Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

stochastic processes and their applications

www.elsevier.com/locate/spa

Large deviations of Markov chains with multiple time-scales

Lea Popovic

Department of Mathematics and Statistics, Canada

Received 12 April 2018; received in revised form 28 August 2018; accepted 7 September 2018
Available online xxxx

Abstract

For Markov processes evolving on multiple time-scales a combination of large component scalings and averaging of rapid fluctuations can lead to useful limits for model approximation. A general approach to proving a law of large numbers to a deterministic limit and a central limit theorem around it have already been proven in Kang and Kurtz (2013) and Kang et al. (2014). We present here a general approach to proving a large deviation principle in path space for such multi-scale Markov processes. Motivated by models arising in systems biology, we apply these large deviation results to general chemical reaction systems which exhibit multiple time-scales, and provide explicit calculations for several relevant examples.

MSC: 60F10; 60J75; 92C45; 92C37; 80A30; 60F17; 60J27; 60J28; 60J60

Keywords: Large deviation principle; Multiple time-scales; Reaction networks; Markov chains; Jump diffusions; Piecewise deterministic Markov process; Comparison principle

1. Introduction

In recent years, continuous-time Markov chain models have found extensive use in systems biology. The complexity of the models introduced has led to interest in a variety of model reduction techniques. Some of these techniques result in what are effectively laws of large numbers giving approximations of the model or subsets of the model by systems of ordinary differential equations. Corresponding central limit theorems for the deviations of the stochastic

E-mail address: lpopovic@mathstat.concordia.ca.
URL: http://www.mathstat.concordia.ca/faculty/lpopovic.

https://doi.org/10.1016/j.spa.2018.09.009

model from the approximating ordinary differential equation have also been given. In addition to the laws of large numbers and central limit theorems it is both natural and of some biological interest to consider the corresponding large deviation behaviour of these models.

Models with what we will refer to as the "classical" scaling fit naturally into classical large deviation results going back to Wentzell [27], and we will review these briefly; however, our primary interest is in models with multiple time-scales. These models arise from non-standard scalings of Markov chains with density-dependent rates. Dependence of the transition rates on all variables implies a full coupling of, and an interaction between all the components. We consider arbitrary scalings of Markov chains that lead to dynamics on two dominant time-scales: a fast one — on which rapid fluctuations for a subset of components leads to geometric ergodicity; and a slow one — on which the remaining subset of components converge to a solution of a system of ordinary differential equations.

Perhaps the simplest example in the realm of chemical reactions is a model of enzyme kinetics

$$S + E \underset{\kappa_2'}{\overset{\kappa_1'}{\rightleftharpoons}} ES \overset{\kappa_3'}{\rightharpoonup} P + E, \tag{1.1}$$

where S is the substrate, E the enzyme, ES the enzyme–substrate complex, and P the product. Under appropriate scaling of the parameters, we can write the model as the solution of the system (see (5.1) for the generator of the process)

$$\begin{aligned}
Z_1^N(t) &= Z_1^N(0) - N^{-1}Y_1(N\int_0^t \kappa_1 Z_1^N(s)Z_2^N(s)ds) + N^{-1}Y_2(N\int_0^t \kappa_2 Z_3^N(s)ds)\\
Z_2^N(t) &= Z_2^N(0) - Y_1(N\int_0^t \kappa_1 Z_1^N(s)Z_2^N(s)ds) + Y_2(N\int_0^t \kappa_2 Z_3^N(s)ds)\\
&\quad + Y_3(N\int_0^t \kappa_3 Z_3^N(s)ds)\\
Z_3^N(t) &= Z_2^N(0) + Y_1(N\int_0^t \kappa_1 Z_1^N(s)Z_2^N(s)ds) - Y_2(N\int_0^t \kappa_2 Z_3^N(s)ds)\\
&\quad - Y_3(N\int_0^t \kappa_3 Z_3^N(s)ds)\\
Z_4^N(t) &= N^{-1}Y_3(N\int_0^t \kappa_3 Z_3^N(s)ds),
\end{aligned}$$

where Y_1, Y_2, Y_3 are independent unit Poisson processes, and Z_1^N, Z_2^N, Z_3^N, Z_4^N are the scaled amounts of substrate, free enzyme, enzyme–substrate complex, and product, respectively. Note that $M \equiv Z_2^N(t) + Z_3^N(t)$ is constant in time and we will also assume independent of the scaling parameter N. The amount of substrate is an order of magnitude larger than the amount of enzymes, and hence assumed to be proportional to the scaling parameter N. Due to the relatively small fluctuations of the scaled amount of substrate the process Z_1^N can be approximated by a deterministic one. The law of large numbers for this system goes back to Darden [9] and is derived from the above system of equations in [20]. Specifically, it is shown that as $N \to \infty$, Z_1^N converges to the solution of

$$\dot{x}(t) = -\frac{M\kappa_1\kappa_3 x(t)}{\kappa_2 + \kappa_3 + \kappa_1 x(t)}, \tag{1.2}$$

which, of course, is simply the Michaelis–Menten equation. The corresponding central limit theorem for the scaled deviations $N^{1/3}(Z_1^N - x(t))$ is given in [21].

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

A less straightforward example is a model of packaged virus particle production

$$\text{stuff} \xrightarrow{\kappa_1'} G, \quad G \xrightarrow{\kappa_2'} T, \quad T + \text{stuff} \xrightarrow{\kappa_3'} T + S \tag{1.3}$$

$$T \xrightarrow{\kappa_4'} \emptyset, \quad S \xrightarrow{\kappa_5'} \emptyset, \quad G + T + (S) \xrightarrow{\kappa_6'} V,$$

where T is the viral template, G the viral genome, S the viral structural protein that uses up resources from the cell, and V is the pre-packaged material necessary for further proliferation of the virus in another cell (the structural protein is packaged, but it affects the packaging rate only in its order of magnitude). Under the appropriate scaling of the component amounts and chemical rate constants we can write the model as the solution of (see (5.2) for the generator of the process)

$$\begin{aligned}
Z_1^N(t) &= Z_1^N(0) + Y_2\Big(\int_0^t N^{2/3}\kappa_2 Z_2^N(s)ds\Big) - Y_4\Big(\int_0^t N^{2/3}\kappa_4 Z_1^N(s)ds\Big) \\
&\quad - Y_6\Big(\int_0^t N^{2/3}\kappa_6 Z_1^N(s)Z_2^N(s)ds\Big) \\
Z_2^N(t) &= Z_2^N(0) + N^{-2/3}Y_1\Big(\int_0^t N^{2/3}\kappa_1 ds\Big) - N^{-2/3}Y_2\Big(\int_0^t N^{2/3}\kappa_2 Z_2^N(s)ds\Big) \\
&\quad - N^{-2/3}Y_6\Big(\int_0^t N^{2/3}\kappa_6 Z_1^N(s)Z_2^N(s)ds\Big) \\
Z_3^N(t) &= Z_3^N(0) + N^{-1}Y_3\Big(\int_0^t N^{5/3}\kappa_3 Z_1^N(s)ds\Big) - N^{-1}Y_5\Big(\int_0^t N^{5/3}\kappa_5 Z_3^N(s)ds\Big) \\
&\quad - N^{-1}Y_6\Big(\int_0^t N^{2/3}\kappa_6 Z_1^N(s)Z_2^N(s)ds\Big) \\
Z_4^N(t) &= Z_4^N(0) + N^{-2/3}Y_6\Big(\int_0^t N^{2/3}\kappa_6 Z_1^N(s)Z_2^N(s)ds\Big),
\end{aligned}$$

where Y_k, $k = 1, \ldots, 6$ are independent unit Poisson processes, and Z_1^N, Z_2^N, Z_3^N, Z_4^N are the scaled amounts of template, genome, structural protein, and viral package, respectively. The fast fluctuating components are essentially evolving as a piecewise deterministic Markov process as defined by [11] with Z_1 a discrete component and Z_3 continuous. Note that the scaled amount of template and structural protein have large fluctuations relative to their amounts, while the fluctuations of the scaled amount of genome are relatively small so the process Z_2^N can be approximated by a deterministic one. The law of large numbers, obtained for the above system of equations by adapting the results in [4], shows that as $N \to \infty$, Z_2^N converges to a solution of

$$\dot{x}(t) = \kappa_1 - \kappa_2 x(t) - \kappa_6 \frac{\kappa_3}{\kappa_5} \frac{\kappa_2 x(t)}{\kappa_4 + \kappa_6 x(t)} x(t), \tag{1.4}$$

and the scaled deviations $N^{1/3}(Z_2^N - x(t))$, as can be shown by adapting the example in [21], converge to a Gaussian process. In addition to two other examples, we provide large deviations for the enzyme kinetics and viral production models.

Our results allow a great deal of generality for the original Markov process, requiring only that it satisfies necessary technical assumptions on: the existence and uniqueness of limiting processes on both time-scales and some control on their exponential growth; geometric ergodicity of the occupation measure for the rapidly fluctuating subset of components, and uniqueness for the limiting exponential operator of the remaining subset of components. They

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

are general enough to allow the original Markov process to be a multi-scale jump–diffusion with density dependent (non-Lévy) jump measure. Earlier results for such processes with a Lévy measure driving the jump terms were given in [23]. We use the same proof methodology, which relies on the general method for Markov processes developed in [17] based on non-linear semigroups and viscosity methods, and a generalization of Barles and Perthame limit arguments for PDEs given in [18].

Part of the motivation for this work was to develop results that can be useful in the context of modelling chemical reaction networks. Results on large deviations for such models on a single time scale can be found in the recent work of [1,2]. An application of the use of large deviation results within a basic model of enzyme catalysis can be found in [10]. Some results for Markov chain models of chemical kinetics on two (well-separated) time-scales were recently proved in [25], using different techniques based on the approximation and change-of-measure approach. Our results fully cover the extent of their conclusions and further extend them to more general reaction systems with two time-scale effective dynamics. In particular, our results allow the effective dynamics of the fast fluctuating component to be a combination of discrete and continuous variables fully linked by piecewise deterministic Markovian dynamics (PDMP) as defined in [11]. Most importantly our results do not assume the fast fluctuating component in the original Markov process to be limited to a finite or bounded state space. This is an assumption that has so far been assumed for all large deviation results on chemical reaction systems in the literature. The generality of the exponential weak convergence approach combined with the power of the viscosity solution technique allows all components to live in a non-compact subset of $\mathbb{R}^d$.

Our large deviation principle (LDP) for general multi-scale Markov chains is given in Theorem 3.6. Our other goal was to verify its conditions for multi-scale chemical reaction systems. Propositions 4.4 and 4.6 give a way to verify two technically challenging conditions: exponential compact containment, and existence of a solution to an eigenvalue problem, respectively. We also illustrate how they apply to give the LDP in the discussed examples of enzymatic kinetics and viral production, as well as in two others.

Outline. Section 2 contains the terminology for large deviations and the relevant general tools. Section 3 specifies a sequence of Conditions 3.1–3.4 that need to be verified and the statements of the large deviation Theorem 3.6 and its Corollary 3.8. Section 4 identifies specific aspects of the reaction network context that allows one to verify (or relax) the needed conditions for multi-scale chemical reaction systems. Section 5 provides several examples of biologically relevant reaction systems and explicitly verifies the conditions and obtains the LDP result. The Appendix contains proofs of Theorem 3.6, Corollary 3.8, and Lemma 3.7.

2. Large deviations

For details and proofs regarding the following discussion see [17]. Let (S, d) be a metric space. For $N = 1, 2, \ldots$, let X^N be a S-valued random variable. $\{X^N\}$ satisfies the *large deviation principle* with *rate function* I if for each open set A

$$\liminf_{N\to\infty} \frac{1}{N} \log P\{X^N \in A\} \geq -\inf_{x\in A} I(x), \tag{2.1}$$

and for each closed set B

$$\limsup_{N\to\infty} \frac{1}{N} \log P\{X^N \in B\} \leq -\inf_{x\in B} I(x). \tag{2.2}$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Using lower semicontinuous functions I, that is $\{x : I(x) \leq c\}$ is closed $\forall c \in \mathbb{R}$, this definition is equal to the requirement that

$$-I(x) = \lim_{\epsilon \to 0} \liminf_{N \to \infty} \frac{1}{N} \log P\{X^N \in B_\epsilon(x)\} = \lim_{\epsilon \to 0} \limsup_{N \to \infty} \frac{1}{N} \log P\{X^N \in \overline{B}_\epsilon(x)\}. \quad (2.3)$$

Typically, $\{x : I(x) \leq c\}$ is compact $\forall c \in \mathbb{R}$, and then I is called a *good rate function.*

The notion of *exponential tightness* plays the same role in large deviation theory that *tightness* plays in the theory of weak convergence.

Definition 2.1 (*Exponential Tightness*)**.** A sequence of probability measures $\{\mu_N\}$ on S is *exponentially tight* if for each $a > 0$, there exists a compact set $K_a \subset S$ such that

$$\limsup_{N \to \infty} \frac{1}{N} \log \mu_N(K_a^c) \leq -a.$$

A sequence $\{X^N\}$ of S-valued random variables is exponentially tight if the corresponding sequence of distributions is exponentially tight.

The approach we will take to proving our large deviation results is based on the following theorem of Varadhan and Bryc (see [17] Proposition 3.8 or [12] Theorem 4.3.1 and 4.4.2). Let $C_b(S)$ be the space of all continuous bounded functions on S.

Theorem 2.2. *Let $\{X^N\}$ be a sequence of S-valued random variables.*

(a) *(Varadhan Lemma) Suppose that $\{X^N\}$ satisfies the large deviation principle with a good rate function I. Then for each $f \in C_b(S)$,*

$$\lim_{N \to \infty} \frac{1}{N} \log E[e^{Nf(X^N)}] = \sup_{x \in S} \{f(x) - I(x)\}. \quad (2.4)$$

(b) *(Bryc formula) Suppose that the sequence $\{X^N\}$ is exponentially tight and that the limit*

$$\Lambda(f) = \lim_{N \to \infty} \frac{1}{N} \log E[e^{Nf(X^N)}] \quad (2.5)$$

exists for each $f \in C_b(S)$. Then $\{X^N\}$ satisfies the large deviation principle with good rate function

$$I(x) = \sup_{f \in C_b(S)} \{f(x) - \Lambda(f)\}. \quad (2.6)$$

We are interested in time-homogeneous Markov processes $\{X^N(t)\}_{t \geq 0}$ which will have sample paths in the Skorohod space $S = D_E[0, \infty)$. Assuming the limit (2.5) exists for sufficiently many functions f, we can apply Theorem 2.2(b) to

$$\Lambda_t(f, x) = \lim_{N \to \infty} \frac{1}{N} \log E[e^{Nf(X^N(t))} | X^N(0) = x], \quad (2.7)$$

to obtain the large deviation principle for the one dimensional distributions. But if we can show exponential tightness for the distributions of $\{X^N\}_{t \geq 0}$ on $D_E[0, \infty)$, then the Markov property gives the large deviation principle for the finite dimensional distributions which in turn gives the large deviation principle for the processes in $D_E[0, \infty)$, see Chapter 5 of [17] Theorem 4.28.

Suppose X^N is a Markov processes with generator A_N. Define

$$V_N(t) f(x) = \frac{1}{N} \log E[e^{Nf(X^N(t))} | X^N(0) = x].$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Then by the Markov property, $\{V_N(t)\}_{t\geq 0}$ is a nonlinear contraction semigroup, that is,

$$V_N(t+s)f(x) = V_N(t)V_N(s)f(x), \quad s,t \geq 0$$

and

$$\sup_x |V_N(t)f_1(x) - V_N(t)f_2(x)| \leq \sup_x |f_1(x) - f_2(x)|,$$

and we can define a nonlinear *exponential generator* by

$$H_N f(x) = \lim_{t\to 0} \frac{1}{t}(V_N(t)f(x) - f(x)) = \frac{1}{N} e^{-Nf(x)} A_N e^{Nf}(x),$$

provided e^{Nf} is in $\mathcal{D}(A_N)$. Since it is H_N that we typically know how to compute explicitly, it is natural to ask for conditions on the sequence of generators $\{H_N\}$ that imply convergence of $\{V_N\}$. Observe that (2.7) is just the convergence of the semigroup V_N. We define the "limit" as $N \to \infty$ of the sequence H_N to be the set of $\{(f, g_*, g^*) \in C_b(E) \times B(E) \times B(E)\}$ for which there exists $f_N \in \mathcal{D}(H_N)$ such that for $x_N \in E$ satisfying $x_N \to x$, $f_N(x_N) \to f(x)$ and

$$g_*(x) \leq \liminf_{n\to\infty} H_N f_N(x_N) \leq \limsup_{n\to\infty} H_N f_N(x_N) \leq g^*(x).$$

With the two examples from the introduction in mind, let us separate component-wise the notation for the multi-scale Markov process $Z^N = (X^N, Y^N)$ so that X^N satisfies a law of large numbers while Y^N has fast ergodic fluctuations. There are several complications to overcome. Since it is only $\{X^N(t)\}_{t\geq 0}$ that converges to a deterministic limit, we are really only interested in the large deviation behaviour for that sequence. Also, since the fluctuations of $\{Y^N(t)\}_{t\geq 0}$ average out, the "limit" of the sequence $\{H_N\}_{N\to\infty}$ will typically be a multi-valued operator. One way to deal with identifying this limit is to select the functions $f_N(x, y) = f_0(x) + \frac{1}{N} f_1(x, y)$ in such a way that $\lim_{N\to\infty} H_N f_N(x, y) = g(x)$ for some function g that is independent of y. For geometrically ergodic processes Y^N this can typically be accomplished by solving an eigenvalue problem based on a perturbed operator for Y^N. However, technical challenges still remain in order to prove existence and uniqueness of the limiting semi-group by verifying the "range condition" for the limiting non-linear operator.

For the processes in this paper, the state space of Z^N will always be a subset of a Euclidean space $E^N \subseteq \mathbb{R}^d$ which converges, in the sense that $E^N \subseteq E = E_X \times E_Y \subseteq \mathbb{R}^d$ is asymptotically dense in $\mathbb{R}^d$ so that for each compact $K \subset \mathbb{R}^d$,

$$\lim_{N\to\infty} \sup_{(x,y)\in E\cap K} \inf_{(x_N,y_N)\in E^N} |(x, y) - (x_N, y_N)| = 0.$$

This fact allows us to approach the problem of convergence of generators $\{H_N\}$ and semi-groups $\{V_N\}$ by using the sequence of viscosity solutions of the associated Cauchy problems. Namely, for each $h \in C_b(E_X)$, the function

$$u_N^h(t, x, y) := V_N(t)h(x) = \frac{1}{N} \log E[e^{Nh(X^N(t))} | (X^N, Y^N)(0) = (x, y)]$$

satisfies the non-linear partial integro-differential equation

$$\begin{aligned} \partial_t u_N(t, x, y) &= H_N u_N(t, x, y), \text{ in } (0, T] \times E_X \times E_Y; \\ u_N(0, x) &= h(x), \text{ for } (x, y) \in E_X \times E_Y. \end{aligned} \tag{2.8}$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

The goal is to show that viscosity solutions of (2.8) converge to a viscosity solution $u_0^h(t,x)$ of the limiting equation

$$\begin{aligned} \partial_t u_0(t,x) &= \overline{H}_0 u_0(t,x), \text{ in } (0,T]\times E_X; \\ u_0(0,x) &= h(x), \text{ for } x\in E_X, \end{aligned} \tag{2.9}$$

where the non-linear operator $\overline{H}_0$ is to be identified from the limit of the non-linear generators $\{H_N\}$. In the viscosity method, existence will follow by construction, while uniqueness will be obtained via the comparison principle. Thus, to show existence and uniqueness of the semi-group limit (2.7) using this technique, one only needs to verify the convergence of u_N^h to u_0^h for sufficiently many initial value functions h, and to check the comparison principle for the limiting Cauchy problem (2.9).

Convergence of u_N^h to u_0^h can be proved based on a general construction of subsolutions and supersolutions to two families of operators: $\{H_0(\cdot;\alpha)\}_{\alpha\in\Lambda}$ and $\{H_1(\cdot;\alpha)\}_{\alpha\in\Lambda}$, which are meant to "sandwich" the limiting operator $\overline{H}_0$ (see [18]). A comparison principle between viscosity subsolutions of $\inf_{\alpha\in\Lambda}\{H_0(\cdot;\alpha)\}$ and viscosity supersolutions of $\sup_{\alpha\in\Lambda}\{H_0(\cdot;\alpha)\}$ in conjunction with the "operator inequality" between $\inf_{\alpha\in\Lambda}\{H_0(\cdot;\alpha)\}$ and $\sup_{\alpha\in\Lambda}\{H_0(\cdot;\alpha)\}$ will imply the desired convergence. The proof of uniqueness of the solution u_0 to the limiting problem then has to be shown by the weak comparison principle for the Cauchy problem (2.9). Knowledge of the eigenvalue characterization of $\overline{H}_0$ in the limiting problem (2.9) can be used in this process. The remainder of the proof of the large deviations result comes from showing exponential tightness of $\{X^N\}$ (Definition 2.1) and using Bryc formula (Theorem 2.2(b)).

