Electronic Journal of Probability

Electron. J. Probab. **19** (2014), no. 59, 1–53.
ISSN: 1083-6489 DOI: 10.1214/EJP.v19-3246

Sensitivity analysis for stochastic chemical reaction networks with multiple time-scales

Ankit Gupta* Mustafa Khammash*

Abstract

Stochastic models for chemical reaction networks have become very popular in recent years. For such models, the estimation of parameter sensitivities is an important and challenging problem. Sensitivity values help in analyzing the network, understanding its robustness properties and also in identifying the key reactions for a given outcome. Most of the methods that exist in the literature for the estimation of parameter sensitivities, rely on Monte Carlo simulations using Gillespie's stochastic simulation algorithm or its variants. It is well-known that such simulation methods can be prohibitively expensive when the network contains reactions firing at different time-scales, which is a feature of many important biochemical networks. For such networks, it is often possible to exploit the time-scale separation and approximately capture the original dynamics by simulating a "reduced" model, which is obtained by eliminating the fast reactions in a certain way. The aim of this paper is to tie these model reduction techniques with sensitivity analysis. We prove that under some conditions, the sensitivity values for the reduced model can be used to approximately recover the sensitivity values for the original model. Through an example we illustrate how our result can help in sharply reducing the computational costs for the estimation of parameter sensitivities for reaction networks with multiple time-scales. To prove our result, we use coupling arguments based on the random time change representation of Kurtz. We also exploit certain connections between the distributions of the occupation times of Markov chains and multi-dimensional wave equations.

Keywords: parameter sensitivity; chemical reaction network; time-scale separation; multiscale network; reduced models; random time change; coupling.
AMS MSC 2010: 60J10; 60J22; 60J27; 60H35; 65C05.
Submitted to EJP on January 8, 2014, final version accepted on June 7, 2014.
Supersedes arXiv:1310.1729v1.

1 Introduction

Chemical reaction networks have traditionally been studied using deterministic models that express the dynamics as a set of ordinary differential equations. Such models ignore the randomness in the dynamics which is caused by the discrete nature of molecular interactions. It is now widely accepted that this randomness can have a significant

*Department of Biosystems Science and Engineering, ETH Zurich, Switzerland.
E-mail: ankit.gupta,mustafa.khammash@bsse.ethz.ch

impact on the macroscopic properties of the system [15, 26, 24], when the molecules are present in low copy numbers. To account for this randomness and study its effects, a stochastic formulation of the dynamics is necessary, and the most common choice is to model the dynamics as a continuous time Markov process. Such stochastic models have been extensively used in many recent articles [8, 3, 23, 25, 26, 19] to understand the biological implications of random dynamics. For a detailed survey of Markov models for chemical reaction networks we refer the readers to [2].

Typically, a chemical reaction network depends on various kinetic parameters whose values are uncertain or suffer from measurement error. To determine the effects of inaccuracies in the parameter values, one needs to estimate the sensitivities of a given output with respect to the parameter values. If an output is highly sensitive to a specific parameter value, then greater time and effort may be invested in determining that parameter precisely. Such sensitivity values can also be useful in fine-tuning a certain output (see [11]) and understanding the robustness properties of a system (see [34]).

Estimation of parameter sensitivities is fairly straightforward for deterministic models, but it poses a major challenge for stochastic models. Many methods have been proposed in the literature for tackling this problem [16, 30, 33, 1, 17]. However all these methods rely on extensive simulations of the stochastic model, which is usually carried out using Gillespie's Stochastic Simulation Algorithm [14] or its variants [12, 13]. These simulation methods account for each and every reaction event, which makes them prohibitively expensive, when the network consists of reactions firing at different time-scales. In such a scenario, the "fast" reactions take up most of the computational time causing the simulation method to become very inefficient. Since time-scale separation is a feature of many important biochemical networks [28], a new class of methods have been designed to exploit this feature and efficiently simulate the stochastic model [5, 36, 6]. These methods simulate a "reduced" model which is obtained by eliminating the fast components of the dynamics through a quasi-steady state approximation [18, 29]. Such reduced models capture the original dynamics in an approximate sense and the error in approximation disappears as the time-scale separation gets larger and larger. In [22], Kang and Kurtz develop a systematic theoretical framework for constructing these reduced models. As discussed in [5] and elsewhere, simulations of reduced models are generally much faster than the original model. Since most sensitivity estimation algorithms are simulation-based, it is of interest to determine if the parameter sensitivities for the original model can be approximated by the parameter sensitivities for the reduced model. Our aim in this paper is to present a theoretical result which shows that can indeed be done under certain conditions. Therefore one can obtain enormous savings in the computational costs required for the estimation of parameter sensitivities for stochastic models of multiscale reaction networks. From now on, the term "multiscale network" refers to a chemical reaction network which consist of reactions firing at different time-scales.

It is observed in [22] that variations in the reaction time-scales could be both due to variation in species numbers and due to variation in rate constants. However in this paper we will only consider the latter source of variation. We now describe our stochastic model of a multiscale chemical reaction network. Suppose we have a well-stirred system consisting of d chemical species. Its state at any time can be described by a vector in $\mathbb{N}_0^d$ whose i-th component is the non-negative integer corresponding to the number of molecules of the i-th species. These chemical species interact through K predefined reaction channels and every time the k-th reaction fires, the state of the system is displaced by the d-dimensional stoichiometric vector $\zeta_k \in \mathbb{Z}^d$. If the state of the system is x, the rate at which the k-th reaction fires is given by $N_0^{\beta_k}\lambda_k(x)$, where N_0 is assumed to be a "large" normalization parameter and $\lambda_k : \mathbb{N}_0^d \to [0,\infty)$ is the *propensity* *function* for the k-th reaction. The powers of N_0 in front of the *propensity functions*, determine the various time-scales at which different reactions act. In a stochastic setting, such a chemical reaction network can be modeled as a continuous time Markov process $\{X^{N_0}(t) : t \geq 0\}$ over $\mathbb{N}^d_0$. Given such a reaction network, we have the flexibility of selecting our reference time-scale γ. This means that we observe the reaction dynamics at times that are scaled by the factor N_0^γ. In other words, we observe the process $\{X^{N_0}_\gamma(t) : t \geq 0\}$ defined by

$$X^{N_0}_\gamma(t) = X^{N_0}(tN_0^\gamma) \quad \text{for } t \geq 0.$$

Note that in the process $X^{N_0}_\gamma$, each reaction k fires at a rate of order $N_0^{\beta_k+\gamma}$. Hence reactions can be termed as "fast", "slow" or "natural" according to whether $\beta_k + \gamma > 0$, $\beta_k + \gamma < 0$ or $\beta_k + \gamma = 0$ respectively. Note that as the value of N_0 increases, the slow reactions get slower and the fast reactions get faster. On the other hand, the natural reactions remain unaffected by the increase in N_0. If we simulate the process $X^{N_0}_\gamma$ using Gillespie's Stochastic Simulation Algorithm, then the fast reactions take up most of the computational time, making the simulation procedure extremely cumbersome.

Fortunately in certain situations, we can obtain a fairly good approximation of the dynamics by simulating a reduced model which does not contain any fast reactions. The state variables in this reduced model correspond to *linear combinations* of species numbers that are unaffected by the fast reactions (see [5, 36]). As described in [22], such model reductions can be derived by replacing N_0 by N and showing that for a certain projection map Π on $\mathbb{R}^d$, the sequence of processes $\{\Pi X^N_\gamma : N \in \mathbb{N}\}$ has a well-defined limit as $N \to \infty$. The limiting process $\hat{X}$ corresponds to the stochastic model of a reduced reaction network made up of only those reactions that are "natural" for the reference time-scale γ, making its simulation far less computationally demanding than the original model. In Section 2 we present these model reduction results in greater detail. Now suppose that the output of interest is given by a real-valued function f and we would like to estimate the expectation $\mathbb{E}\left(f(X^{N_0}_\gamma(t))\right)$ for some observation time $t \geq 0$. If f is invariant under the projection Π (that is, $f(x) = f(\Pi x)$ for all $x \in \mathbb{N}^d_0$) then we would expect that

$$\lim_{N\to\infty} \mathbb{E}\left(f(X^N_\gamma(t))\right) = \lim_{N\to\infty} \mathbb{E}\left(f(\Pi X^N_\gamma(t))\right) = \mathbb{E}\left(f(\hat{X}(t))\right). \tag{1.1}$$

This limit implies that for large values of N_0, the quantity $\mathbb{E}(f(X^{N_0}_\gamma(t)))$ is "close" to $\mathbb{E}(f(\hat{X}(t))$. Hence instead of estimating the former quantity directly we can estimate the latter quantity through simulations of the reduced model, and save a significant amount of computational effort.

As stated before, our aim in this paper is to tie these model reduction results with sensitivity analysis. Suppose that the propensity functions $\lambda_1, \dots, \lambda_K$ depend on a scalar parameter θ. Now when the state is x, the k-th reaction fires at rate $N_0^{\beta_k}\lambda_k(x,\theta)$. With these propensity functions, we can define the processes $X^{N_0}_{\gamma,\theta}$ and $X^N_{\gamma,\theta}$ as before, where the subscript θ is introduced to make the parameter dependence explicit. For an output function f chosen as above, we would like to estimate the sensitivity of the expectation $\mathbb{E}(f(X^{N_0}_\gamma(t)))$ with respect to θ. In other words, we are interested in estimating

$$S^{N_0}_{\gamma,\theta}(f,t) = \frac{\partial}{\partial\theta}\mathbb{E}\left(f(X^{N_0}_{\gamma,\theta}(t))\right). \tag{1.2}$$

We remarked before that most direct methods that estimate this quantity are simulation-based. Since simulations of the process $X^{N_0}_{\gamma,\theta}$ are very expensive, it is worthwhile to explore the possibility of using reduced models to obtain a close approximation for

$S^{N_0}_{\gamma,\theta}(f,t)$. Suppose that for each θ we have a process $\hat{X}_\theta$ which corresponds to the reduced model. Moreover there exists a projection Π (independent of θ) such that $\Pi X^N_{\gamma,\theta}$ converges in distribution to $\hat{X}_\theta$ as $N \to \infty$. Then similar to (1.1) we would get

$$\lim_{N\to\infty} \mathbb{E}\left(f(X^N_{\gamma,\theta}(t))\right) = \mathbb{E}\left(f(\hat{X}_\theta(t))\right).$$

However this relation does not ensure that

$$\lim_{N\to\infty} \frac{\partial}{\partial\theta}\mathbb{E}\left(f(X^N_{\gamma,\theta}(t))\right) = \frac{\partial}{\partial\theta}\left(\lim_{N\to\infty} \mathbb{E}\left(f(X^N_{\gamma,\theta}(t))\right)\right) = \frac{\partial}{\partial\theta}\mathbb{E}\left(f(\hat{X}_\theta(t))\right), \tag{1.3}$$

because in general, limits and derivatives do not commute. Note that if (1.3) holds then for large values of N_0, the quantity $S^{N_0}_{\gamma,\theta}(f,t)$ is close to the value

$$\hat{S}_\theta(f,t) = \frac{\partial}{\partial\theta}\mathbb{E}\left(f(\hat{X}_\theta(t))\right),$$

which can be easily estimated using any of the sensitivity estimation methods [16, 30, 33, 1, 17], since simulations of the reduced model is computationally much easier than the original model. This motivates the main result of the paper which essentially shows that (1.3) holds under certain conditions. In the above discussion we had assumed that the output function f is invariant under the projection Π, which is a highly restrictive assumption. Therefore we will prove a relation analogous to (1.3) for a general function f.

Even though our result is easy to state, its proof is quite technical. The main complication comes from the fact that the dynamics at different time-scales, may interact with each other in non-linear ways. Due to this problem, the proof of our main result involves several steps which are loosely described below. We mentioned above that for a certain projection Π, the process $\Pi X^N_{\gamma,\theta}$ may have a well-defined limit as $N \to \infty$. In such a situation, the *left-over* part of the process, $(I-\Pi)X^N_{\gamma,\theta}$[1], does not converge in the *functional* sense but it converges in the sense of occupation measures (see [22] or Section 2). As reported in [31], the distribution of occupation measures of Markov processes is related to the evolution of a system of multi-dimensional wave equations. Using this relation we construct another process W^N_θ whose distribution has some regularity properties with respect to θ. The process W^N_θ captures the single-time distribution of the process $X^N_{\gamma,\theta}$, which means that for any function f and time t, we can find a function g such that

$$\mathbb{E}\left(f(X^N_{\gamma,\theta}(t))\right) = \mathbb{E}\left(g(W^N_\theta(t))\right).$$

Furthermore, the fast components of the dynamics are *averaged* out in the process W^N_θ, making it simpler to analyze than the original process $X^N_{\gamma,\theta}$. Next we couple the processes W^N_θ and $W^N_{\theta+h}$ (for a small h) in such a way, that it allows us to take the limits $h \to 0$ and $N \to \infty$ (in this order) of an appropriate quantity and prove our main result. This coupling is constructed using the random time change representation of Kurtz (see Chapter 7 in [9]).

As a corollary of our main result we obtain an important relationship which can be useful in estimating steady-state parameter sensitivities. Let X_θ be a stochastic process which models the dynamics of the reaction network described above, with $\beta_k = 0$ for each k and $\gamma = 0$. Assume that this process is ergodic with stationary distribution π_θ and this distribution is difficult to compute analytically. Ergodicity implies that for any *output* function f we have

$$\lim_{t\to\infty} \mathbb{E}\left(f(X_\theta(t))\right) = \left(\int f(y)\pi_\theta(dy)\right),$$

[1] Here I is the identity projection

where the integral is taken over the state space of X_θ. Suppose we are interested in computing the steady-state parameter sensitivity given by

$$\frac{d}{d\theta}\left(\int f(y)\pi_\theta(dy)\right).$$

Since π_θ is unknown, this quantity cannot be computed directly and one has to estimate it using simulations. This can be problematic because simulations can only be performed until a finite time, and in general one is not sure if the sensitivity value estimated at a finite (but large) time is close to the steady-state value. However using our main result, we can conclude that under certain conditions we have

$$\lim_{t\to\infty}\frac{\partial}{\partial\theta}\mathbb{E}\left(f(X_\theta(t))\right)=\frac{d}{d\theta}\left(\int f(y)\pi_\theta(dy)\right). \tag{1.4}$$

The details are given in Section 3.1. Relation 1.4 proves that for a large (but finite) t, the steady-state parameter sensitivity is well-approximated by

$$\frac{\partial}{\partial\theta}\mathbb{E}\left(f(X_\theta(t))\right)$$

which can be estimated using known simulation-based methods [16, 30, 33, 1, 17]. Note that (1.4) is sometimes implicitly assumed (see [35] for example) without proof.

Our main result gives rise to an approximation relationship between the parameter sensitivity value for the original model and the corresponding value for a reduced model which is derived by a *single* model reduction step, exploiting a *single* time-scale separation. It is natural to ask if such an approximation relationship is preserved with reduced models that are obtained using *multiple* model reduction steps. We argue that this is indeed the case in Section 3.2. Hence for a given multiscale network, we may perform many steps of model reduction until we obtain a reduced model which is *simple enough* to allow for extensive simulations, that are required for sensitivity estimation. Then we can estimate the parameter sensitivity value for this highly reduced model, and our result guarantees that the estimated value will be close to the original parameter sensitivity value.

