master equation (CME) describing the process. The CME is a set of ordinary differential equations with enormously high dimension if the number of the states of the corresponding network is large. Although SSA gives us a Monte Carlo approach to compute the CME, it becomes less efficient when applied to the so called stiff systems, in which reaction channels and reacting species with two or more different time and concentration scales coexist. There have been many approaches overcoming this difficulty, such as the celebrated Tau-Leaping method [6, 14] and Nested Stochastic Simulation Algorithm (NSSA) [7, 8].

NSSA relies only on the disparity of the rates, and makes no a priori assumption on the form of the slow and fast variables, nor upon the analytic form of the rate functions.

*Corresponding author. *Email addresses:* canhuang@math.msu.edu (C. Huang), richardl@math.msu.edu (D. Liu)

The main idea is to capture the effective dynamics on the slow time scale by assuming the fast processes to reach a quasi-equilibrium in a sufficiently short time. Weak convergence of NSSA can be found in [8]. Strong convergence of NSSA is still open and this paper tries to fill in the gap.

In this paper, we will further investigate the speed up of NSSA. One key component of implementation is the approximation of the invariant measure of the fast process. Several authors studied the ergodic approximation of a stationary distribution in the case of Brownian diffusions, see [26,29,30] among others. Computation of the invariant measure of stochastic differential equation driven by a Lévy process is investigated in [32], where the intensity of the Lévy process, different from SSA, is independent of the state. In this paper, we will adopt the Tau-Leaping method to sample the quasi-equilibrium of the fast processes in NSSA, which will enable us to use larger time steps when concentrations of reacting species in the fast reactions are high. Note that strong convergence analysis of multiscale schemes for standard SDE is performed in [9,24]. However, the extension in this work is not trivial due to the necessity of establishment of invariant measure for fast processes described by the τ-leaping method and the analysis of the associated generator in the form of a difference operator.

In the following sections, we will first provide the notations for SSA, NSSA and Tau-Leaping method. Then we will prove the strong convergence of NSSA, when either direct SSA or Tau-Leaping method is used as the Inner solver for the fast processes. Finally, numerical examples will be provided to test the error estimates.

2 Stochastic simulation algorithms

As a model taking into account of stochastic effects at the molecular level, SSA considers an isothermal, spatially homogeneous mixture of chemically reacting network in a fixed volume V. Suppose there are N_S species of molecules, with M_R reactions. Let $x_i \in N$ denote the number of molecules of species S_i. Then, each reaction R_j can be characterized by a propensity function $a_j(x)$ where $x=(x_1,x_2,\cdots,x_{N_S})$ and a state change vector $\nu_j \in N^{N_S}$. We write $R_j=(a_j,\nu_j)$. Given state x, the probability that reaction R_j fires on an infinitesimal time interval dt is independent from other reactions and is given by $a_j(x)dt$. The state of the network after the jth reaction is $x+\nu_j$. $a_j(x)$ usually takes the form of polynomials or rational functions in terms of x. From [23], SSA can be formulated as a stochastic differential equation in the following form:

$$dX_t=\sum_j \int_0^\infty \nu_j \mathbf{A}_j(q,X_t)\mathcal{P}(dt,dq), \tag{2.1}$$

where

$$\mathbf{A}_j(q,X_t)=\begin{cases} 1, & q\in\Big(\sum\limits_{i=1}^{j-1}a_i(X_t),\sum\limits_{i=1}^{j}a_i(X_t)\Big), \\ 0, & \text{otherwise}, \end{cases} \tag{2.2}$$

and $\mathcal{P}(dt,dq)$ is the Poisson random measure with Lebesgue intensity.

Numerically, SSA constructs realizations of the time evolution of the state vector x_t. Assume that the current time is $t=t_n$, and the state of the system is at $x=X_n$. One version of SSA called the Direct Method performs the following steps:

1. Let $a_0(x)=\sum_j a_j(x)$. Generate independent random numbers r_1 and r_2 with uniform distribution on the unit interval $[0,1]$. Let
$$\delta t=\frac{1}{a_0(x)}\ln\left(\frac{1}{r_1}\right),$$
and k be the positive integer such that
$$\sum_{j=1}^{k-1}a_j(x)<r_2a_0(x)\leq\sum_{j=1}^{k}a_j(x). \tag{2.3}$$
2. Update time and state of the system by
$$t_{n+1}=t_n+\delta t,\qquad X_{n+1}=X_n+\nu_k.$$
Goto Step 1, unless a certain stopping criterion is met.

Nested SSA (NSSA) deals with the situation when the system has two or more disparate time scales due to separation in the magnitudes of reaction rates. Assume the rate functions of a chemical reaction network $R=(a,\nu)$ have the following form
$$R^s(x)=(a^s(x),\nu^s),\quad R^f(x)=\left(\frac{1}{\epsilon}a^f(x),\nu^f\right), \tag{2.4}$$
where $\epsilon\ll 1$ measures the time scale separation. To overcome the numerical stiffness, NSSA uses two SSA solvers with one nested in the other:

1. **Inner SSA.** Run M-independent replicas of SSA with the fast reactions $R^f=(\frac{1}{\epsilon}a^f,\ \nu^f)$ only, for a time interval of $[T_0,T_0+T_f]$. During this calculation, compute the modified slow rates with
$$\tilde{a}_j^s=\frac{1}{M}\sum_{i=1}^{M}\frac{1}{T_f}\int_{T_0}^{T_0+T_f}a_j^s(X_\tau^i)d\tau, \tag{2.5}$$
where X_τ^i is the ith replica of the fast process with initial value $X_{\tau=0}^i=X_n$.
2. **Outer SSA.** Run one step of SSA for the modified slow reactions $(R^s(\tilde{a}^s,\nu^s))$ to generate (t_{n+1},X_{n+1}) from (t_n,X_n) unless a certain criterion is satisfied.

Generalizations to systems with more than two time scales can be found in [8]. The justification of the Nested SSA is that, viewed as a singular perturbation problem, the slow-fast chemical reacting network can be proved [8] to have the effective dynamics on the slow time scale up to order $\mathcal{O}(\epsilon)$ in the following form:
$$\bar{R}=(\bar{a}^s(z),\nu^s), \tag{2.6}$$

where z is the effective slow variables defined to be linearly independent functions conserved in the fast reactions and

$$\bar{a}_j^s(z) = \langle a_j^s(x) \rangle_z \equiv \sum_{x \in X} a_j^s(x) \mu_z(x), \tag{2.7}$$

with $\mu_z(x)$ being the quasi-equilibrium distribution of the fast reactions. But the perturbation method essentially gives only the weak convergence of the effective dynamics. In Section 3, we will investigate the strong convergence of NSSA.

Another important issue needs to be addressed on Nested SSA is the efficient evaluation of slow reaction rates $\bar{a}_j^s$, which takes the averaged form of (2.7). In the original NSSA, it is approximated with direct simulation of the fast processes and a time-ensemble average of the result. However, the algorithm will be very time consuming if the involved reacting species are in high concentration, which will entail a very small time step for direct SSA. Here we consider using the Tau-Leaping method to speed up the Inner loop of NSSA. The Tau-leaping method, proposed by Gillespie [14] for simulating stiff chemical networks, greatly improves the efficiency of simulation at the expense of sacrificing an affordable accuracy. The explicit Tau-Leaping algorithm can be formulated as the following [6]. Suppose $x = X_n$ is the current state at time t. Then for a time step of $\tau > 0$, the state at $t+\tau$ is given by

$$X_{n+1} = x + \sum_{j=1}^{M_R} \nu_j P_j(a_j(x), \tau),$$

where $P_j(a_j(x),\tau), j = 1, \cdots M_R$ are independent Poisson random variables with mean and variance $a_j(x)\tau$. We will study of effect of using Tau-Leaping method as the Inner solver in NSSA. In Section 4, we will give the strong convergence of the modified scheme.

To proceed, we first reformulate the equation for the multiscale stochastic dynamics to facilitate our proof. Some similar techniques and results used in this section can be found in [3,9,24]. Let M_s be the number of slow reactions and M_f be the number of fast reactions such that $M_R = M_s + M_f$. If a function $b(x)$ does not change over fast reactions, i.e. for any x and any state change vector ν_j^f, we have

$$b(x + \nu_j^f) = b(x),$$

then we define $b(x)$ as a slow observable [8]. Clearly, for a constant vector $b \in R^{N_s}$, linear function $b(x) = b \cdot x$ is a slow variable if and only if

$$b \cdot \nu_j^f = 0, \quad j = 1, \cdots, M^f.$$

Hence, the complementary orthogonal subspace of $\{\nu_j^f\}$ forms a linear subspace of slow observables. Let $b_1, b_2, \cdots, b_J$ be a set of basis vectors of this subspace, then we can define the state exchange vectors associated with the slow variables as

$$\nu_j^b = (b_1 \cdot \nu_j^s, \cdots, b_J \cdot \nu_j^s), \quad j = 1, \cdots, M_s.$$

Now with the transformation

$$\begin{cases} Z_t^\epsilon = (b_1 \cdot X_t, \cdots, b_J \cdot X_t), \\ Y_t^\epsilon = (\nu_1^f \cdot X_t, \cdots, \nu_{M_f}^f \cdot X_t), \end{cases}$$

Eq. (2.1) when time scale separation (2.4) happens can be decomposed into

$$\begin{cases} dZ_t^\epsilon = \sum_{j=1}^{M_s} \int_0^\infty \nu_j^b \mathbf{A}_j^s(q, Z_t^\epsilon, Y_t^\epsilon) \mathcal{P}(dt, dq), \\ dY_t^\epsilon = \frac{1}{\epsilon} \sum_{j=1}^{M_f} \int_0^\infty \nu_j^{ff} \mathbf{A}_j^f(q, Z_t^\epsilon, Y_t^\epsilon) \mathcal{P}(dt, dq) + \sum_{j=1}^{M_s} \int_0^\infty \nu_j^{sf} \mathbf{A}_j^s(q, Z_t^\epsilon, Y_t^\epsilon) \mathcal{P}(dt, dq), \end{cases} \tag{2.8}$$

with

$$\begin{aligned} \nu_j^{ff} &= (\nu_1^f \cdot \nu_j^f, \cdots, \nu_{M_f}^f \cdot \nu_j^f), \quad j = 1, \cdots, M_f, \\ \nu_j^{sf} &= (\nu_1^f \cdot \nu_j^s, \cdots, \nu_{M_f}^f \cdot \nu_j^s), \quad j = 1, \cdots, M_s. \end{aligned}$$

and

$$\mathbf{A}_j^s(q, Z_t^\epsilon, Y_t^\epsilon) = \begin{cases} 1, & q \in \Big(\sum_{i=1}^{j-1} a_i^s(Z_t^\epsilon, Y_t^\epsilon), \sum_{i=1}^{j} a_i^s(Z_t^\epsilon, Y_t^\epsilon) \Big), \\ 0, & \text{otherwise;} \end{cases}$$

$$\mathbf{A}_j^f(q, Z_t^\epsilon, Y_t^\epsilon) = \begin{cases} 1, & q \in \Big(\sum_{i=1}^{j-1} a_i^f(Z_t^\epsilon, Y_t^\epsilon), \sum_{i=1}^{j} a_i^f(Z_t^\epsilon, Y_t^\epsilon) \Big), \\ 0, & \text{otherwise.} \end{cases}$$

Throughout the paper, we impose three assumptions for our process.