3. Markov processes on multiple time-scales

The Markov processes we are interested in have generators of the form

$$Af(z) = \sum_k \lambda_k(z)(f(z+\zeta_k) - f(z)), \tag{3.1}$$

with k indexing the different jumps of size ζ_k which occur at density-dependent rates $\lambda_k(z)$. We assume the rates are non-negative, locally Lipschitz and locally bounded. Although, we could have allowed the jump sizes to be state dependent as well, for simplicity we let ζ_k be constant. We use powers of a parameter N to scale individual component sizes $Z_i^N = N^{-\alpha_i} Z_i$ and scale density-dependent jump rates $\lambda_k(z) = N^{\beta_k}\lambda_k^N(z_N)$ leading to generators of the form, for $f\in\mathcal{D}(A_N)\subset C(E^N)$,

$$A_N f(z) = \sum_k N^{\beta_k}\lambda_k^N(z)(f(z+N^{-\underline{\alpha}}\zeta_k^N) - f(z)), \tag{3.2}$$

where $N^{-\underline{\alpha}}$ is the diagonal matrix with entries $N^{-\alpha_i}$. The nonlinear generator has the form, for $e^{Nf}\in\mathcal{D}(A_N)$,

$$H_N f(z) = \frac{1}{N}\sum_k N^{\beta_k}\lambda_k^N(z)(e^{N(f(z+N^{-\underline{\alpha}}\zeta_k^N)-f(z))} - 1). \tag{3.3}$$

3.1. Model assumptions

Separating the components into $Z^N = (X^N, Y^N)$ we need to make some assumptions on the dynamics of the rescaled Markov process on two time-scales. We start with a general look at the needed conditions.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Let L_0, L_1 be the linear operators defined on $\mathcal{D}(L_0) = C_c^2(E_X)$, $\mathcal{D}(L_1) = C_c^2(E_X \times E_Y)$ respectively, given by

$$L_0 f(x) = \sum_k \overline{\lambda}_k(x)\widetilde{\zeta}_k^X \cdot \nabla_x f(x), \tag{3.4}$$

$$L_1 f(z) = \sum_{k:\beta_k=1} \widetilde{\lambda}_k(z)(f(x, y + \widetilde{\zeta}_k^Y) - f(x, y)) + \sum_{k:\beta_k>1} \widetilde{\lambda}_k(z)\widetilde{\zeta}_k^Y \cdot \nabla_y f(x, y), \tag{3.5}$$

with non-negative, locally Lipschitz, locally bounded functions $\overline{\lambda}_k$ and $\widetilde{\lambda}_k$, and with $\mathbb{R}^{|E_X|}$-valued and $\mathbb{R}^{|E_Y|}$-valued vectors $\widetilde{\zeta}_k^X$ and $\widetilde{\zeta}_k^Y$ respectively. Here $\overline{\lambda}_k$, $\widetilde{\lambda}_k$ and $\widetilde{\zeta}_k^X$, $\widetilde{\zeta}_k^Y$ are determined by the jump rates $\lambda_k^N(z)$ and sizes ζ_k^N from (3.2) by the convergence conditions we impose next. Suppose the rescaled Markov process Z^N satisfies the following.

Condition 3.0 (*General Conditions*)**.** *The generator A_N given by (3.2) satisfies*

$$\lim_{N\to\infty} \sup_{z\in E^N} |A_N f(z) - L_0 f(z)| = 0, \quad \forall f \in \mathcal{D}(L_0)$$

$$\lim_{N\to\infty} \sup_{z\in E^N} |\frac{1}{N} A_N f(z) - L_1 f(z)| = 0, \quad \forall f \in \mathcal{D}(L_1)$$

For some generator $\overline{H}_0$ on $C_c^2(E_X)$ the exponential generator H_N given by (3.3) satisfies

$$\lim_{N\to\infty} \sup_{z\in E^N} |H_N f_N(z) - \overline{H}_0 f(z)| = 0,$$

$\forall f \in \mathcal{D}(\overline{H}_0)$ *and* $e^{Nf_N} \in \mathcal{D}(A_N)$ *chosen so that* $\lim_{N\to\infty}\sup_{z\in E^N} |f_N(z) - f(z)| = 0$.

The first convergence condition essentially insures that the slow component X^N has a deterministic limit. The second convergence condition insures that the fluctuations of the fast component Y^N on the time-scale tN have a limit that is either deterministic, a Markov chain or a piecewise-deterministic Markov chain. The latter is a Markov process with a discrete and a continuous component, where the discrete component jumps while the continuous component follows deterministic dynamics, with the jump rate and the deterministic flow determined from the value of both the discrete and the continuous component (see [11]). These two convergence conditions describe a separation of time scales which is a common occurrence in stochastic models of intracellular dynamics: fast fluctuations on the time scale tN are described by a limiting generator L_1 and slow fluctuations on the time scale t are described by L_0 in the limit.

We next provide a sequence of explicit conditions for a large class of rescaled Markov chains with effective dynamics on two time-scales, and show that the above general conditions hold.

Condition 3.1 (*Scaling Parameters*)**.** *Let $\mathcal{I}_X$ and $\mathcal{I}_Y$ denote indices of components belonging to X^N and Y^N respectively. Recalling the scaling parameters α_i of the components Z_i^N and scaling parameters β_k of the jump rates λ_k^N, let $\beta(i) = \max_k\{\beta_k : \zeta_{ik}^N \neq 0\}$ be the maximal jump rate that affects the component $i \in \mathcal{I}_X \cup \mathcal{I}_Y$. Then*

$$\forall i \in \mathcal{I}_X : \ \alpha_i \geq 1 \ \text{ and } \ \beta(i) \leq \alpha_i,$$

$$\forall j \in \mathcal{I}_Y : \ \alpha_j \geq 0 \ \text{ and } \ \beta(j) \leq 1 + \alpha_j.$$

There is at least one $i \in \mathcal{I}_X$ with $\alpha_i = 1$ and at least one $j \in \mathcal{I}_Y$ with $\beta(j) = 1 + \alpha_j$.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

The first condition above requires that all the components of X^N are of size at least N, and that jumps that change their amounts occur at rate at most N. The second condition insures that the separation of time-scales between fluctuations of X^N and Y^N is at least of order N. The last condition implies that our choice of the scaling parameter N reflects both the smallest size of the slow components and the largest separation to the time-scale of fast fluctuations. Under this condition on the scaling parameter it is reasonable to consider a large deviations result at "speed" N.

We can describe the *effective dynamics* on each time-scale if we consider the *effective change* $(\widetilde{\zeta}^X, \widetilde{\zeta}^Y)$ to each component due to jumps of the process. For each $i \in \mathcal{I}_X$ let

$$\widetilde{\zeta}^X_{ik} = \zeta^N_{ik} \text{ if } \beta_k = \alpha_i \text{ and } \widetilde{\zeta}^X_{ik} = 0 \text{ if } \beta_k < \alpha_i, \tag{3.6}$$

and for each $j \in \mathcal{I}_Y$ let

$$\widetilde{\zeta}^Y_{jk} = \zeta^N_{jk} \text{ if } \beta_k = 1 + \alpha_j \text{ and } \widetilde{\zeta}^Y_{jk} = 0 \text{ if } \beta_k < 1 + \alpha_j. \tag{3.7}$$

Likewise we let $\widetilde{\lambda}_k(z) = \lim_{N\to\infty} \lambda^N_k(z)$ denote the *effective rates* of change, though in many instances we will simply have $\widetilde{\lambda}_k = \lambda_k$. Using effective changes and rates we will see that Conditions 3.1, together with subsequent Conditions 3.3–3.4 on the behaviour of Y^N, will imply the convergence of the sequence of generators A_N from (3.2) to the generator L_0 given in (3.4) operating on the slow variables only, as well as the convergence of the sequence of scaled versions of these generators $\frac{1}{N}A_N$ to a generator L_1 given in (3.5). Stochastic averaging will provide the *averaged* effective rates in L_0 to be $\overline{\lambda}_k(x) = \int \widetilde{\lambda}_k(z)\pi^{x,0}(dy)$, where $\pi^{x,0}(\cdot)$ is the stationary distribution for the process $Y^{x,0}$ with generator $L^{x,0}_1$ in (3.9).

Standard assumptions on the effective processes on both time-scales are as follows. Let

$$b_0(z) = \sum_k \widetilde{\lambda}_k(z)\widetilde{\zeta}^X_k, \quad b_1(z) = \sum_{k:\beta_k>1} \widetilde{\lambda}_k(z)\widetilde{\zeta}^Y_k, \quad c(z) = \sum_{k:\beta_k=1} \widetilde{\lambda}_k(z)|\widetilde{\zeta}^Y_k|.$$

Condition 3.2 (*Lipschitz and Growth*). *There exists* $K_1 > 0$ *such that* $\forall z, z' \in E$

$$|b_0(z) - b_0(z')| + |b_1(z) - b_1(z')| + |c(z) - c(z')| \le K_1|z - z'|,$$

and there exists $K_2 > 0$ *such that* $\forall z \in E$

$$|b_0(z)| + |b_1(z)| \le K_2|z|, \quad \sup_y c(x, y) < \infty, \ \forall x \in E_X.$$

These conditions insure existence and uniqueness of the deterministic process on time-scale t given by the differential operator (3.4), and of the piecewise deterministic Markov process on time-scale Nt given by the mixed operator (3.5). *However*, if existence and uniqueness of processes defined by (3.4) and (3.5) can be established by other means, for example by conditions that control their overall growth, then we can drop the above conditions and *assume only* that the drift coefficients b_0, b_1 are *locally Lipschitz and locally bounded* (as is implied by same assumptions on λ_k). We will provide conditions for dropping the global Lipschitz conditions on models of chemical reaction systems in the next Section.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

We next consider the form of the exponential generator (3.3) applied to functions of the form $f_N(x, y) = f(x) + \frac{1}{N} g(x, y)$

$$
\begin{aligned}
H_N f_N(x, y) &= \frac{1}{N} \sum_k N^{\beta_k} \lambda_k^N(z) (e^{N(f_N(z+N^{-\underline{\alpha}}\zeta_k^N) - f_N(z))} - 1) \\
&= \sum_{k:\beta_k=1} \lambda_k^N(z) \big(e^{N(f(x+N^{-\underline{\alpha}}\zeta_k^N) - f(x)) + (g(z+N^{-\underline{\alpha}}\zeta_k^N) - g(z))} - 1\big) \\
&+ \sum_{k:\beta_k>1} N^{\beta_k} \lambda_k^N(z) N^{-\underline{\alpha}} \zeta_k \cdot \nabla f(x) + \sum_{k:\beta_k>1} N^{\beta_k - 1} \lambda_k^N(z) N^{-\underline{\alpha}} \zeta_k \cdot \nabla g(x, y) \\
&+ \sum_{k:\beta_k>1} N^{\beta_k - 1} \lambda_k^N(z) \big(e^{N(f(x+N^{-\underline{\alpha}}\zeta_k^N) - f(x)) + (g(z+N^{-\underline{\alpha}}\zeta_k^N) - g(z))} - N^{1-\underline{\alpha}} \zeta_k^N \cdot \nabla f(x) \\
&- N^{-\underline{\alpha}} \zeta_k^N \cdot \nabla g(x, y) - 1\big).
\end{aligned}
$$

Using constraints from Conditions 3.1 we have that the last row above has zero limit since: $\beta_k + 1 \le 2\alpha_i$ for any $i \in \mathcal{I}_X$, and $\beta_k + 1 = 2\alpha_i$ holds only if $\alpha_i = \beta_k = 1$ (as $\beta_k \le \alpha_i, \alpha_i \ge 1$); $\beta_k - 1 \le 2\alpha_j$ for any $j \in \mathcal{I}_Y$, and $\beta_k - 1 = 2\alpha_j$ implies $\alpha_j = 0$, $\beta_k = 1$ (as $\beta_k \le \alpha_j + 1$). In the limit the effective changes $(\widetilde{\zeta}^X, \widetilde{\zeta}^Y)$ that we defined in (3.6)–(3.7) appear as

$$
\begin{aligned}
\lim_{N\to\infty} H_N f_N(x, y) = &\sum_{k:\beta_k=1} \widetilde{\lambda}_k(z) \big(e^{\widetilde{\zeta}_k^X \cdot \nabla f(x)} - 1\big) + \sum_{k:\beta_k=1} \widetilde{\lambda}_k(z) e^{\widetilde{\zeta}_k^X \cdot \nabla f(x)} \big(e^{g(x,y+\widetilde{\zeta}_k^Y) - g(x,y)} - 1\big) \\
&+ \sum_{k:\beta_k>1} \widetilde{\lambda}_k(z) \widetilde{\zeta}_k^X \cdot \nabla f(x) + \sum_{k:\beta_k>1} \widetilde{\lambda}_k(z) \widetilde{\zeta}_k^Y \cdot \nabla_y g(x, y).
\end{aligned}
$$

We let

$$
V(y; x, p) = \sum_{k:\beta_k=1} \widetilde{\lambda}_k(z) \big(e^{\widetilde{\zeta}_k^X \cdot p} - 1\big) + \sum_{k:\beta_k>1} \widetilde{\lambda}_k(z) \widetilde{\zeta}_k^X \cdot p \tag{3.8}
$$

denote the (Hamiltonian) "potential" from the effective slow process, and the operator

$$
L_1^{x,p} f(x, y) = \sum_{k:\beta_k=1} \widetilde{\lambda}_k(z) e^{\widetilde{\zeta}_k^X \cdot p} (f(x, y + \widetilde{\zeta}_k^Y) - f(x, y)) + \sum_{k:\beta_k>1} \widetilde{\lambda}_k(z) \widetilde{\zeta}_k^Y \cdot \nabla_y f(x, y) \tag{3.9}
$$

denote the "perturbed" version of the effective dynamics of the fast process. Then we can formulate the limit of the sequence of exponential generators on this type of functions as

$$
\lim_{N\to\infty} H_N f_N(x, y) = V(y; x, \nabla_x f(x)) + e^{-g(x,y)} L_1^{x, \nabla_x f(x)} e^{g(x,y)}. \tag{3.10}
$$

In order for the limiting operator to be a function of x only, we aim to select g in such a way that the right-hand side does not depend on y. Let $p = \nabla_x f(x)$, and suppose for some positive function $\overline{g}$ (representing e^g) the limiting operator in (3.10) solves the eigenvalue problem for the operator $V(y; x, p) + L_1^{x,p}$: for all $x \in E_X$, $p \in \mathbb{R}$

$$
\big(V(y; x, p) + L_1^{x,p}\big) \overline{g}(x, y) = \overline{H}_0(x, p) \overline{g}(x, y). \tag{3.11}
$$

Let $\overline{H}_0(x, p)$ be the principal ("largest") eigenvalue for the operator $V(y; x, p) + L_1^{x,p}$. In that case $\overline{H}_0(x, p)$ does not depend on $\overline{g}$ (it is a function of the dominant term f of f_N only) and the convergence assumption in Condition 3.0 provides a single valued limit

$$
\lim_{N\to\infty} H_N f_N(x, y) := \overline{H}_0(x, \nabla_x f(x))
$$

operating on the slow variables only. We will use a method that relies on solutions of partial integro-differential equations (PIDEs) to show the sequence of exponential generators H_N when applied to functions of the form f_N converge to $\overline{H}_0$ applied to f.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Having established a candidate for the limiting operator $\overline{H}_0$ one further needs to establish exponential tightness for the slow process $\{X^N\}$ and for the occupation measures of the fast process $\{\Gamma^N(\cdot, C)\}$, where

$$\Gamma^N(t, C) = \int_0^t \mathbf{1}_C(Y^N(s))ds.$$

The next two conditions insure multiplicative ergodicity of the occupation measures and exponential stability of the dynamics for the effective fast process perturbed by the directional changes of the slow process. We will show that exponential tightness will follow when we combine these conditions with control of the growth of the slow process and with convergence of the exponential generators.

Condition 3.3 (*Transition Density*). *For each $x \in E_X$, $p \in \mathbb{R}$ the process $Y^{x,p}$ defined by the generator (3.9) is Feller continuous with transition probability $p_t^{x,p}(y, dy)$ which at $t = 1$ has a positive density with respect to some reference measure $\alpha(dy)$ on E_Y.*

Condition 3.4 (*Lyapunov I*). *There exists a positive function $\varphi(\cdot) \in C^1(E_Y)$ with compact level sets and such that for each compact $K \subset \mathbb{R}$, $\theta \in (0, 1]$ and $l \in \mathbb{R}$ there exists a compact set $A_{l,\theta,K} \subset E_Y$ satisfying*

$$\{y \in E_Y : -\theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y) - |V(y; x, p)| \le l\} \subset A_{l,\theta,K}, \quad \forall p \in K, \forall x \in E_X, \tag{3.12}$$

and for each $x \in E_X$, $p \in \mathbb{R}^{|E_X|}$ there exists $K_{x,p} > -\infty$ such that $\forall z \in E$

$$V(y; x, p) \ge K_{x,p}, \ \forall y \in E_Y.$$

The transition density condition was first given in [14] to insure a large deviation result for the occupation measures of a Markov process, which we impose on the process $Y^{x,p}$ determined by the generator $L_1^{x,p}$ from (3.9). Together with the Lyapunov condition it implies that for each (x, p) there exists a unique stationary distribution for $Y^{x,p}$ (see [17] Lemma 11.23, also [22]). Different versions of such conditions are given in [17] Condition 11.21 (also, see Appendix B for other related references). We will discuss conditions for piecewise deterministic Markov processes given by a perturbation of (3.5) which can be used to verify multiplicative ergodicity and exponential stability in the next Section.

In some cases it will suffice to verify exponential stability of the effective fast process using an (x, p)-dependent function for the Lyapunov condition. This condition together with a version of a condition on the transition kernels implies Condition (DV3+) in [22] for verifying multiplicative ergodicity and establishing a large deviation principle for the occupation measure of single time scale Markov processes.

Condition 3.5 (*Lyapunov II*). *For each p in a compact set $K \subset \mathbb{R}$ and for each $x \in E_X$ there is a positive function $\varphi_{x,p}(\cdot) \in C^1(E_Y)$ with compact level sets such that there exist $c > 1$ and $d_{x,p} < \infty$ satisfying*

$$e^{-\varphi_{x,p}} L_1^{x,p} e^{\varphi_{x,p}}(y) \le -c|V(y; x, p)| + d_{x,p}, \quad \forall y \in E_Y. \tag{3.13}$$

where we assume that $|V(y; x, p)|$ has compact level sets.

Finally, uniqueness of the limiting operator $\overline{H}_0$ needs to be either assumed or established. We will consider the first option in our main (upcoming) Theorem and the second in its Corollary.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

3.2. Large deviation principle for the two time-scale model

We can now present the large deviation result for the Markov process on two time-scales.

Theorem 3.6. *Assume Conditions 3.1–3.4 hold as well as the weak comparison principle for the Cauchy problem:*

$$\begin{aligned} \partial_t u_0(t,x) &= \overline{H}_0(x, \partial_x u_0(t,x)), \quad \text{for } (t,x) \in (0,T] \times E_X \\ u_0(0,x) &= f(x), \quad \text{for } x \in E_X, \end{aligned} \tag{3.14}$$

with $\overline{H}_0$ defined by (3.10). Then the sequence $\{X^N(t)\}$ is exponentially tight and satisfies a large deviation principle with speed N and good rate function I given by the variational principle:

$$I(x, x_0, t) = \sup_{f \in C_b(E_X)} \{f(x) - u_0^f(t, x_0)\}, \tag{3.15}$$

where u_0^f is the unique continuous viscosity solution of (3.14).

The proof of Theorem 3.6 mainly follows the argument in [23]. We show that this argument can be carried through for processes generated by more general (non-Lévy) jump terms and without diffusions. Key ingredients in this proof are: (1) multiplicative ergodicity of the effective fast process (perturbed by the value of the slow process and its drift) which insures exponential control on the behaviour of its occupation measure; and (2) exponential stability for the dynamics of the effective fast process (in a potential generated by the slow process) which insures exponential compact containment and tightness. The proof is given in Appendix A.1 in the Appendix.

Much of the proof concentrates on proving the convergence of solutions to Cauchy problem (2.8) via a sequence of sub- and super-operators which sandwich the limiting operator $\overline{H}_0$. A part of this proof can be significantly simplified when we know the principle eigenvalue problem with $\overline{H}_0$ has a solution and we make use of its associated positive eigenfunction. Being able to solve the eigenvalue problem explicitly relies on the form of the rates λ_k as well as on some of the structural properties of the Markov process itself. For polynomial λ_k there is a class of models (see Condition 4.5) for which we can explicitly solve for the eigenvalue and eigenfunction of (3.11). We will discuss a procedure for solving $\overline{H}_0$ for models of chemical reaction systems in the next Section.

Having an expression for $\overline{H}_0(x, p)$ also allows for a relatively easy verification of the comparison principle for the limiting Cauchy problem (3.14) by using the following result. The proof of this result is based on the theory of discontinuous viscosity solutions (see either [5] Chapter V, or [19] Chapter VII) and is contained in Appendix A.3 of the Appendix. For definitions of lower semicontinuous viscosity sub- and upper semicontinuous super-solutions for discontinuous functions and of the weak comparison principle we refer to [19] VII Definition 4.2 and VII Definition 7.1.

Lemma 3.7. *Suppose u_1 and u_2 are, respectively, a bounded upper semicontinuous (USC) viscosity sub-solution and a bounded lower semicontinuous (LSC) viscosity super-solution of (3.14) for some $T > 0$ and $E_X \subset \mathbb{R}^d$. Either of the following conditions are sufficient for the weak comparison principle for (3.14) to hold:*

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

(a) *$\overline{H}_0$ is such that for all $\lambda \geq 1$, $R > 0$ and for all $|p|, |q| \leq 1$, $|x|, |y| \leq R$ and for some continuous non-decreasing functions ω_R, $\widetilde{\omega}_1 : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with $\omega_R(0) = \widetilde{\omega}_1(0) = 0$*

$$\overline{H}_0(y, \lambda(x-y)+p) - \overline{H}_0(x, \lambda(x-y)+q) \leq \omega_R(|x-y| + \lambda|x-y|^2) + \widetilde{\omega}_1(|p-q|); \quad (3.16)$$

(b) *$\overline{H}_0$ is such that for all $\lambda \geq 1, R, \ell > 0$ and for all $|p|, |q| \leq 1$, $|x|, |y| \leq R$ with $\lambda|x-y| < \ell$, and for continuous non-decreasing functions $\omega_{R,\ell}$, $\widetilde{\omega}_1 : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with $\omega_{R,\ell}(0) = \widetilde{\omega}_1(0) = 0$ the inequality (3.16) holds with $\omega_{R,\ell}$ in place of ω_R; and $\overline{H}_0$ satisfies a coercivity condition in p uniformly for bounded x, i.e. for any $R > 0$*

$$\inf_{|x| \leq R} |\overline{H}_0(x, p)| \to \infty \text{ as } |p| \to \infty. \quad (3.17)$$

If $\overline{H}_0$ is convex in p and $\lim_{r\to\infty} |\overline{H}_0(x, rp)| \to \infty$ for each x, p, then (3.17) holds.

We now have a more easily verified large deviation result for multi-scale Markov processes.

Corollary 3.8. *Assume Conditions 3.1–3.3 and 3.5 hold, without necessarily a lower bound on $V(y; x, p)$ but assuming $|V(y; x, p)|$ has compact level sets. Suppose the principal eigenvalue problem (3.11) for $\overline{H}_0(x, p)$ can be explicitly solved, and either of the conditions (a) or (b) from Lemma 3.7 hold. Then the large deviation principle for $\{X^N\}$ as stated in Theorem 3.6 holds.*

The proof of Corollary 3.8 is centred on simplifying the construction of the approximating sequence of operators from the proof of Theorem 3.6. In order to do so it uses the positive eigenfunction in the principle eigenvalue problem for $\overline{H}_0$. The proof is given in Appendix A.2 in the Appendix. Its practical advantages are that it allows one to use a Lyapunov function dependent on the variable for the slow process; and it simplifies establishing uniqueness of the limiting operator $\overline{H}_0$. Note that if the state space for the fast process is compact one can trivially take $\varphi = 0$ in either (3.12) or (3.13). Moreover, if the state space is non-compact, the process of finding the explicit solution for the eigenvalue problem can be adapted to find the Lyapunov function $\varphi_{x,p}$ as well. We will discuss finding candidates for the Lyapunov functions in the next Section.