Our main result proves a relation like (1.3) in the situation where the output function f only depends on the state value at a *single* time point t. However using the underlying Markov property it is possible to extend this result to cover situations where the output function f depends on the state values at *several* time points $t_1,\dots,t_m$. We discuss this issue in Section 3.3.

All the results in the paper are stated for a scalar parameter θ, but the extension of these results for vector-valued parameters is relatively straightforward. Finally we would like to mention that even though our paper is written in the context of chemical reaction networks, our main result can be applied to any continuous time Markov process over a discrete lattice with time-scale separation in the transition rates. Other than reaction networks, such processes arise naturally in queuing theory and population modeling.

This paper is organized as follows. In Section 2 we discuss the model reduction results for multiscale networks. The results stated there are simple adaptations of the results in [22]. Our main result is presented in Section 3 and its proof is given in Section 4. In Section 5 we provide an illustrative example to show how our result can be useful.

Notation

We now introduce some notation that we will use throughout this paper. Let $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{N}_0$ denote the sets of all reals, nonnegative reals, integers, positive integers and nonnegative integers respectively. For any $a, b \in \mathbb{R}$, their minimum is given by $a \wedge b$. The positive and negative parts of a are indicated by a^+ and a^- respectively. The number of elements in any finite set E is denoted by $|E|$. By Unif$(0,1)$ we refer to the uniform distribution on $(0,1)$. If Π is a *projection map* on $\mathbb{R}^n$ then we write Πx instead of $\Pi(x)$ for any $x \in \mathbb{R}^n$ and for any $S \subset \mathbb{R}^n$, the set ΠS is given by

$$\Pi S = \{\Pi x : x \in S\}.$$

For any $n \in \mathbb{N}$, $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbb{R}^n$. Moreover for any $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, $\|v\|$ is the 1-norm defined by $\|v\| = \sum_{i=1}^n |v_i|$. The vectors of all zeros and all ones in $\mathbb{R}^n$ are denoted by $\bar{0}_n$ and $\bar{1}_n$ respectively. Let $\mathbb{M}(n,n)$ be the space of all $n \times n$ matrices with real entries. For any $M \in \mathbb{M}(n,n)$, the entry at the i-th row and the j-th column is indicated by M_{ij}. The transpose and inverse of M are indicated by M^T and M^{-1} respectively. The symbol I_n refers to the identity matrix in $\mathbb{M}(n,n)$. For any $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, Diag$(v)$ refers to the matrix in $\mathbb{M}(n,n)$ whose non-diagonal entries are all 0 and whose diagonal entries are $v_1, \dots, v_n$. A matrix in $\mathbb{M}(n,n)$ is called *stable* if all its eigenvalues have strictly negative real parts. While multiplying a matrix with a vector we always regard the vector as a column vector.

Let (S,d) be a metric space. Then by $\mathcal{B}(S)$ we refer to the set of all bounded real-valued Borel measurable functions on S and by $\mathcal{B}_c(S)$ we refer to the set of all those functions in $\mathcal{B}(S)$ that are supported on a compact subset of S. By $\mathcal{P}(S)$ we denote the space of all Borel probability measures on S. This space is equipped with the weak topology. The space of cadlag functions (that is, right continuous functions with left limits) from $[0,\infty)$ to S is denoted by $D_S[0,\infty)$ and it is endowed with the Skorohod topology (for details see Chapter 3, Ethier and Kurtz [9]). For any $f \in D_S[0,\infty)$ and $t > 0$, $f(t-)$ refers to the left-limit $\lim_{s \to t^-} f(s)$.

An operator A on $\mathcal{B}(S)$ is a linear mapping that maps any function in its domain $\mathcal{D}(A) \subset \mathcal{B}(S)$ to a function in $\mathcal{B}(S)$. The notion of the *martingale problem* associated to an operator A is introduced and developed in Chapter 4, Ethier and Kurtz [9]. In this paper, by a solution of the martingale problem for A we mean a measurable stochastic process X with paths in $D_S[0,\infty)$ such that for any $f \in \mathcal{D}(A)$,

$$f(X(t)) - \int_0^t Af(X(s))ds$$

is a martingale with respect to the filtration generated by X. For a given initial distribution $\pi \in \mathcal{P}(S)$, a solution X of the martingale problem for A is a solution of the martingale problem for (A, π) if $\pi = \mathbb{P}X(0)^{-1}$. If such a solution X exists uniquely for all $\pi \in \mathcal{P}(S)$, then we say that the martingale problem for A is well-posed. Additionally, we say that A is the generator of the process X.

Throughout the paper $\Rightarrow$ denotes convergence in distribution.

2 Model Reduction results for multiscale networks

In this section we present the model reduction results for multiscale networks. Recall the definition of the process X_γ^N from Section 1. We shall soon see that this process is well-defined under some assumptions on the propensity functions. Our primary goal in this section, is to find the values of the reference time-scale γ such that the process X_γ^N has a well-behaved limit as $N \to \infty$. This limit may not exist for the whole process but only for a suitable projection of the process. When the limit exists, the limiting process can be viewed as the stochastic model of a reduced reaction network, which only has reactions firing at a *single* time-scale. The results mentioned in this section are derived from the more general results in [22]. Before we proceed we define a property of real-valued functions.

Definition 2.1. *Let U be a subset of $\mathbb{R}^m$, f be a real-valued function on U and Π be a projection map on $\mathbb{R}^m$. We say that the function f is polynomially growing with respect to projection Π if there exist constants $C, r > 0$ such that*

$$|f(x)| \leq C(1 + \|\Pi x\|^r) \text{ for all } x \in U. \tag{2.1}$$

We say that a function f in linearly growing with respect to projection Π if (2.1) *is satisfied for $r = 1$. A sequence of real-valued functions $\{f^N : N \in \mathbb{N}\}$ on U is said to be polynomially (linearly) growing with respect to projection Π if for some $C > 0$ and $r > 0$ ($r = 1$), the relation* (2.1) *holds for each f^N. A function (or a sequence of functions) is called polynomially (linearly) growing if it is polynomially (linearly) growing with respect to the identity projection I.*

Our first task is to ensure that there is a well-defined process which describes the stochastic dynamics of our multiscale reaction network. For this purpose we make certain assumptions.

Assumption 2.2. *The propensity functions $\lambda_1, \dots, \lambda_K$ satisfy the following conditions.*

(A) For any k and $x \in \mathbb{N}_0^d$, if $\lambda_k(x) > 0$ then $(x + \zeta_k)$ has all non-negative components.

(B) Let P be the set of those reactions which have a net positive affect on the total population, that is,

$$P = \{k = 1, \dots, K : \langle \bar{1}_d, \zeta_k \rangle > 0\}. \tag{2.2}$$

Then the function $\lambda_P : \mathbb{N}_0^d \to \mathbb{R}_+$ defined by $\lambda_P(x) = \sum_{k \in P} \lambda_k(x)$ is linearly growing.

Part (A) of this assumption prevents the reaction dynamics from leaving the state space $\mathbb{N}_0^d$. The significance of part (B) will become clear in the next paragraph. Informally, part (B) says that all the reactions that add molecules into the system have orders 0 or 1. If there is a compact set S such that for each k, $\lambda_k(x) = 0$ for all $x \notin S$, then part (B) is trivially satisfied.

Let x_0 be a vector in $\mathbb{N}_0^d$. Throughout the paper, the initial state of the reaction dynamics is fixed to be $x_0 \in \mathbb{N}_0^d$ and the corresponding *stoichiometric compatibility class* is given by

$$\mathcal{S} = \left\{ x_0 + \sum_{k=1}^{K} \eta_k \zeta_k \in \mathbb{N}_0^d : \eta_1, \dots, \eta_K \in \mathbb{N}_0 \right\}.$$

Part (A) of Assumption 2.2 ensures that the reaction dynamics is always inside $\mathcal{S}$. From the description of the multiscale network with reference time-scale γ (see Section 1), it is clear that the generator of the reaction dynamics should be given by the operator $\mathbb{A}_\gamma^N$ whose domain is $\mathcal{D}(\mathbb{A}_\gamma^N) = \mathcal{B}_c(\mathcal{S})$ and its action on any $f \in \mathcal{B}_c(\mathcal{S})$ is given by

$$\mathbb{A}_\gamma^N f(x) = \sum_{k=1}^{K} N^{\beta_k + \gamma} \lambda_k(x)(f(x + \zeta_k) - f(x)). \tag{2.3}$$

From Lemma A.1 we can argue that under Assumption 2.2, the martingale problem for $\mathbb{A}_\gamma^N$ is well-posed. Hence we can define X_γ^N as the Markov process with generator $\mathbb{A}_\gamma^N$ and initial state x_0. The random time change representation (see Chapter 7 in [9]) of this process is given by

$$X_\gamma^N(t) = x_0 + \sum_{k=1}^{K} Y_k \left(N^{\beta_k + \gamma} \int_0^t \lambda_k(X_\gamma^N(s)) ds \right) \zeta_k, \tag{2.4}$$

where $\{Y_k : k = 1, \dots, K\}$ is a family of independent unit rate Poisson processes.

2.1 Convergence at the first time-scale

From (2.4), it is immediate that if the reference time-scale γ is such that $\beta_k+\gamma \leq 0$ for each k, then all the reactions are either “slow" or “natural" at this time-scale[2]. Therefore we would expect the dynamics to converge as $N \to \infty$ and the limiting dynamics will only consist of the natural reactions.

To make this precise, define

$$\gamma_1 = -\max\{\beta_k : k = 1, \dots, K\} \text{ and } \Gamma_1 = \{k = 1, \dots K : \beta_k = -\gamma_1\}. \tag{2.5}$$

Then γ_1 is the first time-scale for which the process $X^N_{\gamma_1}$ has a non-trivial limit as $N \to \infty$ and Γ_1 is the set of natural reactions for this time-scale. Note that

$$\beta_k + \gamma_1 \begin{cases} = 0 & \text{if } k \in \Gamma_1 \\ < 0 & \text{if } k \notin \Gamma_1, \end{cases}$$

and hence using (2.4) we can show that $X^N_{\gamma_1} \Rightarrow \hat{X}$ as $N \to \infty$, where the process $\hat{X}$ satisfies

$$\hat{X}(t) = x_0 + \sum_{k \in \Gamma_1} Y_k \left(\int_0^t \lambda_k(\hat{X}(s)) ds \right) \zeta_k. \tag{2.6}$$

In other words, $\hat{X}$ is the process with initial state x_0 and generator $\mathbb{C}_0$ given by

$$\mathbb{C}_0 f(x) = \sum_{k \in \Gamma_1} \lambda_k(x) \left(f(x+\zeta_k) - f(x) \right) \text{ for } f \in \mathcal{D}(\mathbb{C}_0) = \mathcal{B}_c(\mathcal{S}). \tag{2.7}$$

The well-posedness of the martingale problem for $\mathbb{C}_0$ can be verified from Lemma A.1 and therefore the process $\hat{X}$ is well-defined. The precise statement of this convergence result is given below.

Proposition 2.3. *Suppose that the propensity functions $\lambda_1, \dots, \lambda_K$ satisfy Assumption 2.2. Then we have $X^N_{\gamma_1} \Rightarrow \hat{X}$ as $N \to \infty$ where the limiting process $\hat{X}$ satisfies* (2.6).

Proof. The proof follows easily from Theorem 4.1 in [22]. □

Observe that this proposition can be viewed as a model reduction result, which says that at the time-scale γ_1, the dynamics of the original model (given by $X^{N_0}_{\gamma_1}$) is well-approximated by the dynamics of a reduced model (given by $\hat{X}$) for large values of N_0. This reduced model is obtained by simply dropping the “slow" reactions from the network. Such a model reduction result is *trivial* because one can easily see from the reaction time-scales that the slow reactions will not participate in the limiting dynamics. In the next section we describe a non-trivial model reduction result which is more useful from the point of view of applications.

2.2 Convergence at the second time-scale

As discussed in several recent papers [4, 22], there may be a second time-scale γ_2 ($> \gamma_1$) such that a certain projection Π_2 of the process $X^N_{\gamma_2}$ has a well-behaved limit as $N \to \infty$. At this second time-scale, the network has “fast" reactions in addition to the “slow" and “natural" reactions. The projection Π_2 is such, that the fast reactions do not affect the projected process $\Pi_2 X^N_{\gamma_2}$. Assuming quasi-stationarity for the fast sub-network [18, 29] we can have a well-defined limit $\hat{X}$ for the process $\Pi_2 X^N_{\gamma_2}$. Moreover the limiting process $\hat{X}$ corresponds to the stochastic model of a reduced reaction network which only contains those reactions that are natural for the time-scale γ_2.

[2] The jargon of “slow" , “fast" and “natural" reactions was introduced in Section 1

We now describe this convergence result formally. Suppose that the set

$$\mathbb{S}_2 = \{v \in \mathbb{R}^d_+ : \langle v, \zeta_k \rangle = 0 \ \text{ for all } \ k \in \Gamma_1\}$$

is non-empty. Then for any $v \in \mathbb{S}_2$, the process $\{\langle v, X^N_{\gamma_2}(t)\rangle : t \geq 0\}$ is unaffected by the reactions in Γ_1. Let $\gamma_v = -\max\{\beta_k : k = 1, \dots, K \ \text{ and } \ \langle v, \zeta_k\rangle \neq 0\}$ and define

$$\gamma_2 = \inf\{\gamma_v : v \in \mathbb{S}_2\} \ \text{ and } \ \Gamma_2 = \{k = 1, \dots . K : \beta_k = -\gamma_2\}. \tag{2.8}$$

Then $\gamma_2 > \gamma_1$ by definition and note that the reactions in Γ_1 are fast at the time-scale γ_2. Let $\mathbb{L}_2$ be the subspace spanned by the vectors in $\mathbb{S}_2$ and let Π_2 be the projection map from $\mathbb{R}^d$ to $\mathbb{L}_2$. The definition of $\mathbb{L}_2$ implies that

$$\Pi_2 \zeta_k = \bar{0}_d \ \text{ for all } k \in \Gamma_1, \tag{2.9}$$

which means that the fast reactions would leave the process $\Pi_2 X^N_{\gamma_2}$ unchanged. Let $\mathbb{L}_1$ be the space spanned by the vectors in $(I - \Pi_2)\mathcal{S} = \{(I - \Pi_2)x : x \in \mathcal{S}\}$, where I is the identity map. For any $v \in \Pi_2 \mathcal{S}$ let

$$\mathbb{H}_v = \{y \in \mathbb{L}_1 : y = (I - \Pi_2)x, \ \Pi_2 x = v \ \text{ and } x \in \mathcal{S}\} \tag{2.10}$$

and define the operator $\mathbb{C}^v$ by

$$\mathbb{C}^v f(z) = \sum_{k \in \Gamma_1} \lambda_k(v + z) \left(f(z + \zeta_k) - f(z)\right) \ \text{ for } f \in \mathcal{D}(\mathbb{C}^v) = \mathcal{B}_c(\mathbb{H}_v). \tag{2.11}$$

The operator $\mathbb{C}^v$ can be seen as the generator of a Markov process with state space $\mathbb{H}_v$.