Assumption 1: There exist $\alpha, \beta > 0$ such that

$$2\Big(y, \sum_j \nu_j^{ff} a_j^f(z,y)\Big) + \sum_j |\nu_j^{ff}|^2 a_j^f(z,y) \le \alpha - \beta(|z|^2 + |y|^2).$$

Assumption 2: For any $y_1, y_2 \in R^{M_f}, z \in R^J$, there exists γ sufficiently large such that

$$\begin{aligned} &2(y_1 - y_2) \cdot \sum_{j=1}^{M_f} \nu_j^{ff} \left[a_j^f(z, y_1) - a_j^f(z, y_2) \right] \\ &\quad + \sum_{j=1}^{M_f} |\nu_j^{ff}|^2 \left(a_j^f(z, y_1) - a_j^f(z, y_2) \right) \le -\gamma |y_1 - y_2|^2. \end{aligned}$$

Assumption 3. There exists $\theta>0$ such that

$$2\Big(z,\sum_{j=1}^{M_s}\nu_j^s a_j^b(z,y)\Big)+\sum_{j=1}^{M_s}|\nu_j^b|^2 a_j^s(z,y)\leq\theta(|z|^2+|y|^2).$$

Assumption 1 is a mean-reverting condition which means that the fast process moves inward on average if it is sufficiently far from the origin. Assumption 2 is a dissipative condition on the fast dynamics, which enables us to employ some coupling techniques. Assumption 3 is on slow process, which guarantees that the slow process is regular, i.e. it does not explode in finite time w.p.1. by choosing Lyapunov function $V(z)=|z|^2$.

3 Strong convergence of nested SSA

We start this section by proving the existence of invariant measure for fast processes under Assumptions 1-2.

3.1 Existence of invariant measure

Note that the explicit form of the invariant measure μ for the fast process is not known in most circumstances. Hence, we approximate

$$\bar{a}_j^s(z)=\int a_j^s(z,y)\mu_z(dy)$$

by the ensemble averaging $\tilde{a}_j^s$. If the slow process Z_t^ϵ is fixed as a constant vector z, then for any $\phi\in\mathcal{C}(\mathcal{R}^{M_f})$, we can define

$$(P_t\phi)(\xi)=\mathbb{E}_{z,\epsilon}[\phi(Y_t)|Y_0=\xi],$$

where Y_t is the solution of the fast equation with (z,ϵ) fixed as parameters:

$$dY_t=\frac{1}{\epsilon}\sum_{j=1}^{M_f}\int_0^\infty \nu_j^{ff}\mathbf{A}_j^f(q,z,Y_t)\mathcal{P}(dt,dq)+\sum_{j=1}^{M_s}\int_0^\infty \nu_j^{sf}\mathbf{A}_j^s(q,z,Y_t)\mathcal{P}(dt,dq).$$

It is clear that the process $\{Y_t\}$ is Markovian and has càdlàg (right continuous with left limits) trajectories. In this subsection, we will prove that for each z, Y_t has an invariant measure μ_z^ϵ under our assumptions. For the rest of this section, we denote the stochastic process $\{Y_t\}$ with initial condition ξ by $\{Y_t(\xi)\}$.

Lemma 3.1. *Assume Assumptions* 1-2 *for the process* $\{Y_t\}$. *Then,* $\{Y_t\}$ *is tight and the operator* P_t *has Feller property, i.e. if* $\phi\in C_b(\mathcal{R}^{M_f})$, *then* $P_t\phi\in C_b(\mathcal{R}^{M_f})$, *where* $C_b(\mathcal{R}^{M_f})$ *is the space of continuous bounded functions on* R^{M_f}.

Proof. Let $V(Y_t)=|Y_t|^2$ be a Lyapunov function and $\mathcal{L}$ be the infinitesimal generator for the fast process,

$$\begin{aligned}\mathcal{L}V(Y_t)=&\frac{2}{\epsilon}\mathbb{E}\Big(\sum_j \nu_j^{ff}a_j^f(z,Y_t),Y_t\Big)+\mathbb{E}\frac{1}{\epsilon}\sum_j|\nu_j^{ff}|^2\mathbb{E}a_j^f(z,Y_t)\\&+\mathbb{E}\Big(\sum_j \nu_j^{sf}a_j^f(z,Y_t),Y_t\Big)+\sum_j|\nu_j^{sf}|^2\mathbb{E}a_j^f(z,Y_t)\\\leq&\frac{1}{\epsilon}\big[\alpha-\beta\mathbb{E}(|z|^2+|Y_t|^2)+\mathcal{O}(\epsilon)\big],\end{aligned}$$

which implies [9],

$$\sup_t\mathbb{E}|Y_t|^2\leq\alpha,\quad t\in[T_0,T_0+T_f]. \tag{3.1}$$

From the well-known Markov inequality, for each $R>0$, there is a $K_R>0$ satisfying $K_R>\sqrt{(1/R)}$ such that

$$\mathbb{P}(|Y_t|>R)\leq\frac{\sup\limits_{T_0\leq t\leq T_0+T_f}\mathbb{E}|Y_t|^2}{K_R^2}\leq\frac{\alpha}{R}.$$

Therefore, $\{Y_t\}$ is tight.

Let $\tilde{a}_j^f$ be a functional that keeps a_j^f on the ball $B(0,R)$. Denote $Y_{T,R}(\xi)$ be the solution to the second equation of (2.8) with a_j^f replaced by $\tilde{a}_j^f$. Then for $\xi,\xi_n\in B(0,R)$,

$$\begin{aligned}&\mathbf{P}(|Y_{T_0+T_f}(\xi)-Y_{T_0+T_f}(\xi_n)|>\delta)\\=&\mathbf{P}\Big(|Y_{T_0+T_f}(\xi)-Y_{T_0+T_f}(\xi_n)|>\delta,\sup_{T_0\leq s\leq T_0+T_f}|Y_s(\xi)|\wedge|Y_s(\xi_n)|\leq R\Big)\\&+\mathbf{P}\Big(|Y_{T_0+T_f}(\xi)-Y_{T_0+T_f}(\xi_n)|>\delta,\sup_{T_0\leq s\leq T_0+T_f}|Y_s(\xi)|\wedge|Y_s(\xi_n)|\geq R\Big)\\\leq&\frac{\alpha}{R}+\mathbf{P}\Big(|Y_{T_0+T_f,R}(\xi)-Y_{T_0+T_f,R}(\xi_n)|>\delta,\sup_{T_0\leq s\leq T_0+T_f}|Y_{s,R}(\xi)|\wedge|Y_{s,R}(\xi_n)|\leq R\Big)\\\leq&\frac{\alpha}{R}+\frac{1}{\delta^2}\mathbb{E}|Y_{T_0+T_f,R}(\xi)-Y_{T_0+T_f,R}(\xi_n)|^2.\end{aligned}$$

Hence, by applying the Ito formula and Assumption 2,

$$\begin{aligned}&\mathbb{E}|Y_{T_0+T_f,R}(\xi)-Y_{T_0+T_f,R}(\xi_n)|^2\\\leq&\mathbb{E}|\xi-\xi_n|^2+\mathbb{E}\frac{1}{\epsilon}\int_{T_0}^{T_0+T_f}\sum_j|\nu_j^{ff}|^2\big[a_j^f(z,Y_{t,R}(\xi))-a_j^f(z,Y_{t,R}(\xi_n))\big]dt\end{aligned}$$

$$
\begin{aligned}
&+\frac{2}{\epsilon}\mathbb{E}\int_{T_0}^{T_0+T_f}\left(Y_{t,R}(\xi)-Y_{t,R}(\xi_n),\sum_j \nu_j^{ff}\left(a_j^f(z,Y_{t,R}(\xi))-a_j^f(z,Y_{t,R}(\xi_n))\right)\right)dt\\
&+\mathbb{E}\int_{T_0}^{T_0+T_f}\sum_j|\nu_j^{sf}|^2\left[a_j^f(z,Y_{t,R}z(\xi))-a_j^f(z,Y_{t,R}(\xi_n))\right]dt\\
&+\mathbb{E}\int_{T_0}^{T_0+T_f}\left(Y_{t,R}(\xi)-Y_{t,R}(\xi_n),\sum_j \nu_j^{sf}\left(a_j^f(z,Y_{t,R}(\xi))-a_j^f(z,Y_{t,R}(\xi_n))\right)\right)dt\\
\le&\mathbb{E}|\xi-\xi_n|^2-\frac{\beta}{\epsilon}\int_{T_0}^{T_0+T_f}\mathbb{E}|Y_{t,R}(\xi)-Y_{t,R}(\epsilon)|^2dt,\\
\le&\mathbb{E}|\xi-\xi_n|^2,
\end{aligned}
$$

where the $\mathcal{O}(1)$ terms are absorbed into the $\mathcal{O}(\epsilon^{-1})$ terms. Therefore, we obtain

$$
\mathbf{P}(|Y_{T_0+T_f}(\xi)-Y_{T_0+T_f}(\xi_n)|\ge\delta)\le\frac{\alpha}{R}+\frac{\mathbb{E}|\xi-\xi_n|^2}{\delta^2}. \tag{3.2}
$$

We want to show that $(P_{T_0+T_f}\phi)(\xi_n)\to(P_{T_0+T_f}\phi)(\xi))$. It is equivalent to show that from any subsequence of $\{\xi_n\}$, we can choose sub-subsequence $\{\xi_n^{'}\}$, such that $(P_{T_0+T_f}\phi)(\xi_n^{'})\to(P_{T_0+T_f}\phi)(\xi))$. Since ϕ is continuous and bounded, we only need to show $Y_{T_0+T_f}(\xi_n^{'})\to Y_{T_0+T_f}(\xi)$ almost surely by the dominant convergence theorem. Therefore, we need to show $Y_{T_0+T_f}(\xi_n)\to Y_{T_0+T_f}(\xi)$ in probability, which is obvious from (3.2). By diagonal process, we can obtain such sub-sub sequence $Y_{T_0+T_f}(\xi')$.

Finally, By the Krylov-Bogoliubov theorem (see Theorem 3.1.1 and its corollary of [34]), we conclude that there exists an invariant measure μ_z^ϵ of the fast process $\{Y_t\}$ such that

$$
\lim_{t\to\infty}\mathbb{E}\phi(Y_t)=\int\phi(y)d\mu_z^\epsilon(y),
$$

for any function $\phi(y)$ with polynomial growth at infinity. □

Hence, it is valid for us to define

$$
\bar{a}_j^{s,\epsilon}(z)=\int a_j^s(z,y)\mu_z^\epsilon(dy),
$$

and

$$
\bar{a}_j^s(z)=\lim_{\epsilon\to0}\bar{a}_j^{s,\epsilon}(z).
$$

Furthermore, under Assumptions 1-2, the process Y_t^ϵ is exponentially mixing [27], i.e. for every functional f with polynomial growth at infinity and $T>0$, there exists a C such that

$$
\left|\int f(y)d\mu_z^\epsilon(y)-\mathbb{E}_z f(Y_T^\epsilon)\right|\le Ce^{-\lambda T},
$$

where λ is a positive constant. The smoothness of $\bar{a}_j^{s,\epsilon}(z)$ follows the smoothness of the coefficients $a^s(z,y)$ and the exponential mixing. Therefore the effective system for (2.8) can be written as

$$d\bar{Z}_t = \sum_j \int_0^\infty \nu_j^b \bar{\mathbf{A}}_j^s(q,\bar{Z}_t)\mathcal{P}(dt,dq), \tag{3.3}$$

where

$$\bar{\mathbf{A}}_j^s(q,\bar{Z}_t) = \begin{cases} 1, & q \in \Big(\sum_{i=1}^{j-1} \bar{a}_i^s(\bar{Z}_t), \sum_{i=1}^{j} \bar{a}_i^s(\bar{Z}_t)\Big), \\ 0, & \text{otherwise.} \end{cases}$$

3.2 Strong convergence of slow process

In this subsection, we will derive a strong convergence theorem for the slow process of NSSA.

Lemma 3.2. *For any $T>0$, and initial value x, there exists $C>0$ such that,*

$$\mathbb{E}|\tilde{a}_j^s(x) - \bar{a}_j^s(x)| \le C\left(\epsilon + \frac{e^{-\alpha T_0/\epsilon}}{(1+T_f/\epsilon)} + \frac{1}{\sqrt{M(1+T_f/\epsilon)}}\right),$$

where M, T_0, T_f are defined in our algorithm.