To extend the above results, from either Theorem 3.6 or Corollary 3.8, from finite dimensional distributions of $\{X_N(t)\}$ to pathwise large deviation principle for $\{X^N\}$ can be done by the argument from Theorem 4.28 of [17]. This will require using a variational representation of the operator $\overline{H}_0(x, p)$ and characterizing an expression for the rate function from its Fenchel–Legendre transform ([17] Section 8.6.1, [16]). This is the characterization of the rate functions given in [25].

4. Chemical reaction networks on multiple time-scales

We now apply our large deviation results to models of multi-scale chemical reaction networks. Continuous-time Markov chains have found an important application for modelling chemical reactions describing cellular metabolic, gene regulatory and signal transduction processes, since variability in genetically identical cells was in the past two decades shown to be due to inherent noise of biochemical reactions within each cell [26,15]. A stochastic model of an intra-cellular chemical reaction network treats the system as a continuous-time Markov chain with generator of the form (3.1) whose state Z is a vector giving the number of molecules of different types of chemical species that are relevant. Each reaction is modelled as a possible transition for the state. The model for the kth reaction, for each k, is determined by a vector of inputs ν'_k specifying the number of molecules of each chemical species that are consumed in the reaction, and a vector of

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

outputs ν_k specifying the number of molecules of each species that are produced in the reaction. Transition rate for the kth reaction $\lambda_k(z)$ is a function of the state z, and the state change of kth reaction is given by $\zeta_k = \nu_k' - \nu_k$.

Reaction rates in chemical networks are most commonly modelled by dynamics of *mass-action* type: in the stochastic version of the law of mass action, the rate function is proportional to the number of ways of selecting the molecules that are consumed in the reaction:

$$\lambda_k(z) = \kappa_k' \prod_i \nu_{ik}! \prod_i \binom{z_i}{\nu_{ik}} = \kappa_k' \prod_i z_i(z_i - 1)\cdots(z_i - \nu_{ik} + 1), \tag{4.1}$$

making $\lambda_k(z)$ a product of abundances of all the species going into the reaction and of a *chemical reaction constant* κ'. Physically, $|\nu_k| = \sum_i \nu_{ik}$ is usually assumed to be less than or equal to two, which makes the task of controlling the growth of coordinates much easier. In some models a few reaction rates can be given in terms of a sigmoid function. This is often the result of approximating the overall outcome of a subnetwork of reactions through a model reduction procedure. Nonetheless, in all cases encountered in the literature the jump rates are locally Lipschitz and locally bounded, though, as is the case in binary reactions, they are not always globally Lipschitz.

We next discuss satisfying the conditions needed for Theorem 3.6 and for Corollary 3.8 to apply in models of chemical reaction networks, and specify when they can be relaxed. Multiple scalings of intra-cellular chemical reaction processes arise naturally due to low copy numbers of various key chemical species types. In other words, species i can be in abundance of order N^{α_i} in the system, where typically N is the order of magnitude of the most abundant species and $\alpha_i \in [0, 1]$. Each species is then represented by a component with its rescaled size $Z_i^N = N^{-\alpha_i} Z_i$. In addition, varying strengths in chemical bonds formed or broken by different reactions lead to different orders of magnitude, that can be expressed in terms of N, for the different chemical reaction constants κ_i'. We combine the scaling of constants κ_i' together with the effect of rescaling species amounts into a single rescaling of the reaction rates as $\lambda_k(z) = N^{\beta_k}\lambda_k(z^N)$, where typically $\beta_k > 0$. The rescaled chemical reaction network then becomes a multi-scale Markov process on $E^N \subset \mathbb{R}_+^d$ with generator of the form (3.2).

We are interested in models of intercellular reactions whose dynamics has a multi-scale behaviour with two separated time-scales as in Conditions 3.1. This means that in the reaction networks there is a group of *discrete* species present in small counts $O(1)$ and the rest, *continuous* species, are present in larger amounts approximated by $O(N)$. Their intertwined reactions dynamics consists of fast $O(N)$ fluctuations for the discrete species and slow $O(1)$ changes of the continuous species. The effective changes $\widetilde{\zeta}_k$ are typically the same as the net reaction changes ζ_k, except in instances when the order α_i of some species i with $\zeta_{ki} \neq 0$ is smaller than the order β_k of λ_k and the effective change $\widetilde{\zeta}_{ki} = 0$ (see, for example, species V in reaction (6) of the viral production model from the Introduction).

A reaction network often involves a few binary reactions with a quadratic rate, so for the effective processes to exist and be unique one needs the overall depletion rates for each species to balance out their overall production rates, insuring that all amounts are globally stable. Many multi-scale Markov models have global existence and uniqueness with only local Lipschitz and growth of jump rates: the deterministic dynamics of the effective slow process (governed by growth of the drift $b_0(z)$) may be globally controlled; and the piecewise deterministic Markov process (which is often simply a Markov chain) may have global stability (determined by the local drift $b_1(z)$ of the continuous part and the overall jump rates $c(z)$ of the discrete part) for each fixed value of the slow process. In that case Condition 3.2 is not necessary and instead

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Condition 4.1 on the binary rates (discussed in the next subsection) should be verified. We prove in Proposition 4.4 a truncation argument with which the requirement of globally Lipschitz coefficients can be relaxed (both the enzyme kinetics and the viral production model have some binary reaction rates that can be accommodated by this result).

The effective fast process perturbed by the direction of change of the slow process (given by generator (3.9)) is in general a piecewise deterministic Markov process (though sometimes it is simply a Markov chain). Exponential stability and positivity of its transition density assumed in Condition 3.3 can be verified using recent results on piecewise deterministic processes [3] (see also [7] Ch 4). To summarize the conclusions, suppose the discrete component of this piecewise deterministic Markovian dynamics (PDMP) has jump rates such that their infimum over the continuous component yields an irreducible and positive recurrent Markov chain. Moreover, suppose that for all values of the discrete component the generator of the continuous component satisfies a drift condition with respect to the same Lyapunov function. Finally, suppose that the flow satisfies a Hoermander-type bracket (hypoellipticity) condition at a point of the continuous component. Then, at this point, the process has a positive transition probability with non-trivial absolutely continuous part with respect to Lebesgue measure. These results are particularly useful when the state space for the effective fast process is non-compact and having a positive transition density is non-trivial.

In models where the state space E_Y for the fast process is compact verifying the exponential Lyapunov Condition 3.4 is unnecessary. In the non-compact case we need to look for a candidate function φ based on both the dynamics of perturbed effective fast process with generator $L_1^{x,p}$, and on the potential function $|V(y; x, p)|$. For multi-scale processes for which we can explicitly solve the eigenvalue problem (3.11) we can relax this condition and instead verify the Lyapunov Condition 3.5, and look for a candidate function $\varphi_{x,p}$ dependent on the value and direction of change of the slow process. Moreover, such a function can be found by a similar procedure according to which one can find the positive eigenfunction of the EVP (see, for example, the models of down-regulation and of viral production in Examples section, both of which have an unbounded fast variable). For practical purposes, with Condition 4.5 and Proposition 4.6 we provide a procedure for solving the EVP for principal eigenvalue and positive eigenfunction explicitly.

4.1. Truncating the jump rates

Many chemical reaction networks involve a few binary reactions whose rates have quadratic growth. Our goal is to truncate these jump rates and prove the large deviation result using Theorem 3.6 on a process with truncated rates. Let us illustrate this on the examples from the Introduction. In the model of enzymatic kinetics (1.1) the sum of enzyme and enzyme–substrate abundances together is conserved by the system, so $Z_2(t) + Z_3(t) \equiv M$ is constant in time. Consequently, the rate of enzyme–substrate production $S + E \rightharpoonup ES$ satisfies $\lambda_1(z) = \kappa_1 z_1 z_2 \leq \kappa_1 M z_1$, and this quadratic rate is in fact at most linear. In the model of viral production (1.3), there is a quadratic reaction rate $\lambda_6(z) = \kappa_6 z_1 z_2$ for the viral packaging $G + T + (S) \rightharpoonup V$. There is no conservation law in the system, but the viral genome G is slowly varying, its rate of increase is given only by $\emptyset \rightharpoonup G$ and is bounded by a constant. This allows us to have exponential control on how large the factor G can get. As we will see, from the point of view of large deviations, this quadratic reaction rate can consequently be replaced with versions that are at most linear.

We say that a "general conservation law" holds for a subset of species, if a linear combination of those species is left unchanged by all the reactions in the network (for example, in the model

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

of enzyme kinetics (1.1) the enzyme and the enzyme–substrate are in one such conservation law). We make the following assumption on the jump rates.

Condition 4.1. *Suppose that for each k there are non-negative constants $\theta_{k,0}$ and $\theta_k = \{\theta_{k,i}\}$ so that*

$$\lambda_k(z) \le (\theta_{k,0} + \sum_i \theta_{k,i} z_i)(\sum_i z_i), \tag{4.2}$$

and for each k we have either:

(i) *all species i for which $\theta_{k,i} > 0$ are part of a general conservation law, that is, $\theta_{k,i} > 0$ implies $\widetilde{\theta}_i > 0$, where $\widetilde{\theta} = \{\widetilde{\theta}_i\}$ is such that all net changes satisfy*

$$\zeta_k \cdot \widetilde{\theta} = 0, \quad \forall k;$$

or:

(ii) *all species i for which $\theta_{k,i} > 0$ are part of the slow process, and rates of all those reactions with a net increase of species combinations given by θ_k satisfy*

$$\sum_{k:\zeta_k \cdot \theta_{k,i} > 0} \lambda_k(z) \le C', \quad \text{for some } C' < \infty.$$

For rates that are specified in "mass-action" form the Condition (4.2) requires that in all binary reactions at least one of the reactants S_i must satisfy $\theta_{k,i} > 0$. Also, if that reactant is part of the fast process it must be part of a conservation law (there can be multiple conservation laws in the system), or if it is part of the slow process it is created only by reactions with bounded rates, for example, reactions such as $\emptyset \rightharpoonup S_i$ or reactions $S_j \rightharpoonup S_i$ for some species S_j of bounded abundance.

We need to introduce a notion that will describe what happens to the process once we truncate its jump rates, and justify why it is useful in the context of large deviation results (this is Theorem 4.2.16 in [12]).

Definition 4.2 (*Exponential Approximation*)**.** The sequence $\{X^{N,M}\}$ is an exponentially good approximation of $\{X^N\}$ if for every $\epsilon > 0$

$$\lim_{M\to\infty} \limsup_{N\to\infty} \frac{1}{N} \log P\{|X^N - X^{N,M}| > \epsilon\} = -\infty.$$

Theorem 4.3. *Suppose for each M the sequence $\{X^{N,M}\}$ satisfies a large deviation principle with rate function I_M and suppose $\{X^{N,M}\}$ is an exponentially good approximation of $\{X^N\}$. Then*

(a) $\{X^N\}$ satisfies a weak large deviation principle (meaning that (2.1) holds for each open set A while (2.2) only holds for each compact set B) with rate function

$$I(y) = \sup_{\epsilon>0} \liminf_{M\to\infty} \inf_{z\in B_\epsilon(y)} I_M(x).$$

(b) If $I(\cdot)$ is a good rate function (sublevel sets are compact) and for each closed set B

$$\inf_{y\in B} I(y) \le \limsup_{M\to\infty} \inf_{y\in B} I_M(y),$$

then $\{X^N\}$ satisfies the large deviation principle with rate function I.

The following result is very useful in some models of chemical reaction systems.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Proposition 4.4. *Assume Condition 4.1 holds and that the initial value satisfies* $P\{X^N(0)\cdot\theta_k \le C\} = 1$ *for some* $C < \infty$ *and all* θ_k *as in (4.2). Then for all large enough* $M < \infty$ *replacing the jump rates* $\lambda_k(z)$ *of the Markov process* X^N *by* $\lambda_k(z) \wedge (\theta_{k,0} + M)(\sum_i z_i)$ *produces a sequence of processes* $\{X^{N,M}\}$ *that is an exponentially good approximation of* $\{X^N\}$*, that is*

$$\limsup_{N\to\infty} \frac{1}{N} \log P\{\sup_{s\le t}|X^N(t) - X^{N,M}(t)| > 0\} \le c(t) - M.$$

Proof. In case (i) holds for reaction k the assumption on $\widetilde{\theta}$ implies that $\widetilde{\theta}_i = a_{k,i}\theta_{k,i}$ for some $a_{k,i} > 0$. The conservation law implies that $X^N(t)\cdot\widetilde{\theta} = X^N(0)\cdot\widetilde{\theta}$ for $t > 0$. The assumption on the initial value implies

$$X^N(t)\cdot\theta_k \le (\max_{k,i}\frac{1}{a_{k,i}})X^N(t)\cdot\widetilde{\theta} = (\max_{k,i}\frac{1}{a_{k,i}})X^N(0)\cdot\widetilde{\theta} \le \frac{\max_{k,i} a_{k,i}}{\min_{k,i} a_{k,i}}\, C.$$

So as soon as $M \ge \frac{\max_{k,i} a_{k,i}}{\min_{k,i} a_{k,i}}\, C$ we have that $P\{\sum_i \theta_{k,i} X_i^N(t) \le M\} = 1$, and we can replace the rates $\lambda_k(z)$ by $\lambda_k(z) \wedge (\theta_{k,0} + M)(\sum_i z_i)$ without altering the process.

In case (ii) holds we make use of the martingales

$$\exp\Big\{N\big(X^N(t)\cdot\theta_k - X^N(0)\cdot\theta_k - \int_0^t \sum_k N^{\beta_k-1}\lambda_k^N(X^N(s))\big(e^{N^{1-\alpha}\zeta_k\cdot\theta_k} - 1\big)ds\big)\Big\}$$

to obtain for any $M < \infty$ and stopping time τ^N for X^N that

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N} \log P\{\sup_{s\le t} X^N(s)\cdot\theta_k \ge M\}\\
&\le \limsup_{N\to\infty} \frac{1}{N} \log \frac{E[e^{NX^N(t\wedge\tau^N)\cdot\theta_k}]}{e^{NM}}\\
&\le \limsup_{N\to\infty} \frac{1}{N} \log \frac{E[\exp\{N\big(X^N(0)\cdot\theta_k + \int_0^{t\wedge\tau^N} \lambda_k(X^N(s))\sum_{k:\zeta_k\cdot\theta_k>0} N^{\beta_k-1}(e^{N^{1-\alpha}\zeta_k\cdot\theta_k} - 1\big)ds)\}]}{e^{NM}}\\
&\le C + C't(e^{\max_{k:\beta_k=1,\ \zeta_k\cdot\theta_k>0}\widetilde{\zeta}_k^X\cdot\theta_k} + \max_{k:\beta_k>1,\zeta_k\cdot\theta_k>0}\widetilde{\zeta}_k^X\cdot\theta_k) - M = c(t) - M.
\end{aligned}$$

In the last inequality we used the assumption $P\{\sum_{k:\zeta_k\cdot\theta_k>0}\lambda_k(X^N(t)) \le C'\} = 1$ for $\forall t \ge 0$, and that since $\zeta_k\cdot\theta_k = \zeta_k^X\cdot\theta_k$

$$\lim_{N\to\infty}\sum_{k:\zeta_k\cdot\theta_k>0} N^{\beta_k-1}(e^{N^{1-\alpha}\zeta_k\cdot\theta} - 1) = \sum_{k:\beta_k=1,\ \zeta_k\cdot\theta>0}(e^{\widetilde{\zeta}_k^X\cdot\theta} - 1) + \sum_{k:\beta_k>1,\ \zeta_k\cdot\theta>0}\widetilde{\zeta}_k^X\cdot\theta.$$

Let $X^{N,M}$ be the process with the same chemical reaction network properties (same ζ_k, α_i, β_k, Y_k) except we replace the rates $\lambda_k(z)$ by $\lambda_k(z) \wedge (\theta_{k,0} + M)(\sum_i z_i)$. On the event $\{\sup_{s\le t} X^N(s)\cdot\theta_k < M\}$ the processes X^N and $X^{N,M}$ are the same, hence

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N} \log P\{\sup_{s\le t}|X^N(s) - X^{N,M}(s)| > 0\}\\
&\le \limsup_{N\to\infty} \frac{1}{N} \log P\{\sup_{s\le t} X^N(s)\cdot\theta_k \ge M\} \le c(t) - M. \quad \square
\end{aligned}$$

4.2. Solving the eigenvalue problem

In order to apply Theorem 3.6 we also need to show the comparison principle for the limiting Cauchy problem (3.14). It insures that we can identify the limiting semi-group (2.7) in Bryc

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

formula from our convergence arguments. Theory for uniqueness of viscosity solutions says that the comparison principle will hold if the conditions provided by Lemma 3.7 are satisfied. They are similar to the conditions for proving comparison principle for operator themselves (see Lemma 9.2 [17] or Proposition 7.7 [8]). There is an advantage in having an operator which is convex and coercive in p, as then we only need to check these conditions on compact subsets of $E_X \times \mathbb{R}$ (see option (b) in Lemma 3.7). This can easily be verified if $\overline{H}_0$ can be explicitly calculated.

In order to explicitly calculate $\overline{H}_0$ we need to solve the EVP (3.11) for the principal eigenvalue, finding the positive eigenfunction $\overline{g} = e^{u_1(x,y)}$ in the process. Let us illustrate the conditions we need on the examples from the Introduction. The key feature in both the enzyme kinetics network (1.1) and in the viral production network (1.3) is that the dynamics of the effective processes has rates that are linear in the fast variables, and that the effective changes of the fast process take values in $\{-1, 0, 1\}$.

In the model of enzyme kinetics (1.1), the effective dynamics of the fast process is a pure Markov chain. The fast variables are E and ES, but due to a conservation law between them, there is effectively only one fast variable, say E, in amount y. Since the rates in the effective process are linear in y, and since we need a positive eigenfunction $\overline{g}$, we can look for an eigenfunction whose logarithm is linear in it as well. That is, we let $\overline{g} = e^{a(x)y}$, where the function $a(x)$ of the amount x of the slow variable S is to be determined. We need to insure that $V(y; x, p) + e^{-a(x)y} L_1^{x,p} e^{a(x)y}$ is a function of x only (see Section 5.1 for the evaluation of this expression and others in this explanation). Since the rates appearing in V and $L_1^{x,p}$ are linear in y, this function is linear in y as well. After grouping all the terms we set the coefficient of y to zero in order to determine $a(x)$. From (3.9) we see that $a(x)$ appears in the function $e^{-a(x)y} L_1^{x,p} e^{a(x)y}$ only in the form of $e^{\widetilde{\zeta}_k^Y a(x)}$, and because the effective changes of the fast process satisfy $\widetilde{\zeta}_k^Y \in \{-1, 0, 1\}$, $a(x)$ appears only as $\{e^{a(x)}, e^0, e^{-a(x)}\}$. Setting the coefficient next to y equal to zero for all x leads to a quadratic equation in $e^{a(x)}$, which under appropriate conditions will have a unique positive solution. Solving for $a(x) > 0$ consequently provides the eigenfunction $\overline{g} = e^{a(x)}$ and provides the expression for $\overline{H}_0(x, p) = V(y; x, p) + e^{-a(x)y} L_1^{x,p} e^{a(x)y}$.

In general, for network with multiple fast variables, we can similarly solve the EVP if we impose some conditions. To illustrate them on the above example consider the amounts y_1, y_2 of fast variables E, ES as a process in two variables. Let $a(x) = (a_1(x), a_2(x))$, $\overline{g} = e^{a(x)y} = e^{a_1(x)y_1 + a_2(x)y_2}$, and in $e^{-a(x)y} L_1^{x,p} e^{a(x)y}$ we have $a(x)$ appearing in the forms $\{e^{a_1(x)}, e^{a_2(x)}, e^0, e^{-a_1(x)}, e^{-a_2(x)}\}$. Consider the two coefficients in $V(y; x, p) + e^{-a(x)y} L_1^{x,p} e^{a(x)y}$ next to the terms y_1 and y_2 that are to be set equal to zero and thus determine $a_1(x)$, $a_2(x)$. When a term $e^{-a_1(x)}$ appears it must appear in the coefficient next to y_1 (since E is a source species for this reaction), and we require that, in that case, if a term $e^{a_2(x)}$ also appears in the coefficient next to y_1, we must have $y_1 + y_2 = \text{const}$. This is in fact the case in our example of enzyme kinetics (from $E + S \rightharpoonup ES$), and it insures that we can isolate a quadratic equation to be solved for $a_1(x)$ from the coefficient next to y_1. We make the analogous requirement for coefficient next to y_2 as the term $e^{-a_2(x)}$ appears in it (from $E + S \leftharpoondown ES$ and $ES \rightharpoonup P + E$), in order to allow us to solve for $a_2(x)$. In isolating equations for $a_1(x)$ and $a_2(x)$, we could have also allowed for a term e^0 to appear in coefficient next to either y_1 or y_2, from reactions where a fast species produces a slow one (for example, $E \rightharpoonup E + S$).