We now define the *occupation measure* of the process $(I - \Pi_2)X^N_{\gamma_2}$. This is a random measure on $\mathbb{L}_1 \times [0, \infty)$ given by

$$V^N_{\gamma_2}(C \times [0, t]) = \int_0^t \mathbb{1}_C \left((I - \Pi_2) X^N_{\gamma_2}(s)\right) ds,$$

where C is any Borel measurable subset of $\mathbb{L}_1$ and $\mathbb{1}_C$ is its indicator function. Note that for any k

$$\int_0^t \lambda_k(X^N_{\gamma_2}(s)) ds = \int_0^t \int_{\mathbb{L}_1} \lambda_k(\Pi_2 X^N_{\gamma_2}(s) + y) V^N_{\gamma_2}(dy \times ds).$$

Therefore using (2.4) and (2.9), we can write the random time change representation for the process $\Pi_2 X^N_{\gamma_2}$ as

$$\begin{aligned}
\Pi_2 X^N_{\gamma_2}(t) &= \Pi_2 x_0 + \sum_{k \in \Gamma_1} Y_k \left(N^{\beta_k + \gamma} \int_0^t \lambda_k(X^N_{\gamma_2}(s)) ds \right) \Pi_2 \zeta_k \\
&\quad + \sum_{k \in \Gamma_2} Y_k \left(N^{\beta_k + \gamma} \int_0^t \lambda_k(X^N_{\gamma_2}(s)) ds \right) \Pi_2 \zeta_k \\
&\quad + \sum_{k \notin \Gamma_1 \cup \Gamma_2} Y_k \left(N^{\beta_k + \gamma} \int_0^t \lambda_k(X^N_{\gamma_2}(s)) ds \right) \Pi_2 \zeta_k \\
&= \Pi_2 x_0 + \sum_{k \in \Gamma_2} Y_k \left(N^{\beta_k + \gamma} \int_0^t \int_{\mathbb{L}_1} \lambda_k(\Pi_2 X^N_{\gamma_2}(s) + y) V^N_{\gamma_2}(dy \times ds) \right) \Pi_2 \zeta_k \qquad (2.12) \\
&\quad + \sum_{k \notin \Gamma_1 \cup \Gamma_2} Y_k \left(N^{\beta_k + \gamma} \int_0^t \int_{\mathbb{L}_1} \lambda_k(\Pi_2 X^N_{\gamma_2}(s) + y) V^N_{\gamma_2}(dy \times ds) \right) \Pi_2 \zeta_k.
\end{aligned}$$

Suppose that $V^N_{\gamma_2} \Rightarrow V$ as $N \to \infty$. In other words, for any $f \in \mathcal{B}(\mathcal{S})$ and $t > 0$

$$\int_0^t \int_{\mathbb{L}_1} f(x) V^N_{\gamma_2}(dx \times ds) \Rightarrow \int_0^t \int_{\mathbb{L}_1} f(x) V(dx \times ds) \text{ as } N \to \infty.$$

Since

$$\beta_k + \gamma_2 \begin{cases} = 0 & \text{if } k \in \Gamma_2 \\ < 0 & \text{if } k \notin \Gamma_1 \cup \Gamma_2, \end{cases}$$

we can expect from (2.12) that $\Pi_2 X^N_{\gamma_2} \Rightarrow \hat{X}$ as $N \to \infty$ where the process $\hat{X}$ satisfies

$$\hat{X}(t) = \Pi_2 x_0 + \sum_{k \in \Gamma_2} Y_k \left(\int_0^t \int_{\mathbb{L}_1} \lambda_k(\hat{X}(s) + y) V(dy \times ds) \right) \Pi_2 \zeta_k.$$

It can be seen that between consecutive jump times of the process $\Pi_2 X^N_{\gamma_2}$, if the state of the process $\Pi_2 X^N_{\gamma_2}$ is v, then the process $(I - \Pi_2) X^N_{\gamma_2}$ evolves like a Markov process with generator $\mathbb{C}^v$. If the generator $\mathbb{C}^v$ corresponds to an ergodic Markov process with the unique stationary distribution as $\pi^v \in \mathcal{P}(\mathbb{H}_v)$, then the limiting measure V has the form

$$V(dy \times ds) = \pi^{\hat{X}(s)}(dy) ds. \tag{2.13}$$

Therefore the random time change representation of the process $\hat{X}$ becomes

$$\hat{X}(t) = \Pi_2 x_0 + \sum_{k \in \Gamma_2} Y_k \left(\int_0^t \hat{\lambda}_k(\hat{X}(s)) ds \right) \Pi_2 \zeta_k, \tag{2.14}$$

where $\hat{\lambda}_k(v) = \int_{\mathbb{H}_v} \lambda_k(v+z) \pi^v(dz)$. Before we state the convergence result, we need to make some assumptions.

Assumption 2.4. *(A) For any $v = \Pi_2 \mathcal{S}$, the space $\mathbb{H}_v$ (given by (2.10)) is finite.*

(B) The Markov process with generator $\mathbb{C}^v$ is ergodic and its unique stationary distribution is $\pi^v \in \mathcal{P}(\mathbb{H}_v)$.

(C) Let P be the set of reactions given by

$$P = \{k = 1, \dots, K : \langle \bar{1}_d, \Pi_2 \zeta_k \rangle > 0\}. \tag{2.15}$$

Then the function $\lambda_P : \mathbb{N}_0^d \to \mathbb{R}_+$ defined by $\lambda_P(x) = \sum_{k \in P} \lambda_k(x)$ is linearly growing with respect to projection Π_2 (see Definition 2.1).

Observe that part (C) implies that the functions $\{\hat{\lambda}_k : k \in \Gamma_2\}$ satisfy part (B) of Assumption 2.2. Therefore the process $\hat{X}$ satisfying (2.14) is well-defined due to Lemma A.1. Note that the set $\mathbb{H}_v$ can either be finite or countably infinite. Our main result (Theorem 3.2) should hold in both the cases, but to simplify the proof we assume that $\mathbb{H}_v$ is finite (part (A) of Assumption 2.4). We later indicate how the proof changes when this is not the case (see Remark 4.19). In many important biochemical multiscale networks, the fast reactions conserve some quantity that only depends on the natural dynamics (see [5, 36, 28]). In such a scenario, the set $\mathbb{H}_v$ will be finite. We now state the convergence result at the second time-scale.

Proposition 2.5. *Suppose that Assumption 2.2 and 2.4 hold. Then $(\Pi_2 X^N_{\gamma_2}, V^N_{\gamma_2}) \Rightarrow (\hat{X}, V)$ as $N \to \infty$, where the process $\hat{X}$ satisfies (2.14) and V satisfies (2.13).*

Proof. The proof follows from Theorem 5.1 in [22]. □

2.3 Convergence at higher time-scales

In Section 2.2 we outlined a systematic procedure to obtain a *single-step* model reduction for a multiscale reaction network. The main idea was to assume ergodicity for the “fast" sub-network and incorporate its steady-state information in the propensities of the “natural" reactions. Moreover the “slow" reactions can be ignored completely. This single-step reduction process can be carried over multiple steps to construct a hierarchy of reduced models. This is useful because many biochemical networks have reactions spanning several time-scales (see [21], for example). Hence for a given reference time-scale, many steps of model reduction may be required to a obtain a model which is *simple enough*, to be amenable for extensive simulations that are required for sensitivity estimation.

For our main result, we will assume that we are in the situation of Proposition 2.5, which describes a single-step model reduction. In Section 3.2, we shall discuss how our result can be used to estimate parameter sensitivity using reduced models that are obtained after many steps of model reduction.

3 The Main Result

In this section we present our main result on sensitivity analysis of multiscale networks. Suppose that the propensity functions $\lambda_1, \dots, \lambda_K$ depend on a real-valued parameter θ and Assumption 2.2 is satisfied for each value of θ. If the reference time-scale is γ, then the reaction dynamics will be captured by the generator

$$\mathbb{A}^N_{\gamma,\theta} f(x) = \sum_{k=1}^{K} N^{\beta_k+\gamma} \lambda_k(x,\theta)(f(x+\zeta_k) - f(x)) \text{ for any } f \in \mathcal{D}(\mathbb{A}^N_{\gamma,\theta}) = \mathcal{B}_c(\mathcal{S}). \tag{3.1}$$

Using Lemma A.1 we can argue that the martingale problem corresponding to $\mathbb{A}^N_{\gamma,\theta}$ is well-posed. Let $X^N_{\gamma,\theta}$ be the process with generator $\mathbb{A}^N_{\gamma,\theta}$ and initial state x_0.

We use the same notation as in Section 2.2. Note that the definitions of $\gamma_i, \Gamma_i, \mathbb{S}_i$ and $\mathbb{L}_i$, for $i = 1$ and 2, only depend on the stoichiometry of the reaction network and are hence independent of θ. Similarly the projection map Π_2 and the space $\mathbb{H}_v$ (see (2.10)) do not depend on θ. The definition of the operator $\mathbb{C}^v$ (see (2.11)) changes to

$$\mathbb{C}^v_\theta f(z) = \sum_{k\in\Gamma_1} \lambda_k(v+z,\theta)\left(f(z+\zeta_k) - f(z)\right) \quad \text{for } f \in \mathcal{D}(\mathbb{C}^v_\theta) = \mathcal{B}_c(\mathbb{H}_v). \tag{3.2}$$

For our main result we require the following assumptions.

Assumption 3.1. *(A) Parts (A) and (C) of Assumption 2.4 are satisfied. In addition, the mapping $v \mapsto |\mathbb{H}_v|$ is polynomially growing (see Definition 2.1).*

(B) A Markov process with generator $\mathbb{C}^v_\theta$ is ergodic and its unique stationary distribution is $\pi^v_\theta \in \mathcal{P}(\mathbb{H}_v)$.

(C) Let $x \in \mathcal{S}$ be fixed. Then for any $k = 1,\dots,K$, the function $\lambda_k(x,\cdot)$ is twice-continuously differentiable in a neighbourhood of θ.

(D) For each $k \in \Gamma_2$, the functions $\lambda_k(\cdot,\theta)$ and $\partial\lambda_k(\cdot,\theta)/\partial\theta$ are polynomially growing with respect to projection Π_2. Moreover there exists an $\epsilon > 0$ such that the function

$$\sup_{\xi\in(\theta-\epsilon,\theta+\epsilon)} \left| \frac{\partial^2 \lambda_k(\cdot,\xi)}{\partial\theta^2} \right|$$

is also polynomially growing with respect to projection Π_2.

(E) The functions $\{\lambda_k(\cdot,\theta) : k \in \Gamma_2\}$ satisfy part (B) of Assumption 2.2.

Note that if Assumption 3.1 holds then Assumption 2.4 will also hold. Hence Proposition 2.5 ensures that $\Pi_2 X^N_{\gamma_2,\theta} \Rightarrow \hat{X}_\theta$ as $N \to \infty$. The process $\hat{X}_\theta$ has initial state $\Pi_2 x_0$ and generator $\hat{\mathbb{A}}_\theta$ given by

$$\hat{\mathbb{A}}_\theta f(x) = \sum_{k \in \Gamma_2} \hat{\lambda}_k(x,\theta)(f(x + \Pi_2 \zeta_k) - f(x)) \text{ for any } f \in \mathcal{D}(\hat{\mathbb{A}}_\theta) = \mathcal{B}_c(\Pi_2 \mathcal{S}), \tag{3.3}$$

where the function $\hat{\lambda}_k(\cdot,\theta) : \Pi_2 \mathcal{S} \to \mathbb{R}_+$ is defined by

$$\hat{\lambda}_k(x,\theta) = \int_{\mathbb{H}_x} \lambda_k(x+y,\theta)\pi^x_\theta(dy). \tag{3.4}$$

We now state our main result whose proof is given in Section 4.3.

Theorem 3.2. *Suppose that Assumption 3.1 holds and the function $f : \mathcal{S} \to \mathbb{R}$ is polynomially growing with respect to projection Π_2. Then for any $t > 0$ we have*

$$\lim_{N \to \infty} \frac{\partial}{\partial \theta} \mathbb{E}\left(f(X^N_{\gamma_2,\theta}(t))\right) = \frac{\partial}{\partial \theta} \mathbb{E}\left(f_\theta(\hat{X}_\theta(t))\right), \tag{3.5}$$

where $f_\theta : \Pi_2 \mathcal{S} \to \mathbb{R}$ is given by

$$f_\theta(x) = \int_{\mathbb{H}_x} f(x+y)\pi^x_\theta(dy). \tag{3.6}$$

Remark 3.3. *This theorem will also hold if the function f depends on the parameter θ, as long as the dependence is continuously differentiable. Moreover we can even replace f by f^N in relation* (3.5)*, where the sequence of functions $\{f^N : N \in \mathbb{N}\}$ is polynomially growing with respect to projection Π_2, and satisfies*

$$\lim_{N \to \infty} f^N(x) = f(x)$$

for each $x \in \mathcal{S}$. These conclusions will be evident from the proof of the theorem.

Recall that the reaction dynamics for the original model in the reference time-scale γ_2 is given by $X^{N_0}_{\gamma_2,\theta}$. If the output of interest is captured by function f, then we are interested in estimating the parameter sensitivity $S^{N_0}_{\gamma_2,t}(f,t)$ defined by (1.2). As explained in Section 1, direct estimation of $S^{N_0}_{\gamma_2,t}(f,t)$ is often infeasible because simulations of the process $X^{N_0}_{\gamma_2,\theta}$ are prohibitively expensive. However simulations of the reduced model dynamics $\hat{X}_\theta$ is much cheaper, allowing us to easily estimate the right side of (3.5), using known methods [16, 30, 33, 1, 17]. The main message of Theorem 3.2 is that for large values of N_0

$$S^{N_0}_{\gamma_2,t}(f,t) \approx \hat{S}_\theta(f_\theta,t) := \frac{\partial}{\partial \theta} \mathbb{E}\left(f_\theta(\hat{X}_\theta(t))\right), \tag{3.7}$$

which allows us to approximately estimate $S^{N_0}_{\gamma_2,t}(f,t)$, in a computationally efficient way.

Observe that in (3.5), the function f_θ may depend on θ even if the function f does not. If the stationary distribution π^x_θ is known for each $x \in \Pi_2 \mathcal{S}$, then the function f_θ and the propensities $\hat{\lambda}_k$ can be computed analytically. In this case, the simulations of the process $\hat{X}_\theta$ that are needed for estimating $\hat{S}_\theta(f_\theta,t)$, can be carried out using the slow-scale Stochastic Simulation Algorithm [5]. If π^x_θ is unknown, then one can use nested schemes [36, 6] to estimate f_θ and $\hat{\lambda}_k$ during the simulation runs. In many applications, the “fast" reactions are uninteresting [28, 29, 18] and they do not alter the output function f. In such a scenario we can expect f to be invariant under the projection Π_2 (that is, $f(x) = f(\Pi_2 x)$ for all $x \in \mathcal{S}$) which would imply that the functions f_θ and f are the same on the space $\Pi_2 \mathcal{S}$. Hence we recover (1.3) from Theorem 3.2.