Proof. We follow the lines of proof in [8]. Since the state space is finite almost surely, it is easy to show that the fast process is φ-irreducible and satisfies the stability condition in [27]. Hence, for any test function g, there exists positive constants C and σ such that

$$\sup_{x\in\mathbb{X}} \left| e^{\tilde{\mathcal{L}}t} g(x) - \sum_y g(y)\mu_{b\cdot x}(y) \right| \le Ce^{-\sigma t}, \tag{3.4}$$

where $\mathbb{X} \in R^{N_s}$ is the state space, b is the vector which defines the slow process, and $\tilde{\mathcal{L}}/\epsilon$ is the infinitesimal generator for the fast process given by

$$\tilde{\mathcal{L}}V(y) = \sum_{j=1}^{M_f} a_j^f(z,y)\big(V(y+\nu_j^{ff}) - V(y)\big) + \epsilon \sum_{j=1}^{M_s} a_j^s(z,y)\big(V(y+\nu_j^{sf}) - V(y)\big).$$

We have

$$\begin{aligned} |\mathbb{E}\tilde{a}_j^s(x) - \bar{a}_j^s(x)|^2 &= \frac{1}{M^2}\mathbb{E}\left|\frac{1}{T_f}\sum_k \int_{T_0}^{T_0+T_f} (a_j^s(X_t^k) - \bar{a}_j^s(x))dt\right|^2 \\ &= \frac{1}{T_f^2 M^2}\sum_k \mathbb{E}\left|\int_{T_0}^{T_0+T_f} (a_j^s(X_t^k) - \bar{a}_j^s(x))dt\right|^2 \\ &\quad + \frac{1}{T_f^2 M^2}\sum_{k\neq l} \mathbb{E}\int_{T_0}^{T_0+T_f} (a_j^s(X_t^k) - \bar{a}_j^s(x))dt \cdot \int_{T_0}^{T_0+T_f} (a_j^s(X_t^l) - \tilde{a}_j^s(x))dt \\ &=: I_1 + I_2, \end{aligned}$$

where $\mathbb{E}$ denotes conditional expectation on $X^k_{t=0}=x$. Applying (3.4), we have

$$\begin{aligned}
I_1 &= \frac{2}{T_f^2M^2}\sum_k\left(\mathbb{E}\int_{T_0}^{T_0+T_f}(a_j^s(X_t^k)-\bar{a}_j^s(x))\cdot E_{X_t^k}\int_t^{T_0+T_f}(a_j^s(X_\tau^k)-\bar{a}_j^s(x))d\tau dt\right)\\
&\le \frac{2}{T_f^2M^2}\sum_k\mathbb{E}_x\int_{T_0}^{T_0+T_f}|a_j^s(X_t^k)-\bar{a}_j^s(x)|\int_{T_0}^{T_0+T_f}R|a_j^s|e^{-\sigma(\tau-t)/\epsilon}d\tau dt\\
&\le \frac{4R|a_j^s|^2(e^{-\sigma T_f/\epsilon-1+\sigma T_f/\epsilon})}{M(\sigma T_f/\epsilon)^2}\\
&\le \frac{C}{M(1+T_f/\epsilon)},
\end{aligned}$$

whereas,

$$\begin{aligned}
I_2 &\le \frac{1}{T_f^2M^2}\sum_{k\neq l}\mathbb{E}\int_{T_0}^{T_0+T_f}(a_j^s(X_t^k)-\bar{a}_j^s(x))dt\int_{T_0}^{T_0+T_f}(a_j^s(X_t^l)-\bar{a}_j^s(x))dt\\
&\le \frac{1}{T_f^2}\left|\mathbb{E}\int_{T_0}^{T_0+T_f}(a_j^s(X_t^k)-\bar{a}_j^s(x))dt\right|^2\\
&\le \frac{R^2|a_j^s|^2e^{-2\sigma T_0/\epsilon}(1-e^{-\sigma T_f/\epsilon})^2}{(\sigma T_f/\epsilon)^2}\\
&\le \frac{Ce^{-2\sigma T_0/\epsilon}}{(1+T_f/\epsilon)^2}.
\end{aligned}$$

Therefore,

$$\mathbb{E}|\tilde{a}_j^s(x)-\bar{a}_j^s(x)|\le C\left(\epsilon+\frac{e^{-\sigma T_0/\epsilon}}{1+T_f/\epsilon}+\frac{1}{\sqrt{M(1+T_f/\epsilon)}}\right).$$

The proof is complete. □

Proposition 3.1. Assume Assumptions 1-3. Then for any $T>0$, the following equation holds for a constant $C>0$ independent of ϵ

$$\sup_{0\le t\le T}\mathbb{E}|Z_t^\epsilon-\bar{Z}_t|^2\le C\epsilon, \tag{3.5}$$

where $\bar{Z}_t$ is defined in (3.3).

Proof. First, we partition the interval $[0,T]$ into subintervals of the same length $\delta=\epsilon$. Then, we construct an auxiliary processes $(\tilde{Z}_t^\epsilon,\tilde{Y}_t^\epsilon)$ such that for t in each subinterval $[t_n,t_{n+1})$,

$$\begin{cases}
d\tilde{Z}_t^\epsilon=\sum_j\int_0^\infty \nu_j^s\mathbf{A}_j^s(q,Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)\mathcal{P}(dt,dq), \quad \tilde{Z}_0^\epsilon=z,\\
d\tilde{Y}_t^\epsilon=\frac{1}{\epsilon}\sum_j\int_0^\infty \nu_j^{ff}\mathbf{A}_j^f(q,Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)\mathcal{P}(dt,dq)+\sum_j\int_0^\infty \nu_j^{sf}\mathbf{A}_j^s(q,Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)\mathcal{P}(dt,dq),\\
\tilde{Y}_{t_n}^\epsilon=Y_{t_n}^\epsilon.
\end{cases}$$

Therefore, by the Ito formula for jump-diffusion [4,31],

$$\begin{aligned}d\mathbb{E}|Y_t^\epsilon-\tilde{Y}_t^\epsilon|^2=&\frac{2}{\epsilon}\mathbb{E}(Y_t^\epsilon-\tilde{Y}_t^\epsilon)\cdot\sum_j\nu_j^{ff}(a_j^f(Z_t^\epsilon,Y_t^\epsilon)-a_j^f(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon))dt\\&+\frac{1}{\epsilon}\mathbb{E}\sum_j|\nu_j^{ff}|^2(a_j^f(Z_t^\epsilon,Y_t^\epsilon)dt-a_j^f(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)dt)\\&+\mathbb{E}\sum_j|\nu_j^{sf}|^2(a_j^f(Z_t^\epsilon,Y_t^\epsilon)dt-a_j^f(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)dt)\\&+\mathbb{E}(Y_t^\epsilon-\tilde{Y}_t^\epsilon)\cdot\sum_j\nu_j^{sf}(a_j^f(Z_t^\epsilon,Y_t^\epsilon)-a_j^f(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon))dt.\end{aligned}$$

Noting that the first two terms are dominant and absorb the last two terms, we have

$$\begin{aligned}d\mathbb{E}|Y_t^\epsilon-\tilde{Y}_t^\epsilon|^2\leq&\frac{2}{\epsilon}\mathbb{E}(Y_t^\epsilon-\tilde{Y}_t^\epsilon)\cdot\sum_j\nu_j^{ff}(a_j^f(Z_t^\epsilon,Y_t^\epsilon)dt-a_j^f(Z_t^\epsilon,\tilde{Y}_t^\epsilon)dt)\\&+\frac{2}{\epsilon}\mathbb{E}(Y_t^\epsilon-\tilde{Y}_t^\epsilon)\cdot\sum_j\nu_j^{ff}(a_j^f(Z_t^\epsilon,\tilde{Y}_t^\epsilon)dt-a_j^f(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)dt)\\&+\frac{1}{\epsilon}\mathbb{E}\sum_j|\nu_j^{ff}|^2(a_j^f(Z_t^\epsilon,Y_t^\epsilon)dt-a_j^f(Z_t^\epsilon,\tilde{Y}_t^\epsilon)dt)\\&+\frac{1}{\epsilon}\mathbb{E}\sum_j|\nu_j^{ff}|^2(a_j^f(Z_t^\epsilon,\tilde{Y}_t^\epsilon)dt-a_j^f(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)dt)\\\leq&\frac{2\sum_j|\nu_j^{ff}|L_j+|\nu_j^{ff}|^2L_j}{\epsilon}\cdot\mathbb{E}\Big(|Y_t^\epsilon-\tilde{Y}_t^\epsilon||Z_t^\epsilon-Z_{t_n}^\epsilon|+|Z_t^\epsilon-Z_{t_n}^\epsilon|\Big)-\frac{\gamma}{\epsilon}\mathbb{E}|Y_t^\epsilon-\tilde{Y}_t^\epsilon|^2.\end{aligned}$$

By virtue of the Young's inequality,

$$\begin{aligned}&|Y_t^\epsilon-\tilde{Y}_t^\epsilon||Z_t^\epsilon-Z_{t_n}^\epsilon|+|Z_t^\epsilon-Z_{t_n}^\epsilon|\\\leq&\frac{\gamma}{2C'}|Y_t^\epsilon-\tilde{Y}_t^\epsilon|^2+\frac{C'}{2\gamma}|Z_t^\epsilon-Z_{t_n}^\epsilon|^2+|Z_t^\epsilon-Z_{t_n}^\epsilon|,\end{aligned}$$

where $C'=2\sum_j|\nu_j^{ff}|L_j+|\nu_j^{ff}|^2L_j$ and L_j is the Lipschitz constant for a_j^s [23]. It exists w.p.1 because of the fact that $a_j^s(Z_t^\epsilon,Y_t^\epsilon)$ is a polynomial of both Z_t^ϵ and Y_t^ϵ and that Z_t^ϵ and Y_t^ϵ are bounded w.p.1. Hence, for $t\in[t_n,t_{n+1})$, there exists one jump on (t_n,t_{n+1}) with probability of order δ,

$$\mathbb{E}|Z_t^\epsilon-Z_{t_n}^\epsilon|^2\leq C\delta,\quad \mathbb{E}|Z_t^\epsilon-Z_{t_n}^\epsilon|\leq C\delta.$$

Therefore,

$$d\mathbb{E}|Y_t^\epsilon-\tilde{Y}_t^\epsilon|^2\leq\left(\frac{C'^2}{2\gamma\epsilon}+\frac{C'}{\epsilon}\right)C\delta-\frac{\gamma}{2\epsilon}d\mathbb{E}|Y_t^\epsilon-\tilde{Y}_t^\epsilon|^2.$$

The Gronwall inequality implies

$$\mathbb{E}|Y_t^\epsilon - \tilde{Y}_t^\epsilon|^2 \le C\delta. \tag{3.6}$$

Next, we move to the estimation of macro time scales. Since

$$Z_t^\epsilon - \tilde{Z}_t^\epsilon = \sum_j \nu_j \int_0^t \int_0^\infty (\mathbf{A}_j(q, Z_t^\epsilon, Y_t^\epsilon) - \mathbf{A}_j(q, Z_{t_n}^\epsilon, \tilde{Y}_t^\epsilon))\mathcal{P}(dt,dq),$$

by Ito formula for jump SDE again,

$$\begin{aligned}
d\mathbb{E}|Z_t^\epsilon - \tilde{Z}_t^\epsilon|^2 &= 2\mathbb{E}\left((Z_t^\epsilon - \tilde{Z}_t^\epsilon)\cdot \sum_j \nu_j^b [a_j^s(Z_t^\epsilon, Y_t^\epsilon) - a_j^s(Z_{t_n}^\epsilon, \tilde{Y}_t^\epsilon)]dt\right) \\
&\quad + \mathbb{E}\sum_j |\nu_j^b|^2 [a_j^s(Z_t^\epsilon, Y_t^\epsilon) - a_j^s(Z_{t_n}^\epsilon, \tilde{Y}_t^\epsilon)]dt \\
&\le 2\mathbb{E}\sum_j |\nu_j^b| L_j |Z_t^\epsilon - \tilde{X}_t^\epsilon||Z_t^\epsilon - Z_{t_n}^\epsilon|dt \\
&\quad + 2\mathbb{E}\sum_j |\nu_j^b| L_j |X_t^\epsilon - \tilde{Z}_t^\epsilon||Y_t^\epsilon - \tilde{Y}_t^\epsilon|dt \\
&\quad + \mathbb{E}\sum_j |\nu_j^b|^2 L_j [|Z_t^\epsilon - Z_{t_n}^\epsilon| + |Y_t^\epsilon - \tilde{Y}_t^\epsilon|]dt \\
&\le \mathbb{E}\sum_j |\nu_j^b| L_j \left[|Z_t^\epsilon - \tilde{Z}_t^\epsilon|^2 + |Z_t^\epsilon - Z_{t_n}^\epsilon|^2\right]dt \\
&\quad + \mathbb{E}\sum_j |\nu_j^b| L_j \left[|Z_t^\epsilon - \tilde{Z}_t^\epsilon|^2 + |Y_t^\epsilon - \tilde{Y}_t^\epsilon|^2\right]dt \\
&\quad + \mathbb{E}\sum_j |\nu_j^b|^2 L_j \left[|Z_t^\epsilon - Z_{t_n}^\epsilon|^2 + |Y_t^\epsilon - \tilde{Y}_t^\epsilon|^2\right]dt. \\
&\le \left(C\delta + 2\mathbb{E}\sum_j |\nu_j^b| L_j |Z_t^\epsilon - \tilde{Z}_t^\epsilon|^2\right)dt.
\end{aligned}$$