We formalize the above requirements and make the following assumptions on the jump rates in case the effective dynamics of the fast process is a pure Markov chain.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Condition 4.5. *Suppose that* $\widetilde{\zeta}^Y_{k,i} \in \{-1, 0, 1\}$ $\forall k, \forall i$*; there are constants* $\theta_{k,0}, \{\theta_{k,i}\} \in \{0, 1\}$ *satisfying* $\theta_{k,0} + \sum_i \theta_{k,i} = 1$ *such that*

$$\lambda_k(x, y) = \lambda_k(x)(\theta_{k,0} + \sum_i \theta_{k,i} y_i);$$

and for each species j *for which* $\exists k, i$ *such that* $\theta_{k,i} = 1$ *and* $\widetilde{\zeta}^Y_{k,j} = -1$ *we have that either:*

(i) *a species* $j' \neq j$ *for which* $\exists k'$ *(possibly* k*) such that* $\theta_{k',i} = 1$ *and* $\widetilde{\zeta}^Y_{k',j'} \neq 0$ *is part of a conservation law together with species* j*, that is, both* $\widetilde{\theta}_j > 0$ *and* $\widetilde{\theta}_{j'} > 0$ *where* $\widetilde{\theta} = \{\widetilde{\theta}_i\}$ *is such that effective net changes for the fast process satisfy*

$$\widetilde{\zeta}^Y_k \cdot \widetilde{\theta} = 0, \quad \forall k;$$

or:

(ii) *a species* $j' \neq j$ *for which* $\exists k'$ *(possibly* k*) such that* $\theta_{k',i} = 1$ *and* $\widetilde{\zeta}^Y_{k',j'} \neq 0$ *is part of the slow process.*

This condition is simplest to explain in case the jump rates on the time scale of the fast dynamics are given by mass-action functions, that is, a reaction k acting on fast species $\{S_i\}$:

$$\text{(slow species)} \; + \sum_i \theta_{k,i} S_i \rightharpoonup \sum_i (\theta_{k,i} + \widetilde{\zeta}^Y_{k,i}) S_i + \; \text{(slow species)}$$

has jump rate of the form:

$$\lambda_k(z) = \lambda_k(x) \prod\nolimits_i y^{\theta_{k,i}}.$$

Linearity of the jump rates (in the fast variables) is implied by $\theta_{k,i} \in \{0, 1\}, \forall k, i$. When a reaction's jump change for species i is $\widetilde{\zeta}^Y_{k,i} = -1$ then this is a reaction which is either simply using up species i or converting species i into some other species $j \neq i$:

$$\text{(slow species)} \; + S_i \rightharpoonup \; \text{(slow species)} + \widetilde{\zeta}^Y_{k,j} S_j,$$

with $\widetilde{\zeta}^Y_{k,j} = 0$ in the former and $\widetilde{\zeta}^Y_{k,j} > 0$ in the latter case. Since species i goes into this reaction, it implies that also $\theta_{k,i} = 1$. There may be other reactions k' with $\theta_{k',i} = 1$ which only use species i as a catalyst and then $\widetilde{\zeta}^Y_{k',i} \geq 0$:

$$\text{(slow species)} \; + S_i \rightharpoonup \; \text{(slow species)} + \widetilde{\zeta}^Y_{k',i} S_i + \widetilde{\zeta}^Y_{k',j} S_j.$$

Our condition says that any species $j \neq i$ created by such a reaction from i must be either in a conservation relation with it: $\sum_i \widetilde{\theta}_i Y^N_i(t) = \sum_i \widetilde{\theta}_i Y^N_i(0)$ for all $t > 0$, or is a slow species. For simplicity we assume that if species i is part of a conservation law $\widetilde{\theta}$ with $\widetilde{\theta}_i > 0$ then there is a unique species i' such that $\widetilde{\theta}_{i'} > 0$ as well, we denote this relationship by $i' \sim i$. Let $M_i = \widetilde{\theta}_i Y^N_i(t) + \widetilde{\theta}_{i'} Y^N_{i'}(t), \forall t \geq 0$, and in such pairs apply a change of variables $y_{i'} = M_i/\widetilde{\theta}_{i'} - (\widetilde{\theta}_i/\widetilde{\theta}_{i'}) y_i$.

We can now solve the eigenvalue problem when the fast process is purely a Markov chain.

Proposition 4.6. *Assume the effective dynamics of the fast process is a Markov chain and Condition 4.5 holds. Let* $\mathcal{I} = \{i : \exists k, j \; \theta_{k,i} = 1, \widetilde{\zeta}^Y_{k,j} = -1\}, \mathcal{I}_j = \{i : \exists k \; \theta_{k,i} = 1, \widetilde{\zeta}^Y_{k,j} = -1\}$*, and* $\mathcal{J} = \{j : \exists k \; \widetilde{\zeta}^Y_{k,j} = -1\}$*. Let* $\{A_{ij}, B_i, C_{ij}\}$ *be the functions given in (4.4)–(4.6). If* $\forall x \in E$, $p \in \mathbb{R}$*:* ***(a)*** *for each* $j \in \mathcal{J}$ *the quadratic equation*

$$z_j^2 \sum_{i \in \mathcal{I}_j} A_{ij}(x, p) + z_j \sum_{i \in \mathcal{I}_j} B_{ij}(x, p) + \sum_{i \in \mathcal{I}_j} C_{ij}(x, p) = 0$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

has a unique positive solution z_j; and: ***(b)*** *over the set $\mathcal{I}^c = \{i : \nexists k,\ j\ \theta_{k,i} = 1, \widetilde{\zeta}^Y_{k,j} = -1\}$ the system of linear equations*

$$\Big\{\sum_{j\in\mathcal{J}^c} z_j A_{ij}(x,\,p) + \sum_{j\in\mathcal{J}^c} B_{ij}(x,\,p) = 0\Big\}_{i\in\mathcal{I}^c}$$

has a unique positive solution of variables $\{z_j\}_{j\in\mathcal{J}^c}$ over the set $\mathcal{J}^c = \{j : \nexists k\ \widetilde{\zeta}^Y_{k,j} = -1\}$. Then, for any $j \in \mathcal{J}$ for which $\sum_{i\in\mathcal{I}_j} A_{ij}(x,\,p) \neq 0$ only ***(i)*** *in Condition 4.5 is possible. Moreover, the eigenvalue problem (3.11) has a unique solution given by*

$$\begin{aligned}\overline{H}_0(x,\,p) = &\sum_{k:\theta_{k,0}=1} \lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}^X_k\cdot p}e^{\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} - 1) + \mathbf{1}_{\beta_k>1}\widetilde{\zeta}^X_k\cdot p\Big)\\ &+ \sum_i \mathbf{1}_{\exists i'\sim i}(M_i/\widetilde{\theta}_{i'}) \sum_{k:\theta_{k,i}=1} \lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}^X_k\cdot p}e^{\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} - 1) + \mathbf{1}_{\beta_k>1}\widetilde{\zeta}^X_k\cdot p\Big),\end{aligned} \tag{4.3}$$

where $e^{\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} = \prod_i z_i^{\widetilde{\zeta}^Y_{k,i}}$ in the formula is determined by the solutions to the quadratic and linear equations above, and the associated eigenfunction is given by $e^{h(x,y)} = \prod_i z_i^{y_i}$.

Proof. We rewrite (3.11) with $h(x,\,y) = \sum_i a_i(x)y_i$

$$\begin{aligned}&\big(V(y;\,x,\,p) + e^{-h(x,y)}L_1^{x,p}e^{h(x,y)}\big)\\ &= \sum_{k:\beta_k=1} \lambda_k(z)(e^{\widetilde{\zeta}^X_k\cdot p+\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} - 1) + \sum_{k:\beta_k>1} \lambda_k(z)\widetilde{\zeta}^X_k\cdot p\\ &= \sum_{k:\beta_k=1} \lambda_k(x)(\theta_{k,0} + \sum_i \theta_{k,i}y_i)(e^{\widetilde{\zeta}^X_k\cdot p+\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} - 1) + \sum_{k:\beta_k>1} \lambda_k(x)(\theta_{k,0}\\ &\quad + \sum_i \theta_{k,i}y_i)\widetilde{\zeta}^X_k\cdot p\\ &= \sum_{k:\theta_{k,0}=1} \lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}^X_k\cdot p+\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} - 1) + \mathbf{1}_{\beta_k>1}\widetilde{\zeta}^X_k\cdot p\Big)\\ &\quad + \sum_{i\in\mathcal{I}^c} y_i \sum_{k:\theta_{k,i}=1} \lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}^X_k\cdot p+\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} - 1) + \mathbf{1}_{\beta_k>1}\widetilde{\zeta}^X_k\cdot p\Big)\\ &\quad + \sum_{i\in\mathcal{I}} y_i \sum_{k:\theta_{k,i}=1} \lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}^X_k\cdot p+\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} - 1) + \mathbf{1}_{\beta_k>1}\widetilde{\zeta}^X_k\cdot p\Big).\end{aligned}$$

In order to get a result which is independent of y we need to set all of the terms next to y_i to zero, which will leave the first row intact, except for adding to it all the M_i dependent terms from the change of variables $y_{i'} = M_i/\widetilde{\theta}_{i'} - (\widetilde{\theta}_i/\widetilde{\theta}_{i'})y_i$ where $i' \sim i$ exists

$$\sum_i \mathbf{1}_{\exists i'\sim i}(M_i/\widetilde{\theta}_{i'}) \sum_{k:\theta_{k,i}=1} \lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}^X_k\cdot p+\sum_i a_i(x)\widetilde{\zeta}^Y_{k,i}} - 1) + \mathbf{1}_{\beta_k>1}\widetilde{\zeta}^X_k\cdot p\Big).$$

Using Condition 4.5 we can rewrite the middle row as

$$\sum_{i\in\mathcal{I}^c} y_i \sum_{j\in\mathcal{J}^c} \big(e^{a_j(x)}A_{ij}(x,\,p) + B_{ij}(x,\,p)\big),$$

and likewise the last row as

$$\sum_{i\in\mathcal{I}} y_i \sum_{j\in\mathcal{J}} \big(e^{a_j(x)}A_{ij}(x,\,p) + B_{ij}(x,\,p) + e^{-a_j(x)}C_{ij}(x,\,p)\big).$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

where we used the following functions

$$A_{ij}(x,p) = \sum_{k:\theta_{k,i}=1} \mathbf{1}_{\widetilde{\zeta}^Y_{k,j}=1}\lambda_k(x)\mathbf{1}_{\beta_k=1}e^{\widetilde{\zeta}^X_k\cdot p} - \mathbf{1}_{i'\sim i}(\widetilde{\theta}_i/\widetilde{\theta}_{i'}) \times \sum_{k:\theta_{k,i'}=1} \mathbf{1}_{\widetilde{\zeta}^Y_{k,j}=1}\lambda_k(x)\mathbf{1}_{\beta_k=1}e^{\widetilde{\zeta}^X_k\cdot p}, \tag{4.4}$$

$$\begin{aligned} B_{ij}(x,p) = {} & \sum_{k:\theta_{k,i}=1} \mathbf{1}_{\widetilde{\zeta}^Y_{k,j}=0}\lambda_k(x)\mathbf{1}_{\beta_k=1}e^{\widetilde{\zeta}^X_k\cdot p} - \mathbf{1}_{i'\sim i}(\widetilde{\theta}_i/\widetilde{\theta}_{i'}) \\ & \times \sum_{k:\theta_{k,i'}=1} \mathbf{1}_{\widetilde{\zeta}^Y_{k,j}=0}\lambda_k(x)\mathbf{1}_{\beta_k=1}e^{\widetilde{\zeta}^X_k\cdot p} \\ & + \sum_{k:\theta_{k,i}=1} \lambda_k(x)\big(-\mathbf{1}_{\beta_k=1} + \mathbf{1}_{\beta_k>1}\widetilde{\zeta}^X_k\cdot p\big) - \mathbf{1}_{i'\sim i}(\widetilde{\theta}_i/\widetilde{\theta}_{i'}) \\ & \times \sum_{\theta_{k,i'}=1} \lambda_k(x)\big(-\mathbf{1}_{\beta_k=1} + \mathbf{1}_{\beta_k>1}\widetilde{\zeta}^X_k\cdot p\big), \end{aligned} \tag{4.5}$$

$$C_{ij}(x,p) = \sum_{k:\theta_{k,i}=1} \mathbf{1}_{\widetilde{\zeta}^Y_{k,j}=-1}\lambda_k(x)\mathbf{1}_{\beta_k=1}e^{\widetilde{\zeta}^X_k\cdot p} - \mathbf{1}_{i'\sim i}(\widetilde{\theta}_i/\widetilde{\theta}_{i'}) \times \sum_{k:\theta_{k,i'}=1} \mathbf{1}_{\widetilde{\zeta}^Y_{k,j}=-1}\lambda_k(x)\mathbf{1}_{\beta_k=1}e^{\widetilde{\zeta}^X_k\cdot p}. \tag{4.6}$$

For each species i from the middle row we get one equation in a system of $|\mathcal{I}^c|$ linear equations in the variables $z_j = e^{a_j(x)}$ over the set $j \in \mathcal{J}^c$. For each i from the last row we get a single quadratic equation in the variables $z_j = e^{a_j(x)}$, note that Condition 4.5 implies the map $i \mapsto j$ is unique. The set $\mathcal{I}_j = \{i : \exists k\ \theta_{k,i} = 1, \widetilde{\zeta}_{k,j} = -1\}$ may not be of size 1, so these quadratic equations will combine to identify a single solution z_j to the quadratic equation

$$\sum_{i\in\mathcal{I}_j} A_{ij}(x)e^{a_j(x)} + \sum_{i\in\mathcal{I}_j} B_{ij}(x) + \sum_{i\in\mathcal{I}_j} C_{ij}(x)e^{-a_j(x)} = 0.$$

Note that this equation can in fact have the coefficient $\sum_{i\in\mathcal{I}_j} A_{ij} = 0$ in which case option **(ii)** of the Condition 4.5 is possible. However, if the equation has $\sum_{i\in\mathcal{I}_j} A_{ij} \neq 0$ then the quadratic equation has a unique positive solution iff $\sum_{i\in\mathcal{I}_j} A_i(x)\sum_{i\in\mathcal{I}_j} C_i(x) < 0$. This implies that only option **(i)** of the Condition 4.5 is possible and species $j \in \mathcal{J}$ has to be in a conservation law with some other species, as otherwise all of the functions $A_{ij}(x,p)$, $C_{ij}(x,p)$ are positive and only $B_{ij}(x,p)$ can be negative. □

Note that $|\mathcal{I}| \geq |\mathcal{J}|$ so $|\mathcal{J}^c| \geq |\mathcal{I}^c|$ and in general there may be variables that are not defined by the system (e.g. $\emptyset \rightharpoonup S$ with no $S \rightharpoonup$ in system). This is the reason we include in out state space Z only *active* species in the reaction network which are defined as species that appears as an input in at least one reaction. An example of an inactive species is the product species P in the enzymatic kinetics example, or the packaged virus particle species V in the viral production example.

We next consider the case when the effective dynamics of the fast process is a piecewise-deterministic Markov chain and a dynamical system (a PDMP). Recall that the state space of a PDMP separates into a discrete and a continuous component, where the discrete component jumps according to a Markov chain and the continuous component performs deterministic dynamics. In the model of viral production (1.3), the fast variables are T and S, in amounts y_1 and

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

y_2 respectively. Their effective dynamics are a birth–death Markov chain for y_1, and deterministic dynamics for y_2 whose gradient depends on the state y_1. Since y_1 is discrete and y_2 is continuous from (3.9) we see that now $a(x)$ appears in the function $e^{-a(x)y}L_1^{x,p}e^{a(x)y}$ in the forms of $e^{\widetilde{\zeta}_k^{y_1}a_1(x)}$ and $\widetilde{\zeta}_k^{y_2}a_2(x)$, that are effectively $\{e^{a_1(x)}, e^0, e^{-a_1(x)}\}$ and $\{a_2(x), 0, -a_2(x)\}$ (see Section 5.4 for evaluation of expressions). When a term $e^{-a_1(x)}$ appears in the coefficient next to y_1 (which it does from $T \rightharpoonup \emptyset$ and $G + T + (S) \rightharpoonup V$), it may happen that the term $a_2(x)$ also appears next to y_1 (as it does from $T \rightharpoonup T + S$). Then, in order to be able to isolate an equation for $a_1(x)$ from the coefficient next to y_1, we require that $a_2(x)$ can be solved independently, which means that in the coefficient next to y_2 we only allow appearance of $a_2(x)$ (from $S \rightharpoonup \emptyset$). In general, we could have had two discrete fast variables in the network, in which case we would need to impose the same requirements as we had in the pure Markov chain case. The only additional requirement made in the PDMP case here comes from reactions where a discrete fast species produces a continuous fast species (for example, $T \rightharpoonup T + S$).

In case the fast process is a PDMP, sufficient conditions for the existence of an explicit solution to the eigenvalue problem is slightly more difficult to state, but equally straightforward to solve. No restrictions are necessary on its continuous component, and conditions on its discrete component are similar to the case when the fast process is just a Markov chain. The only difference is that, here we have an extra option (iii) in which species created from reactions using fast discrete species are allowed to be fast continuous species (similarly to the option (ii) where they are allowed to be slow species, see the explanation after Condition 4.5). We formalize this as follows.

Condition 4.7. *Suppose there are constants $\theta_{k,0}, \{\theta_{k,i}\} \in \{0, 1\}$ satisfying $\theta_{k,0} + \sum \theta_{k,i} = 1$ such that*

$$\lambda_k(x, y) = \lambda_k(x)(\theta_{k,0} + \sum_i \theta_{k,i} y_i);$$

*for each i in the **discrete** component $\widetilde{\zeta}_{k,i}^Y \in \{-1, 0, 1\}\ \forall k$; and for each j in the **discrete** component for which $\exists k, i$ such that $\theta_{k,i} = 1$ and $\widetilde{\zeta}_{k,j}^Y = -1$ we have that either **(i)** or **(ii)** of Condition 4.5 hold or:*

***(iii)** a species $j' \neq j$ for which $\exists k'$ (possibly k) such that $\theta_{k',i} = 1$ and $\widetilde{\zeta}_{k',j'}^Y \neq 0$ is in the continuous component of the fast process, and is such that: $\exists i^* \neq i, \exists k^* \neq k'$ such that $\theta_{k^*,i^*} = 1$ and $\widetilde{\zeta}_{k^*,j'}^Y \neq 0$ and also $\widetilde{\zeta}_{k^*,j^*}^Y \neq 0$ only if j^* is a slow species.*

We note that if **(i)** holds for species i than the conserved species $i' \sim i$ must also be in the discrete component of the fast process.

Corollary 4.8. *Assume the effective dynamics of the fast process is a piecewise-deterministic Markov chain and Condition 4.7 holds. Let $\mathcal{I}, \mathcal{I}_j, \mathcal{J}$ be as in Proposition 4.6. If $\forall x \in E$, $p \in \mathbb{R}$ for each discrete component species $j \in Y_d \cap \mathcal{J}$ the quadratic equation*

$$z_j^2 \sum_{i \in \mathcal{I}_j} A_{jj}(x, p) + z_j \sum_{i \in \mathcal{I}_j} B_{ij}(x, p) + \sum_{i \in \mathcal{I}_j} C_{ij}(x, p) = 0$$

has a unique positive solution z_j, and over the set $\mathcal{I}^c$ the system of linear equations

$$\Big\{ \sum_{j \in Y_d \cap \mathcal{J}^c} z_j A_{ij}(x, p) + \sum_{j \in Y_c} u_j A_{ij}^c(x, p) + \sum_{j \in (Y_d \cap \mathcal{J}^c) \cup Y_c} B_{ij}(x, p) = 0 \Big\}_{i \in \mathcal{I}^c}$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

has a unique positive solution of variables $\{z_j\}_{j\in Y_d\cap\mathcal{J}^c}, \{u_j\}_{j\in Y_c}$, *then for any* $j \in Y_d \cap \mathcal{J}$ *for which* $\sum_{i\in\mathcal{I}_j} A_{ij}(x,p) \neq 0$ *only* ***(i)*** *in Condition 4.7 is possible. Moreover, the eigenvalue problem (3.11) has a unique solution given by*

$$\begin{aligned}\overline{H}_0(x,p) = &\sum_{k:\theta_{k,0}=1}\lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}_k^X\cdot p}e^{\sum_i a_i(x)\widetilde{\zeta}_{k,i}^Y}-1)+\mathbf{1}_{\beta_k>1}(\widetilde{\zeta}_k^X\cdot p+\widetilde{\zeta}_k^Y\cdot a(x))\Big)\\ &+\sum_i \mathbf{1}_{\exists i'\sim i}(M_i/\widetilde{\theta}_{i'})\\ &\times\sum_{k:\theta_{k,i}=1}\lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}_k^X\cdot p}e^{\sum_i a_i(x)\widetilde{\zeta}_{k,i}^Y}-1)+\mathbf{1}_{\beta_k>1}(\widetilde{\zeta}_k^X\cdot p+\widetilde{\zeta}_k^Y\cdot a(x))\Big), \quad (4.7)\end{aligned}$$

where in the above formula the terms $e^{\sum_i a_i(x)\widetilde{\zeta}_{k,i}^Y}=\prod_{i\in Y_d} z_i^{\widetilde{\zeta}_{k,j}^Y}$ *and* $\widetilde{\zeta}_k^Y\cdot a(x)=\sum_{j\in Y_c}\widetilde{\zeta}_{k,j}^Y u_j(x)$ *are determined by the solutions to the stated quadratic and linear equations, and the associated eigenfunction is given by* $e^{h(x,y)}=\prod_{i\in Y_d} z_i^{y_i}\prod_{i\in Y_c} e^{u_j(x)y_j}$.

Proof. The generator of the fast process has an additional term from the continuous component

$$L_1^{x,p}f(z)=\sum_{k:\beta_k=1}\lambda_k(z)(f(x,y+\widetilde{\zeta}_k^Y)-f(x,y))+\sum_{k:\beta_k>1}\lambda_k(z)\widetilde{\zeta}_k^Y\cdot\nabla_Y f(x,y),$$

so if $a(x)=\{a_i(x)\}$ using $h(x,y)=\sum_i a_i(x)y_i$ as before Eq. (3.11) now becomes

$$\begin{aligned}&\big(V(y;x,p)+e^{-h(x,y)}L_1^{x,p}e^{h(x,y)}\big)\\ &=\sum_{k:\theta_{k,0}=1}\lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}_k^X\cdot p+\sum_i a_i(x)\widetilde{\zeta}_{k,i}^Y}-1)+\mathbf{1}_{\beta_k>1}(\widetilde{\zeta}_k^X\cdot p+\widetilde{\zeta}_k^Y\cdot a(x))\Big)\\ &\quad+\sum_{i\in\mathcal{I}^c}y_i\sum_{k:\theta_{k,i}=1}\lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}_k^X\cdot p+\sum_i a_i(x)\widetilde{\zeta}_{k,i}^Y}-1)+\mathbf{1}_{\beta_k>1}(\widetilde{\zeta}_k^X\cdot p+\widetilde{\zeta}_k^Y\cdot a(x))\Big)\\ &\quad+\sum_{i\in\mathcal{I}}y_i\sum_{k:\theta_{k,i}=1}\lambda_k(x)\Big(\mathbf{1}_{\beta_k=1}(e^{\widetilde{\zeta}_k^X\cdot p+\sum_i a_i(x)\widetilde{\zeta}_{k,i}^Y}-1)+\mathbf{1}_{\beta_k>1}(\widetilde{\zeta}_k^X\cdot p+\widetilde{\zeta}_k^Y\cdot a(x))\Big),\end{aligned}$$

where we rewrite the middle row, using Y_d and Y_c to denote discrete and continuous components respectively, as

$$\sum_{i\in\mathcal{I}^c}y_i\Big(\sum_{j\in Y_d\cap\mathcal{J}^c}e^{a_j(x)}A_{ij}(x,p)+\sum_{j\in Y_c}a_j(x)A_{ij}^c(x,p)+\sum_{j\in(Y_d\cap\mathcal{J}^c)\cup Y_c}B_{ij}(x,p)\Big),$$

and likewise the last row as

$$\sum_{i\in\mathcal{I}}y_i\sum_{j\in Y_d\cap\mathcal{J}}\big(e^{a_j(x)}A_{jj}(x,p)+B_{ij}(x,p)+e^{-a_j(x)}C_{ij}(x,p)\big),$$

where the functions $A_{ij}(x)$, $B_i(x)$, $C_i(x)$ are the same as in (4.4),(4.5),(4.6), and the only new contribution is from the function

$$A_{ij}^c(x,p)=\sum_{k:\theta_{k,i}=1}\lambda_k(x)\mathbf{1}_{\beta_k>1}\widetilde{\zeta}_{k,j}^Y-\mathbf{1}_{i'\sim i}(\widetilde{\theta}_i/\widetilde{\theta}_{i'})\sum_{k:\theta_{k,i'}=1}\lambda_k(x)\mathbf{1}_{\beta_k>1}\widetilde{\zeta}_{k,j}^Y. \quad (4.8)$$

Otherwise the rest of the process of solving for the result is the same. □

The condition precludes systems with reactions that are bi-molecular in fast species, which is an assumption we made for the sake of simplicity of the proposition. It is in principle clear

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

how one can try to extend the above derivation in case of multi-molecular reactions between fast species to get a solution to the eigenvalue problem.