3.1 Estimation of steady-state parameter sensitivities

We now discuss how relation (1.4) can be derived using our main result. In Section 1 we mentioned the importance of this relation in the context of estimating steady-state parameter sensitivities. Let $\{X_\theta(t) : t \geq 0\}$ be an ergodic $\mathcal{S}$-valued Markov process with generator

$$\mathbb{C}_\theta f(x) = \sum_{k=1}^{K} \lambda_k(x,\theta)(f(x+\zeta_k) - f(x)) \text{ for any } f \in \mathcal{D}(\mathbb{C}_\theta) = \mathcal{B}_c(\mathcal{S}),$$

and stationary distribution π_θ. If we define another process X_θ^N by

$$X_\theta^N(t) = X_\theta(Nt) \text{ for } t \geq 0, \tag{3.8}$$

then X_θ^N represents the dynamics of a multiscale network with $\beta_k = 1$ for each $k = 1,\dots,K$. For this network, clearly $\gamma_2 = 0, \Gamma_2 = \emptyset$ and $\Pi_2\mathcal{S} = \{0\}$. From Theorem 3.2 we obtain

$$\lim_{N\to\infty} \frac{\partial}{\partial\theta}\mathbb{E}\left(f(X_\theta^N(t))\right) = \frac{d}{d\theta}\left(\int_{\mathcal{S}} f(x)\pi_\theta(dx)\right),$$

for any $t > 0$. Hence (1.4) immediately follows from (3.8).

3.2 Sensitivity estimation with multiple reduction steps

We have presented Theorem 3.2 in the setting of Section 2.2, where a *single-step* reduction procedure was described to obtain a "reduced" model (with dynamics $\hat{X}_\theta$) from the original model (with dynamics $X_{\gamma,\theta}^{N_0}$), in the reference time-scale $\gamma = \gamma_2$. As mentioned in Section 2.3, there are examples of multiscale networks where many steps of model reduction may be required to arrive at a *sufficiently* simple model. It is interesting to know that even in such cases, the main approximation relationship (3.7) that falls out of Theorem 3.2, will continue to hold. To illustrate this point, we now consider an example where two-steps of model reduction are needed for sensitivity estimation.

Recall the description of a multiscale network from Section 1. Let γ_1, γ_2 and γ_3 be real numbers such that $\gamma_3 > \gamma_2 > \gamma_1$. Suppose that the sets Γ_1, Γ_2 and Γ_3 form a partition of the reaction set $\{1,\dots,K\}$, and for each $k \in \Gamma_i$, we have $\beta_k = -\gamma_i$ for $i = 1,2,3$. The dynamics of the model in the reference time-scale γ is given by the process $X_{\gamma,\theta}^{N_0}$ whose random time change representation is

$$\begin{aligned} X_{\gamma,\theta}^{N_0}(t) = &\sum_{k\in\Gamma_1} Y_k\left(N_0^{\gamma-\gamma_1}\int_0^t \lambda_k\left(X_{\gamma,\theta}^{N_0}(s),\theta\right)ds\right)\zeta_k \\ &+ \sum_{k\in\Gamma_2} Y_k\left(N_0^{\gamma-\gamma_2}\int_0^t \lambda_k\left(X_{\gamma,\theta}^{N_0}(s),\theta\right)ds\right)\zeta_k \\ &+ \sum_{k\in\Gamma_3} Y_k\left(N_0^{\gamma-\gamma_3}\int_0^t \lambda_k\left(X_{\gamma,\theta}^{N_0}(s),\theta\right)ds\right)\zeta_k, \end{aligned} \tag{3.9}$$

where $\{Y_k : k = 1,\dots,K\}$ is a family of independent unit rate Poisson processes. Clearly this multiscale network has three time-scales γ_1, γ_2 and γ_3. Suppose we want to estimate the sensitivity value $S_{\gamma,t}^{N_0}(f,t)$ (given by (1.2)) at the reference time-scale $\gamma = \gamma_3$. Observe that for this time-scale, the reactions in both the sets Γ_1 and Γ_2 are "fast", but the reactions in Γ_1 are "faster" than those in Γ_2. Ideally we would like to estimate $S_{\gamma_3,t}^{N_0}(f,t)$ using a reduced model which only involves reactions in Γ_3. It is possible to obtain such a reduced model by applying the reduction procedure *twice*. We now demonstrate that even with this *second-order* reduced model, the main approximation relationship (3.7) will still hold.

Replacing $N_0^{\gamma-\gamma_1}$ by $N_0^{\gamma-\gamma_2}N^{\gamma_2-\gamma_1}$ in (3.9), we get another process $X^{N_0,N}_{\gamma,\theta}$ defined by

$$\begin{aligned} X^{N_0,N}_{\gamma,\theta}(t) &= \sum_{k\in\Gamma_1} Y_k\left(N^{\gamma_2-\gamma_1}N_0^{\gamma-\gamma_2}\int_0^t \lambda_k\left(X^{N_0,N}_{\gamma,\theta}(s),\theta\right)ds\right)\zeta_k \\ &+ \sum_{k\in\Gamma_2} Y_k\left(N_0^{\gamma-\gamma_2}\int_0^t \lambda_k\left(X^{N_0,N}_{\gamma,\theta}(s),\theta\right)ds\right)\zeta_k \\ &+ \sum_{k\in\Gamma_3} Y_k\left(N_0^{\gamma-\gamma_3}\int_0^t \lambda_k\left(X^{N_0,N}_{\gamma,\theta}(s),\theta\right)ds\right)\zeta_k. \end{aligned} \tag{3.10}$$

Certainly for large values of N_0 we have

$$S^{N_0}_{\gamma_3,t}(f,t) \approx \lim_{N\to\infty}\frac{\partial}{\partial\theta}\mathbb{E}\left(f\left(X^{N_0,N}_{\gamma_3,\theta}(t)\right)\right). \tag{3.11}$$

Observe that the process $X^{N_0,N}_{\gamma_3,\theta}$ can be treated in the same way as the process $X^N_{\gamma_2,\theta}$ in Theorem 3.2. Suppose that the conditions of this theorem are satisfied. We can construct a projection Π_2 satisfying (2.9) such that the process $\Pi_2 X^{N_0,N}_{\gamma_3,\theta}$ has a well-behaved limit as $N\to\infty$. For any $v\in\Pi_2\mathcal{S}$ let π^v_θ be the stationary distribution for the Markov process with generator $\mathbb{C}^v_\theta$ (see (3.2)). Define $\bar{f}_\theta$ by (3.6) and for each $k\in\Gamma_2\cup\Gamma_3$ let $\bar{\lambda}_k$ be given by (3.4). Using Theorem 3.2 we can conclude that

$$\lim_{N\to\infty}\frac{\partial}{\partial\theta}\mathbb{E}\left(f\left(X^{N_0,N}_{\gamma_3,\theta}(t)\right)\right) = \frac{\partial}{\partial\theta}\mathbb{E}\left(\bar{f}_\theta\left(\bar{X}^{N_0}_{\gamma_3,\theta}(t)\right)\right), \tag{3.12}$$

where $\bar{X}^{N_0}_{\gamma_3,\theta}$ is the $\Pi_2\mathcal{S}$-valued process given by

$$\begin{aligned} \bar{X}^{N_0}_{\gamma_3,\theta}(t) &= \sum_{k\in\Gamma_2} Y_k\left(N_0^{\gamma_3-\gamma_2}\int_0^t \bar{\lambda}_k\left(\bar{X}^{N_0}_{\gamma_3,\theta}(s),\theta\right)ds\right)\Pi_2\zeta_k \\ &+ \sum_{k\in\Gamma_3} Y_k\left(\int_0^t \bar{\lambda}_k\left(\bar{X}^{N_0}_{\gamma_3,\theta}(s),\theta\right)ds\right)\Pi_2\zeta_k. \end{aligned}$$

Substituting N_0 by N we get another process $\bar{X}^N_{\gamma_3,\theta}$ which can again be dealt in the same way as the process $X^N_{\gamma_2,\theta}$ in Theorem 3.2. Moreover for large values of N_0,

$$\frac{\partial}{\partial\theta}\mathbb{E}\left(\bar{f}_\theta\left(\bar{X}^{N_0}_{\gamma_3,\theta}(t)\right)\right) \approx \lim_{N\to\infty}\frac{\partial}{\partial\theta}\mathbb{E}\left(\bar{f}_\theta\left(\bar{X}^{N}_{\gamma_3,\theta}(t)\right)\right). \tag{3.13}$$

Assuming that the conditions of Theorem 3.2 hold, we can construct a projection Π_3, such that $\Pi_3\Pi_2\zeta_k = \bar{0}_d$ for all $k\in\Gamma_2$, and the process $\Pi_3\bar{X}^N_{\gamma_3,\theta}$ has a well-behaved limit as $N\to\infty$. For any $w\in\Pi_3\Pi_2\mathcal{S}$, let μ^w_θ be the stationary distribution for the Markov process with generator

$$\mathbb{C}^w_\theta g(z) = \sum_{k\in\Gamma_2}\bar{\lambda}_k(w+z,\theta)\left(g(w+\Pi_2\zeta_k)-g(z)\right) \quad \text{for } g\in\mathcal{D}(\mathbb{C}^w_\theta)=\mathcal{B}_c(\mathbb{H}_w),$$

where the definition of $\mathbb{H}_w$ is similar to (2.10). Define

$$\hat{f}_\theta(w) = \int_{\mathbb{H}_w}\bar{f}_\theta(w+y)\mu^w_\theta(dy) \quad \text{and} \quad \hat{\lambda}_k(w,\theta) = \int_{\mathbb{H}_w}\bar{\lambda}_k(w+y,\theta)\mu^w_\theta(dy),$$

for each $k\in\Gamma_3$. From Theorem 3.2 we get

$$\lim_{N\to\infty}\frac{\partial}{\partial\theta}\mathbb{E}\left(\bar{f}_\theta\left(\bar{X}^{N}_{\gamma_3,\theta}(t)\right)\right) = \frac{\partial}{\partial\theta}\mathbb{E}\left(\hat{f}_\theta\left(\hat{X}_\theta(t)\right)\right), \tag{3.14}$$

where $\hat{X}_\theta$ is the process given by

$$\hat{X}_\theta(t) = \sum_{k \in \Gamma_3} Y_k \left(\int_0^t \hat{\lambda}_k \left(\hat{X}_\theta(s), \theta \right) ds \right) \Pi_3 \Pi_2 \zeta_k.$$

Combining (3.11), (3.12), (3.13) and (3.14), we get that for large values of N_0

$$S^{N_0}_{\gamma_3,t}(f,t) \approx \frac{\partial}{\partial \theta} \mathbb{E}\left(\hat{f}_\theta \left(\hat{X}_\theta(t) \right) \right). \tag{3.15}$$

This shows that the main approximation relationship (3.7) that arises from Theorem 3.2 will hold even with a reduced model obtained after two steps of model reduction. Observe that the reactions in Γ_3 are “natural" for the time-scale γ_3, and the reduced model corresponding to $\hat{X}_\theta$ only consists of these reactions. Hence the process $\hat{X}_\theta$ is easy to simulate and $S^{N_0}_{\gamma_3,t}(f,t)$ can be easily estimated using (3.15).

3.3 Sensitivity estimation for outputs at multiple time points

In Theorem 3.2 we only consider the situation where the sensitivity is computed for an output at a *single* time point t. However using the Markov property it is possible to extend this result to encompass situations where the output is a function of the state values at *multiple* time points $t_1, \dots, t_m$ (with $0 < t_1 < t_2 < \cdots < t_m$). To be more precise, let $\mathcal{S}^m = \mathcal{S} \times \cdots \times \mathcal{S}$ be the product-space formed by m copies of the state space $\mathcal{S}$ and let $f : \mathcal{S}^m \to \mathbb{R}$ be a function which is polynomially growing with respect to Π_2 in each coordinate. Then we claim that

$$\lim_{N \to \infty} \frac{\partial}{\partial \theta} \mathbb{E}\left(f\left(X^N_{\gamma_2,\theta}(t_1), \dots, X^N_{\gamma_2,\theta}(t_m) \right) \right) = \frac{\partial}{\partial \theta} \mathbb{E}\left(f_\theta \left(\hat{X}_\theta(t_1), \dots, \hat{X}_\theta(t_m) \right) \right), \tag{3.16}$$

where the processes $X^N_{\gamma_2,\theta}$ and $\hat{X}_\theta$ are as in Theorem 3.2 and the function $f_\theta : (\Pi_2 \mathcal{S})^m \to \mathbb{R}$ is given by

$$f_\theta(x_1, \dots, x_m) = \int_{\mathbb{H}_{x_1}} \dots \int_{\mathbb{H}_{x_m}} f(x_1 + y_1, \dots, x_m + y_m) \pi^{x_1}_\theta(dy_1) \dots \pi^{x_m}_\theta(dy_m). \tag{3.17}$$

We shall prove this claim in the case $m = 2$. The proof for a general m simply follows by extending the arguments made in this case.

Pick two time points t_1, t_2 (with $0 < t_1 < t_2$) and a function $f : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ which is polynomially growing with respect to Π_2 in each coordinate. Let $\bar{g}_\theta$ be a real-valued function on $\mathcal{S} \times \Pi_2 \mathcal{S}$ given by

$$\bar{g}_\theta(x_1, x_2) = \int_{\mathbb{H}_{x_2}} f(x_1, x_2 + y) \pi^{x_2}_\theta(dy).$$

For any $x \in \mathcal{S}$ define

$$g^N_\theta(x) = \mathbb{E}\left(f(x, \bar{X}^N_{\gamma_2,\theta}(t_2 - t_1)) \right) \qquad \text{and} \qquad g_\theta(x) = \mathbb{E}\left(\bar{g}_\theta(x, \bar{X}_\theta(t_2 - t_1)) \right),$$

where $\{\bar{X}^N_{\gamma_2,\theta}(t) : t \geq 0\}$ and $\{\bar{X}_\theta(t) : t \geq 0\}$ are processes with initial states $\bar{X}^N_{\gamma_2,\theta}(0) = x$ and $\bar{X}_\theta(0) = \Pi_2 x$, and generators $\mathbb{A}^N_{\gamma_2,\theta}$ (see (3.1)) and $\hat{\mathbb{A}}_\theta$ (see (3.3)) respectively. Due to Proposition 2.5 and Theorem 3.2 we have

$$\lim_{N \to \infty} g^N_\theta(x) = g_\theta(x) \qquad \text{and} \qquad \lim_{N \to \infty} \frac{\partial g^N_\theta(x)}{\partial \theta} = \frac{\partial g_\theta(x)}{\partial \theta}, \tag{3.18}$$

for each $x \in \mathcal{S}$. The Markov property implies that the conditional distribution of $X^N_{\gamma_2,\theta}(t_2)$ ($\hat{X}_\theta(t_2)$) given $X^N_{\gamma_2,\theta}(t_1) = x$ ($\hat{X}_\theta(t_1) = \Pi_2 x$), is same as the distribution of $\bar{X}^N_{\gamma_2,\theta}(t_2 - t_1)$ ($\bar{X}_\theta(t_2 - t_1)$). This shows that

$$\begin{aligned}\mathbb{E}\left(f\left(X^N_{\gamma_2,\theta}(t_1), X^N_{\gamma_2,\theta}(t_2)\right)\right) &= \mathbb{E}\left(\mathbb{E}\left(f\left(X^N_{\gamma_2,\theta}(t_1), X^N_{\gamma_2,\theta}(t_2)\right)\middle| X^N_{\gamma_2,\theta}(t_1)\right)\right) \\ &= \mathbb{E}\left(g^N_\theta(X^N_{\gamma_2,\theta}(t_1))\right), \end{aligned} \tag{3.19}$$

and for any $y_1 \in \mathbb{H}_{\hat{X}_\theta(t_1)}$

$$\begin{aligned} g_\theta(\hat{X}_\theta(t_1) + y_1) &= \mathbb{E}\left(\bar{g}_\theta(\hat{X}_\theta(t_1) + y_1, \hat{X}_\theta(t_2))\middle|\hat{X}_\theta(t_1)\right) \\ &= \mathbb{E}\left(\int_{\mathbb{H}_{\hat{X}_\theta(t_2)}} f(\hat{X}_\theta(t_1) + y_1, \hat{X}_\theta(t_2) + y_2)\pi^{\hat{X}_\theta(t_2)}_\theta(dy_2)\middle|\hat{X}_\theta(t_1)\right). \end{aligned} \tag{3.20}$$