Here, we use the fact [23] that either

$$\begin{aligned}
|Z_t^\epsilon - Z_{t_n}^\epsilon| = 0 \quad \text{or} \quad |Z_t^\epsilon - Z_{t_n}^\epsilon| \ge 1, \\
|Y_t^\epsilon - \tilde{Y}_t^\epsilon| = 0 \quad \text{or} \quad |Y_t^\epsilon - \tilde{Y}_t^\epsilon| \ge 1,
\end{aligned}$$

which implies

$$\mathbb{E}|Z_t^\epsilon - Z_{t_n}^\epsilon| \le \mathbb{E}|Z_t^\epsilon - Z_{t_n}^\epsilon|^2, \quad \mathbb{E}|Y_t^\epsilon - \tilde{Y}_t^\epsilon| \le \mathbb{E}|Y_t^\epsilon - \tilde{Y}_t^\epsilon|^2.$$

By the Gronwall inequality,

$$\mathbb{E}|Z_t^\epsilon - \tilde{Z}_t^\epsilon|^2 \le C\epsilon. \tag{3.7}$$

Next, we focus on the estimation of $\mathbb{E}|\tilde{Z}_t^\epsilon - \bar{Z}_t|$. By the exponential mixing property of the fast process (3.4), see also [7] and the invariant measure μ_z^ϵ is finite, we have

$$\begin{aligned}
&\mathbb{E}\left|\int_{t_n}^{t_{n+1}}(a_j^s(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)-\bar{a}_j^s(\bar{Z}_{t_n}))dt\right| \\
\leq&\mathbb{E}\left|\int_{t_n}^{t_{n+1}}(a_j^s(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)-\bar{a}_j^{s,\epsilon}(Z_{t_n}^\epsilon))dt\right|+\mathbb{E}|\bar{a}_j^{s,\epsilon}(Z_{t_n}^\epsilon)-\bar{a}_j^s(Z_{t_n}^\epsilon)|\delta \\
\leq&C\epsilon\mathbb{E}(|Z_{t_n}^\epsilon|^2+1) \\
\leq&C\epsilon. \qquad (3.8)
\end{aligned}$$

Thus, from Ito formula for jump SDE,

$$\begin{aligned}
\mathbb{E}|\tilde{Z}_t^\epsilon-\bar{Z}_t|^2=&\int_0^t\mathbb{E}(\tilde{Z}_t^\epsilon-\bar{Z}_t,\sum_j\nu_j^b(a_j^s(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)dt-\bar{a}_j^s(\bar{Z}_t))dt) \\
&+\int_0^t\sum_j|\nu_j^b|^2\mathbb{E}(a_j^s(Z_{t_n}^\epsilon,\tilde{Y}_t^\epsilon)dt-\bar{a}_j^s(\bar{Z}_t)dt) \\
=&\sum_{i=0}^n\int_{t_i}^{t_{i+1}}\mathbb{E}(\tilde{Z}_t^\epsilon-\bar{Z}_t,\sum_j\nu_j^b(a_j^s(Z_{t_i}^\epsilon,\tilde{Y}_t^\epsilon)dt-\bar{a}_j^s(\bar{Z}_t))dt) \\
&+\sum_j|\nu_j^b|^2\mathbb{E}(a_j^s(Z_{t_i}^\epsilon,\tilde{Y}_t^\epsilon)dt-\bar{a}_j^s(\bar{Z}_t)dt) \\
=&\sum_{i=0}^n\int_{t_i}^{t_{i+1}}\mathbb{E}(\tilde{Z}_t^\epsilon-\bar{Z}_t,\sum_j\nu_j^b(a_j^s(Z_{t_i}^\epsilon,\tilde{Y}_t^\epsilon)-\bar{a}_j^s(Z_{t_i}^\epsilon))) \\
&+\mathbb{E}(\tilde{Z}_t^\epsilon-\bar{Z}_t,\sum_j\nu_j^b(\bar{a}_j^s(Z_{t_i}^\epsilon)-\bar{a}_j^s(\tilde{Z}_t^\epsilon)) \\
&+\mathbb{E}(\tilde{Z}_t^\epsilon-\bar{Z}_t,\sum_j\nu_j^b(\bar{a}_j^s(\tilde{Z}_t^\epsilon)-\bar{a}_j^s(\bar{Z}_t))dt \\
&+\sum_{i=0}^n\int_{t_i}^{t_{i+1}}\sum_j|\bar{\nu}_j^b|^2\mathbb{E}(a_j^s(Z_{t_i}^\epsilon,\tilde{Y}_t^\epsilon)-\bar{a}_j^s(Z_{t_i}^\epsilon))dt \\
&+\sum_{i=0}^n\int_{t_i}^{t_{i+1}}\mathbb{E}\sum_j|\bar{\nu}_j^b|^2(\bar{a}_j^s(Z_{t_i}^\epsilon)-\bar{a}_j^s(\tilde{Z}_t^\epsilon))dt \\
&+\sum_{i=0}^n\int_{t_i}^{t_{i+1}}\mathbb{E}\sum_j|\bar{\nu}_j^b|^2(\bar{a}_j^s(\tilde{Z}_t^\epsilon)-\bar{a}_j^s(\bar{Z}_t))dt \\
:=&\sum_{i=1}^6 I_i.
\end{aligned}$$

Now let us estimate these six terms one by one.

Using the elementary inequality $(a,b)\leq\frac{|a|^2}{2}+\frac{|b|^2}{2}$, the smoothness of $\bar{a}_j^s$ and (3.8),

$$\begin{aligned}
I_1 &\leq \frac{1}{2}\int_0^t \mathbb{E}|\tilde{Z}_s^\epsilon-\bar{Z}_s|^2ds+\frac{1}{2}\sum_{i=0}^{n}\int_{t_i}^{t_{i+1}}\mathbb{E}|\sum_j \nu_j^b(a_j^s(Z_t^\epsilon,\tilde{Y}_t^\epsilon)-\bar{a}_j^s(\bar{Z}_{t_i}^\epsilon)|^2dt\\
&\leq \frac{1}{2}\int_0^t \mathbb{E}|\tilde{Z}_s^\epsilon-\bar{Z}_s|^2ds+\frac{T}{2}\sum_j|\nu_j^b|^2C\epsilon,\\
I_2 &\leq \frac{1}{2}\int_0^t \mathbb{E}|\tilde{Z}_s^\epsilon-\bar{Z}_s|^2ds+\frac{T}{2}\sum_j|\nu_j^b|^2L_j^2C\delta^2,\\
I_3 &\leq \int_0^t(\frac{1}{2}+\sum_j|\nu_j^b|^2L_j)\mathbb{E}|\tilde{Z}_s^\epsilon-\bar{Z}_s|^2ds,\\
I_4 &\leq \frac{T}{2}\sum_j|\nu_j^b|^2C\epsilon,\\
I_5 &\leq \frac{T}{2}\sum_j|\nu_j^b|^2L_jC\delta,\\
I_6 &\leq \int_0^t\sum_j|\nu_j^b|^2L_j\mathbb{E}|\tilde{Z}_s^\epsilon-\bar{Z}_s|^2ds.
\end{aligned}$$

Combining all estimates above and applying the Gronwall inequality, we conclude

$$\mathbb{E}|\tilde{Z}_t^\epsilon-\bar{Z}_t|^2\leq C\epsilon. \tag{3.9}$$

Since $\mathbb{E}|Z_t^\epsilon-\bar{Z}_t|^2\leq 2\mathbb{E}|Z_t^\epsilon-\tilde{Z}_t^\epsilon|^2+2\mathbb{E}|\tilde{Z}_t^\epsilon-\bar{Z}_t|^2$, combining (3.7) and (3.9), we arrive at (3.5). This completes the proof. □

Let X_n be the output of our algorithm NSSA. Define $Z_n=(b_1\cdot X_n,\cdots,b_J\cdot X_n)$. The following theorem gives the strong convergence of the Nested SSA.

Theorem 3.1. *For any $T>0$, let Z_t^ϵ be the macro-scale solution to (2.8) and Z_n be the macro-scale process yielded by NSSA at t_n. Then,*

$$\mathbb{E}|Z_{t_n}^\epsilon-Z_n|^2\leq C\left(\epsilon+\frac{e^{-\alpha T_0/\epsilon}}{1+T_f/\epsilon}+\frac{1}{\sqrt{M(1+T_f/\epsilon)}}\right),$$

where t_n are jump time points of the process.

Proof. It is clear that

$$\mathbb{E}|Z_t^\epsilon-Z_t|^2\leq 2\mathbb{E}|Z_t^\epsilon-\bar{Z}_t|^2+2\mathbb{E}|\bar{Z}_t-Z_t|^2,$$

where $Z_t=Z_n$ for $t\in[t_n,t_{n+1})$. The first term in the right hand side is estimated by Proposition 3.1. Hence, we only need to estimate the second term. The difference between (3.3)

and our algorithm lies in the difference of intensities of slow process, which are $\bar{a}_j^s$ and $\tilde{a}_j^s$, respectively. Moreover, the difference of these two intensities is estimated in Lemma 3.2 since initial values of associated processes are chosen the same. By the same technique used in the preceding proposition,

$$\begin{aligned}
&\mathbb{E}|Z_n-\bar{Z}_{t_n}|^2\\
=&\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}(Z_i-\bar{Z}_{t_i},\sum_j\nu_j^b(\tilde{a}_j^s-\bar{a}_j^s)dt)\\
&+\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}\sum_j|\nu_j^b|^2\sum_j(\tilde{a}_j^s-\bar{a}_j^s)dt\\
:=&J_1+J_2,
\end{aligned}$$

where

$$\begin{aligned}
J_1\le&\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\frac{1}{2}\mathbb{E}|Z_i-\bar{Z}_{t_i}|^2dt+\frac{TM_s}{2}\sum_j|\nu_j^b|^2\mathbb{E}|\tilde{a}_j^s-\bar{a}_j^s|^2,\\
J_2\le&\frac{T}{2}\sum_j|\nu_j^b|^2\mathbb{E}|\tilde{a}_j^s-\bar{a}_j^s|.
\end{aligned}$$

Lemma 3.2 and the discrete Gronwall's inequality yield the result. □

4 Tau-leaping approximation for invariant measure

When the reacting spices for the fast reactions are in high concentrations, direct SSA simulations for the inner loop will require very small time steps therefore become very inefficient. We want to use the Tau-Leaping method to circumvent this difficulty. In this section, we will give the strong convergence of NSSA when the Tau-Leaping method with step size τ is adopted as the Inner solver. For simplicity of analysis, we choose the number of replicas $M=1$ in the algorithm. At each slow time step of NSSA, denote by X_n the state of the system and the fast process produced by our modified NSSA by $Y_{j\tau}$. Our modified algorithm reads as follows.