We will show how the procedure for solving the eigenvalue problem works in a variety of examples, including the two mentioned in the Introduction.

5. Examples

We assume that reaction rates have mass-action form throughout the following examples, as in (4.1) although in one of the examples (Section 5.2) we will in addition allow some of the chemical reaction constants κ_i to be a function of a specific species of interest in the system. We consider the LDP on $E_X = (0, \infty)$ and E_Y is either $[0, K]$ for some K or $(0, \infty)$ depending on existence of conservation laws in each example.

5.1. Enzymatic kinetics (MM)

We recall the model for enzyme kinetics (Michaelis–Menten), with an inflow of the substrate

$$(0) \quad \emptyset \xrightarrow{\kappa_0'} S \qquad (1,2) \quad S + E \underset{\kappa_2'}{\overset{\kappa_1'}{\rightleftharpoons}} ES \qquad (3) \quad ES \xrightarrow{\kappa_3'} P + E.$$

The scaling of the amounts is implied by the fact that molecular amount of the substrate S is an order of magnitude greater than the amount of enzyme E and of the enzyme–substrate complex ES. Let Z_1, Z_2, Z_3, Z_4 represent the amounts of S, E, ES, P molecular species respectively, the orders of magnitude lead to appropriately scaled species amounts $Z_1^N = Z_1/N, Z_2^N = Z_2, Z_3^N = Z_3, Z_4^N = Z_4/N$. The reaction constants κ_i also have different orders of magnitude, with those for the dissolution of the enzyme–substrate complex in reactions (2, 3) being an order of magnitude larger than the forming of the complex. Let $\kappa_0 = \kappa_0'/N, \kappa_1 = \kappa_1', \kappa_2 = \kappa_2'/N, \kappa_3 = \kappa_3'/N$, so the model of the system is

$$\begin{aligned}
Z_1^N(t) &= Z_1^N(0) + -N^{-1}Y_1(N\kappa_0 t) - N^{-1}Y_1(N\int_0^t \kappa_1 Z_1^N(s)Z_2^N(s)ds) \\
&\quad + N^{-1}Y_2(N\int_0^t \kappa_2(M - Z_2^N(s))ds) \\
Z_2^N(t) &= Z_2^N(0) - Y_1(N\int_0^t \kappa_1 Z_1^N(s)Z_2^N(s)ds) + Y_2(N\int_0^t \kappa_2(M - Z_2^N(s))ds) \\
&\quad + Y_3(N\int_0^t \kappa_3(M - Z_2^N(s))ds) \\
Z_3^N(t) &= Z_3^N(0) + Y_1(N\int_0^t \kappa_1 Z_1^N(s)Z_2^N(s)ds) \\
&\quad - Y_2(N\int_0^t \kappa_2 Z_3^N(s)ds) - Y_3(N\int_0^t \kappa_3 Z_3^N(s)ds) \\
Z_4^N(t) &= N^{-1}Y_3(N\int_0^t \kappa_3 Z_3^N(s)ds).
\end{aligned}$$

There is a conservation law between E and ES since $Z_2^N(t) + Z_3^N(t) \equiv M, \forall t > 0$ hence we will use a change of variables $z_3 = M - z_2$. We also note that P is not an "active" species of the system, as it does not enter on the left hand side of any reaction.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

This leaves a system with the slow and the fast process $X^N = Z_1^N$ and $Y^N = Z_2^N$ respectively. The scaling conditions are clearly satisfied with the time-scale separation between slow X^N and Y^N of order N. The effective dynamics of the slow process is given by the ODE in (1.2) while the effective dynamics of the fast process is a birth–death Markov chain with death rate $\kappa_1 z_1 z_2$ and birth rate $(\kappa_2 + \kappa_3)(M - z_2)$.

The rate of reaction (1) is binary, with $\lambda_1(z) = \kappa_1 z_1 z_2$, but since one of the factors, Z_2, is part of a conservation law and hence bounded by the constant M, by Proposition 4.4 the process has an exponentially good approximation in a sequence of processes, indexed by increasing values of M', in which the rate λ_1 is replaced by $\lambda_1'(z) = \kappa_1 z_1 (z_2 \wedge M') \equiv \lambda_1(z)$ for $M' \geq M$.

The generator for the pair is

$$\begin{aligned} A_N f(x,y) &= N\kappa_0(f(x+N^{-1},y) - f(x,y)) \\ &+ N\kappa_1 xy(f(x-N^{-1},y-1) - f(x,y)) \\ &+ N\kappa_2(M-y)(f(x+N^{-1},y+1) - f(x,y)) \\ &+ N\kappa_3(M-y)(f(x,y+N^{-1}) - f(x,y)), \end{aligned} \tag{5.1}$$

so the exponential generator, acting on $f_N(x,y) = f(x) + N^{-1}h(x,y)$, and its limit are

$$\begin{aligned} H_N f_N(x,y) &= \kappa_0(e^{N(f(x+N^{-1})-f(x))+h(x+N^{-1},y)-h(x,y)} - 1) \\ &+ \kappa_1 xy(e^{N(f(x-N^{-1})-f(x))+h(x-N^{-1},y-1)-h(x,y)} - 1) \\ &+ \kappa_2(M-y)(e^{N(f(x+N^{-1})-f(x))+h(x+N^{-1},y+1)-h(x,y)} - 1) \\ &+ \kappa_3(M-y)(e^{h(x,y+1)-h(x,y)} - 1), \end{aligned}$$

$$\begin{aligned} \lim_{N\to\infty} H_N f_N(x,y) &= \kappa_0 e^{f'(x)} + \kappa_1 xy(e^{-f'(x)}e^{h(x,y-1)-h(x,y)} - 1) \\ &\quad + (\kappa_2+\kappa_3)(M-y)(e^{f'(x)}e^{h(x,y+1)-h(x,y)} - 1) \\ &= \overline{H}_0(x, f'(x)), \end{aligned}$$

so that $\overline{H}_0(x,p)$ satisfies the EVP equation (3.11) with

$$\begin{aligned} V(y;x,p) &= \kappa_0 e^p + \kappa_1 xy(e^{-p} - 1) + \kappa_2(M-y)(e^p - 1), \\ L_1^{x,p} e^{g(x,y)} &= \kappa_1 xy e^{-p}(e^{g(x,y-1)} - e^{g(x,y)}) + (\kappa_2+\kappa_3)(M-y)(e^{g(x,y+1)} - e^{g(x,y)}). \end{aligned}$$

Since the perturbed effective fast dynamics is a birth–death Markov chain with birth rate $(\kappa_2 + \kappa_3)(M - y)$ and death rate $\kappa_1 xye^{-p}$ the density Condition 3.3 is satisfied, and there is a unique stationary distribution which is a Binomial$(M, \pi^{x,p})$ with size parameter M and probability parameter $\pi^{x,p} = (\kappa_2 + \kappa_3)/(\kappa_2 + \kappa_3 + \kappa_1 xe^{-p})$. Since the state space for $Y^{x,p}$ is $\{0, 1, \ldots, M\}$ both Lyapunov Conditions 3.4 are trivially satisfied.

In order to solve the EVP for $\overline{H}_0(x,p)$ above let $g(x,y) = a(x)y$, then

$$\begin{aligned} \overline{H}_0(x,y) &= \kappa_1 xy(e^{-p}e^{-a(x)} - 1) + (\kappa_2+\kappa_3)(M-y)(e^{a(x)} - 1) \\ &= y\left(\kappa_1 x(e^{-p}e^{-a(x)} - 1) - \kappa_2(e^p e^{a(x)} - 1) - \kappa_3(e^{a(x)} - 1)\right) \\ &\quad + M\left(\kappa_2(e^p e^{a(x)} - 1) + \kappa_3(e^{a(x)} - 1)\right) + \kappa_0(e^p - 1), \end{aligned}$$

which after setting the coefficient of y to zero gives a single quadratic equation

$$(\kappa_2 e^p + \kappa_3)e^{2a(x)} + (\kappa_1 x - \kappa_2 - \kappa_3)e^{a(x)} - \kappa_1 xe^{-p} = 0.$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Since there was a conservation law in the fast variable the quadratic coefficients can satisfy $A(x)C(x) < 0$ and produce a unique positive solution

$$e^{a(x)} = \frac{-(\kappa_1 x - \kappa_2 - \kappa_3) + \sqrt{(\kappa_1 x - \kappa_2 - \kappa_3)^2 + 4(\kappa_2 e^p + \kappa_3)\kappa_1 x e^{-p}}}{2(\kappa_2 e^p + \kappa_3)},$$

which by Proposition 4.6 implies

$$\overline{H}_0(x, p) = \frac{M}{2}\left(-\kappa_2 - \kappa_3 - \kappa_1 x + \sqrt{(\kappa_2 + \kappa_3 - \kappa_1 x)^2 + 4(\kappa_2 + \kappa_3 e^{-p})\kappa_1 x}\right) + \kappa_0(e^p - 1).$$

From the explicit equation it is clear that on $(0, \infty) \times \mathbb{R}$ the function $\overline{H}_0(x, p)$ is convex and coercive in p and it satisfies the condition (3.16).

5.2. *Self-regulated gene expression (SRG)*

Another common example from systems biology is that of self-regulated gene expression

$$(1,2)\quad G_0 \underset{\kappa_2'(P)}{\overset{\kappa_1'(P)}{\rightleftharpoons}} G_1 \qquad (3,4)\quad G_1 \overset{\kappa_3'}{\rightharpoonup} G_1 + R, \qquad R \overset{\kappa_4'}{\rightharpoonup} R + P \qquad (5,6)\quad R \overset{\kappa_5'}{\rightharpoonup} \emptyset, \qquad P \overset{\kappa_6'}{\rightharpoonup} \emptyset,$$

where G_0, G_1 are inactive and active molecular conformations of a gene, R is the MRNA, and P is a protein expressed by this gene. The self-regulation of the gene comes via the protein it expresses, whose amount affects the rate at which the active conformation becomes an inactive conformation and vice versa.

We assume there is one copy of the gene, and the amount of MRNA and protein in the long run is expressed in terms of a scaling parameter N. The two conformations of the gene are rapidly changing their states, and the active one is rapidly involved in expression, so the rates $\kappa_1', \kappa_2', \kappa_3'$ are assumed to be of order N, while the translation rate and the MRNA and protein degradation rates $\kappa_4', \kappa_5', \kappa_6'$ are of order 1. Let Z_1, Z_2, Z_3, Z_4 represent the amounts of G_0, G_1, R, P respectively, so their rescaled versions are $Z_1^N = Z_1, Z_2^N = Z_2, Z_3^N = Z_3/N, Z_4^N = Z_4/N$. Let $\kappa_1(Z_3/N) = \kappa_1'(Z_3)/N, \kappa_2(Z_3/N) = \kappa_2'(Z_3)/N, \kappa_3 = \kappa_3'/N$ and $\kappa_4 = \kappa_4', \kappa_5 = \kappa_5', \kappa_6 = \kappa_6'$. We assume the self-regulating rates $\kappa_1(\cdot), \kappa_2(\cdot)$ are Lipschitz and grow at most linearly, and to prevent absorption of the system we will assume they are positive.

There is a conservation law between G_0 ad G_1 since $G_0(t) + G_1(t) \equiv 1, \forall t$ so we use $z_2 = 1 - z_1$. The model of the system is

$$\begin{aligned}
Z_1^N(t) &= Z_1^N(0) - Y_1\Big(\int_0^t N\kappa_1(Z_4^N(s))Z_1^N(t)ds\Big) + Y_2\Big(\int_0^t N\kappa_2(Z_4^N(s))(1 - Z_1^N(s))ds\Big)\\
Z_3^N(t) &= N^{-1}Y_3\Big(\int_0^t N\kappa_3(1 - Z_1^N(s))ds\Big) - N^{-1}Y_5\Big(\int_0^t N\kappa_5 Z_3^N(s)ds\Big)\\
Z_4^N(t) &= N^{-1}Y_4\Big(\int_0^t N\kappa_4 Z_3^N(s)ds\Big) - N^{-1}Y_6\Big(\int_0^t N\kappa_6 Z_4^N(s)ds\Big).
\end{aligned}$$

The slow and fast process are $(X_1^N, X_2^N) = (Z_3^N, Z_4^N)$ and $Y^N = Z_1^N$ respectively, with the time-scale separation of order N, and the generator of the pair is

$$\begin{aligned}
&A_N f(x, y)\\
&= N\kappa_1(x_2)y(f(x, y-1) - f(x, y)) + N\kappa_2(x_2)(1-y)(f(x, y+1) - f(x, y))\\
&+ N\kappa_3(1-y)(f(x_1 + N^{-1}, x_2, y) - f(x, y)) + N\kappa_4 x_1(f(x_1, x_2 + N^{-1}, y) - f(x, y))\\
&+ N\kappa_5 x_1(f(x_1 - N^{-1}, x_2, y) - f(x, y)) + N\kappa_6 x_2(f(x_1, x_2 - N^{-1}, y) - f(x, y)).
\end{aligned}$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

The effective dynamics of the fast process is a Markov chain on $\{0, 1\}$ with $1 \mapsto 0$ rate $\kappa_1(x_2)y$ and $0 \mapsto 1$ rate $\kappa_2(x_2)(1-y)$ and a unique stationary distribution that is Bernoulli with probability $\pi^x = \kappa_2(x_2)/(\kappa_2(x_2)+\kappa_1(x_2))$. This implies that the effective dynamics of the slow process is given by the system of ODEs

$$\begin{aligned}\dot{x}_1(t) &= \frac{\kappa_3\kappa_1(x_2(t))}{\kappa_2(x_2(t))+\kappa_1(x_2(t))} - \kappa_5 x_1(t)\\ \dot{x}_2(t) &= \kappa_4 x_1(t) - \kappa_6 x_2(t).\end{aligned}$$

The exponential generator acting on $f_N(x, y) = f(x) + N^{-1}g(x, y)$, and its limit are

$$\begin{aligned}H_N f_N(x, y) &= \kappa_1(x_2)y(e^{g(x,y-1)-g(x,y)} - 1)\\ &+ \kappa_2(x_2)(1-y)(e^{g(x,y+1)-g(x,y)} - 1)\\ &+ \kappa_3(1-y)(e^{N(f(x_1+N^{-1},x_2)-f(x))+g(x_1+N^{-1},x_2,y)-g(x,y)} - 1)\\ &+ \kappa_4 x_1(e^{N(f(x_1,x_2+N^{-1})-f(x))+g(x_1,x_2+N^{-1},y)-g(x,y)} - 1)\\ &+ \kappa_5 x_1(e^{N(f(x_1-N^{-1},x_2)-f(x))+g(x_1-N^{-1},x_2,y)-g(x,y)} - 1)\\ &+ \kappa_6 x_2(e^{N(f(x_1,x_2-N^{-1})-f(x))+g(x_1,x_2-N^{-1},y)-g(x,y)} - 1),\\ \lim_{N\to\infty} H_N f_N(x, y) &= \kappa_1(x)y(e^{g(x,y-1)-g(x,y)} - 1) + \kappa_2(x)(1-y)(e^{g(x,y+1)-g(x,y)} - 1)\\ &+ \kappa_3(1-y)(e^{\partial_{x_1} f(x)} - 1) + \kappa_4 x_1(e^{\partial_{x_2} f(x)} - 1)\\ &+ \kappa_5 x_1(e^{-\partial_{x_1} f(x)} - 1) + \kappa_6 x_2(e^{-\partial_{x_2} f(x)} - 1)\end{aligned}$$

and we can identify V and $L_1^{x,p}$ as

$$\begin{aligned}V(y; x, p) &= \kappa_3(1-y)(e^{p_1} - 1) + \kappa_4 x_1(e^{p_2} - 1) + \kappa_5 x_1(e^{-p_1} - 1) + \kappa_6 x_2(e^{-p_2} - 1),\\ L_1^{x,p} e^{g(x,y)} &= \kappa_1(x_2)y(e^{g(x,y-1)} - e^{g(x,y)}) + \kappa_2(x_2)(1-y)(e^{g(x,y+1)} - e^{g(x,y)}).\end{aligned}$$

Since the perturbed effective fast dynamics is a simple $\{0, 1\}$ Markov chain with positive transition rates (same as in the unperturbed case) the density Condition 3.3 is satisfied, and both Lyapunov Conditions 3.4 and 3.5 are trivially satisfied.

Letting $g(x, y) = a(x)y$, implies

$$\begin{aligned}\overline{H}_0(x, p) &= \kappa_1(x_2)y(e^{-a(x)} - 1) + \kappa_2(x_2)(1-y)(e^{a(x)} - 1) + \kappa_3(1-y)(e^{p_1} - 1)\\ &+ \kappa_4 x_1(e^{p_2} - 1) + \kappa_5 x_1(e^{-p_1} - 1) + \kappa_6 x_2(e^{-p_2} - 1)\\ &= y\big(\kappa_1(x_2)(e^{-a(x)} - 1) - \kappa_2(x_2)(e^{a(x)} - 1) - \kappa_3(e^{p_1} - 1)\big)\\ &+ \kappa_2(x_2)(e^{a(x)} - 1) + \kappa_3(e^{p} - 1) + \kappa_4 x_1(e^{p_2} - 1) + \kappa_5 x_1(e^{-p_1} - 1)\\ &+ \kappa_6 x_2(e^{-p_2} - 1),\end{aligned}$$

which after setting the coefficient of y to 0 gives the quadratic equation

$$-\kappa_2(x_2)e^{2a(x)} + (\kappa_2(x_2) - \kappa_1(x_2) - \kappa_3(e^{p_1} - 1))e^{a(x)} + \kappa_1(x_2) = 0,$$

with $A(x)C(x) < 0$ and one positive solution (regardless of the functions $\kappa_1(\cdot)$, $\kappa_2(\cdot)$ or the values of reaction coefficients κ_i, $i = 3, 4, 5, 6$)

$$e^{a(x)} = \frac{\kappa_2(x_2) - \kappa_1(x_2) - \kappa_3(e^{p_1} - 1) + \sqrt{(\kappa_2(x_2) - \kappa_1(x_2) - \kappa_3(e^{p_1} - 1))^2 + 4\kappa_2(x_2)\kappa_1(x_2)}}{2\kappa_2(x_2)}.$$

By Proposition 4.6

$$\begin{aligned}\overline{H}_0(x, p) &= -\kappa_2(x_2) - \kappa_1(x_2) + \sqrt{(\kappa_2(x_2) - \kappa_1(x_2) - \kappa_3(e^{p_1} - 1))^2 + 4\kappa_2(x_2)\kappa_1(x_2)}\\ &+ \kappa_4 x_1(e^{p_2} - 1) + \kappa_5 x_1(e^{-p_1} - 1) + \kappa_6 x_2(e^{-p_2} - 1).\end{aligned}$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Since the rates $\kappa_1(x), \kappa_2(x)$ are assumed Lipschitz in x, with at most linear growth, $\overline{H}_0(x, p)$ is convex and coercive in p and satisfies the condition (3.16) in $(0, \infty) \times \mathbb{R}$ for arbitrary $\epsilon > 0$. In [25] authors obtain the same large deviation principle for this example (note that the slow and fast variables are labelled differently there).