Theorem 3.2, Remark 3.3 and (3.18) imply that

$$\lim_{N\to\infty} \frac{\partial}{\partial\theta}\mathbb{E}\left(g^N_\theta(X^N_{\gamma_2,\theta}(t_1))\right) = \mathbb{E}\left(\int_{\mathbb{H}_{\hat{X}_\theta(t_1)}} g_\theta(\hat{X}_\theta(t_1)) + y_1)\pi^{\hat{X}_\theta(t_1)}_\theta(dy_1)\right). \tag{3.21}$$

However using (3.20) we obtain

$$\begin{aligned} &\mathbb{E}\left(\int_{\mathbb{H}_{\hat{X}_\theta(t_1))}} g_\theta(\hat{X}_\theta(t_1)) + y_1)\pi^{\hat{X}_\theta(t_1)}_\theta(dy_1)\right) \\ &= \mathbb{E}\left(\int_{\mathbb{H}_{\hat{X}_\theta(t_1))}} \mathbb{E}\left(\int_{\mathbb{H}_{\hat{X}_\theta(t_2)}} f(\hat{X}_\theta(t_1) + y_1, \hat{X}_\theta(t_2) + y_2)\pi^{\hat{X}_\theta(t_2)}_\theta(dy_2)\middle|\hat{X}_\theta(t_1)\right)\pi^{\hat{X}_\theta(t_1)}_\theta(dy_1)\right) \\ &= \mathbb{E}\left(\int_{\mathbb{H}_{\hat{X}_\theta(t_1))}}\int_{\mathbb{H}_{\hat{X}_\theta(t_2)}} f(\hat{X}_\theta(t_1) + y_1, \hat{X}_\theta(t_2) + y_2)\pi^{\hat{X}_\theta(t_2)}_\theta(dy_2)\pi^{\hat{X}_\theta(t_1)}_\theta(dy_1)\right) \\ &= \mathbb{E}\left(f_\theta(\hat{X}_\theta(t_1), \hat{X}_\theta(t_2))\right), \end{aligned}$$

where the function f_θ is defined by (3.17) for $m = 2$. This relation along with (3.19) and (3.21) gives us

$$\lim_{N\to\infty} \frac{\partial}{\partial\theta}\mathbb{E}\left(f\left(X^N_{\gamma_2,\theta}(t_1), X^N_{\gamma_2,\theta}(t_2)\right)\right) = \frac{\partial}{\partial\theta}\mathbb{E}\left(f_\theta(\hat{X}_\theta(t_1), \hat{X}_\theta(t_2))\right),$$

which proves (3.16) for $m = 2$.

4 Proofs

We mentioned in Section 1 that the proof of our main result, Theorem 3.2, will require many steps. We now describe these steps in detail. In Section 4.1 we show some regularity properties of the distributions of weighted occupation times for finite Markov chains with *fast* parameter-dependent rates. For this, we exploit certain connections between the distribution of weighted occupation times and multi-dimensional wave equations (see [31]). These regularity properties allow us to later argue that the distribution of the weighted occupation times for the “fast" sub-network of our multiscale network, is differentiable with respect to θ, and the derivative operation commutes with the limit $N \to \infty$. In Section 4.2, we construct a “new" process W^N_θ, which captures the single-time distribution of the process $X^N_{\gamma_2,\theta}$, in the sense described in Section 1. The main difference between $X^N_{\gamma_2,\theta}$ and W^N_θ, is that the dynamics of the fast sub-network is *averaged* out in the process W^N_θ, making it easier to work with. In particular the process W_θ^N is well-behaved in the limit $N \to \infty$ (see Proposition 4.16), unlike the process $X_{\gamma_2,\theta}^N$. The proof of Theorem 3.2 is given in Section 4.3. The main idea of the proof is to couple the processes W_θ^N and $W_{\theta+h}^N$, in such a way, that it allows us to compute a double-limit of the form

$$\lim_{N\to\infty} \lim_{h\to 0} \frac{\mathbb{E}\left(f_{\theta+h}^N(W_{\theta+h}^N(t))\right) - \mathbb{E}\left(f_\theta^N(W_\theta^N(t))\right)}{h},$$

for some functions f_θ^N and $f_{\theta+h}^N$ that depend on our output function f. The results from Section 4.2 will imply that this quantity is equal to the left-hand side of (3.5). On the other hand, using Dynkin's formula (see Lemma 19.21 in [20]) and some coupling arguments, we will show that this quantity is also equal to the right-side of (3.5), thereby proving Theorem 3.2.

4.1 Weighted occupation times of finite Markov chains

Let $\{Z(t) : t \geq 0\}$ be a continuous time Markov chain on a finite state space $\mathcal{E} = \{e_1, \dots, e_m\}$ and with generator

$$\mathbb{A} f(z) = \sum_{k=1}^{K} \lambda_k(z) \left(f(z+\zeta_k) - f(z) \right) \quad \text{for all } f \in \mathcal{D}(\mathbb{A}) = \mathcal{B}(\mathcal{E}).$$

Here $\lambda_1, \dots, \lambda_K$ are positive functions on $\mathcal{E}$. For this Markov chain the Q-matrix (matrix of transition rates) is given by

$$Q_{ij} = \begin{cases} \lambda_k(e_i) & \text{if } i \neq j \text{ and } e_j = e_i + \zeta_k \\ -\sum_{k=1}^{K} \lambda_k(e_i) & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

For a function $\Lambda : \mathcal{E} \to [0, \infty)$ define

$$V(t) = \int_0^t \Lambda(Z(s)) ds. \tag{4.1}$$

Then $V(t)$ is essentially the weighted occupation time of the process Z, where the weight is given by the function Λ. For each $i = 1, \dots, m$ define $p_i, \beta_i : \mathbb{R}_+ \to [0,1]$ by

$$\beta_i(t) = \mathbb{E}\left(\mathbb{1}_{\{Z(t)=e_i\}} \exp(-V(t))\right) \quad \text{and} \quad p_i(t) = \mathbb{P}(Z(t) = e_i).$$

Note that $\beta_i(t)$ can be seen as the *Laplace Transform* of the distribution of $V(t)$ on the event $Z(t) = e_i$. Let $p(t)$ and $\beta(t)$ denote the vectors

$$p(t) = (p_1(t), \dots, p_m(t)) \quad \text{and} \quad \beta(t) = (\beta_1(t), \dots, \beta_m(t)).$$

The definition of matrix Q implies that

$$\frac{dp(t)}{dt} = Q^T p(t). \tag{4.2}$$

The next proposition describes the dynamics of β.

Proposition 4.1. *The function β satisfies the following ordinary differential equation*

$$\frac{d\beta(t)}{dt} = \left(Q^T - D\right) \beta(t),$$

where D is the $m \times m$ diagonal matrix with entries $\Lambda(e_1), \dots, \Lambda(e_m)$.

Proof. Let $r_1, \dots, r_l$ be l distinct values in the set $\{\Lambda(e_1), \dots, \Lambda(e_m)\}$, arranged in ascending order. For each $i = 1, \dots, (l-1)$ let $B_i = \{e \in \mathcal{E} : \Lambda(e) = r_i\}$, and for each $i = 1, \dots, m$ define $F_i : \mathbb{R}_+ \times \mathbb{R} \to [0,1]$ by

$$F_i(t,x) = \mathbb{P}\left(Z(t) = e_i, V(t) > x\right).$$

The random variable $V(t)$ (given by (4.1)) can only take values between $r_1 t$ and $r_l t$. Hence

$$F_i(t, r_l t) = 0 \ \text{ and } \ F_i(t, r_1 t-) = \lim_{h \to 0^-} F_i(t, r_1 t + h) = \mathbb{P}(Z(t) = e_i). \tag{4.3}$$

It has been shown in [31] that the distribution of the real-valued random variable $V(t)$ is continuous in the interval $[r_1 t, r_l t]$, except at points $r_1 t, \dots, r_l t$. Whenever $x = r_j t$ for some $j = 1, \dots, l$, the function F_i has a discontinuity of size

$$F_i(x, r_j t) - F_i(x, r_j t-) = -\mathbb{P}\left(Z(t) = e_i, V(t) = r_j t\right).$$

Moreover, the event $\{V(t) = r_j t\}$ can only happen if $Z(s) \in B_j$ for all $s \in [0,t]$. Therefore $\mathbb{P}\left(Z(t) = e_i, V(t) = r_j t\right)$ is non-zero only if $e_i \in B_j$ and hence

$$\begin{aligned}\sum_{j=1}^{l} g(r_j)(F_i(t, r_j t-) - F_i(t, r_j t)) &= \sum_{j=1}^{l} g(r_j)\mathbb{P}\left(Z(t) = e_i, V(t) = r_j t\right) \\ &= g(\Lambda(e_i))\mathbb{P}\left(Z(t) = e_i, V(t) = \Lambda(e_i) t\right),\end{aligned} \tag{4.4}$$

for any $g : \mathbb{R}_+ \to \mathbb{R}_+$. It is shown in [31] that on the set $\mathcal{R} = \{(t,x) : t > 0 \text{ and } x \in (r_{j-1}t, r_j t),\ j = 2, \dots, l\}$, each F_i is continuously differentiable and the family of functions $\{F_i : i = 1, \dots, m\}$ satisfies the following system of multi-dimensional wave equations

$$\frac{\partial F_i(t,x)}{\partial t} = -\Lambda(e_i)\frac{\partial F_i(t,x)}{\partial x} + \sum_{k=1}^{m} F_k(t,x) Q_{ki}, \ \text{ for } i = 1, \dots, m. \tag{4.5}$$

For each $i = 1, \dots, m$ we can write $\beta_i(t)$ as

$$\begin{aligned}\beta_i(t) = \mathbb{E}\left(\mathbb{1}_{\{Z(t)=e_i\}} e^{-V(t)}\right) &= e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t) = e_i, V(t) = \Lambda(e_i)t\right) \\ &\quad - \sum_{j=2}^{l} \int_{r_{j-1}t}^{r_j t} e^{-x}\left(\frac{\partial F_i(t,x)}{\partial x}\right) dx.\end{aligned} \tag{4.6}$$

Using integration by parts, (4.3) and (4.4) we get

$$
\begin{aligned}
&\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_jt} e^{-x}\left(\frac{\partial F_i(t,x)}{\partial x}\right)dx \\
&= \sum_{j=2}^{l}\left(e^{-r_jt}F_i(t,r_jt-) - e^{-r_{j-1}t}F_i(t,r_{j-1}t)\right) + \sum_{j=2}^{l}\int_{r_{j-1}t}^{r_jt} e^{-x}F_i(t,x)dx \\
&= \sum_{j=2}^{l}\left(e^{-r_jt}F_i(t,r_jt) - e^{-r_{j-1}t}F_i(t,r_{j-1}t)\right) + \sum_{j=2}^{l}\int_{r_{j-1}t}^{r_jt} e^{-x}F_i(t,x)dx \\
&+ \sum_{j=2}^{l} e^{-r_jt}(F_i(t,r_jt-) - F_i(t,r_jt)) \\
&= -e^{-r_1t}F_i(t,r_1t) + \sum_{j=2}^{l}\int_{r_{j-1}t}^{r_jt} e^{-x}F_i(t,x)dx + \sum_{j=2}^{l} e^{-r_jt}(F_i(t,r_jt-) - F_i(t,r_jt)) \\
&= e^{-r_1t}\left(F_i(t,r_1t-) - F_i(t,r_1t)\right) - e^{-r_1t}F_i(t,r_1t-) \\
&+ \sum_{j=2}^{m}\int_{r_{j-1}t}^{r_jt} e^{-x}F_i(t,x)dx + \sum_{j=2}^{l} e^{-r_jt}(F_i(t,r_jt-) - F_i(t,r_jt)) \\
&= -e^{-r_1t}\mathbb{P}(Z(t)=e_i) + \sum_{j=2}^{l}\int_{r_{j-1}t}^{r_jt} e^{-x}F_i(t,x)dx \\
&+ e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t)=e_i, V(t)=\Lambda(s_i)t\right).
\end{aligned}
$$

Substituting the above expression in (4.6) we obtain

$$
\beta_i(t) = e^{-r_1t}p_i(t) - \sum_{j=2}^{l}\int_{r_{j-1}t}^{r_jt} e^{-x}F_i(t,x)dx, \tag{4.7}
$$

where $p_i(t) = \mathbb{P}(Z(t)=e_i)$.