1. **Inner SSA** Run tau-leaping method with the fast reactions $R^f=(\frac{1}{\epsilon}a^f,\ \nu^f)$ only, for a time interval of $[T_0,T_0+T_f]$.

$$X_{n,i+1}=X_{n,i}+\sum_{j=1}^{M_f}\nu_j^f P_j(a_j^f(X_{n,i}),\tau),\quad X_{n,0}=X_n,\quad i=0,\cdots,N, \tag{4.1}$$

where $N\cdot\tau=T_f$. During this calculation, compute the modified slow rates $\tilde{a}_j^{s,\prime}=\frac{1}{N}\sum_{i=1}^N a_j^s(X_{n,i})$.

2. **Outer SSA** Run one step of SSA for the modified slow reactions $(R^s(\tilde{a}^{s,\prime},\nu^s))$ to generate $(t_{n+1},\ X_{n+1})$ from (t_n,X_n) unless a certain criterion is satisfied. Specifically,

$$t_{n+1}=t_n+\delta t,\qquad X_{n+1}=X_n+\nu_k,$$

where k is chosen based upon the criterion (2.3) associated with the computed slow rates.

Under the transformation

$$Z_n=(b_1\cdot X_n,\cdots,b_J\cdot X_n),\quad Y_n=(\nu_1^f\cdot X_n,\cdots,\nu_{M_f}^f\cdot X_n),$$

the tau-leaping scheme for the fast process reads as

$$Y_{n,i+1}=Y_{n,i}+\sum_{j=1}^{M_f}\nu_j^f P_j(a_j^f(Z_n,Y_{n,i}),\tau),\quad Y_{n,0}=Y_n.$$

Based upon $Y_{n,i}$, we compute the modified slow rates $\tilde{a}_j^{s,\prime}=\frac{1}{N}\sum_{i=0}^{N-1}a_j^s(Z_n,Y_{n,i})$. The effective dynamics in outer SSA is then specified as $Z_{n+1}=Z_n+\nu_k^s$ under $(R^s(\tilde{a}^{s,\prime},\nu^s))$.

Evidently, our inner solver is the time-averaging for the stationary measure via tau-leaping method. For any bounded functional ϕ and any macro-process state Z_n, we need to estimate

$$\left|\int\phi(y)d\mu_{Z_n}(y)-\mathbb{E}\frac{1}{N}\sum_{i=1}^{N}\phi(Y_{n,i})\right|.$$

Following the idea of [25], we define the stationary average of $\phi:R^{M_f}\to R$ by

$$\bar{\phi}^\epsilon=\int_{R^{M_f}}\phi(y)\mu_{Z_n}^\epsilon(dy),\qquad \bar{\phi}=\int_{R^{M_f}}\phi(y)\mu_{Z_n}(dy).$$

Let u and u^ϵ solve

$$\mathcal{L}u=\phi-\bar{\phi},\tag{4.2}$$

and

$$\mathcal{L}u^\epsilon=\phi-\bar{\phi}^\epsilon,\tag{4.3}$$

respectively, where

$$\mathcal{L}V(y)=\frac{1}{\epsilon}\sum_{j=1}^{M_f}a_j^f(Z_n,y)\big(V(y+\nu_j^{ff})-V(y)\big)+\sum_{j=1}^{M_s}a_j^s(Z_n,y)\big(V(y+\nu_j^{sf})-V(y)\big),$$

is the generator of the fast process. Notice that both Eq. (4.3) and Eq. (4.2) are finite-difference equations.

Next, we prove that the solution is bounded if the right-hand side satisfies some certain conditions. To this end, we introduce some preliminaries on difference equations. Consider a homogeneous difference system

$$U(k+1)=AU(k),\tag{4.4}$$

where U is a vector, A is a nonsingular matrix. The general solution can be written as $U(k)=A^k\eta$, where η is an arbitrary constant vector [2]. For the inhomogeneous system

$$U(k+1)=AU(k)+G(k,U(k)), \tag{4.5}$$

we have the following result

Lemma 4.1. ([2, Theorem 5.3.2]) *Assume (1): G satisfies the condition $\|G(k,U)\|\leq h(k)\|U\|$ with $h(k)$ a nonnegative function; (2): Solution of* (4.4) *are uniformly bounded. Then all solution of* (4.5) *are bounded.*

We now proceed to explore (4.2).

Lemma 4.2. *Let ϕ be a bounded function. Then there exists a unique bounded solution u to* (4.2). *Furthermore,*

$$\begin{cases} u(y+\nu_j^{ff})-u(y)=\mathcal{O}(\epsilon), & j=1,\cdots,M_f; \\ u(y+\nu_j^{sf})-u(y)=\mathcal{O}(1), & j=M_f+1,\cdots,M_f+M_s. \end{cases} \tag{4.6}$$

Proof. We rewrite (4.2) in a compact form

$$\frac{1}{\epsilon}\sum_{j=1}^{M_f}a_j^f(Z_n,y)\tilde{u}(j)+\sum_{j=1}^{M_s}a_j^s(Z_n,y)\tilde{u}(j)+a_0(Z_n,y)\tilde{u}(0)=\tilde{\phi}(y), \tag{4.7}$$

where $\tilde{\phi}(y)=\phi-\bar{\phi}^\epsilon$, and $\tilde{u}(k)=u(y+\nu_k)$, $k=0,\cdots,M_f+M_s$, and

$$a_0(Z_n,y)=-\frac{1}{\epsilon}\sum_{j=1}^{M_f}a_j^f(Z_n,y)-\sum_{j=1}^{M_s}a_j^s(Z_n,y).$$

Then, it is clear that (4.7) is a difference equation with order M_f+M_s+1. From Theorem 2 of [28, p. 134], there exists one and only one solution to (4.7). Next, we prove this solution is bounded. Denote $U_i=\tilde{u}(i)$, $i=1,\cdots,M_f+M_s$ and

$$A=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & & \cdots & & \cdots \\ -\frac{a_1^f(Z_n,y)}{a_0(Z_n,y)} & -\frac{a_2^f(Z_n,y)}{a_0(Z_n,y)} & \cdots & \cdots & -\frac{a_{M_s-1}(Z_n,y)}{a_0(Z_n,y)} \end{bmatrix},\quad G=\left[0,0,\cdots,\tilde{\phi}(y)/a_0(Z_n,y)\right]^T.$$

We obtain

$$U(k+1)=AU(k)+G. \tag{4.8}$$

The Gerschgorin theorem implies that the modulus of all eigenvalues of A are less than or equal to 1. Therefore, $\|A^k\|$ is uniformly bounded, which implies the solution of $U(k+1)=AU(k)$ is uniformly bounded. Clearly, G satisfies Assumption (1) in Lemma 4.1.

Hence, by the lemma, every $\tilde{u}(k)$ is bounded. Since ϵ is arbitrary, by equating both sides of

$$\frac{1}{\epsilon}\sum_{j=1}^{M_f} a_j^f(Z_n,y)(u(y+\nu_j^f)-u(y))+\sum_{j=1}^{M_s} a_j^s(Z_n,y)(u(y+\nu_j^s)-u(y))=\tilde{\phi}(y) \tag{4.9}$$

on ϵ, (4.6) follows. Similarly, the solution of (4.3) has the same property. □

We return to our own system. Applying the Ito formula for jump SDE,

$$\begin{aligned} u(Y_t^\epsilon)-u(y)=&\int_0^t \mathcal{L}u(s)ds+\int_0^t\int_0^\infty [u(Y_s^\epsilon+\frac{1}{\epsilon}\sum_{j=1}^{M_f}\nu_j^{ff}A_j(q,Z_s^\epsilon,Y_s^\epsilon)\\ &+\sum_{j=1}^{M_s}\nu_j^{sf}A_j(q,Z_s^\epsilon,Y_s^\epsilon))-u(Y_s^\epsilon)]\tilde{\mathcal{P}}(ds,dq), \end{aligned}$$

where $\tilde{\mathcal{P}}(ds,dq)$ is the compensated measure of $\mathcal{P}$. Clearly, the second term is a martingale and we denote it by $M(t)$. Following [25], and applying (4.2), we obtain

$$\frac{1}{t}\int_0^t \phi(Y_s^\epsilon)ds-\bar{\phi}=\frac{u(Y_t^\epsilon)-u(y)}{t}-\frac{1}{t}M(t).$$

Clearly, the fist term of the right hand side is bounded by $\mathcal{O}(\frac{1}{T_f})$. For the martingale term, Lemma 4.2 yields

$$\begin{aligned} &\frac{1}{t^2}\mathbb{E}(M(t))^2\\ =&\frac{1}{t^2}\int_0^t\frac{1}{\epsilon}\sum_{j=1}^{M_f}|\nu_j^{ff}|^2a_j^f(Z_s^\epsilon,Y_s^\epsilon)\left[u\left(Y_s^\epsilon+\frac{1}{\epsilon}\sum_{j=1}^{M_f}\nu_j^{ff}A_j(q,Z_s^\epsilon,Y_s^\epsilon)\right)-u(Y_s^\epsilon)\right]^2ds\\ &+\frac{1}{t^2}\int_0^t\sum_{j=M_f+1}^{M_f+M_s}|\nu_j^{sf}|^2a_j^s(Z_s^\epsilon,Y_s^\epsilon)\left[u\left(Y_s+\sum_{j=M_f+1}^{M_f+M_s}\nu_j^{sf}A_j(q,Z_s^\epsilon,Y_s^\epsilon)\right)-u(Y_s^\epsilon)\right]^2ds\\ \leq&\frac{C}{t}. \end{aligned}$$

As a result,

$$\mathbb{E}\left(\frac{1}{T_f}\int_0^{T_f}\phi(Y_s^\epsilon)ds-\bar{\phi}\right)^2\leq\frac{C}{T_f}.$$

We can also show that

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t\phi(Y_s^\epsilon)ds=\bar{\phi}\qquad\text{a.s.}$$

and the reasoning is the same as that in [25].

Lemma 4.3. *Assume Assumption (1) for the fast process. Then for any bounded ϕ and each Z_n, there exists a $C>0$ such that*

$$\left|\int \phi(y)d\mu_{Z_n}(y)-\mathbb{E}\frac{1}{N}\sum_{i=1}^{N}\phi(Y_{n,i})\right|\leq C\Big(\tau+\frac{1}{T_f}\Big).$$

Proof. Since $\left|\int \phi(y)d\mu_{Z_n}(y)-\int \phi(y)d\mu^{\epsilon}_{Z_n}(y)\right|$ is arbitrarily small if ϵ is sufficiently small by definition. We focus on the estimate of $\left|\int \phi(y)d\mu^{\epsilon}_{Z_n}(y)-\mathbb{E}\frac{1}{N}\sum_{i=1}^{N}\phi(Y_{n,i})\right|$.