5.3. Down-regulation (DR)

We consider a simple model of a negative self-regulation mechanism

$$(0) \quad \emptyset \xrightarrow{\kappa_0'} A \qquad (1) \quad A + B \xrightarrow{\kappa_1'} \emptyset \qquad (2) \quad A \xrightarrow{\kappa_2'} A + B \qquad (3) \quad B \underset{\kappa_4'}{\overset{\kappa_3'}{\rightleftharpoons}} \emptyset,$$

where A is a species of interest and B is a species used to down-regulate it, namely, A controls its own molecular amount by producing more of the regulating species B. Suppose the molecular amounts and rates satisfy

$$|A| = O(N) \mapsto X^N = |A|/N, \ |B| = O(1), \quad \kappa_1', \kappa_2' \sim O(1), \ \kappa_0', \kappa_3', \kappa_4' \sim O(N).$$

Let Z_1, Z_2 represent the amounts of A and B molecules respectively, so that the rescaled versions are $Z_1^N = Z_1/N, Z_2^N = Z_2$. Let $\kappa_1' = \kappa_1, \kappa_2' = \kappa_2$ and $\kappa_3 = \kappa_3' N, \kappa_4 = \kappa_4' N$. The model for the system is

$$\begin{aligned} Z_1^N(t) &= Z_1^N(0) + N^{-1} Y_0(\kappa_0 t) - N^{-1} Y_1\Big(\int_0^t N\kappa_1 Z_1^N(s) Z_2^N(s) ds\Big) \\ Z_2^N(t) &= Z_2^N(0) - Y_1\Big(\int_0^t N\kappa_1 Z_1^N(s) Z_2^N(s) ds\Big) + Y_2\Big(\int_0^t N\kappa_2 Z_1^N(s) ds\Big) \\ &\quad - Y_3\Big(\int_0^t N\kappa_3 Z_2^N(s) ds\Big) + Y_4\Big(\int_0^t N\kappa_4 ds\Big). \end{aligned}$$

The slow and the fast process are $X^N = Z_1^N$ and $Y^N = Z_2^N$ respectively, and the generator of the pair is

$$\begin{aligned} A_N f(x, y) &= N\kappa_0 (f(x + N^{-1}, y) - f(x, y)) + N\kappa_1 yx (f(x - N^{-1}, y - 1) - f(x, y)) \\ &\quad + N\kappa_2 x (f(x, y + 1) - f(x, y)) + N\kappa_3 y (f(x, y - 1) - f(x, y)) \\ &\quad + N\kappa_4 (f(x, y + 1) - f(x, y)). \end{aligned}$$

The effective dynamics of the fast process is a simple birth and death Markov chain on $\mathbf{Z}_+$ with birth rate $\kappa_2 x + \kappa_4$ and death rate $\kappa_1 xy + \kappa_3 y$ with a unique stationary distribution that is Poisson(μ^x) with parameter $\mu^x = (\kappa_2 x + \kappa_4)/(\kappa_1 x + \kappa_3)$. The effective dynamics of the slow process x is a solution to the ODE

$$\dot{x}(t) = \kappa_0 - \kappa_1 x(t) \frac{\kappa_2 x(t) + \kappa_4}{\kappa_1 x(t) + \kappa_3}.$$

The rate of reaction (1) is binary, $\lambda_1 = \kappa_1 xy$, but one of the factors, x, is the variable for the slow process which has constant rate of increase. Hence, by Proposition 4.4, this model has an exponentially good approximation in a sequence of processes, indexed by increasing values of M', where only the rate of reaction (1) is replaced by $\lambda_1'(z) = \kappa_1 y(x \wedge M')$.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

The exponential generator for $f_N(x, y) = f(x) + N^{-1}g(x, y)$ is

$$\begin{aligned} H_N f_N(x, y) &= \kappa_0(e^{N(f(x+N^{-1})-f(x))} - 1) \\ &+ \kappa_1 xy(e^{N(f(x-N^{-1})-f(x))+g(x-N^{-1},y-1)-g(x,y)} - 1) \\ &+ \kappa_2 x(e^{g(x,y+1)-g(x,y)} - 1) \\ &+ \kappa_3 y(e^{g(x,y-1)-g(x,y)} - 1) \\ &+ \kappa_4(e^{g(x,y+1)-g(x,y)} - 1), \end{aligned}$$

so that its limit $\overline{H}_0(x, p)$ solves the eigenvalue problem (3.11) with

$$\begin{aligned} V(y; x, p) &= \kappa_0(e^p - 1) + \kappa_1 xy(e^{-p} - 1), \\ L_1^{x,p} e^{g(x,y)} &= \kappa_1 xye^{-p}(e^{g(x,y-1)} - e^g) + \kappa_2 x(e^{g(x,y+1)} - e^g) + \kappa_3 y(e^{g(x,y-1)} - e^g) \\ &\quad + \kappa_4(e^{g(x,y+1)} - e^g). \end{aligned}$$

Letting $g(x, y) = a(x)y$ we get

$$\begin{aligned} \overline{H}_0(x, p) &= \kappa_0(e^p - 1) + \kappa_1 xy(e^{-p-a(x)} - 1) + \kappa_2 x(e^{a(x)} - 1) + \kappa_3 y(e^{-a(x)} - 1) \\ &\quad + \kappa_4(e^{a(x)} - 1) \\ &= y\big(\kappa_1 x(e^{-p-a(x)} - 1) + \kappa_0(e^p - 1) + \kappa_3(e^{-a(x)} - 1)\big) + \kappa_2 x(e^{a(x)} - 1) \\ &\quad + \kappa_4(e^{a(x)} - 1). \end{aligned}$$

In order to solve the EVP we set the coefficient of y to 0 which gives the equation

$$e^{a(x)} = \frac{\kappa_1 xe^{-p} + \kappa_3}{\kappa_1 x + \kappa_3} > 0, \forall x \geq 0, \forall p,$$

and hence

$$\overline{H}_0(x, p) = \kappa_0(e^p - 1) + \frac{(\kappa_2 x + \kappa_4)(\kappa_1 xe^{-p} + \kappa_3)}{\kappa_1 x + \kappa_3},$$

which is convex and coercive in p and satisfies (3.16) in $(0, \infty) \times \mathbb{R}$. The effect of the needed truncation will be the replacement of $\kappa_1 x$ by $\kappa_1(x \wedge M')$ in the above formula for $\overline{H}_0$.

To satisfy Condition 3.5 we note that $|V(y; x, p)| \to \infty$ as $y \to \infty$. For any $c > 1$ we let $\varphi_{x,p} = a_{x,p}y$, with $a_{x,p}$ to be chosen. Then calculating as above

$$\begin{aligned} &e^{-\varphi_{x,p}(y)} L_1^{x,p} e^{\varphi_{x,p}(y)} + c|V(y; x, p)| \\ &\quad = \kappa_1 xye^{-p}(e^{-a_{x,p}} - 1) + \kappa_2 x(e^{a_{x,p}} - 1) + \kappa_3 y(e^{-a_{x,p}} - 1) + \kappa_4(e^{a_{x,p}} - 1) \\ &\qquad + c\kappa_1 xy|e^{-p} - 1| \\ &\quad = y\big((\kappa_1 xe^{-p} + \kappa_3)(e^{-a_{x,p}} - 1) + c\kappa_1 x|e^{-p} - 1|\big) + (\kappa_2 x + \kappa_4)(e^{a_{x,p}} - 1), \end{aligned}$$

and choosing $a_{x,p}$ which sets the coefficient of y to 0

$$e^{a_{x,p}} = \frac{\kappa_1 xe^{-p} + \kappa_3}{\kappa_1 xe^{-p} + \kappa_3 - c\kappa_1 x|e^{-p} - 1|} \geq 1$$

implies $\varphi_{x,p} = a_{x,p}y \to \infty$ as $y \to \infty$ and

$$e^{-\varphi_{x,p}(y)} L_1^{x,p} e^{\varphi_{x,p}(y)} + c|V(y; x, p)| = (\kappa_2 x + \kappa_4)(e^{a_{x,p}} - 1), \ \forall y \in E_Y,$$

with the right hand side $d = (\kappa_2 x + \kappa_4)(e^{a_{x,p}} - 1) \in [0, \infty)$ being independent of y, as needed for (3.13).

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

5.4. Viral production (VP)

A somewhat more complicated model is one for production of packaged virus particles

$$\begin{array}{llll} (1) & \text{stuff} \xrightarrow{\kappa_1'} G & (2) & G \xrightarrow{\kappa_2'} T \\ (3) & T \xrightarrow{\kappa_3'} T + S & (4) & T \xrightarrow{\kappa_4'} \emptyset \\ (5) & S \xrightarrow{\kappa_5'} \emptyset & (6) & G + T + (S) \xrightarrow{\kappa_6'} V, \end{array}$$

where T is the viral template, G the viral genome, S the viral structural protein and V the packaged virus. The virus has very few templates from which it manages to co-opt the cell's MRNA to make a relatively large copy number of its genomes, and an order of magnitude larger number of viral structural proteins. Letting Z_1, Z_2, Z_3, Z_4 denote the amounts of species T, G, S, V respectively, the appropriate rescaling gives $Z_1^N = Z_1, Z_2^N = Z_2/N^{2/3}, Z_3^N = Z_3/N, Z_4^N = Z_4/N^{2/3}$. The chemical rates also have a range of orders of magnitude, relative to species rescaling they are best expressed by $\kappa_1 = \kappa_1', \kappa_2 = \kappa_2' N^{2/3}, \kappa_3 = \kappa_3'/N, \kappa_4 = \kappa_4', \kappa_5 = \kappa_5', \kappa_6 = \kappa_6' N^{5/3}$. The only reaction that is not in standard mass-action form is (6) where the effect of viral proteins is felt only in terms of their order of magnitude, and the usual dependence on the amounts is binary in the amounts of viral templates and genomes $\lambda_6(z) = \kappa_6 z_2 z_1$. The model for this system is

$$\begin{aligned} Z_1^N(t) &= Z_1^N(0) + Y_2\Big(\int_0^t N^{2/3}\kappa_2 Z_2^N(s)ds\Big) - Y_4\Big(\int_0^t N^{2/3}\kappa_4 Z_1^N(s)ds\Big) \\ &\quad - Y_6\Big(\int_0^t N^{2/3}\kappa_6 Z_1^N(s)Z_2^N(s)ds\Big) \\ Z_2^N(t) &= Z_2^N(0) + N^{-2/3}Y_1\Big(\int_0^t N^{2/3}\kappa_1 ds\Big) - N^{-2/3}Y_2\Big(\int_0^t N^{2/3}\kappa_2 Z_2^N(s)ds\Big) \\ &\quad - N^{-2/3}Y_6\Big(\int_0^t N^{2/3}\kappa_6 Z_1^N(s)Z_2^N(s)ds\Big) \\ Z_3^N(t) &= Z_3^N(0) + N^{-1}Y_3\Big(\int_0^t N^{5/3}\kappa_3 Z_1^N(s)ds\Big) - N^{-1}Y_5\Big(\int_0^t N^{5/3}\kappa_5 Z_3^N(s)ds\Big) \\ &\quad - N^{-1}Y_6\Big(\int_0^t N^{2/3}\kappa_6 Z_1^N(s)Z_2^N(s)ds\Big) \\ Z_4^N(t) &= Z^N(0) + N^{-2/3}Y_6\Big(\int_0^t N^{2/3}\kappa_6 Z_1^N(s)Z_2^N(s)ds\Big). \end{aligned}$$

The packaged virus V is the final product but not an "active" species in the system, and can be tracked from knowledge on the behaviour of Z_2^N. The slow and fast processes are respectively $X^N = Z_2^N$ and $Y^N = (Z_1^N, Z_3^N)$ and the time-scale separation is now $N^{2/3}$ which will also give the scaling for the large deviation asymptotics as $N^{2/3}$, rather than N.

This is a modified version of the viral production model considered in [4], where we showed that the effective dynamics of the fast process Y^N is a piecewise deterministic Markov process with discrete component Y_1^N and continuous component Y_2^N. The discrete component is a birth–death Markov chain with birth rate $\kappa_2 x$ and death rate $\kappa_4 y_1 + \kappa_6 x y_1$. The continuous component follows the ODE $\dot{y_2}(t) = \kappa_3 y_1 - \kappa_5 y_2$ which depends on the value y_1 of the discrete component

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Y_1^N. This process has a unique stationary distribution $\mu^x(y_1, y_2)$ which satisfies

$$\int \Big[\kappa_2 x\,(g(y_1+1, y_2) - g(y_1, y_2)) + (\kappa_4 y_1 + \kappa_6 x y_1)\,(g(y_1-1, y_2) - g(y_1, y_2)) \\ + (\kappa_3 y_1 - \kappa_5 y_2)\partial_{y_2} g(y_1, y_2)\Big]\mu_x(y_1, y_2) = 0.$$

In particular for the discrete component μ^x has a Poisson(m^x) distribution with parameter $m^x = \kappa_2 x/(\kappa_4 + \kappa_6 x)$ so $E_{\mu^x}[Y_1] = V_{\mu^x}[Y_1] = m^x$. Moreover, the mean and variance of the continuous component satisfy $E_{\mu^x}[Y_2] = (\kappa_3/\kappa_5)m^x$. Using the above results and averaging techniques it can be shown (in the same way as in [4]) that the effective dynamics of the slow process is given by the ODE

$$\dot{x}(t) = \kappa_1 - \kappa_2 x(t)dt - \kappa_6 \frac{\kappa_3}{\kappa_5}\frac{\kappa_2 x(t)}{\kappa_4 + \kappa_6 x(t)} x(t)dt.$$

The rate of reaction (6) is binary, $\lambda_6 = \kappa_6 x y_1$, but one of the factors is the variable for the slow process which has a constant rate of increase. By Proposition 4.4, this model has an exponentially good approximation in a sequence of processes, indexed by increasing values of M', where only the rate of reaction (6) is replaced by $\lambda_6'(z) = \kappa_6 y_2 (x \wedge M')$. The effect of the needed truncation will be the replacement of $\kappa_6 x$ by $\kappa_6(x \wedge M')$ in the final formula for $\overline{H}_0$.

The generator for the triple is

$$\begin{aligned} &A_N f(x, y_1, y_2) = \\ &\quad N\kappa_1\left(f(x + N^{-2/3}, y_1, y_2) - f(x, y_1, y_2)\right) + N^{2/3}\kappa_2 x\left(f(x - N^{2/3}, y_1 + 1, y_2) - f(x, y_1, y_2)\right) \\ &+ N^{5/3}\kappa_3 y_1\left(f(x, y_1, y_2 + N^{-1}) - f(x, y_1, y_2)\right) + N^{2/3}\kappa_4 y_1\,(f(x, y_1 - 1, y_2) - f(x, y_1, y_2)) \\ &+ N^{5/3}\kappa_5 y_2\left(f(x, y_1, y_2 - N^{-1}) - f(x, y_1, y_2)\right) + N^{2/3}\kappa_6 x y_1\left(f(x - N^{2/3}, y_1 - 1, y_2) - f(x, y_1, y_2)\right), \end{aligned} \tag{5.2}$$

so the exponential generator, on functions of the form $f_N(x, y) = f(x) + N^{-2/3}g(x, y_1, y_2)$ is

$$\begin{aligned} &H_N f_N(x, y_1, y_2) = \\ &\quad \kappa_1\left(e^{N^{2/3}(f(x+N^{-2/3}) - f(x)) + g(x+N^{-2/3}, y_1, y_2) - g(x, y_1, y_2)} - 1\right) \\ &+ \kappa_2 x\left(e^{N^{2/3}(f(x-N^{2/3}) - f(x)) + g(x-N^{-2/3}, y_1+1, y_2) - g(x, y_1, y_2)} - 1\right) \\ &+ \kappa_3 y_1\left(e^{g(x, y_1, y_2+N^{-1}) - g(x, y_1, y_2)} - 1\right) + \kappa_4 y_1\left(e^{g(x, y_1-1, y_2) - g(x, y_1, y_2)} - 1\right) \\ &+ \kappa_5 y_2\left(e^{g(x, y_1, y_2-N^{-1}) - g(x, y_1, y_2)} - 1\right) \\ &+ \kappa_6 x y_1\left(e^{N^{2/3}(f(x-N^{2/3}) - f(x)) + g(x, y_1-1, y_2) - g(x, y_1, y_2)} - 1\right), \end{aligned}$$

and its limit is

$$\begin{aligned} &\lim_{N\to\infty} H_N f_N(x, y_1, y_2) \\ &\quad = \kappa_1\left(e^{f'(x)} - 1\right) + \kappa_2 x\left(e^{-f'(x) + g(x, y_1+1, y_2) - g(x, y_1, y_2)} - 1\right) + \kappa_3 y_1 \partial_{y_2} g(x, y_1, y_2) \\ &\quad + \kappa_4 y_1\left(e^{g(x, y_1-1, y_2) - g(x, y_1, y_2)} - 1\right) - \kappa_5 y_2 \partial_{y_2} g(x, y_1, y_2) \\ &\quad + \kappa_6 x y_1\left(e^{-f'(x) + g(x, y_1-1, y_2) - g(x, y_1, y_2)} - 1\right). \end{aligned}$$

We can identify the potential function V as

$$V(y_1, y_2; x, p) = \kappa_1(e^p - 1) + \kappa_2 x(e^{-p} - 1) + \kappa_6 x y_1(e^{-p} - 1),$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

and perturbed generator $L_1^{x,p}$ for effective fast process as

$$L_1^{x,p}e^{g(x,y_1,y_2)} = \kappa_2 x e^{-p}(e^{g(x,y_1+1,y_2)} - e^{g(x,y_1,y_2)}) + \kappa_4 y_1(e^{g(x,y_1-1,y_2)} - e^{g(x,y_1,y_2)}) + (\kappa_3 y_1 - \kappa_5 y_2)\partial_{y_2} e^{g(x,y_1,y_2)} + \kappa_6 x y_1 e^{-p}(e^{g(x,y_1-1,y_2)} - e^{g(x,y_1,y_2)}).$$

To solve the eigenvalue problem (3.11) with this V and $L_1^{x,p}$, we let $g(x, y) = a_1(x)y_1 + a_2(x)y_2$ which would imply

$$\overline{H}_0(x, p) = \kappa_1(e^p - 1) + \kappa_2 x(e^{-p} - 1) + \kappa_6 x y_1(e^{-p} - 1) + \kappa_2 x e^{-p}(e^{a_1(x)} - 1) + \kappa_4 y_1(e^{-a_1(x)} - 1) + (\kappa_3 y_1 - \kappa_5 y_2)a_2(x) + \kappa_6 x y_1 e^{-p}(e^{-a_1(x)} - 1).$$

In the reactions which have linear rates in y_1, reactions (4) and (6) have $\widetilde{\zeta}_{4,1}^{Y} = -1$ in the discrete variable y_1, while reaction (3) changes only the continuous variable y_2. Hence we get an equation which is potentially quadratic in $e^{a_1(x)}$, and involves $a_2(x)$ as well. However, Condition 4.7 is satisfied since there exist only reaction (5) which has linear rate in y_2 and it only changes y_2. Thus the equation for $a_2(x)$ can be solved independently of $e^{a_1(x)}$, and the one equation for $e^{a_1(x)}$ potentially quadratic in fact has the quadratic coefficient equal to 0. Setting coefficients of y_1 and y_2 to zero we get

$$a_2(x) = 0,\ e^{-a_1(x)} = \frac{\kappa_4 + \kappa_6 x - \kappa_3 a_2(x)}{\kappa_4 + \kappa_6 x e^{-p}} > 0\ \forall x \geq 0, \forall p$$

$$\overline{H}_0(x, p) = \kappa_1(e^p - 1) + \kappa_2 x(e^{-p} - 1)(1 + \frac{\kappa_6 x e^{-p}}{\kappa_4 + \kappa_6 x}),$$

which is convex and coercive in p and satisfies (3.16) in $(0, \infty) \times \mathbb{R}$. Recall that one needs to replace $\kappa_6 x$ by $\kappa_6(x \wedge M')$ in the above formula for $\overline{H}_0$.

Since the state space for the fast variables in not compact we need to check the Lyapunov Condition 3.5. Note that $|V(y_1, y_2; x, p)| \to \infty$ as $|(y_1, y_2)| \to \infty$. Let y_1^* be large enough so that $(\kappa_6 x y_1^* + \kappa_2 x)|e^{-p} - 1| > \kappa_1(e^p - 1)$. For any $c > 1$ we let $\varphi_{x,p} = a_{x,p} y_1$, then $\forall y_1 > y_1^*, \forall y_2$

$$e^{-\varphi_{x,p}(y)} L_1^{x,p} e^{\varphi_{x,p}(y)} + c|V(y; x, p)| = \kappa_2 x e^{-p}(e^{a_{x,p}} - 1) + \kappa_4 y_1(e^{-a_{x,p}} - 1) + \kappa_6 x y_1 e^{-p}(e^{a_{x,p}} - 1) + c((\kappa_6 x y_1^* + \kappa_2 x)|e^{-p} - 1| - \kappa_1(e^p - 1)),$$

and choosing $a_{x,p}$ to set the coefficients of y_1 to 0

$$e^{-a_{x,p}} = \frac{\kappa_4 + \kappa_6 x - c\kappa_6 x|e^{-p} - 1|}{\kappa_4 + \kappa_6 x} < 1$$

implies $\varphi_{x,p} = a y_1 \to \infty$ as $|y_1| \to \infty$, and

$$e^{-\varphi_{x,p}(y)} L_1^{x,p} e^{\varphi_{x,p}(y)} + c|V(y; x, p)| = \kappa_2 x e^{-p}(e^{a_{x,p}} - 1) + c\kappa_2 x|e^{-p} - 1| - c\kappa_1(e^p - 1),$$

the right-hand side being independent of (y_1, y_2) as needed for (3.13).

Acknowledgements

This work was supported by the Natural Sciences and Engineering Council of Canada (NSERC) Discovery Grant. The author would like to thank an anonymous referee for pointing out a flaw in the statement and the proof of the comparison principle Lemma 3.7 in an earlier version of this manuscript. The author is grateful to Rohini Kumar and to Tom Kurtz for very useful discussions about the material.

Appendix

We adapt the proof of the large deviation principle for the two time-scale jump–diffusions given in [23] to the context of our multiscale Markov chain Z_N. Since the steps of the proof are the same, we give a fairly terse outline and explain carefully the details only where the proof is modified.

A.1. Proof of the LDP Theorem 3.6

Proof. Our proof is based on the viscosity solution to the Cauchy problem for each $h \in C_b(E)$

$$\partial_t u_N = H_N u_N, \text{ in } (0,T] \times E; \quad u_N(0,\cdot) = h(\cdot) \text{ in } E, \tag{A.1}$$

where the non-linear operator is the exponential generator $H_N f = \frac{1}{N} e^{-Nf} A_N e^{Nf}$ for $e^{Nf} \in \mathcal{D}(A_N)$ given by

$$H_N f(z) = \frac{1}{N} \sum_k N^{\beta_k} \lambda_k^N(z) (e^{N(f(z+N^{-\alpha}\zeta_k^N) - f(z))} - 1).$$

The definition of viscosity solutions for these types of non-local partial integro-differential equations (PIDEs) and their properties were given in [24], and various results can be found in [6]. In order to establish the convergence of u_N we need to use a family of integro-differential operators and a sequence of viscosity sub- and super-solutions to associated Cauchy problems.