For $i = 1,\dots,m$, the functions p_i and $F_i(\cdot,x)$ are differentiable (see (4.2) and (4.5)). Hence the function β_i is also differentiable. Taking derivative with respect to t in (4.7) yields

$$
\begin{aligned}
\frac{d\beta_i(t)}{dt} &= -\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_jt} e^{-x}\frac{\partial F_i(t,x)}{\partial t}dx - \sum_{j=2}^{l}\left(r_je^{-r_jt}F_i(t,r_jt-) - r_{j-1}e^{-r_{j-1}t}F_i(t,r_{j-1}t)\right) \\
&\quad - r_1e^{-r_1t}p_i(t) + e^{-r_1t}\frac{dp_i(t)}{dt}.
\end{aligned}
$$

From (4.3) and (4.4) it follows that

$$\begin{aligned}
&\sum_{j=2}^{l}\left(r_j e^{-r_j t}F_i(t,r_j t-) - r_{j-1}e^{-r_{j-1}t}F_i(t,r_{j-1}t)\right)\\
&=\sum_{j=2}^{l}\left(r_j e^{-r_j t}F_i(t,r_j t) - r_{j-1}e^{-r_{j-1}t}F_i(t,r_{j-1}t)\right)\\
&+\sum_{j=2}^{l} r_j e^{-r_j t}\left(F_i(t,r_j t-) - F_i(t,r_j t)\right)\\
&=-r_1 e^{-r_1 t}F_i(t,r_1 t) + \sum_{j=2}^{l} r_j e^{-r_j t}\mathbb{P}\left(Z(t)=e_i, V(t)=r_j t\right)\\
&=-r_1 e^{-r_1 t}p_i(t) + \sum_{j=1}^{l} r_j e^{-r_j t}\mathbb{P}\left(Z(t)=e_i, V(t)=r_j t\right)\\
&=-r_1 e^{-r_1 t}p_i(t) + \Lambda(e_i)e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t)=e_i, V(t)=\Lambda(e_i)t\right). \qquad (4.8)
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{d\beta_i(t)}{dt} &= -\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_j t} e^{-x}\frac{\partial F_i(t,x)}{\partial t}dx - \Lambda(e_i)e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t)=e_i, V(t)=\Lambda(e_i)t\right)\\
&\quad + e^{-r_1 t}\frac{dp_i(t)}{dt}.
\end{aligned}$$

From (4.5) we get

$$\begin{aligned}
\frac{d\beta_i(t)}{dt} &= \Lambda(e_i)\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_j t} e^{-x}\frac{\partial F_i(t,x)}{\partial x}dx - \sum_{k=1}^{m}\left(\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_j t} e^{-x}F_k(t,x)dx\right)Q_{ki}\\
&\quad - \Lambda(e_i)e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t)=e_i, V(t)=\Lambda(e_i)t\right) + e^{-r_1 t}\frac{dp_i(t)}{dt}\\
&= \Lambda(e_i)\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_j t} e^{-x}\frac{\partial F_i(t,x)}{\partial x}dx + \sum_{k=1}^{m}\left(\beta_k(t) - e^{-r_1 t}p_k(t)\right)Q_{ki}\\
&\quad - \Lambda(e_i)e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t)=e_i, V(t)=r_i t\right) + e^{-r_1 t}\frac{dp_i(t)}{dt}\\
&= \Lambda(e_i)\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_j t} e^{-x}\frac{\partial F_i(t,x)}{\partial x}dx + \sum_{k=1}^{m}\beta_k(t)Q_{ki}\\
&\quad - \Lambda(e_i)e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t)=e_i, V(t)=r_i t\right) + e^{-r_1 t}\left(\frac{dp_i(t)}{dt} - \sum_{k=1}^{m}p_k(t)Q_{ki}\right).
\end{aligned}$$

Due to (4.2), the last term is 0 and hence

$$\begin{aligned}
\frac{d\beta_i(t)}{dt} &= \Lambda(e_i)\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_j t} e^{-x}\frac{\partial F_i(t,x)}{\partial x}dx + \sum_{k=1}^{m}\beta_k(t)Q_{ki} \qquad (4.9)\\
&\quad - \Lambda(e_i)e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t)=e_i, V(t)=\Lambda(e_i)t\right).
\end{aligned}$$

Using integration by parts, (4.8) and (4.7) we obtain

$$\begin{aligned}\sum_{j=2}^{l}\int_{r_{j-1}t}^{r_j t} e^{-x}\frac{\partial F_i(t,x)}{\partial x}dx &= \sum_{j=2}^{l}\left(e^{-r_j t}F_i(t,r_j t-) - e^{-r_{j-1}t}F_i(t,r_{j-1}t)\right)\\ &\quad + \sum_{j=2}^{l}\int_{r_{j-1}t}^{r_j t} e^{-x}F_i(t,x)dx\\ &= e^{-\Lambda(e_i)t}\mathbb{P}\left(Z(t)=e_i, V(t)=\Lambda(e_i)t\right) - \beta_i(t).\end{aligned}$$

Substituting this expression in (4.9) yields

$$\frac{d\beta_i(t)}{dt} = -\Lambda(e_i)\beta_i(t) + \sum_{k=1}^{m}\beta_k(t)Q_{ki}.$$

This completes the proof of the proposition. □

Using the above proposition, we now establish some regularity properties of the distributions of weighted occupation times for finite Markov chains with *fast* parameter-dependent rates. Let $\{Z_\theta^N(t) : t \geq 0\}$ be a continuous time Markov chain on $\mathcal{E} = \{e_1,\dots,e_m\}$ with generator given by

$$\mathbb{C}_\theta^N f(z) = N\sum_{k=1}^{K}\lambda_k(z,\theta)\left(f(z+\zeta_k) - f(z)\right) \quad \text{for all } f \in \mathcal{D}(\mathbb{C}_\theta^N) = \mathcal{B}(\mathcal{E}),$$

where the function $\theta \mapsto \lambda_k(z,\theta)$ is continuously differentiable for each k and $z \in \mathcal{E}$. For this Markov chain, the matrix of transition rates is given by NQ_θ where

$$Q_{\theta,ij} = \begin{cases} \lambda_k(e_i,\theta) & \text{if } i \neq j \text{ and } e_j = e_i + \zeta_k \\ -\sum_{k=1}^{K}\lambda_k(e_i,\theta) & \text{if } i = j \\ 0 & \text{otherwise.}\end{cases}$$

We assume that this Markov chain is ergodic. Then its unique stationary distribution π_θ is a left eigenvector for Q_θ corresponding to the eigenvalue 0. Hence

$$\pi_\theta Q_\theta = \bar{0}_m \quad \text{and} \quad \langle \bar{1}_m, \pi_\theta\rangle = \bar{1}_m^T\pi_\theta = 1. \tag{4.10}$$

Remark 4.2. *Due to the ergodicity assumption, the matrix Q_θ has 0 as a simple eigenvalue and all its other eigenvalues have strictly negative real parts.*

For a function $\Lambda : \mathcal{E}\times\mathbb{R} \to [0,\infty)$ define

$$V_\theta^N(t) = \int_0^t \Lambda(Z_\theta^N(s),\theta)ds \tag{4.11}$$

and let

$$\beta_{\theta,i}^N(t) = \mathbb{E}\left(\mathbb{1}_{\{Z_\theta^N(t)=e_i\}}\exp(-V_\theta^N(t))\right),$$

for each $i = 1,\dots,m$. From Proposition 4.1 it follows that the function
$\beta_\theta^N(t) = (\beta_{\theta,1}^N(t),\dots,\beta_{\theta,m}^N(t))$ satisfies

$$\frac{d\beta_\theta^N(t)}{dt} = \left(NQ_\theta^T - D_\theta\right)\beta_\theta^N(t), \tag{4.12}$$

where D_θ is the $m\times m$ diagonal matrix with entries $\Lambda(e_1,\theta),\dots,\Lambda(e_m,\theta)$. We now define a condition on sequences of functions on $\mathbb{R}_+$.

Condition 4.3. *For each $N \in \mathbb{N}$, let f^N be a function from $\mathbb{R}_+$ to $\mathbb{R}^m$ and let $\epsilon_N = 1/\sqrt{N}$. Then the sequence of functions $\{f^N : N \in \mathbb{N}\}$ satisfies this condition if for any $T > 0$*

$$\lim_{N\to\infty} \sup_{t\in[\epsilon_N,T]} \|f^N(t)\| = 0 \text{ and } \lim_{N\to\infty} \int_0^T \|f^N(t)\|dt = 0.$$

The main result of this section is given as the next proposition.

Proposition 4.4. *Define $\hat{\beta}^N_\theta : [0,\infty) \to \mathbb{R}^m$ by*

$$\hat{\beta}^N_\theta(t) = \beta^N_\theta(t) - e^{-\lambda_\theta t}\pi_\theta,$$

where

$$\lambda_\theta = \bar{1}^T_m D_\theta \pi_\theta. \tag{4.13}$$

Then the functions $\hat{\beta}^N_\theta$ and $\partial\hat{\beta}^N_\theta/\partial\theta$ satisfy Condition 4.3.

Remark 4.5. *Here $\partial\beta^N_\theta/\partial\theta$ should be interpreted as the map $t \mapsto \partial\beta^N_\theta(t)/\partial\theta$. Of course this proposition can only be true if $\partial\beta^N_\theta(t)/\partial\theta$ and $\partial\pi_\theta/\partial\theta$ exist. Note that entries of the matrices Q_θ and D_θ are differentiable with respect to θ. Hence* (4.12) *implies the existence of $\partial\beta^N_\theta(t)/\partial\theta$. Moreover due to the implicit mapping theorem and the relation $\pi_\theta Q_\theta = \bar{0}_d$ (see* (4.10)*) one can also conclude that $\partial\pi_\theta/\partial\theta$ exists.*

Proof. We start by defining some notation that will be useful in the proof. We say that a $\mathbb{R}^m$-valued sequence $\{a_N : N \in \mathbb{N}\}$ belongs to class $O(N^{-m})$ for some $m \in \mathbb{N}_0$, if and only if

$$\sup_{N\in\mathbb{N}} N^m \|a_N\| < \infty.$$

For two such sequences $\{a_N : N \in \mathbb{N}\}$ and $\{b_N : N \in \mathbb{N}\}$, we will say that $a_N = b_N + O(N^{-m})$ when the sequence $\{(a_N - b_N) : N \in \mathbb{N}\}$ belongs to class $O(N^{-m})$.

For the proof, we can assume without loss of generality, that for each N, $Z^N_\theta(0) = e_{i_0}$ for some $i_0 = 1, \dots, m$. This implies that $\beta^N_\theta(0) = (0, \dots, 0, 1, 0, \dots, 0)$, where the 1 is in place i_0. Hence

$$\langle \bar{1}_m, \beta^N_\theta(0) - \pi_\theta\rangle = \langle \bar{1}_m, \beta^N_\theta(0)\rangle - \langle \bar{1}_m, \pi_\theta\rangle = 0. \tag{4.14}$$

Define a function $h^N_\theta : \mathbb{R}_+ \to \mathbb{R}^m$ by

$$h^N_\theta(t) = e^{\lambda_\theta t}\beta^N_\theta(t) - \pi_\theta. \tag{4.15}$$

To prove the proposition it is sufficient to show that both h^N_θ and $\partial h^N_\theta/\partial\theta$ satisfy Condition 4.3.

From (4.12) we obtain

$$\frac{dh^N_\theta(t)}{dt} = \left(NQ^T_\theta - D_\theta + \lambda_\theta I_m\right) h^N_\theta(t) - D_\theta\pi_\theta + \lambda_\theta\pi_\theta, \tag{4.16}$$

where I_m is the $m \times m$ identity matrix. Consider the matrix $B^N_\theta = Q^T_\theta - N^{-1}D_\theta$, which can be seen as a small perturbation of Q^T_θ for large values of N. The eigenvalues of B^N_θ is slightly perturbed with respect to the eigenvalues of Q^T_θ (see [32]). We know that matrix Q^T_θ has 0 as a simple eigenvalue (see Remark 4.2) and the corresponding left eigenvector is $\bar{1}_m$. From Theorem 2.7 in [32], we can conclude that B^N_θ has an eigenvalue λ^N_θ with the corresponding left eigenvector as v^N_θ, where λ^N_θ and v^N_θ have the form

$$\lambda^N_\theta = -\frac{\lambda_\theta}{N} + O(N^{-2}) \text{ and } v^N_\theta = \bar{1}_m + O(N^{-1}). \tag{4.17}$$

Therefore

$$\begin{aligned}(v^N_\theta)^T \left(NQ^T_\theta - D_\theta + \lambda_\theta I_m\right) &= N(v^N_\theta)^T B^N_\theta + \lambda_\theta (v^N_\theta)^T \\ &= N\lambda^N_\theta (v^N_\theta)^T + \lambda_\theta (v^N_\theta)^T \\ &= \left(N\lambda^N_\theta + \lambda_\theta\right)(v^N_\theta)^T.\end{aligned}$$

Let $S^N_\theta = \langle v^N_\theta, h^N_\theta(t)\rangle$. Taking inner product with v^N_θ in (4.16) we get

$$\frac{dS^N_\theta(t)}{dt} = \left(N\lambda^N_\theta + \lambda_\theta\right) S^N_\theta(t) + (v^N_\theta)^T \left(-D_\theta \pi_\theta + \lambda_\theta \pi_\theta\right).$$

Note that $a^N_\theta := N\lambda^N_\theta + \lambda_\theta = O(N^{-1})$ due to (4.17). From (4.13) and (4.10) we can see that $b^N_\theta := (v^N_\theta)^T \left(-D_\theta \pi_\theta + \lambda_\theta \pi_\theta\right) = \bar{1}^T_m \left(-D_\theta \pi_\theta + \lambda_\theta \pi_\theta\right) + O(N^{-1}) = O(N^{-1})$. Therefore we can write

$$\frac{dS^N_\theta(t)}{dt} = a^N_\theta S^N_\theta(t) + b^N_\theta, \tag{4.18}$$

where $\{a^N_\theta\}, \{b^N_\theta\}$ are sequences in $O(N^{-1})$. Using (4.14) we obtain

$$\begin{aligned}S^N_\theta(0) = \langle v^N_\theta, h^N_\theta(0)\rangle = \langle \bar{1}_m, h^N_\theta(0)\rangle + O(N^{-1}) &= \langle \bar{1}_m, \beta^N_\theta(0)\rangle - \langle \bar{1}_m, \pi_\theta\rangle + O(N^{-1}) \\ &= O(N^{-1}).\end{aligned} \tag{4.19}$$

Pick any $T > 0$. From (4.18), (4.19) and Gronwall's inequality it follows that

$$\sup_{t\in[0,T]} |S^N_\theta(t)| = \sup_{t\in[0,T]} |\langle v^N_\theta(t), h^N_\theta(t)\rangle| = O(N^{-1}),$$

which also implies that

$$\sup_{t\in[0,T]} |\langle \bar{1}_m, h^N_\theta(t)\rangle| = O(N^{-1}). \tag{4.20}$$

This allows us to write

$$h^N_{\theta,m}(t) = -\sum_{i=1}^{m-1} h^N_{\theta,i}(t) + O(N^{-1}).$$

Let C_θ be the $(m-1)\times(m-1)$ matrix whose ij-th entry is given by

$$C_{\theta,ij} = Q_{\theta,ji} - Q_{\theta,mi}.$$

If we define

$$P = \begin{bmatrix} I_{m-1} & \bar{1}_{m-1} \\ \bar{0}^T_{m-1} & 1 \end{bmatrix} \quad \text{and} \quad P^{-1} = \begin{bmatrix} I_{m-1} & -\bar{1}_{m-1} \\ \bar{0}^T_{m-1} & 1 \end{bmatrix}$$

then using $\bar{1}^T_m Q^T_\theta = \bar{0}^T_m$ we can write

$$P^T Q^T_\theta (P^T)^{-1} = \begin{bmatrix} C_\theta & v \\ \bar{0}^T_{m-1} & 0 \end{bmatrix}, \tag{4.21}$$

where v is some vector in $\mathbb{R}^{m-1}$. The matrix Q_θ has a simple eigenvalue at 0 and all its other eigenvalues have strictly negative real parts (see Remark 4.2). This shows that matrix C_θ is stable.