Note that the fast process is bounded with probability 1. We assume that the reaction rates $a_j^f(Z_t,Y_t)$ is Lipschitz continuous with constant L for each Z_t. Here $Z_t=Z_n, t\in[t_n,t_{n+1})$. Divide the interval $[T_0,T_0+T_f]$ into equidistant subintervals with length τ. Denoting $s_i=T_0+i\tau$, we define $Y_t=Y_{n,i}, t\in[s_i,s_{i+1})$. Then, by Dynkin's formula

$$\mathbb{E}u^{\epsilon}(Y_{n,i+1})-\mathbb{E}u^{\epsilon}(Y_{n,i})=\mathbb{E}\int_{s_i}^{s_{i+1}}\tilde{\mathcal{L}}u^{\epsilon}(Y_s)ds, \tag{4.10}$$

where $\tilde{\mathcal{L}}$ is the generator of the virtual process Y_t [1],

$$\tilde{\mathcal{L}}V(y)=\frac{1}{\epsilon}\sum_{j=1}^{M_f}a_j^f(Z_n,y_{n,i})(V(y+\nu_j^{ff})-V(y))+\sum_{j=1}^{M_s}a_j^s(Z_n,y_{n,i})(V(y+\nu_j^{sf})-V(y)),\ \ t\in[s_i,s_{i+1}).$$

Since on each $[T_0,T_0+T_f]$, the macro-process Z_t does not change, we suppress the notation Z_n in the following lines. Summing (4.10) over the first N terms, dividing by $N\cdot\tau=T_f$, we have

$$\begin{aligned}
&\frac{1}{N\tau}(\mathbb{E}u^{\epsilon}(Y_{n,N})-\mathbb{E}u^{\epsilon}(Y_{n,0}))\\
=&\sum_{i=1}^{N}\mathbb{E}(\tilde{\mathcal{L}}-\mathcal{L})u^{\epsilon}(Y_{n,i})+\frac{1}{N}\mathbb{E}\sum_{i=1}^{N}[\phi(Y_{n,i})-\bar{\phi}^{\epsilon}]\\
=&\underbrace{\tau\mathbb{E}\sum_{i=1}^{N}\sum_{j=1}^{M_f}\nu_j^{ff}\frac{1}{\epsilon}(a_j^f(Y_{n,i})-a_j^f(Y_t^{\epsilon}))\mathbf{1}_{t\in(s_i,s_{i+1})}(u^{\epsilon}(Y_{n,i}+\nu_j^{ff})-u^{\epsilon}(Y_{n,i}))}_{I_1}\\
&\underbrace{+\tau\mathbb{E}\sum_{i=1}^{N}\sum_{j=M_f+1}^{M_f+M_s}\nu_j^{sf}(a_j^s(Y_{n,i})-a_j^s(Y_t^{\epsilon}))\mathbf{1}_{t\in(s_i,s_{i+1})}(u^{\epsilon}(Y_{n,i}+\nu_j^{sf})-u^{\epsilon}(Y_{n,i}))}_{I_2}\\
&+\frac{1}{N}\mathbb{E}\sum_{i=1}^{N}[\phi(Y_{n,i})-\bar{\phi}^{\epsilon}].
\end{aligned}$$

By the fact that u^ϵ is bounded, the left-hand side can be bounded by $\mathcal{O}(1/T_f)$. From the Lipschtiz assumption on $a(y)$ and Lemma 4.2, we easily have

$$|I_1+I_2|\leq\sum_{j=1}^{N}C\tau^2$$

since the probability that reaction R_j fires on $[s_i,s_{i+1}]$ is $a_j^f(Y_{n,i})\tau$. Therefore,

$$\frac{1}{N}\mathbb{E}\sum_{i=1}^{N}[\phi(Y_{n,i})-\bar{\phi}^\epsilon]\leq C\Big(\tau+\frac{1}{T_f}\Big).$$

The result follows. □

Theorem 4.1. *Assume Assumptions 1-3 and the modified NSSA is applied for the chemical network. Then, for any jump time point t_n on $[0,T]$, the following inequality holds for a constant $C>0$ independent of ϵ and τ*

$$\mathbb{E}|Z_{t_n}^\epsilon-Z_n|^2\leq C\Big(\epsilon+\tau+\frac{1}{T_f}\Big),\tag{4.11}$$

where Z_t is the process yielded by the algorithm, ϵ is the scale separation parameter, τ is the step size for the tau-leaping method, and T_f is defined in our NSSA algorithm.

Proof. Since

$$\mathbb{E}|Z_{t_n}^\epsilon-Z_n|^2\leq 2\mathbb{E}|Z_{t_n}^\epsilon-\bar{Z}_{t_n}|^2+2\mathbb{E}|\bar{Z}_{t_n}-Z_n|^2.$$

The first term in the right hand side is estimated in Proposition 3.1. We only need to estimate the second term. From $\tilde{a}_j^{s,\prime}=\frac{1}{N}\sum_{i=0}^{N-1}a_j^s(Z_n,Y_{n,i})$ with $N\cdot\tau=T_f$

$$\begin{aligned}\mathbb{E}|Z_n-\bar{Z}_{t_n}|^2&=\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}(Z_i-\bar{Z}_{t_i},\sum_j\nu_j^b(\tilde{a}_j^{s,\prime}-\bar{a}_j^s)dt\\&\quad+\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}\sum_j|\nu_j^b|^2\sum_j(\tilde{a}_j^{s,\prime}-\bar{a}_j^s)dt\\&:=J_1+J_2,\end{aligned}$$

where

$$\begin{aligned}J_1&\leq\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\frac{1}{2}\mathbb{E}|Z_i-\bar{Z}_{t_i}|^2dt+\frac{T}{2}\sum_j|\nu_j^b|^2\mathbb{E}|\tilde{a}_j^{s,\prime}-\bar{a}_j^s|^2,\\J_2&\leq\frac{T}{2}\sum_j|\nu_j^b|^2\mathbb{E}|\tilde{a}_j^{s,\prime}-\bar{a}_j^s|.\end{aligned}$$

By Lemma 4.3, $\mathbb{E}|\tilde{a}_j^{s,\prime}-\bar{a}_j^s|\leq C(\tau+\frac{1}{T_f})$. The discrete Gronwall's inequality and Proposition 3.1 yield the result. □

5 Numerical examples

Example 5.1. Consider a simple isomerization:

$$\star \underset{c_{10}}{\overset{c_{01}}{\rightleftharpoons}} S_1,$$

where c_{01} implies the constant rate of the production of S_1 from source and c_{10} means the degradation rate coefficient of S_1. Suppose that there are x_0 molecules of S_1 at time $t=0$. Let $P(t,x)$ be the probability that at time t there are x molecules of S_1. Then, from [37, Sec. 8.4],

$$P(t,x)=\sum_{k=0}^{\min\{x_0,x\}}\binom{x_0}{k}p^k(t)(1-p(t))^{x_0-k}\cdot\frac{\lambda^{x-k}(t)}{(x-k)!}e^{-\lambda(t)},$$

where $p(t)=e^{-c_{10}t}$ and $\lambda(t)=c_{01}(1-e^{-c_{10}t})/c_{10}$ are solutions of the equations,

$$\begin{aligned} &dp(t)/dt=-c_{10}p(t), \quad p(0)=1,\\ &d\lambda(t)/dt=-c_{10}\lambda(t)+c_{01}, \quad \lambda(0)=0. \end{aligned}$$

Note that $\lim_{t\to\infty}p(t)=0$ and $\lim_{t\to\infty}=c_{01}/c_{10}:=\bar{\lambda}$. Hence, the limit distribution

$$\lim_{t\to\infty}P(t,x)=\frac{\bar{\lambda}^x}{x!}e^{-\bar{\lambda}}:=\mu(x).$$

Set initial value of S_1 to be 900 and $c_{01}=c_{10}=10$, and choose the functional $f(x)=1/(1+x^2)$.

The time averaging error for τ-leaping method is shown in Fig. 1. To test the dependency of error on τ, we simulate the system on $[0,10]$ with an increasing τ from 0.01 to

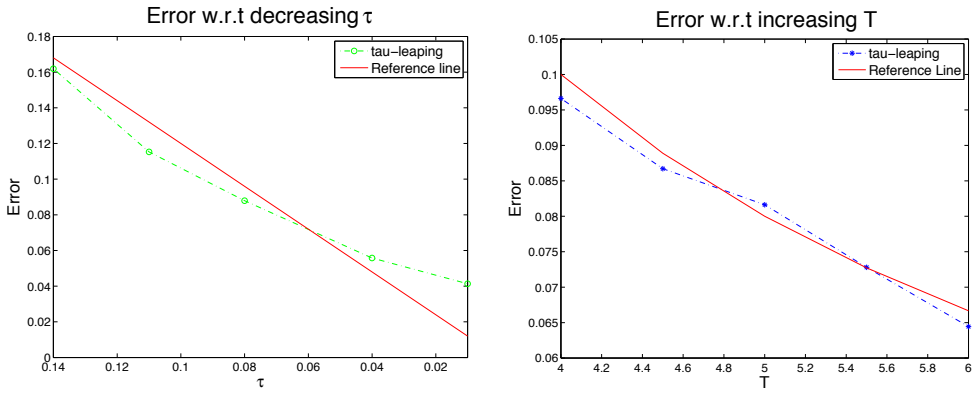

Figure 1: Time averaging error of τ-leaping method for Example 5.1. Left: $\tau=0.01,0.04,0.08,0.11,0.14$ with fixed $\tilde{T}=5$. Right: $\tilde{T}=4,4.5,5,5.5,6$ with fixed $\tau=0.02$.

0.14. For each τ, we collect the arithmetic average of 1000 replicas of $\frac{1}{N}\sum_{i=1}^{N} f(S_{1,i})$, where $N=\frac{10}{\tau}$ to approximate $\mathbb{E}\frac{1}{N}\sum_{i=1}^{N} f(S_{1,i})$. We find that the error is approximately proportional to τ with a coefficient 1.2, see Fig. 1 (left). Similarly, we fix $\tau=0.02$ and vary T to test the dependency of error on the time length T. We also run 1000 replicas to approximate the expectation. From Fig. 1 (right), we can conclude that the error is approximately inversely proportional to T with a coefficient 0.4. Therefore, this simple example shows the sharpness of our Lemma 4.3.

We observe that even we choose $\tau=0.1$ in the tau-leaping method and repeat for 1000 times, the distribution of S_1 at time 10 is close to that of limit distribution (with a relative error 22%) and it only takes 5.1 CPU seconds. Nevertheless, 1000 repetition of direct SSA takes 9.7 seconds. Hence, the speedup of tau-leaping in this example is 1.9.

Example 5.2. Consider a virus infection model from [16]. The reaction channels are listed in Table 1. In the model, reactants *genome, template, virus* and *struct* need to be simulated and nucleotide and amino acids remain constants. When $template>0$, the production and degradation of *struct*, which are marked with (*), are much faster than others. Hence, the system is stiff. One efficient technique to solve such system is nested stochastic simulation algorithm (NSSA) [7, 8]. We choose ratio of time scale separation $\epsilon=1\times10^{-5}$. The initial condition is chosen to be

$$(struct,genome,template,virus)=(0,0,1000,0).$$

In this way, the entire network is divided into fast reaction group and slow reaction group. In NSSA, direct SSA is applied to simulate the fast reaction group on a certain time period T_f. The result is then used to approximate slow reaction rates under the limit measure induced by the fast process. In this example, instead of direct SSA, we apply tau-leaping method to simulate the fast process.

To test the result of Theorem 4.1, we obtain 500 relative error of *genome* at time 1 with respect to T_f (τ frozen as 1×10^{-5}) and take an arithmetic average to approximate

Table 1: The virus infection model.