For $\theta \in (0,1), \xi \in C_c^1(E_Y)$, using φ satisfying (3.12) of the Lyapunov Condition 3.4 we define two sequences (over N) of functions:

$$\begin{aligned} f_{0,N}(x,y) &= f_0(x) + \frac{1}{N} g_0(y), \\ g_0(y) &= (1-\theta)\xi(y) + \theta\varphi(y), \ \ f_0(x) = \phi(x) + \gamma \log(1+x^2); \end{aligned}$$

for some $\gamma > 0$ and $\phi \in C_c^1(E_X)$, and

$$\begin{aligned} f_{1,N}(x,y) &= f_1(x) + \frac{1}{N} g_1(y) \\ g_1(y) &= (1+\theta)\xi(y) - \theta\varphi(y), \ \ f_1(x) = \phi(x) - \gamma \log(1+x^2) \end{aligned}$$

Then,

$$\begin{aligned} &H_N f_{0,N}(x,y) \\ &= \sum_{k:\beta_k=1} \lambda_k^N(z) \big(e^{N(f_0(x+N^{-\alpha}\zeta_k^N) - f_0(x)) + (g_0(y+N^{-\alpha}\zeta_k^N) - g_0(y))} - 1 \big) \\ &+ \sum_{k:\beta_k>1} N^{\beta_k - 1} \lambda_k^N(z) \big(e^{N(f_0(x+N^{-\alpha}\zeta_k^N) - f_0(x)) + (g_0(y+N^{-\alpha}\zeta_k^N) - g_0(y))} \\ &- NN^{-\alpha}\zeta_k^N \cdot \nabla f_0(x) - N^{-\alpha}\zeta_k^N \cdot \nabla g_0(y) - 1 \big) \\ &+ \sum_{k:\beta_k>1} N^{\beta_k} \lambda_k^N(z) N^{-\alpha}\zeta_k \cdot \nabla f_0(x) + \sum_{k:\beta_k>1} N^{\beta_k - 1} \lambda_k^N(z) N^{-\alpha}\zeta_k \cdot \nabla g_0(y) \\ &= \sum_{k:\beta_k=1} \lambda_k^N(z) \big(e^{\widetilde{\zeta}_k^X \cdot \nabla f_0} - 1 \big) + \sum_{k:\beta_k=1} \lambda_k^N(z) e^{\widetilde{\zeta}_k^X \cdot \nabla f_0} \big(e^{g_0(y+\widetilde{\zeta}_k^Y) - g_0(y)} - 1 \big) \\ &+ \sum_{k:\beta_k>1} \lambda_k^N(z) \widetilde{\zeta}_k^X \cdot \nabla f_0 + \sum_{k:\beta_k>1} \lambda_k^N(z) \widetilde{\zeta}_k^Y \cdot \nabla g_0(y) + N^{-1} \varepsilon_N(x,y) \end{aligned}$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

$$= V(y; x, \nabla f_0) + \sum_{k:\beta_k=1} \lambda_k^N(z) e^{\widetilde{\zeta}_k^X \cdot \nabla f_0} \left(e^{g_0(y+\widetilde{\zeta}_k^Y)-g_0(y)} - 1\right) + \sum_{k:\beta_k>1} \lambda_k^N(z) \widetilde{\zeta}_k^Y \cdot \nabla g_0(y) + N^{-1}\varepsilon_N(x, y),$$

with $|\varepsilon_N|$ bounded on compact subsets of $E_X \times E_Y$ (note $\sup_x |\nabla f_0(x)| < \infty$).

Letting $p = \nabla f_0$ we have (by convexity of e)

$$\begin{aligned} &H_N f_{0,N}(x, y) \\ &\quad \le V(y; x, p) + \sum_{k:\beta_k=1} \lambda_k^N(z) e^{\widetilde{\zeta}_k^X \cdot p} \left((1-\theta)(e^{\xi(y+\widetilde{\zeta}_k^Y)-\xi(y)} - 1) + \theta(e^{\varphi(y+\widetilde{\zeta}_k^Y)-\varphi(y)} - 1)\right) \\ &\qquad\qquad + \sum_{k:\beta_k>1} \lambda_k^N(z) \widetilde{\zeta}_k^Y \cdot \left((1-\theta)\nabla\xi(y) + \theta\nabla\varphi(y)\right) + N^{-1}\varepsilon_N(x, y) \\ &\quad = V(y; x, p) + (1-\theta)e^{-\xi} L_1^{x,p} e^{\xi}(y) + \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y) + N^{-1}\varepsilon_N(x, y), \end{aligned}$$

so

$$\limsup_{N\to\infty} H_N f_{0,N}(x, y) \le V(y; x, p) + (1-\theta)e^{-\xi} L_1^{x,p} e^{\xi}(y) + \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y).$$

Similarly

$$\begin{aligned} &H_N f_{1,N}(x, y) \\ &\quad = V(y; x, \nabla f_1) + \sum_{k:\beta_k=1} \lambda_k^N(z) e^{\widetilde{\zeta}_k^X \cdot \nabla f_1} \left(e^{g_1(y+\widetilde{\zeta}_k^Y)-g_1(y)} - 1\right) \\ &\qquad + \sum_{k:\beta_k>1} \lambda_k^N(z) \widetilde{\zeta}_k^Y \cdot \nabla g_1(y) + N^{-1}\varepsilon_N(x, y) \\ &\quad \ge V(y; x, p) + (1+\theta)e^{-\xi} L_1^{x,p} e^{\xi}(y) - \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y) + N^{-1}\varepsilon_N(x, y), \end{aligned}$$

where now $p = \nabla f_1$ (and we used inequalities $e^x - 1 \ge x \ge 1 - e^{-x}$). So

$$\liminf_{N\to\infty} H_N f_{1,N}(x, y) \ge V(y; x, p) + (1+\theta)e^{-\xi} L_1^{x,p} e^{\xi}(y) - \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y).$$

This implies that for any $\{x_N, y_N\}$ contained in a compact subset of $E_X \times E_Y$ with $x_N \to x$

$$\limsup_{N\to\infty} H_N f_{0,N}(x_N, y_N) \le H_0(x, p; \xi, \theta)$$

and

$$\liminf_{N\to\infty} H_N f_{1,N}(x_N, y_N) \ge H_1(x, p; \xi, \theta),$$

where the Lyapunov Condition 3.4 allows us to define two families (over $\theta \in (0, 1)$, $\xi \in C_c^1(E_Y)$) of operators

$$\begin{aligned} H_0(x, p; \xi, \theta) &:= \sup_{y\in E_Y} \{V(y; x, p) + (1-\theta)e^{-\xi} L_1^{x,p} e^{\xi}(y) + \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y)\}, \\ H_1(x, p; \xi, \theta) &:= \inf_{y\in E_Y} \{V(y; x, p) + (1+\theta)e^{-\xi} L_1^{x,p} e^{\xi}(y) - \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y)\}. \end{aligned}$$

It is easily seen (again using Condition 3.4 and the fact that $\xi \in C_c^1(E_Y)$) that for $c > 0$ $\{H_{N,0} f_{0,N} \ge -c\} \cap \{f_{0,N} \le c\}$ and $\{H_{N,0} f_{1,N} \le c\} \cap \{f_{1,N} \ge -c\}$ are contained in compact subsets of $E_X \times E_Y$ and by construction $\{f_{0,N}\}$ and $\{f_{1,N}\}$ converge uniformly on compact subsets of $E_X \times E_Y$ to f_0 and f_1 respectively.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

This technique of constructing upper and lower limits for H^N was developed in [17] for stochastic models with multiple time scales, see Chapter 11 for example of its use in the context of random evolutions.

The defined sequence of functions thus verifies conditions needed to establish the following result (Conditions 3.1 and 3.2 for Lemma 6 in [23]): suppose for a sequence of uniformly bounded (over N) viscosity solutions u_N^h of the Cauchy problem (A.1) we construct the upper semicontinuous regularization $\overline{u}^h$ of the function

$$u_\uparrow^h := \sup_{y_N}\{\limsup_{N\to\infty} u_N^h(t_N, x_N, y_N) : (t_N, x_N) \to (t, x)\}$$

and, similarly, we construct the lower semicontinuous regularization $\underline{u}^h$ of the function

$$u_\downarrow^h := \inf_{y_N}\{\liminf_{N\to\infty} u_N^h(t_N, x_N, y_N) : (t_N, x_N) \to (t, x)\};$$

then (see Lemma 6 in [23] for details of this conclusion) $\overline{u}^h$ is a sub-solution of $\partial_t u \le H_0(x, \nabla u)$ and $\underline{u}^h$ is a super-solution of $\partial_t u \ge H_1(x, \nabla u)$ with the same initial conditions $u(0, \cdot) = h$, where the two non-linear operators above are defined from the two earlier constructed families of operators by

$$H_0(x, p) := \inf_{\xi,\theta} H_0(x, p; \xi, \theta), \qquad H_1(x, p) := \sup_{\xi,\theta} H_1(x, p; \xi, \theta).$$

By definition of $u_\uparrow^h$ and $u_\downarrow^h$ we immediately have $\overline{u}^h \ge \underline{u}^h$, so if we establish the reverse inequality, we will have proved (Lemma 7 [23]) uniform convergence (over compact subsets of $[0, T] \times E_X \times E_Y$) of u_N to u_0 the viscosity solution to the Cauchy problem

$$\partial_t u_0 = \overline{H}_0 u_0, \text{ in } (0, T] \times E; \quad u_0(0, \cdot) = h(\cdot), \text{ in } E. \tag{A.2}$$

To establish $\overline{u}^h \le \underline{u}^h$ we need to verify that the comparison principle holds between sub-solutions of $\partial_t \overline{u}^h \le H_0(x, \nabla \overline{u}^h)$ and super-solutions of $\partial_t \underline{u}^h \ge H_1(x, \nabla \underline{u}^h)$, which will follow if we establish the operator inequality

$$H_0(x, p) \le \overline{H}_0(x, p) \le H_1(x, p), \tag{A.3}$$

where $\overline{H}_0(x, p)$ is defined as the solution to the eigenvalue problem

$$(V(y; x, p) + L_1^{x,p})e^{h(x,y)} = \overline{H}_0(x, p)e^{h(x,y)}. \tag{A.4}$$

To prove the left-hand side we can use the Donsker–Vardhan (see [13]) variational representation of the principal eigenvalue $\overline{H}_0(x, p)$ in (A.4) which is given by

$$\overline{H}_0(x, p) = \sup_{\mu\in\mathcal{P}(E_Y)} \Big(\int_{E_Y} V(y; x, p) d\mu(y) + \inf_{g\in D^{++}(L_1^{x,p})} \int_{E_Y} \frac{L_1^{x,p} g}{g}(y) d\mu(y) \Big) \tag{A.5}$$

(with $D^{++}(L_1^{x,p}) \subset C_b(E_Y)$ denoting functions strictly bounded below by a positive constant). We then repeat the same argument given in Lemma 11.35 of [17] that proves

$$\begin{aligned} H_0(x, p) &= \inf_{\xi,\theta} \sup_{\mu\in\mathcal{P}(E_Y)} \int_{E_Y} \Big(V(y; x, p) + (1-\theta)e^{-\xi} L_1^{x,p} e^{\xi}(y) + \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y) \Big) d\mu(y) \\ &\le \sup_{\mu\in\mathcal{P}(E_Y)} \Big(\int_{E_Y} V(y; x, p) d\mu(y) + \inf_{g\in D^{++}(L_1^{x,p})} \int_{E_Y} \frac{L_1^{x,p} g}{g}(y) d\mu(y) \Big) = \overline{H}_0(x, p) \end{aligned}$$

(note that in [17] operators we denoted as $H_0(x, p)$, $H_1(x, p)$ are indexed by $H_1(x, p)$, $H_2(x, p)$, respectively). The only difference is that we use the density Condition 3.3 and the Lyapunov

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Condition 3.4 on φ to replace their analogous Condition 11.21, and to insure the finiteness of the sum of integrals in the above expression.

To prove the right-hand side we first use Lemma B.10 of [17] which proves the inequality

$$H_1(x,p) = \sup_{\xi,\theta}\inf_{y\in E_Y}\{V(y;x,p) + (1+\theta)e^{-\xi}L_1^{x,p}e^{\xi}(y) - \theta e^{-\varphi}L_1^{x,p}e^{\varphi}(y)\}$$
$$\geq \inf_{\mu\in\mathcal{P}(E_Y)}\liminf_{t\to\infty}\frac{1}{t}\log E^{\mu}\left[e^{\int_0^t V(Y(s);x,p)ds}\right].$$

We next use Lemma 8 of [23] which uses the density Condition 3.3 (and the fact that $\inf_y V(y;x,p) > -\infty$) to insure a uniform (over initial points $Y(0)$) large deviation principle lower bound for the occupation measure of the effective dynamics of the fast process $(Y(s);x,p)$ (given by the generator $L_1^{x,p}$) to prove that for any initial $\mu\in\mathcal{P}(E_Y)$ for $(Y(s);x,p)$ we have the inequality

$$\liminf_{t\to\infty}\frac{1}{t}\log E^{\mu}\left[e^{\int_0^t V(Y(s);x,p)ds}\right] \geq \overline{H}_0(x,p),$$

with $\overline{H}_0(x,p)$ given by the variational form (A.5) above.

Having proved the Operator Inequality (A.3) the constructed sub- and super-solutions sandwich the sequence of viscosity solutions u_N^h, establishing their convergence to the viscosity solution u_0^h, which is by assumption on the limiting Cauchy problem unique. The next step is to prove exponential tightness of the sequence $\{X_N(t)\}$. Exponential tightness of paths of $\{X_N\}$ follows from the convergence of the exponential generators H_N by a standard argument (see Lemma 2 in [23], or equivalently see Corollary 4.17 and a simple calculation from Lemma 4.22 in [17]). Finally, by Bryc formula (Theorem 2.2) we have established a large deviation principle for $\{X_N(t)\}$ with speed $1/N$ and good rate function I given by $u_0^h(t)$ in terms of a variational principle (3.15). □

We now use the solution to the eigenvalue problem (3.11) to simplify the proof of the above theorem by using the associated eigenfunction in forming the family of operators in the definitions of H_0 and H_1. We also replace the Lyapunov Condition 3.4 by the less stringent Condition 3.5 (sufficient for proving multiplicative ergodicity of Markov processes, [22] of single scale processes). Since for many chemical reaction models one can explicitly solve for $\overline{H}_0(x,p)$ as well as verify (3.13), this result is used in all the examples in this paper.

A.2. *Proof of the LDP Corollary 3.8*

Proof. For each $x\in E_X$, $p\in K\subset\mathbb{R}$ let $e^{\xi_{x,p}}\in\mathcal{D}(\overline{H}_0)$ denote the eigenfunction associated with the eigenvalue problem (3.11), that is,

$$V(y;x,p) + e^{-\xi_{x,p}(y)}L_1^{x,p}e^{\xi_{x,p}(y)} = \overline{H}_0(x,p),\ \forall y E_Y.$$

For $\theta\in(0,1)$ we let $\varphi_{x,p}$ be the function satisfying (3.13) of the Lyapunov Condition 3.4, and define the following two sequences of functions:

$$f_{0,N}(x,y) = f_0(x) + \frac{1}{N}((1-\theta)\xi_{x,p}(y) + \theta\varphi_{x,p}(y)),$$
$$f_0(x) = \phi(x) + \gamma\log(1+x^2)$$

for some $\gamma > 0$ and $\phi\in C_c^1(E_X)$, as well as

$$f_{1,N}(x,y) = f_1(x) + \frac{1}{N}((1+\theta)\xi_{x,p}(y) - \theta\varphi_{x,p}(y))$$
$$f_1(x) = \phi(x) - \gamma\log(1+x^2).$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Then, for $p = \nabla f_0$, we have (using eigenfunction property of $\xi_{x,p}$ and (3.13) for $\varphi_{x,p}$)

$$\begin{aligned} H_N f_{0,N}(x,y) &= V(y;x,p) + \sum_{k:\beta_k=1} \lambda_k^N(z) e^{\widetilde{\zeta}_k^X \cdot p} \left(e^{((1-\theta)\xi_{x,p}+\theta\varphi_{x,p})(y+\widetilde{\zeta}_k^Y) - ((1-\theta)\xi_{x,p}+\theta\varphi_{x,p})(y)} - 1 \right) \\ &\quad + \sum_{k:\beta_k>1} \lambda_k^N(z) \widetilde{\zeta}_k^Y \cdot \nabla((1-\theta)\xi_{x,p} + \theta\varphi_{x,p})(y) + N^{-1}\varepsilon(x,y) \\ &\le V(y;x,p) + (1-\theta)e^{-\xi_{x,p}} L_1^{x,p} e^{\xi_{x,p}}(y) + \theta e^{-\varphi_{x,p}} L_1^{x,p} e^{\varphi_{x,p}}(y) + N^{-1}\varepsilon(x,y) \\ &= (1-\theta)\overline{H}_0(x,p) + \theta V(y;x,p) + \theta e^{-\varphi_{x,p}} L_1^{x,p} e^{\varphi_{x,p}}(y) + N^{-1}\varepsilon(x,y) \\ &\le (1-\theta)\overline{H}_0(x,p) - \theta\big((c-1)V(y;x,p) + d\big) + N^{-1}\varepsilon(x,y). \end{aligned}$$

Hence, for $c' > 0$ we have that $\{H_{N,0} f_{0,N} \ge -c'\} \cap \{f_{0,N} \le c'\}$ is contained in compact subset of $E_X \times E_Y$ (since $V(y;x,p)$ has compact level sets and $\overline{H}_0$ is finite), and for any $\{x_N, y_N\}$ in a compact subset of $E_X \times E_Y$ with $x_N \to x$

$$\limsup_{N\to\infty} H_N f_{0,N}(x_N, y_N) \le \sup_{y\in E_Y} \left\{ (1-\theta)\overline{H}_0(x,p) - \theta\big((c-1)V(y;x,p) + d\big) \right\} =: H_0(x,p;\theta)$$

(which by (3.13) is well defined). Similarly, for $p = \nabla f_1$,

$$\begin{aligned} H_N f_{0,N}(x,y) &\ge V(y;x,p) + (1+\theta)e^{-\xi} L_1^{x,p} e^{\xi}(y) - \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y) + N^{-1}\varepsilon(x,y) \\ &= (1+\theta)\overline{H}_0(x,p) - \theta V(y;x,p) - \theta e^{-\varphi} L_1^{x,p} e^{\varphi}(y) + N^{-1}\varepsilon(x,y) \\ &\ge (1+\theta)\overline{H}_0(x,p) + \theta\big((c-1)V(y;x,p) + d\big) + N^{-1}\varepsilon(x,y), \end{aligned}$$

for $c' > 0$ $\{H_{N,1} f_{1,N} \le c'\} \cap \{f_{1,N} \ge -c'\}$ is contained in compact subsets of $E_X \times E_Y$, and for any $\{x_N, y_N\}$ in a compact subset of $E_X \times E_Y$ with $x_N \to x$

$$\limsup_{N\to\infty} H_N f_{1,N}(x_N, y_N) \ge H_1(x,p;\theta) := \inf_{y\in E_Y} \left\{ (1+\theta)\overline{H}_0(x,p) + \theta\big((c-1)V(y;x,p) + d\big) \right\}.$$

As all the conditions of Lemma 6 in [23] are met, it insures that the two functions $\overline{u}^h$ and $\underline{u}^h$ constructed as in the proof of Theorem 3.6 from the sequence of viscosity solutions of $\partial_t u_N = H_N u$ are, respectively, a sub-solution of $\partial_t u = H_0(x, \nabla u)$ and a super-solution of $\partial_t u = H_1(x, \nabla u)$ with the same initial condition. By construction $\overline{u}^h \ge \underline{u}^h$ and the reverse is immediate once we notice that in this case the operators $H_0(x,p)$ and $H_1(x,p)$ given by

$$H_0(x,p) := \inf_{\theta\in(0,1)} \sup_{y\in E_Y} \left\{ (1-\theta)\overline{H}_0(x,p) - \theta\big((c-1)V(y;x,p) + d\big) \right\}$$

and

$$H_1(x,p) := \sup_{\theta\in(0,1)} \inf_{y\in E_Y} \left\{ (1+\theta)\overline{H}_0(x,p) + \theta\big((c-1)V(y;x,p) + d\big) \right\}$$

coincide and are equal to $\overline{H}_0$ (see also [17] Appendix B, Lemma 11.4). □

A.3. *Proof of the comparison principle Lemma 3.7*

Proof. (a) We first prove the comparison principle under the assumptions on $\overline{H}_0$ given in (a). Let $A = \sup_{\{0\}\times E_X} (u_1 - u_2)^+$, and suppose for some $\widetilde{x}$ and $\widetilde{t}$, $u_1(\widetilde{t},\widetilde{x}) - u_2(\widetilde{t},\widetilde{x}) = A + \delta$ $(*)$ with $\delta > 0$. For $\beta, m, \eta > 0$, $N < \infty$ define

$$\psi_N(t,x,s,y) = u_1(t,x) - u_2(s,y) - \frac{N}{2}\left[|x-y|^2 + |t-s|^2\right] - \beta\left(g(x)^m + g(y)^m\right) - \eta(t+s),$$

where $g(x) = \sqrt{1+|x|^2}$. Now choose β, η such that $2\beta g(\widetilde{x})^m + 2\eta\widetilde{t} \le \frac{\delta}{2}$ holds for all $m \le 1$. Then $(*)$ implies

$$\sup_{[0,T]\times E_X\times[0,T]\times E_X} \psi_N(t,x,s,y) \ge \psi_N(\widetilde{t},\widetilde{x},\widetilde{t},\widetilde{x}) = A + \delta - 2\beta g(\widetilde{x})^m - 2\eta\widetilde{t} \ge A + \frac{\delta}{2}, \quad \text{(A.6)}$$

and we show this gives a contradiction.