Let $\bar{h}^N_\theta(t)$ and $\bar{\pi}_\theta$ be vectors containing the first $(m-1)$ components of $h^N(t)$ and π_θ. Also let $\bar{D}_\theta$ be the $(m-1)\times(m-1)$ diagonal matrix with entries $\lambda(e_2,\theta),\dots,\lambda(e_m,\theta)$. From (4.16) we get

$$\frac{d\bar{h}^N_\theta(t)}{dt} = \left(NC_\theta - \bar{D}_\theta + \lambda_\theta I_{m-1}\right)\bar{h}^N_\theta(t) - \bar{D}_\theta\bar{\pi}_\theta + \lambda_\theta\bar{\pi}_\theta. \tag{4.22}$$

Let C^N_θ be the matrix given by

$$C^N_\theta = C_\theta - \frac{1}{N}\left(\bar{D}_\theta - \lambda_\theta I_{m-1}\ \right). \tag{4.23}$$

The stability of matrix C_θ implies that there exists a $\alpha > 0$ such that for any $t \geq 0$ and N

$$\|\exp(NC^N_\theta t)\| \leq \exp(-N\alpha t). \tag{4.24}$$

The exact solution of (4.22) is

$$\bar{h}^N_\theta(t) = \exp(NC^N_\theta t)\bar{h}^N_\theta(0) - \int_0^t \exp(NC^N_\theta(t-s))\left(\bar{D}_\theta\bar{\pi}_\theta - \lambda_\theta\bar{\pi}_\theta\right)ds,$$

which implies that

$$\begin{aligned}\|\bar{h}^N_\theta(t)\| &\leq e^{-N\alpha t}\|\bar{h}^N_\theta(0)\| + \int_0^t e^{-N\alpha(t-s)}\|\bar{D}_\theta\bar{\pi}_\theta - \lambda_\theta\bar{\pi}_\theta\|ds \\ &\leq e^{-N\alpha t}\|\bar{h}^N_\theta(0)\| + \frac{\|\bar{D}_\theta\bar{\pi}_\theta - \lambda_\theta\bar{\pi}_\theta\|}{N\alpha}.\end{aligned}$$

This along with (4.20) shows that the function h^N_θ satisfies Condition 4.3. In fact for any $T > 0$

$$\sup_{t\in[\epsilon_N,T]}\|h^N_\theta(t)\| = O(N^{-1}) \text{ and } \int_0^T \|h^N_\theta(t)\|dt = O(N^{-1}), \tag{4.25}$$

where $\epsilon_N = 1/\sqrt{N}$.

Let $H^N_\theta : \mathbb{R}_+ \to \mathbb{R}^m$ be defined by

$$H^N_\theta(t) = \frac{\partial h^N_\theta(t)}{\partial\theta}.$$

Differentiating (4.16) with respect to θ we get

$$\begin{aligned}\frac{dH^N_\theta(t)}{dt} &= \left(NQ^T_\theta - D_\theta + \lambda_\theta I_m\right)H^N_\theta(t) + \left(N\frac{\partial Q^T_\theta}{\partial\theta} - \frac{\partial D_\theta}{\partial\theta} + \frac{\partial\lambda_\theta}{\partial\theta}I_m\right)h^N_\theta(t) \\ &\quad - \frac{\partial(D_\theta\pi_\theta)}{\partial\theta} + \frac{\partial(\lambda_\theta\pi_\theta)}{\partial\theta}.\end{aligned}$$

Note that

$$\left\langle v^N_\theta, N\frac{\partial Q^T_\theta}{\partial\theta}\right\rangle = N\left\langle v^N_\theta, \frac{\partial Q^T_\theta}{\partial\theta}\right\rangle = N\left\langle \bar{1}_m, \frac{\partial Q^T_\theta}{\partial\theta}\right\rangle + O(1) = N\frac{\partial(\bar{1}_m Q^T_\theta)}{\partial\theta} + O(1) = O(1),$$

where the last equality is true because $Q_\theta\bar{1}_m = \bar{0}_d$. Let $G^N_\theta(t) = \langle v^N_\theta, H^N_\theta(t)\rangle$. Then G^N_θ satisfies an ordinary differential equation of the form

$$\frac{dG^N_\theta(t)}{dt} = e^N_\theta G^N_\theta(t) + f^N_\theta h^N_\theta(t) + g^N_\theta,$$

where the sequences $\{e^N_\theta\}, \{g^N_\theta\}$ are in $O(N^{-1})$ and the sequence f^N_θ is in $O(1)$. Gronwall's inequality along with (4.25) and (4.17) imply that

$$\sup_{t \in [0,T]} |G^N_\theta(t)| = O(N^{-1}) \text{ and } \sup_{t\in[0,T]} |\langle \bar{1}_m, H^N_\theta(t)\rangle| = O(N^{-1}). \tag{4.26}$$

Let $\bar{H}^N_\theta(t)$ be the first $(m-1)$ components of $H^N_\theta(t)$. Differentiating (4.22) with respect to θ, we see that $\bar{H}^N_\theta$ satisfies an equation of the form

$$\begin{aligned}\frac{d\bar{H}^N_\theta(t)}{dt} &= \left(NC_\theta - \bar{D}_\theta + \lambda_\theta I_{m-1}\right)\bar{H}^N_\theta(t) + \left(N\frac{\partial C_\theta}{\partial\theta} - \frac{\partial \bar{D}_\theta}{\partial \theta} + \frac{\partial\lambda_\theta}{\partial\theta} I_{m-1}\right) h^N_\theta(t) \\ &\quad - \frac{\partial(\bar{D}_\theta\bar{\pi}_\theta)}{\partial\theta} + \frac{\partial(\lambda_\theta\bar{\pi}_\theta)}{\partial\theta}.\end{aligned}$$

If C^N_θ is the matrix given by (4.23), then we can solve for $\bar{H}^N_\theta$ as

$$\begin{aligned}\bar{H}^N_\theta(t) &= \exp(NC^N_\theta t)\bar{H}^N_\theta(0) - \int_0^t \exp(NC^N_\theta(t-s))\left(\frac{\partial(\bar{D}_\theta\bar{\pi}_\theta)}{\partial\theta} - \frac{\partial(\lambda_\theta\bar{\pi}_\theta)}{\partial\theta}\right) ds \\ &\quad + \int_0^t \exp(NC^N_\theta(t-s))\left(N\frac{\partial C_\theta}{\partial\theta} - \frac{\partial\bar{D}_\theta}{\partial\theta} + \frac{\partial\lambda_\theta}{\partial\theta} I_{m-1}\right) h^N_\theta(s) ds.\end{aligned}$$

From (4.24) and (4.25) we can deduce that $\bar{H}^N_\theta$ satisfies Condition 4.3. Using (4.26) it can be seen that H^N_θ also satisfies Condition 4.3. This completes the proof of the proposition. □

Corollary 4.6. *Let $\hat{\beta}^N_\theta$ be the function defined in Proposition 4.4. Then for any $T > 0$*

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \left|\left\langle \bar{1}_m, \hat{\beta}^N_\theta(t)\right\rangle\right| = 0 \text{ and } \lim_{N\to\infty}\sup_{t\in[0,T]}\left|\left\langle \bar{1}_m, \frac{\partial\hat{\beta}^N_\theta(t)}{\partial\theta}\right\rangle\right| = 0$$

Proof. The proof is immediate from (4.20) and (4.26). □

We end this section with an important observation.

Remark 4.7. *To prove Proposition 4.4 we used results from the theory of perturbation of finite matrices. Consider the situation where the state space $\mathcal{E}$ of the Markov chain is countably infinite. Now the matrix of transition rates Q_θ is infinite and it can be seen as a linear operator on $\mathcal{E}$. Proposition 4.1 will still hold in this case and assuming the existence of a suitable* Lyapunov function *(see [27]) for the Markov chain, one can use results from the perturbation theory of linear operators (see [10]) to prove Proposition 4.4 in a similar way.*

4.2 Construction of a new process

In this section we construct a new process W^N_θ and study some of its properties. As mentioned before, this process captures the single-time distribution of $X^N_{\gamma_2,\theta}$ (see Section 1) and its dynamics does not involve any "fast" transitions. We begin by making a remark which will simplify the proof of Theorem 3.2.

Remark 4.8. *From now on we will assume that $\gamma_2 = \gamma_1 + 1$, which can be ensured by redefining N, if necessary. Recall the description of the limiting process $\hat{X}_\theta$ in Theorem 3.2. Note that this process corresponds to a reduced model which does not contain any reactions in the set $(\Gamma_1 \cup \Gamma_2)^c = \{k = 1, \dots, K : k \notin \Gamma_1 \cup \Gamma_2\}$. We will prove Theorem 3.2 under the assumption that $(\Gamma_1\cup\Gamma_2)^c$ is empty. We later explain how the proof changes when this is not the case (see Remark 4.18).*

Since $\gamma_2 = \gamma_1 + 1$ and $(\Gamma_1 \cup \Gamma_2)^c = \emptyset$, the random time change representation of $\{X^N_{\gamma_2,\theta}(t) : t \geq 0\}$ is given by

$$X^N_{\gamma_2,\theta}(t) = x_0 + \sum_{k\in\Gamma_1} Y_k \left(N \int_0^t \lambda_k \left(X^N_{\gamma_2,\theta}(s), \theta \right) ds \right) \zeta_k + \sum_{k\in\Gamma_2} Y_k \left(\int_0^t \lambda_k \left(X^N_{\gamma_2,\theta}(s), \theta \right) ds \right) \zeta_k.$$

For each $k \in \Gamma_1 \cup \Gamma_2$ we let $\zeta^s_k = \Pi_2 \zeta_k$ and $\zeta^f_k = (I - \Pi_2)\zeta_k$. From (2.9) we know that $\zeta^s_k = \bar{0}_d$ for each $k \in \Gamma_1$. If we define two processes $X^N_{S,\theta}$ and $X^N_{F,\theta}$ by

$$X^N_{S,\theta}(t) = \Pi_2 X^N_{\gamma_2,\theta}(t) \quad \text{and} \quad X^N_{F,\theta}(t) = (I - \Pi_2) X^N_{\gamma_2,\theta}(t), \tag{4.27}$$

then their random time change representations are given by

$$X^N_{S,\theta}(t) = \Pi_2 x_0 + \sum_{k\in\Gamma_2} Y_k \left(\int_0^t \lambda_k \left(X^N_{S,\theta}(s) + X^N_{F,\theta}(s), \theta \right) ds \right) \zeta^s_k \tag{4.28}$$

$$\begin{aligned} X^N_{F,\theta}(t) = (I - \Pi_2) x_0 &+ \sum_{k\in\Gamma_2} Y_k \left(\int_0^t \lambda_k \left(X^N_{S,\theta}(s) + X^N_{F,\theta}(s), \theta \right) ds \right) \zeta^f_k \\ &+ \sum_{k\in\Gamma_1} Y_k \left(N \int_0^t \lambda_k \left(X^N_{S,\theta}(s) + X^N_{F,\theta}(s), \theta \right) ds \right) \zeta^f_k. \end{aligned} \tag{4.29}$$

Remark 4.9. *These representations show that between the successive jump times of $X^N_{S,\theta}$, if the state of this process is v, then the process $X^N_{F,\theta}$ evolves like a Markov process with state space $\mathbb{H}_v$ and generator $N\mathbb{C}^v_\theta$, where $\mathbb{C}^v_\theta$ is given by* (3.2).

The above remark motivates the construction of the process W^N_θ. Before we describe this construction we need to define certain quantities. Let $\lambda_0(x,\theta) = \sum_{k\in\Gamma_2} \lambda_k(x,\theta)$ and for any $k \in \Gamma_2$, $v \in \Pi_2\mathcal{S}$, $z \in \mathbb{H}_v$ and $t \geq 0$ define

$$\rho^N_{k,\theta}(t,v,z) = \frac{\mathbb{E}\left(\lambda_k(v + Z^N_\theta(t), \theta) \exp\left(-\int_0^t \lambda_0(v + Z^N_\theta(s), \theta) ds \right) \right)}{\mathbb{E}\left(\exp\left(-\int_0^t \lambda_0(v + Z^N_\theta(s), \theta) ds \right) \right)}, \tag{4.30}$$

where $\{Z^N_\theta(t) : t \geq 0\}$ is an independent Markov process with initial state z and generator $N\mathbb{C}^v_\theta$. For any $e \in \mathbb{H}_v$ define

$$\beta^N_\theta(t,v,z,e) = \mathbb{E}\left(\mathbb{1}_{\{Z^N_\theta(t) = e\}} \exp\left(-\int_0^t \lambda_0(v + Z^N_\theta(s), \theta) ds \right) \right) \tag{4.31}$$

$$\text{and} \quad \Theta^N_{k,\theta}(t,v,z,e) = \frac{\lambda_k(v+e,\theta)\beta^N_\theta(t,v,z,e)}{\rho^N_{k,\theta}(t,v,z) \exp\left(-\int_0^t \rho^N_{0,\theta}(t,v,z) ds \right)}, \tag{4.32}$$

where

$$\rho^N_{0,\theta}(t,v,z) = \sum_{k\in\Gamma_2} \rho^N_{k,\theta}(t,v,z). \tag{4.33}$$

If $\rho^N_{k,\theta}(t,v,z) = 0$ then instead of defining $\Theta^N_{k,\theta}(t,v,z,e)$ by (4.32) we do the following. We set $\Theta^N_{k,\theta}(t,v,z,z) = 1$ and set $\Theta^N_{k,\theta}(t,v,z,e) = 0$ for all $e \in \mathbb{H}_v - \{z\}$.

Recall that the set $\mathbb{H}_v$ is finite due to part (A) of Assumption 3.1. Proposition 4.1 shows that the mapping $t \mapsto \beta^N_\theta(t,v,z,e)$ is continuously differentiable, and hence the mappings $t \mapsto \rho^N_{k,\theta}(t,v,z)$ and $t \mapsto \Theta^N_{k,\theta}(t,v,z,e)$ are also continuously differentiable.

Lemma 4.10. *Fix a $v \in \Pi_2\mathcal{S}$, $z \in \mathbb{H}_v$ and $t \geq 0$.*

(A) Let $\{Z_\theta^N(t) : t \geq 0\}$ *be an independent Markov process with initial state* z *and generator* $N\mathbb{C}_\theta^v$*. Then*

$$\exp\left(-\int_0^t \rho_{0,\theta}^N(s,v,z)ds\right) = \mathbb{E}\left(\exp\left(-\int_0^t \lambda_0(v+Z_\theta^N(s),\theta)ds\right)\right).$$

(B) For any $k \in \Gamma_2$

$$\sum_{e\in\mathbb{H}_v} \Theta_{k,\theta}^N(t,v,z,e) = 1.$$

Proof. Observe that

$$\rho_{0,\theta}^N(s,v,z) = \sum_{k\in\Gamma_2} \rho_{k,\theta}^N(s,v,z) = \frac{\mathbb{E}\left(\lambda_0(v+Z_\theta^N(s),\theta)\exp\left(-\int_0^t \lambda_0(v+Z_\theta^N(s),\theta)ds\right)\right)}{\mathbb{E}\left(\exp\left(-\int_0^t \lambda_0(v+Z_\theta^N(s),\theta)ds\right)\right)}$$
$$= -\frac{d}{dt}\log\left(\mathbb{E}\left(\exp\left(-\int_0^t \lambda_0(x+Z_\theta^N(s),\theta)ds\right)\right)\right).$$

Integrating both sides with respect to t and then exponentiating proves part (A). From (4.30) we get

$$\begin{aligned}
&\rho_{k,\theta}^N(s,v,z)\exp\left(-\int_0^t \rho_{0,\theta}^N(s,v,z)ds\right)\\
&= \mathbb{E}\left(\lambda_k(v+Z_\theta^N(t),\theta)\exp\left(-\int_0^t \lambda_0(v+Z_\theta^N(s),\theta)ds\right)\right)\\
&= \sum_{e\in\mathbb{H}_v} \lambda_k(v+e,\theta)\mathbb{E}\left(\mathbb{1}_{\{Z_\theta^N(t)=e\}}\exp\left(-\int_0^t \lambda_0(v+Z_\theta^N(s),\theta)ds\right)\right)\\
&= \sum_{e\in\mathbb{H}_v} \lambda_k(v+e,\theta)\beta_\theta^N(t,v,z,e). \qquad (4.34)
\end{aligned}$$

Hence

$$\sum_{e\in\mathbb{H}_v} \Theta_{k,\theta}^N(t,v,z,e) = \sum_{e\in\mathbb{H}_v} \frac{\lambda_k(v+e,\theta)\beta_\theta^N(t,v,z,e)}{\rho_{k,\theta}^N(s,v,z)\exp\left(-\int_0^t \rho_{0,\theta}^N(s,v,z)ds\right)} = 1,$$

and this proves part (B). ☐

Part (B) of Lemma 4.10 shows that for any $k \in \Gamma_2$, $v \in \Pi_2\mathcal{S}$, $z \in \mathbb{H}_v$ and $t \geq 0$, we can regard $\Theta_{k,\theta}^N(t,v,z,\cdot)$ as a probability measure on $\mathbb{H}_v$. We know that $\mathbb{H}_v$ is a finite set. From now on, whenever we write $\mathbb{H}_v = \{e_1,\dots,e_m\}$, we will assume that the elements are arranged in the lexicographical order on $\mathbb{R}^d$. For any $u \in (0,1)$ define

$$F_{k,\theta}^N(t,v,z,u) = e_i \text{ where } i = \min\left\{l = 1,\dots,m : u \leq \sum_{n=1}^{l} \Theta_{k,\theta}^N(t,v,z,e_n)\right\}. \qquad (4.35)$$

Then a $\mathbb{H}_v$-valued random variable with distribution $\Theta_{k,\theta}^N(t,v,z,\cdot)$ can be generated by transforming a Unif$(0,1)$ random variable u with the function $F_{k,\theta}^N(t,v,z,\cdot)$. The next lemma will be useful in proving the main result.