$Nucleotides \xrightarrow{a_1=1*template} genome$
$Nucleotides+genome \xrightarrow{a_2=0.025\times genome} template$
$Nucleotides+aminoacids \xrightarrow{a_3=1000\times template} struct$ (*)
$Template \xrightarrow{a_4=0.25\times template} degraded$
$struct \xrightarrow{a_5=1.9985\times struct} Degraded$ (*)
$genome+struct \xrightarrow{a_6=7.5e-6\times genome\times struct} Degraded$

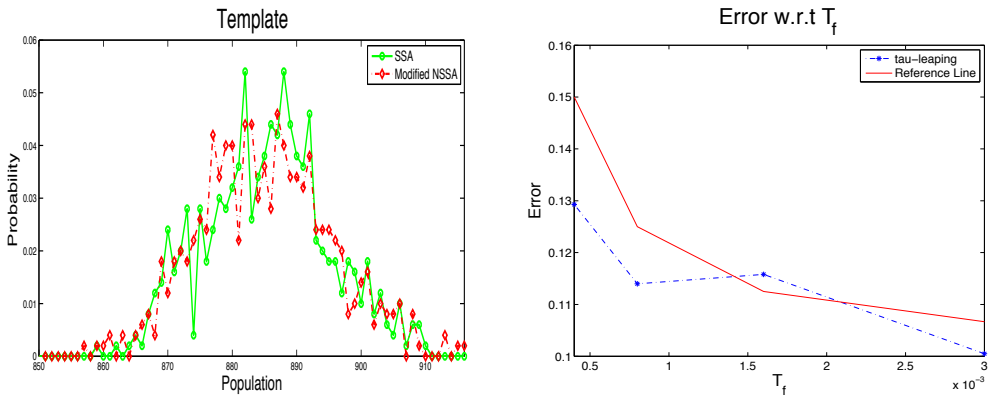

Figure 2: Left: Distribution of template at $T=0.5$ for direct SSA and modified NSSA. Right: Strong convergence error for template at $T=1$ with respect to different T_f's.

the expectation in (4.11). Strong convergence errors are reported in Fig. 2. In the figure, reference curves $error=1.6\times10^{-5}/T_f+0.1$ (Right) is plotted. Furthermore, we compare the efficiency of modified NSSA with original NSSA for the model and find that the speed up of modified NSSA is 1.3 (modified NSSA takes 1105 seconds, and original NSSA takes 850 seconds) to finish the same task above under $T_f=2\times10^{-4}$ and $\tau=5\times10^{-5}$ with relative error 30% on the distribution graph.

Example 5.3 (An Insulin Signaling Model)**.** Our group constructed a systematic mathematical model [17] for insulin signaling network mediated by IRS1 (insulin receptor substrate-1) and IRS2 (insulin receptor substrate-2), incorporating both PKR (double-stranded RNA-dependent kinase)-IRS and MAPK (mitogen-activated protein kinase) pathways, see Fig. 3. We hypothesize that the specificity of the two IRS signaling emerges from the wiring and kinetics of the entire network, and applied discrete dynamic modeling to account for many dynamic features in the system, i.e., complex feedback circuits, cross-talks between pathways, and different regulatory time-scales. Different time-scales of the model are results of the fact that IRS1 is primarily regulated by post-translational modifications such as tyrosine phosphorylation, serine phosphorylation and protein degradation, while IRS2 is regulated at the transcription level [40]. In vivo studies in mice shows that the mRNA level of IRS2 is increased during fasting and reduced after refeeding, while the mRNA level of IRS1 showed no significant change even after refeeding when IRS1 assumes a major role in response to insulin signaling [20]. Generally the regulations at the post-translation level occur quickly and are transient (limited by protein half-life) as compared with regulations at the transcription level, which is on a longer time-scale. Different regulatory time-scales of IRS1 and IRS2 is consistent with their physiological roles: a prompt response in the fed state controlled by IRS1 which is regulated at the post-translational level and the long-term response during fasting controlled by IRS2 which is regulated at the transcriptional level, see [17] for details of the model.

By studying the reaction rates at different time points, we conclude that there exists

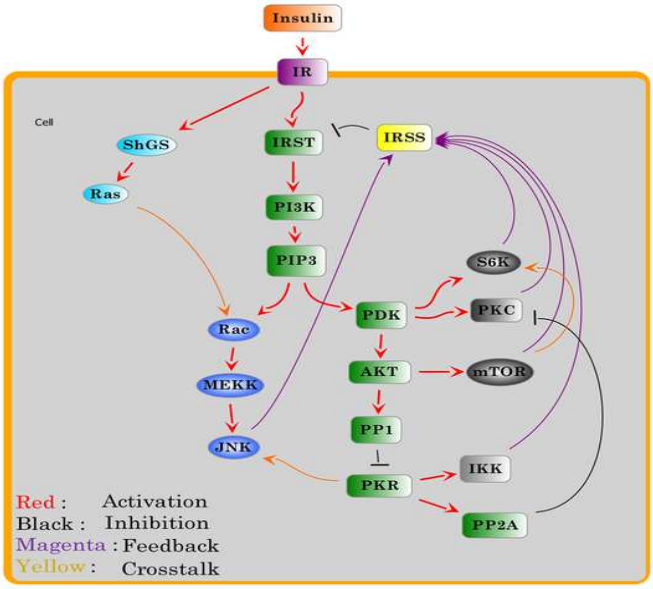

Figure 3: A schematic representation of IRS1 and IRS2 signaling network [17].

three-time scale for this model with a dynamically changing partition between slow-fast-ultra fast reactions (see Fig. 4). Hence, we propose an adaptive three-time scale NSSA for the model. At every macro-step, we dynamically divide all reaction rates into three groups with the following criteria: reactions with rate larger than 10^4 form ultra fast group; reactions with rate less than 10 form slow group, and the remainder forms fast group. As shown by Fig. 4, elements of these three groups may vary along time. Typical examples are the last two reaction channels, whose reaction rates are of magnitude 10^4 at beginning, then decays to 10^0 when the system reaches equilibrium. Hence, these two

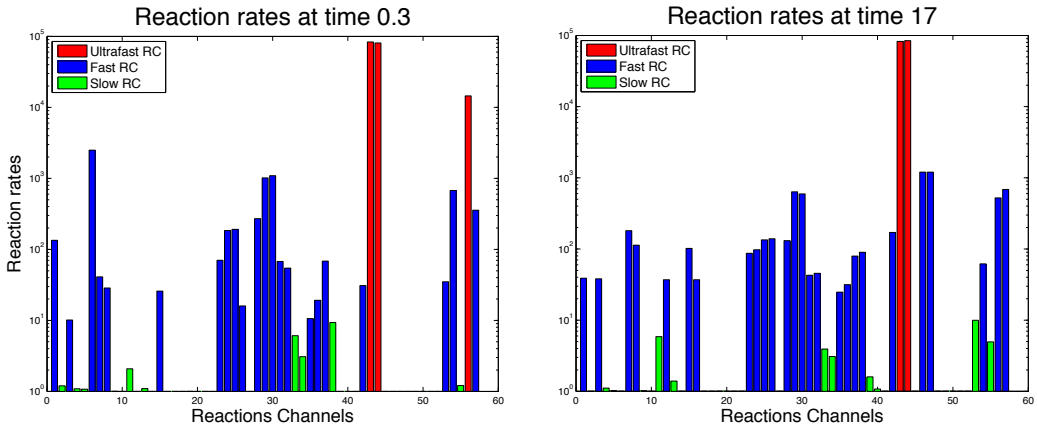

Figure 4: Insulin Model: ultrafast, fast and slow reaction groups at different time points.

reactions switch from ultrafast group to fast group even to slow group. Our algorithm is summarized as follows:

1. Ultra Inner SSA. Run SSA with ultra fast reactions $R^{uf}=(a^{uf},\nu^{uf})$ only, for a time interval T_{uf}. Update fast reaction rates:

$$\bar{a}_j^f=\frac{1}{T_{uf}}\int_{T_0}^{T_0+T_{uf}}a_j^f(x(\tau))d\tau,$$

where $x(\tau)$ is the ultra fast process at time τ.

2. Inner SSA. Run SSA with updated fast reactions $R^f=(a^f,\nu^f)$ for a time interval T_f. Update slow reaction rates:

$$\bar{a}_j^s=\frac{1}{T_f}\int_{T_0}^{T_0+T_f}a_j^s(x(\tau))d\tau,$$

where $x(\tau)$ is the fast process at time τ.

3. Outer SSA Run one step of SSA for the modified slow reactions,

$$\bar{R}^s=(\bar{a}_j^s,\nu^s)$$

to generate (t_{n+1},x_{n+1}) from (t_n,x_n). Then collect all updated reaction rates and divide them into three groups:

1) Ultra fast group: $a^{uf}\geq 10^4$;
2) Fast group: $10<a^f<10^4$;
3) Slow group: $a^s\leq 10$.

Go to 1, until a certain criterion is satisfied.

Fig. 5 shows the time evolution of a key reactant of the model for the three time-scale NSSA. Solution of the associated deterministic system is also plotted as an indicator. Table 2 illustrates efficiency of the algorithm by noting that direct SSA takes 3251 CPU

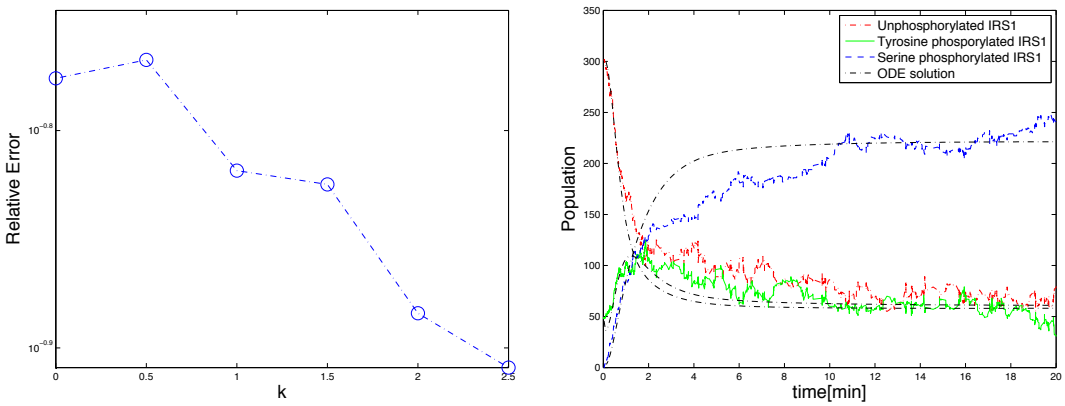

Figure 5: Left: Strong convergence of Rshp. Right: Time evolution of $IRS1$ on $[0,20]$ with $T_{uf}=10^{-4}$ and $T_f=10^{-2}$.

Table 2: Performance of NSSA for the insulin model on [0,20].

T_f	0.01			0.05		
T_{uf}	10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}
CPU time(s)	193	51	28	2623	492	189

seconds to finish the same duty. Clearly, our dynamical NSSA greatly reduces computational cost, but it keeps essential features of the model at the same time (e.g. the distribution of tyrosine phosphorylated IRS1 has an error of 28% for $T_{uf}=10^{-4}$ and $T_f=10^{-2}$ at time 12).

Example 5.4 (The Heat Shock Response (HSR) model of E. Coli)**.** The HSR is a universally evolutionary conserved defense dynamics among almost all eukaryotic and prokaryotic cells to recover protein damage induced by heat and other forms of environmental stress such as oxygen deprivation, herbicides, and ultraviolet irradiation. The process includes the transcriptional up-regulation of genes encoding heat shock proteins (HSPs) as part of the cell's internal repair mechanism. HSPs help new or misfolded proteins to fold into their correct three dimensional conformations, which is essential for their function.

In E. Coli, the transcription of such proteins is controlled by the σ^{32} factor. At normal temperature (37°C), the enzyme RNA polymerase (RNAP) is bound to another σ factor σ^{70}. Then complex RNAP: σ^{70} transcribes the genes necessary for growth under the temperature. When E. Coli are exposed to high temperatures, σ^{32} is rapidly encoded. The RNAP bound to regulatory σ^{32} recognizes the promoter and then transcribes specific genes into messenger RNA. The mRNA is translated by ribosome into protein. The ac-

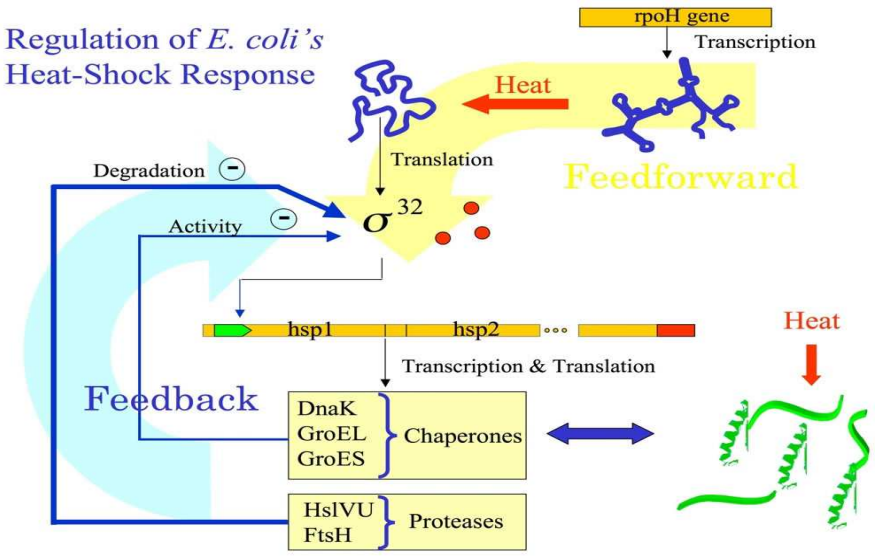

Figure 6: Reaction diagram of the heat shock response model of E. Coli [21].

tivity of σ^{32} itself is regulated through its interaction with chaperones such as Dnak and FtsH [10], see Fig. 6. For more details of the model, readers are referred to Refs. [10,11,22]. We obtain our stochastic HSR model according to the differential-algebraic models in Refs [11,22] and we assume that the heat shock occurs at time 3.