Since ψ_N is USC and tends to $-\infty$ as $|x|+|y| \to \infty$ its maximum in $[0,T]\times E_X\times[0,T]\times E_X$ is achieved at a point $(\overline{t}_N, \overline{x}_N, \overline{s}_N, \overline{y}_N)$. By (A.6) $\psi_N(\overline{t}_N, \overline{x}_N, \overline{s}_N, \overline{y}_N) \ge A + \frac{\delta}{2}$ and u_1, u_2 bounded imply $\beta(g(\overline{x}_N)^m + g(\overline{y}_N)^m) \le \sup(u_1 - u_2) - A - \frac{\delta}{2} =: c_1 \ge 2\beta$ for all N, all $m \le 1$. Hence, $g(\overline{x}_N), g(\overline{y}_N) \le (\frac{c_1}{\beta})^{1/m}$ and for $R := (\frac{c_1}{\beta})^{1/m} > 0$ we have

$$|\overline{x}_N|, |\overline{y}_N| \le R, \ \forall N < \infty. \quad \text{(A.7)}$$

Using $\psi_N(\overline{t}_N, \overline{x}_N, \overline{t}_N, \overline{x}_N) + \psi_N(\overline{s}_N, \overline{y}_N, \overline{s}_N, \overline{y}_N) \le 2\psi_N(\overline{t}_N, \overline{x}_N, \overline{s}_N, \overline{y}_N)$ we have

$$\frac{N}{2}\left[|\overline{x}_N - \overline{y}_N|^2 + |\overline{t}_N - \overline{s}_N|^2\right] \le \frac{1}{2}\left[u_1(\overline{t}_N, \overline{x}_N) - u_1(\overline{s}_N, \overline{y}_N) + u_2(\overline{t}_N, \overline{x}_N) - u_2(\overline{s}_N, \overline{y}_N)\right] \le c_2,$$

where $c_2 := \sup|u_1| + \sup|u_2|$. Therefore, $|\overline{x}_N - \overline{y}_N| + |\overline{t}_N - \overline{s}_N| \le \sqrt{\frac{2}{N}c_2}$, and $|\overline{x}_N - \overline{y}_N|, |\overline{t}_N - \overline{s}_N| \to 0$ as $N \to \infty$. We can also show $\frac{N}{2}\left[|\overline{x}_N - \overline{y}_N|^2 + |\overline{t}_N - \overline{s}_N|^2\right] \to 0$. Let

$$S = \max_{|x|\le R, t\le T}\left[(u_1 - u_2)(x) - 2\beta g(x)^m - 2\eta t\right],$$

then for all N due to (A.7) we have

$$\begin{aligned} S &= \max_{|x|\le R, t\le T} \psi_N(t,x,t,x) \le \max_{|x|,|y|\le R, t,s\le T} \psi_N(t,x,s,y) = \psi_N(\overline{t}_N, \overline{x}_N, \overline{s}_N, \overline{y}_N) \\ &\le u_1(\overline{t}_N, \overline{x}_N) - u_2(\overline{s}_N, \overline{y}_N) - \beta(g(\overline{x}_N)^m + g(\overline{y}_N)^m) - \eta(\overline{t}_N + \overline{s}_N) := S_N. \end{aligned}$$

If we show $\lim_{N\to\infty} S_N \le S$ this will imply that $\lim_{N\to\infty}\frac{N}{2}\left[|\overline{x}_N - \overline{y}_N|^2 + |\overline{t}_N - \overline{s}_N|^2\right] = 0$. Suppose there exists $N_k \to \infty$ such that $\lim_{N_k\to\infty} S_{N_k} > S$. Since $|\overline{x}_N|, |\overline{y}_N| \le R$ and $\overline{t}_N, \overline{s}_N \le T$ we can assume that $\lim_{N_k\to\infty}(\overline{x}_{N_k}, \overline{y}_{N_k}, \overline{t}_{N_k}, \overline{s}_{N_k}) = (\overline{x}, \overline{y}, \overline{t}, \overline{s})$ with $|\overline{x}|, |\overline{y}| \le R$, and $\overline{t}, \overline{s} \le T$; and since $|\overline{x}_N - \overline{y}_N| + |\overline{t}_N - \overline{s}_N| \to 0$ we have $\overline{x} = \overline{y}$ and $\overline{t} = \overline{s}$. Since $u_1 - u_2$ is USC we have

$$\begin{aligned} \lim_{N_k\to\infty} \psi_N(\overline{t}_{N_k}, \overline{x}_{N_k}, \overline{s}_{N_k}, \overline{y}_{N_k}) &\le \lim_{N_k\to\infty} S_{N_k} \\ &= \lim_{N_k\to\infty} u_1(\overline{t}_{N_k}, \overline{x}_{N_k}) - u_2(\overline{s}_{N_k}, \overline{y}_{N_k}) - \beta(g(\overline{x}_{N_k})^m + g(\overline{y}_{N_k})^m) - \eta(\overline{t}_{N_k} + \overline{s}_{N_k}) \\ &\le u_1(\overline{t}, \overline{x}) - u_2(\overline{t}, \overline{x}) - 2\beta g(\overline{x})^m - 2\eta\overline{t} \le S \end{aligned}$$

by definition of S, which contradicts the assumption that $\lim_{N_k\to\infty} S_{N_k} > S$.

We next show that for some $\overline{N} < \infty$ we have $\overline{t}_N, \overline{s}_N > 0$ for all $N > \overline{N}$. Suppose, on the contrary, there exists $N_k \to 0$ such that either $t_{N_k} = 0$ or $s_{N_k} = 0$ for all k. By the same arguments as before we can assume that $\lim_{N_k\to\infty}(\overline{x}_{N_k}, \overline{y}_{N_k}, \overline{t}_{N_k}, \overline{s}_{N_k}) = (\overline{x}, \overline{y}, \overline{t}, \overline{s})$ where $\overline{x} = \overline{y}$ and $\overline{t} = \overline{s} = 0$. Then, if $\overline{s}_{N_k} = 0$,

$$\psi_N(\overline{t}_{N_k}, \overline{x}_{N_k}, \overline{s}_{N_k}, \overline{y}_{N_k}) \le u_1(\overline{t}_{N_k}, \overline{x}_{N_k}) - u_2(0, \overline{y}_{N_k}),$$

while if $\overline{t}_{N_k} = 0$,

$$\psi_N(\overline{t}_{N_k}, \overline{x}_{N_k}, \overline{s}_{N_k}, \overline{y}_{N_k}) \le u_1(0, \overline{x}_{N_k}) - u_2(\overline{s}_{N_k}, \overline{y}_{N_k}).$$

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

Using the fact that $u_1 - u_2$ is USC and the definition of A we have

$$\lim_{N_k\to\infty} \psi_N(\bar{t}_{N_k}, \bar{x}_{N_k}, \bar{s}_{N_k}, \bar{y}_{N_k}) \le u_1(0,\bar{x}) - u_2(0,\bar{x}) \le A,$$

which contradicts $(*)$ according to which $\psi_N(\bar{t}_N, \bar{x}_N, \bar{s}_N, \bar{y}_N) \ge A + \frac{\delta}{2}$ for all N with $\delta > 0$.

We now define two test functions $\varphi_1, \varphi_2 \in \mathcal{C}^1([0,T]\times E_X)$

$$\varphi_1(t,x) := u_2(\bar{s}_N, \bar{y}_N) + \frac{N}{2}\left[|x-\bar{y}_N|^2 + |t-\bar{s}_N|^2\right] + \beta\left(g(x)^m + g(\bar{y}_N)^m\right) + \eta(t+\bar{s}_N),$$
$$\varphi_2(s,y) := u_1(\bar{t}_N, \bar{x}_N) - \frac{N}{2}\left[|\bar{x}_N-y|^2 - |\bar{t}_N-s|^2\right] - \beta\left(g(\bar{x}_N)^m + g(y)^m\right) - \eta(\bar{t}_N+s),$$

so that $(\bar{t}_N, \bar{x}_N)$ is a point of maximum of $u_1(t,x) - \varphi_1(t,x)$ and $(\bar{s}_N, \bar{y}_N)$ is a point of minimum of $u_2(s,y) - \varphi_2(s,y)$. At the extremum their derivatives in time are

$$\partial_t\varphi_1(\bar{t}_N, \bar{x}_N) = N(\bar{t}_N - \bar{s}_N) + \eta,\ \ \partial_t\varphi_2(s,y) = N(\bar{t}_N - \bar{s}_N) - \eta,$$

and in space are

$$D_x\varphi_1(\bar{t}_N, \bar{x}_N) = N(\bar{x}_N - \bar{y}_N) + \gamma\bar{x}_N,\ \ D_y\varphi_2(\bar{s}_N, \bar{y}_N) = N(\bar{x}_N - \bar{y}_N) - \tau\bar{y}_N,$$

with $\gamma = m\beta g(\bar{x}_N)^{m-2}$, $\tau = m\beta g(\bar{y}_N)^{m-2}$. Since $\forall N > \overline{N}$ we have $\bar{t}_N, \bar{s}_N \in (0,T]$, and u_1 and u_2 are sub- and super-solutions of $\partial_t u_0 - \overline{H}_0(x, D_x u_0) = 0$ on $(0,T]\times E_X$, we have

$$\partial_t\varphi_1(\bar{t}_N, \bar{x}_N) - \overline{H}_0\left(\bar{x}_N, D_x\varphi_1(\bar{t}_N, \bar{x}_N)\right) \le 0,$$
$$\partial_s\varphi_2(\bar{s}_N, \bar{y}_N) - \overline{H}_0\left(\bar{y}_N, D_y\varphi_2(\bar{s}_N, \bar{y}_N)\right) \ge 0,$$

so that

$$\begin{aligned} 2\eta &= \partial_t\varphi_1(\bar{t}_N, \bar{x}_N) - \partial_t\varphi_2(\bar{s}_N, \bar{y}_N) \\ &\le \overline{H}_0\left(\bar{x}_N, N(\bar{x}_N - \bar{y}_N) + \gamma\bar{x}_N\right) - \overline{H}_0\left(\bar{y}_N, N(\bar{x}_N - \bar{y}_N) - \tau\bar{y}_N\right). \end{aligned} \tag{A.8}$$

Using (3.16) letting $\lambda = N$, $p = \gamma\bar{x}_N$, $q = -\tau\bar{y}_N$ we get that the right-hand side is bounded above by

$$\begin{aligned} 2\eta &\le \omega_R\left(|\bar{x}_N - \bar{y}_N| + N|\bar{x}_N - \bar{y}_N|^2\right) + \widetilde{\omega}_1\left(m\beta\left[\bar{x}_N g(\bar{x}_N)^{m-2} + \bar{y}_N g(\bar{y}_N)^{m-2}\right]\right) \\ &\le \omega_R\left(|\bar{x}_N - \bar{y}_N| + N|\bar{x}_N - \bar{y}_N|^2\right) + \widetilde{\omega}_1(mc_1) \end{aligned}$$

since $\beta(g(\bar{x}_N)^m + g(\bar{y}_N)^m) \le c_1$. For $m \le \min(1, \frac{1}{c_1})$ small enough so that $\widetilde{\omega}_1(mc_1) < \eta$, and for N large enough so that $\omega_R(|\bar{x}_N - \bar{y}_N| + N|\bar{x}_N - \bar{y}_N|^2) < \eta$, we get a contradiction in the above inequality. This establishes the comparison principle under the conditions given in (a).

(b) We now extend the comparison principle under the more relaxed assumptions on $\overline{H}_0$ given in (b). For many examples the condition (3.16) is too restrictive and needs to be extended to a more local condition in p. This can be done if (3.14) has a coercive behaviour in p uniformly for bounded x as we show next.

Let $A = \sup_{\{0\}\times E_X}(u_1 - u_2)^+$, and $\widetilde{x}, \widetilde{t}$ be such that $u_1(\widetilde{t}, \widetilde{x}) - u_2(\widetilde{t}, \widetilde{x}) = A + \delta$ $(*)$ with $\delta > 0$, as before. For $\beta, m > 0$, $N, M < \infty$ and $g(x) = \sqrt{1+|x|^2}$ now define

$$\psi_{N,M}(t,x,s,y) = u_1(t,x) - u_2(s,y) - \frac{N}{2}|x-y|^2 - \frac{M}{2}|t-s|^2 - \beta\left(g(x)^m + g(y)^m\right) - \frac{\delta t}{4T}.$$

Now chose β so that $2\beta g(\widetilde{x})^m \le \frac{\delta}{4}$ and hence $(*)$ implies (A.6) as before.

Let $(\bar{t}_{N,M}, \bar{x}_{N,M}, \bar{s}_{N,M}, \bar{y}_{N,M})$ be the point at which $\psi_{N,M}$ achieves its maximum in $[0,T]\times E_X \times [0,T]\times E_X$. By (A.6) we again get $|\bar{x}_{N,M}|, |\bar{y}_{N,M}| \le R,\ \ \forall N, M < \infty$. Also

from $\psi_{N,M}(\bar{t}_{N,M}, \bar{x}_{N,M}, \bar{t}_{N,M}, \bar{x}_{N,M}) + \psi_{N,M}(\bar{s}_{N,M}, \bar{y}_{N,M}, \bar{s}_{N,M}, \bar{y}_{N,M}) \leq 2\psi_{N,M}(\bar{t}_{N,M}, \bar{x}_{N,M}, \bar{s}_{N,M}, \bar{y}_{N,M})$ we now get, with $c_2 = \sup|u_1| + \sup|u_2|$,

$$\frac{N}{2}|\bar{x}_N - \bar{y}_N|^2 + \frac{M}{2}|\bar{s}_N - \bar{t}_N|^2 \leq \frac{\delta}{4T}(\bar{s}_{N,M} - \bar{t}_{N,M}) + c_2 \leq \frac{\delta}{4} + c_2 < \infty,$$

hence $|\bar{x}_{N,M} - \bar{y}_{N,M}| \to 0$ as $N \to \infty$ and $|\bar{s}_{N,M} - \bar{t}_{N,M}| \to 0$ as $M \to \infty$.

Furthermore, the same argument as in (a) shows that $\exists \overline{N}, \overline{M} < \infty$ such that, for all β subject to our earlier choice, we have that $\bar{t}_{N,M}, \bar{s}_{N,M} > 0$ for all $N > \overline{N}$ and $M = \overline{M}$. We now fix $M = \overline{M}$.

Defining test functions

$$\varphi_1(t,x) := u_2(\bar{s}_{N,M}, \bar{y}_{N,M}) + \frac{N}{2}|x - \bar{y}_{N,M}|^2 + \frac{M}{2}|t - \bar{s}_{N,M}|^2 + \beta\left(g(x)^m + g(\bar{y}_{N,M})^m\right) + \frac{\delta t}{4T},$$
$$\varphi_2(s,y) := u_1(\bar{t}_{N,M}, \bar{x}_{N,M}) - \frac{N}{2}|\bar{x}_{N,M} - y|^2 - \frac{M}{2}|\bar{t}_N - s|^2 - \beta\left(g(\bar{x}_{N,M})^m + g(y)^m\right) - \frac{\delta \bar{t}_{N,M}}{4T},$$

we now have, with $\gamma = m\beta g(\bar{x}_{N,M})^{m-2}$, $\tau = m\beta g(\bar{y}_{N,M})^{m-2}$,

$$\frac{\delta}{4T} + M(\bar{t}_{N,M} - \bar{s}_{N,M}) \leq \overline{H}_0\left(\bar{x}_{N,M}, N(\bar{x}_{N,M} - \bar{y}_{N,M}) + \gamma \bar{x}_{N,M}\right), \tag{A.9}$$

and

$$M(\bar{t}_{N,M} - \bar{s}_{N,M}) \geq \overline{H}_0\left(\bar{y}_{N,M}, N(\bar{x}_{N,M} - \bar{y}_{N,M}) - \tau \bar{y}_{N,M}\right). \tag{A.10}$$

For fixed M, β as chosen earlier, (A.9) and (A.10) together with (3.17) coercivity assumption on $\overline{H}_0$ in p, imply that

$$\sup_{N > \overline{N}} N(\bar{x}_{N,M} - \bar{y}_{N,M}) =: \ell < \infty$$

since we already have $|\gamma \bar{x}_{N,M}|, |\tau \bar{y}_{N,M}| \leq m\beta, \ \forall N, M < \infty$

Subtracting (A.10) from (A.9) we get

$$\frac{\delta}{4T} \leq \overline{H}_0\left(\bar{x}_N, N(\bar{x}_{N,M} - \bar{y}_N) + \gamma \bar{x}_{N,M}\right) - \overline{H}_0\left(\bar{y}_{N,M}, N(\bar{x}_{N,M} - \bar{y}_{N,M}) - \tau \bar{y}_{N,M}\right), \tag{A.11}$$

and using (3.16) with $\omega_{R,\ell}$ in place of ω_R, with the same choice of $\lambda = N$, $p = \gamma \bar{x}_N$, $q = -\tau \bar{y}_N$ as earlier, gives

$$\begin{aligned}\frac{\delta}{4T} &\leq \omega_{R,\ell}\left(|\bar{x}_{N,M} - \bar{y}_{N,M}| + N|\bar{x}_{N,M} - \bar{y}_{N,M}|^2\right) \\ &\quad + \widetilde{\omega}_1\left(m\beta\left[\bar{x}_{N,M} g(\bar{x}_{N,M})^{m-2} + \bar{y}_{N,M} g(\bar{y}_{N,M})^{m-2}\right]\right) \\ &\leq \omega_{R,\ell}\left(|\bar{x}_{N,M} - \bar{y}_{N,M}| + N|\bar{x}_{N,M} - \bar{y}_{N,M}|^2\right) + \widetilde{\omega}_1(mc_1).\end{aligned}$$

For $m \leq \min(1, \frac{1}{c_1})$ small enough so that $\widetilde{\omega}_1(mc_1) < \delta/8T$, and N large enough so that $\omega_{R,\ell}(|\bar{x}_{N,M} - \bar{y}_{N,M}| + N|\bar{x}_{N,M} - \bar{y}_{N,M}|^2) < \delta/8T$, we get a contradiction in the above inequality, thus establishing $u_1 \leq u_2$ on $[0, T] \times E_X$.

The fact that it suffices to check a non-uniform coercive condition in p for $\overline{H}_0$ when it is also convex in p is shown in Lemma 9.16 in [17] (using $|\overline{H}_0|$ instead of $\overline{H}_0$). □

References

[1] Andrea Agazzi, Amir Dembo, Jean-Pierre Eckmann, Large deviations theory for Markov jump models of chemical reaction networks, Ann. Appl. Probab. 28.3 (2018) 1821–1855.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.

[2] Andrea Agazzi, Amir Dembo, J.-P. Eckmann, On the geometry of chemical reaction networks: Lyapunov function and large deviations, J. Stat. Phys. (2018) 1–32.
[3] Yuri Bakhtin, Tobias Hurth, Invariant densities for dynamical systems with random switching, Nonlinearity 25 (10) (2012) 2937.
[4] Karen Ball, Thomas G. Kurtz, Lea Popovic, Greg Rempala, Asymptotic analysis of multiscale approximations to reaction networks, Ann. Appl. Probab. (ISSN: 1050-5164) 16 (4) (2006) 1925–1961..
[5] Martion Bardi, Italo Capuzzo-Dolcetta, Optimal Control and Viscosity Solutions of Hamilton Jacobi-Bellman Equations, Systems & Control: Foundations & Applications, Birkhauser, Boston, 1997.
[6] Guy Barles, Cyril Imbert, Second-order elliptic integro-differential equations: Viscosity solutions' theory revisited, Annales de l'Institut Henri Poincare (C) Non Linear Analysis 25 (3) (2008). Elsevier Masson.
[7] Bertrand Cloez, Martin Hairer, et al., Exponential ergodicity for Markov processes with random switching, Bernoulli 21 (1) (2015) 505–536.
[8] Michael G. Crandall, Hitoshi Ishii, Pierre-Louis Lions, Users guide to viscosity solutions of second order partial differential equations, Bull. Amer. Math. Soc. 27 (1) (1992) 1–67.
[9] Thomas Darden, A pseudo-steady state approximation for stochastic chemical kinetics, Rocky Mountain J. Math. (ISSN: 0035-7596) 9 (1) (1979) 51–71. Conference on Deterministic Differential Equations and Stochastic Processes Models for Biological Systems (San Cristobal, N.M., 1977).
[10] Biswajit Das, Gautam Gangopadhyay, Large deviation theory for the kinetics and energetics of turnover of enzyme catalysis in a chemiostatic flow, J. Chem. Phys. 148 (17) (2018) 174104.
[11] M.H.A. Davis, Markov models and optimization, in: Monographs on Statistics and Applied Probability, vol. 49, Chapman & Hall, London, ISBN: 0-412-31410-X, 1993.
[12] Amir Dembo, Ofer Zeitouni, Large deviations techniques and applications, Springer-Verlag, New York, 1998.
[13] Monroe D. Donsker, S.R. Srinivasa Varadhan, On a variational formula for the principal eigenvalue for operators with maximum principle, Proc. Natl. Acad. Sci. 72 (3) (1975) 780–783.
[14] Monroe D. Donsker, S.R. Srinivasa Varadhan, Asymptotic evaluation of certain Markov process expectations for large time IV, Comm. Pure Appl. Math. 36 (2) (1983) 183–212.
[15] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, Peter S. Swain, stochastic gene expression in a single cell, Science 297 (5584) (2002) 1183–1186.
[16] Jin Feng, Martingale problems for large deviations of Markov processes, Stochastic Process. Appl. 81 (2) (1999) 165216.
[17] Jin Feng, Thomas G. Kurtz, Large deviations for stochastic processes, in: Mathematical Surveys and Monographs, vol. 131, American Mathematical Society, Providence, RI, 2006. Print ISBN: 978-1-4704-1870-0.
[18] Jin Feng, Jean-Pierre Fouque, Rohini Kumar, Small-time asymptotics for fast mean-reverting stochastic volatility models, Ann. Appl. Probab. 22 (4) (2012) 1541–1575.
[19] Wendell H. Fleming, Halil Mete Soner, Controlled Markov processes and viscosity solutions. Vol. 25, Springer Science & Business Media, 2006.
[20] Hye-Won. Kang, Thomas G. Kurtz, Separation of time-scales and model reduction for stochastic reaction networks, Ann. Appl. Probab. 23 (2) (2013) 529–583, http://dx.doi.org/10.1214/12-AAP841. URL http://projecteuclid.org/euclid.aoap/1360682022.
[21] Hye-Won Kang, Thomas G. Kurtz, Lea Popovic, Central limit theorems and diffusion approximations for multiscale Markov chain models, Ann. Appl. Probab. (ISSN: 1050-5164) 24 (2) (2014) 721–759, http://dx.doi.org/10.1214/13-AAP934.
[22] Ioannis Kontoyiannis, Sean P. Meyn, Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes, Electron. J. Probab. 10 (3) (2005) 61–123.
[23] Rohini Kumar, Lea Popovic, Large deviations for multi-scale jump-diffusion processes, Stochastic Process. Appl. 127 (4) (2017) 1297–1320.
[24] Suzanne M. Lenhart, Naoki Yamada, Perrons method for viscosity solutions associated with piecewise-deterministic processes, Funkcial. Ekvac. 34 (143) (1991) 173–186.
[25] Tiejun Li, Feng Lin, Large deviations for two scale chemical kinetic processes, Commun. Math. Sci. 15 (1) (2017) 123–163, http://dx.doi.org/10.4310/CMS.2017.v15.n1.a6.
[26] Harley H. McAdams, Adam Arkin, Its a noisy business! Genetic regulation at the nanomolar scale, TIG 15 (1999) 65–69.
[27] Alexander D. Wentzell, Rough limit theorems on large deviations for Markov stochastic processes. I, Theory Probab. Appl. 21 (2) (1977) 227–242, http://dx.doi.org/10.1137/1121030.

Please cite this article in press as: L. Popovic, Large deviations of Markov chains with multiple time-scales, Stochastic Processes and their Applications (2018), https://doi.org/10.1016/j.spa.2018.09.009.