Three time-scales are prominently clear in this model. First, reaction channels resulting from binding equations are much faster than others [11], thus we consider all these reactions as ultra fast reactions; Secondly, the rest of the reaction channels can be divided into two groups. From deterministic solutions of the model, the reaction rates of folding and unfolding processes of protein are approximately 10000-fold larger than others in the group and form the fast group, and the remainder forms the slow group.

For this model, time evolution of σ^{32} and $folded\ proteins$ is essential and we plot them in Fig. 7. We choose $T_{uf}=10^{-9}$ and $T_f=2\times10^{-4}$ in our three time-scale NSSA algorithm. Since the association and disassociation rates for binding equations are not known with a great uncertainty, we need to make a guess on these numbers to produce the totally stochastic simulation. Based upon the formula for reaction parameter c_μ in [12],

$$c_\mu = V^{-1}\pi d_{12}^2(8kT/\pi m_{12})^{1/2}\exp(-u_\mu^*/kT), \tag{5.1}$$

where V is the volume of container, d_{12} is the average of diameters of two molecule, m_{12} is the reduced mass $m_1m_2/(m_1+m_2)$, k is the Boltzmann's constant, T is the absolute temperature and u_μ^* is the activation energy. Parameters that we choose are presented in Table 3. We derive that the association rate has a magnitude of $\mathcal{O}(10^{10})$.

Table 3: Parameters in (5.1).

Parameters	Values	Reference
V	$1.5\times10^{-15}L$	[7]
m_1	$24kDa$	[38]
m_2	$150kDa$	[39]
d_1	$5nM$	[5]
d_2	$14nM$	[19]
u_μ^*	-21.6	[35]

In Fig. 7, we plot both results of three time-scale Nested SSA and its modified version. In the modified NSSA, fast groups are simulated by tau-leaping method with step size 10^{-8}. The speedup ratio is 1.33 compared with that of original NSSA. However, our guess on the ultra-fast reaction rates also leads to the difference between nested SSAs and deterministic solutions though the patterns of these three methods overlap. From Fig. 7, we observe that the relative error of folded protein at equilibrium is approximately 5%. We note that an stochastic simulation result was reported [10]. But, the method is not totally stochastic as authors pointed out in their paper, "*the folding of proteins was fixed as constant disturbance which was raised to its high steady-state value at high temperature*."

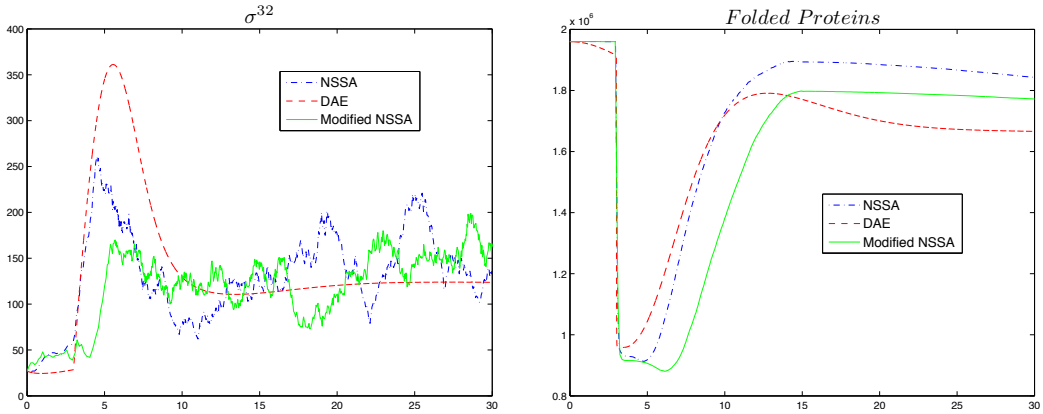

Figure 7: HSR model: time evolution of σ_t (left) and *Folded Proteins* (right) of HSR on $[0,20]$.

Acknowledgments

The research of D. Liu and C. Huang is supported by National Science Foundation DMS-0845061.

References

[1] D.F. Anderson, A. Ganguly and T.G. Kurtz, Error analysis of tau-leap simulation methods, Annal. Appl. Prob., 21 (2011), 2226-2262.
[2] R.P. Agarwal, Difference equations and inequalities, Theory, Methods, and Applications, Marcel Dekker, Inc., New York, 1992.
[3] S. Albeverio, Z. Brzezniak, and J.-L. Wu, Existence of global solutions and invariant measures for stochastic differential equations driven by Poisson type noise with non-Lipschitz coefficients, J. Math. Phys. Appl., 371 (2010), 309-322.
[4] D. Applebaum, Lévy processes and stochastic calculus", Cambridge University Press, 2009.
[5] `http://bionumbers.hms.harvard.edu`.
[6] Y. Cao, D. Gillespie and L. Petzold, The slow-scale stochastic simulation algorithm, J. Chem. Phys., 122(1) (2005), 014116.
[7] W. E, D. Liu, and E. Vanden-Eijnden, Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates, J. Chem. Phys., 123 (2005), 194107.
[8] W. E, D. Liu, and E. Vanden-Eijnden, Nested stochastic simulation algorithm for chemical kinetic systems with multiple time scales, J. Comp. Phys., 221 (2007), 158-180.
[9] W. E, D. Liu, and E. Vanden-Eijnden, Analysis of multiscale methods for stochastic differential equations, Comm. Pure. Appl. Math., 58 (2005), 1544-1585.
[10] H. El-Samad, M. Khammash, H. Kurata, and J.C. Doyle, Feedback regulation of the heat shock response in E. coli, Multidiciplinary Research in Control, LNCIS 289 (2003), 115-128.
[11] H. El-Samad, H. Kurata, J.C. Doyle, C.A. Gross, and M. Khammash, Surviving heat shock: Control strategies for robustness and performance, PNAS, 102 (2004), 2736-2741.
[12] D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comp. Phys., 22 (1976), 403-434.

[13] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Chem. Phys., 81 (1977), 2340-2361.
[14] D.T. Gillespie, Approximate accelerated simulation of chemically reaction systems, J. Chem. Phys., 115 (2001), 1716-1733.
[15] F.B. Hanson, Applied Stochastic Processes and Control for Jump-Diffusions, SIAM, Philadelphia, 2007.
[16] E.L. Haseltine and J.B. Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic kinetics, J. Chem. Phys., 117 (2002), 6959-6969.
[17] C. Huang, M. Wu, D. Liu and C. Chan, Systematic modeling for insulin signaling network mediated by IRS1 and IRS2, to appear.
[18] R. Khasminskii, Stochastic Stability of Differential Equations, Netherlands, Sijthoff & Noordhoff, 1980.
[19] Y. Kitano and T. Kameyama, Molecular structure of RNA polymerase and its complex with DNA, J. Biochem., 65 (1969), 1-16.
[20] N. Kubota, T. Kubota, S. Itoh, H. Kumagai, H. Kozono, I. Takamoto, T. Mineyama, H. Ogata, K. Tokuyama, M. Ohsugi, T. Sasako, M. Moroi, K. Sugi, S. Kakuta, Y. Iwakura, T. Noda, S. Ohnishi, and R. Nagai, Dynamic functional relay between insulin receptor substrate 1 and 2 in hepatic insulin signaling during fasting and feeding, Cell Mepsilonb., 8 (2008), 49-64.
[21] H. Kurata, H. El-Samad, T.M. Yi, M. Khammash, and J.C. Doyle, Feedback regulation of the heat shock resonse in E. coli, Proceedings of the 40th IEEE Conference on Decision and Control, Orlando, 2001.
[22] H. Kurata, H. El-Samad, R. Iwasaki, H. Ohtake, J.C. Doyle, I. Grigorova, C.A. Gross and M. Khammash, Module-based analysis of robustness tradeoffs in the heat shock response system, PLOS Computational Biology, 2 (2006), 663-675.
[23] T. Li, Analysis of explicit tau-leaping schemes for simulating chemically reacting systems, Multiscale Model. Simul., 6 (2007), 417-436.
[24] D. Liu, Analysis of multiscale methods for stochastic dynamical systems with multiple time scales, SIAM Multiscale Model. Simul., 8 (2010), 944-964.
[25] J.C. Mattingly, A.M. Stuart and M.V. tretyakov, Convergence of numerical time-averaging and stationary measures via Poisson equations, SIAM J. Numer. Anal., 48 (2010), 552-577.
[26] J.C. Mattingly, A.M. Stuart, and D.J. Higham, Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise, Stochastic Processes and Their Applications, 101 (2002), 185-232.
[27] S.P. Meyn and R.L. Tweedie, Stability of Markovian processes III. Foster-Lyapunov criteria for continuous-time processes, Adv. in Appl. Probab., 25 (1993), 518-548.
[28] K.S. Miller, An Introduction to the Calculus of Finite Differences and Difference Equations, Henry Holt and Company, New York, 1960.
[29] G.N. Milstein and M.V. Tretyakov, Computing ergodic limits for Langevin equations, Physica D, 229 (2007), 81-95.
[30] G.N. Milstein and M.V. Tretyakov, Numerical integration of stochastic differential equations with non-globally Lipschitz coefficients, SIAM J. Numer. Anal., 43 (2005), 1139-1154.
[31] B. Oksendal and A. Sulem, Applied Stochastic Control of Jump Diffusions, Springer, 2004.
[32] F. Panloup, Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process, Annal. Appl. Prob., 18 (2008), 379-426.
[33] R. Pitas, T. InnerarIity, K. Arnold, and R. Mahley, Rate and equilibrium constants for binding of $apo\text{-}EHDL_C$ (a cholesterol-induced lipoprotein) and low density lipoproteins to human fibroblasts: Evidence for multiple receptor binding of $apo\text{-}EHDL_C$, Proc. Natl. Acad. Sci.

USA (Cell Biology), 76 (1979), 2311-2315.
[34] G. Da Prato and J. Zabczyk, Ergodicity for Infinite Dimensional Systems, Cambridge University Press, 1996.
[35] M.A. Shea and G. K. Ackersf, The O_R control system of bacteriophage Lambda A physical-chemical model for gene regulation, J. Mol. Biol., 181 (1985), 211-230.
[36] D. Talay, Second-order discretization schemes of stochastic differential systems for the computation of the invariant law, Stoch. Stoch. Reports, 29 (1990), 13-36.
[37] P. Todorovic, An Introduction to Stochastic Processes and Their Applications, Springer-Verlag, 1992.
[38] Q. Wang and J. M. Kaguni, A novel sigma factor is involved in expression of the rpoH gene of Escherichia coli, J. Bacteriol., 171 (1989), 4248-4253.
[39] `http://www.worthington-biochem.com/rnap/default.html`
[40] X. Yang, A. Nath, M.J. Opperman, and C. Chan, The double-stranded RNA-dependent protein kinase differentially regulates insulin receptor substrates 1 and 2 in HepG2 cells, Mol. Biol. Cell, 21 (2010), 3449-3458.