Lemma 4.11. *Fix a* $v \in \Pi_2\mathcal{S}$, $z \in \mathbb{H}_v$ *and* $t > 0$. *Let* $\mathbb{H}_v = \{e_1,\dots,e_m\}$ *and* u *be a* Unif$(0,1)$ *random variable. Pick* $i,j \in \{1,\dots,m\}$ *such that* $i \neq j$. *Then*

$$\lim_{h\to 0} \frac{\mathbb{P}\left(F_{k,\theta}^N(t,v,z,u) = e_i \text{ and } F_{k,\theta+h}^N(t,v,z,u) = e_j\right)}{h} \leq \sum_{e\in\mathbb{H}_v}\left|\frac{\partial\Theta_{k,\theta}^N(t,v,z,e)}{\partial\theta}\right|$$

Proof. For proving this lemma we can assume that $\Theta^N_{k,\theta}(t,v,z,e) > 0$ for each $e \in \mathbb{H}_v$. Let $\mathbb{H}_v = \{e_1, \dots, e_m\}$ and for any $l = 1, \dots, m$ define

$$A_l(\theta) = \sum_{n=1}^{l} \Theta^N_{k,\theta}(t,v,z,e_n).$$

Note that $A_m(\theta) = 1$ for any θ due to part(B) of Lemma 4.10. For convenience let $A_0(\theta) = 0$ for any θ. For small values of h we can write

$$\begin{aligned}&\mathbb{P}\left(F^N_{k,\theta}(t,v,z,u) = e_i \text{ and } F^N_{k,\theta+h}(t,v,z,u) = e_j\right)\\ &= \mathbb{P}\left(u \in (A_{i-1}(\theta), A_i(\theta)) \text{ and } u \in (A_{j-1}(\theta+h), A_j(\theta+h))\right).\end{aligned}$$

Since $\Theta^N_{k,\theta}(t,v,z,e_l) > 0$ for each $l = 1, \dots, m$, this probability is 0 if $j > i+1$ or $j < i-1$. Assume that $j = i+1$ for $i < m$. Then for small values of h we can write

$$\begin{aligned}\mathbb{P}\left(F^N_{k,\theta}(t,v,z,u) = e_i \text{ and } F^N_{k,\theta+h}(t,v,z,u) = e_j\right) &= \mathbb{P}\left(u \in (A_i(\theta+h), A_i(\theta))\right)\\ &= \left[\frac{\partial A_i(\theta)}{\partial \theta}\right]^- h + o(h).\end{aligned}$$

Therefore

$$\lim_{h\to 0} \frac{\mathbb{P}\left(F^N_{k,\theta}(t,v,z,u) = e_i \text{ and } F^N_{k,\theta+h}(t,v,z,u) = e_j\right)}{h} = \left[\frac{\partial A_i(\theta)}{\partial \theta}\right]^-.$$

Similarly for $j = i-1$ and $i > 1$ we can show that

$$\lim_{h\to 0} \frac{\mathbb{P}\left(F^N_{k,\theta}(t,v,z,u) = e_i \text{ and } F^N_{k,\theta+h}(t,v,z,u) = e_j\right)}{h} = \left[\frac{\partial A_{i-1}(\theta)}{\partial \theta}\right]^+.$$

Combining the last two relations proves the lemma. □

The new process W^N_θ will be a Markov process on state space $\hat{S}$ given by

$$\hat{\mathcal{S}} = \{(t,v,z) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d : v \in \Pi_2\mathcal{S} \text{ and } z \in \mathbb{H}_v\}. \tag{4.36}$$

Let $\Pi_{\hat{\mathcal{S}}}$ be the projection map from $\hat{\mathcal{S}}$ to $\Pi_2\mathcal{S}$ defined by

$$\Pi_{\hat{\mathcal{S}}}(t,v,z) = v. \tag{4.37}$$

We say that a function $f : \hat{\mathcal{S}} \to \mathbb{R}$ is compactly supported with respect to projection $\Pi_{\hat{\mathcal{S}}}$ if there exists a compact set $\mathcal{K} \subset \Pi_2\mathcal{S}$ such that $f(x) = 0$ for any $x \in \hat{\mathcal{S}}$ with $\Pi_{\hat{\mathcal{S}}}(x) \notin \mathcal{K}$. We now define a class of bounded real-valued functions over $\hat{\mathcal{S}}$ by

$$\begin{aligned}\mathcal{C} = \Big\{ &f \in \mathcal{B}(\hat{\mathcal{S}}) : f \text{ is compactly supported with respect to projection } \Pi_{\hat{\mathcal{S}}} \text{ and the mapping}\\ &t \mapsto f(t,v,z) \text{ is continuously differentiable for each } v \in \Pi_2\mathcal{S} \text{ and } z \in \mathbb{H}_v \Big\}. \end{aligned} \tag{4.38}$$

Let $\{W^N_\theta(t) : t \geq 0\}$ be the $\hat{\mathcal{S}}$-valued Markov process with initial state $(0, v_0, z_0) = (0, \Pi_2 x_0, (I - \Pi_2)x_0)$ and generator given by

$$\begin{aligned}\mathbb{B}^N_\theta f(t,v,z) &= \frac{\partial f(t,v,z)}{\partial t} \\ &+ \sum_{k\in\Gamma_2} \rho^N_{k,\theta}(t,v,z) \sum_{e\in\mathbb{H}_v} \left(f(0, v+\zeta^s_k, e+\zeta^f_k) - f(t,v,z)\right)\Theta^N_{k,\theta}(t,v,z,e),\end{aligned} \tag{4.39}$$

for all $f \in \mathcal{D}(\mathbb{B}^N_\theta) = \mathcal{C}$. The existence and uniqueness of the process W^N_θ is a direct consequence of the well-posedness of the martingale problem for $\mathbb{B}^N_\theta$, which is verified in Lemma A.2.

In the rest of this section we study some properties of the process W^N_θ. Observe that the definition of $\hat{\mathcal{S}}$ (see (4.36)) allows us to write

$$W^N_\theta(t) = \left(\tau^N_\theta(t), V^N_\theta(t), Z^N_\theta(t)\right) \text{ for all } t \geq 0, \tag{4.40}$$

where τ^N_θ, V^N_θ and Z^N_θ are processes with state spaces $\mathbb{R}_+, \Pi_2\mathcal{S}$ and $\cup_{v\in\Pi_2\mathcal{S}}\mathbb{H}_v$ respectively. Let σ^N_i denote the i-th jump time of the process W^N_θ for $i = 1, \dots$. We define $\sigma^N_0 = 0$ for convenience. From the form of the generator $\mathbb{B}^N_\theta$ it is immediate that between the jump times, τ^N_θ increases linearly at rate 1 while V^N_θ and Z^N_θ remain constant. Hence

$$\left(\tau^N_\theta(t), V^N_\theta(t), Z^N_\theta(t)\right) = \left(t - \sigma^N_{i-1}, V^N_\theta(\sigma^N_{i-1}), Z^N_\theta(\sigma^N_{i-1})\right) \text{ for any } i \in \mathbb{N} \text{ and } t \in [\sigma^N_{i-1}, \sigma^N_i). \tag{4.41}$$

Let η_i be the Γ_2-valued random variable that denotes the direction of the jump at time σ^N_i and let ξ_i be the random variable given by $Z^N_\theta(\sigma^N_i-)$. The form of $\mathbb{B}^N_\theta$ allows us to compute the distributions of the random variables $(\sigma^N_i - \sigma^N_{i-1})$, η_i and ξ_i from the values of $V^N_\theta(\sigma^N_{i-1})$ and $Z^N_\theta(\sigma^N_{i-1})$. Let $E_i(v,z)$ denote the event

$$E_i(v,z) = \{V^N_\theta(\sigma^N_i) = v, Z^N_\theta(\sigma^N_i) = z\}.$$

Then given $E_{i-1}(v,z)$, $(\sigma^N_i - \sigma^N_{i-1})$ is a $\mathbb{R}_+$-valued random variable with density

$$\rho^N_{0,\theta}(t,v,z) \exp\left(-\int_0^t \rho^N_{0,\theta}(s,v,z)ds\right) dt. \tag{4.42}$$

Given $E_{i-1}(v,z)$ and $(\sigma^N_i - \sigma^N_{i-1}) = t$, η_i is a Γ_2-valued random variable with distribution

$$\mathbb{P}\left(\eta_i = k | E_{i-1}(v,z), (\sigma^N_i - \sigma^N_{i-1}) = t\right) = \frac{\rho^N_{k,\theta}(t,v,z)}{\rho^N_{0,\theta}(t,v,z)}. \tag{4.43}$$

Moreover conditioned on $E_{i-1}(v,z)$, $(\sigma^N_i - \sigma^N_{i-1}) = t$ and $\eta_i = k$, the $\mathbb{H}_v$-valued random variable ξ_i has distribution $\Theta^N_{k,\theta}(t,v,z,\cdot)$. Using (4.42) and (4.43) we can deduce that

$$\lim_{h\to 0} \frac{\mathbb{P}\left(\sigma^N_i \in (\sigma^N_{i-1} + t, \sigma^N_{i-1} + t + h), V^N_\theta(\sigma^N_i) = v + \zeta^s_k, Z^N_\theta(\sigma^N_i) = e + \zeta^f_k \middle| E_{i-1}(v,z)\right)}{h} \tag{4.44}$$

$$= \rho^N_{k,\theta}(t,v,z) \exp\left(-\int_0^t \rho^N_{0,\theta}(u,v,z)du\right) \Theta^N_{k,\theta}(t,v,z,e),$$

for any $i = 1, 2, \dots$.

Remark 4.12. *The preceding discussion suggests a simple scheme to construct the process $\{W^N_\theta(t) = (\tau^N_\theta(t), V^N_\theta(t), Z^N_\theta(t)) : t \geq 0\}$ with generator $\mathbb{B}^N_\theta$ and initial state $(0, v_0, z_0)$. Consider the random time change representation*

$$V^N_\theta(t) = v_0 + \sum_{k\in\Gamma_2} Y_k \left(\int_0^t \rho^N_{k,\theta}(\tau^N_\theta(s), V^N_\theta(s), Z^N_\theta(s))ds\right) \zeta^s_k, \tag{4.45}$$

where $\{Y_k : k \in \Gamma_2\}$ is a family of independent unit rate Poisson processes. The processes $\tau^N_\theta, V^N_\theta$ and Z^N_θ can be constructed as follows. For each $i \in \mathbb{N}_0$ let σ^N_i be the i-th *jump time of process V_θ^N, where $\sigma_0^N = 0$. Defining $(\tau_\theta^N(0), V_\theta^N(0), Z_\theta^N(0)) = (0, v_0, z_0)$ constructs the process W_θ^N until time σ_0^N. Assume that this process is constructed until time σ_{i-1}^N for some $i = 1, 2, \dots$. Then the next jump time σ_i^N can be evaluated from* (4.45) *and the process W_θ^N can be defined in the time interval $[\sigma_{i-1}^N, \sigma_i^N)$ using* (4.41)*. If $V_\theta^N(\sigma_i^N) = v, Z_\theta^N(\sigma_i^N) = z$ and $\sigma_i^N - \sigma_{i-1}^N = t$ then we choose random variables η_i and ξ_i according to distributions* (4.43) *and $\Theta_{\eta_i,\theta}^N(t, v, z, \cdot)$ respectively and define*

$$(\tau_\theta^N(\sigma_i^N), V_\theta^N(\sigma_i^N), Z_\theta^N(\sigma_i^N)) = (0, v + \zeta_{\eta_i}^s, \xi_i + \zeta_{\eta_i}^f).$$

This completes the construction of the process until the next jump time σ_i^N. Proceeding this way we can define $W_\theta^N(t) = (\tau_\theta^N(t), V_\theta^N(t), Z_\theta^N(t))$ for all $t \geq 0$. The relation (4.44) *ensures that the process W_θ^N has generator $\mathbb{B}_\theta^N$.*

In the next proposition we show that the single-time distribution of the process $X_{\gamma_2,\theta}^N$ can be captured with the process W_θ^N.

Proposition 4.13. *For $i \in \mathbb{N}$, let δ_i^N and σ_i^N denote the i-th jump time of the processes $\Pi_2 X_{\gamma_2,\theta}^N$ and W_θ^N respectively. We define $\delta_0^N = \sigma_0^N = 0$ for convenience. Then we have the following.*

(A) Let the processes V_θ^N and Z_θ^N be related to the process W_θ^N by (4.40)*. For each $i = 0, 1, 2, \dots$,*

$$\left(\delta_i^N, \Pi_2 X_{\gamma_2,\theta}^N(\delta_i^N), (I - \Pi_2) X_{\gamma_2,\theta}^N(\delta_i^N)\right) \stackrel{d}{=} \left(\sigma_i^N, V_\theta^N(\sigma_i^N), Z_\theta^N(\sigma_i^N)\right), \tag{4.46}$$

where $\stackrel{d}{=}$ denotes equality in distribution.

(B) Let $f : \mathcal{S} \to \mathbb{R}$ be a polynomially growing function with respect to projection Π_2 (see Definition 2.1). Then for any $t \geq 0$

$$\mathbb{E}\left(f(X_{\gamma_2,\theta}^N(t))\right) = \mathbb{E}\left(f_\theta^N(W_\theta^N(t))\right),$$

where $f_\theta : \hat{\mathcal{S}} \to \mathbb{R}$ is the function given by

$$f_\theta^N(t, v, z) = \frac{\sum_{e \in \mathbb{H}_v} f(v + e) \beta_\theta^N(t, v, z, e)}{\exp\left(-\int_0^t \rho_{0,\theta}^N(s, v, z) ds\right)}. \tag{4.47}$$

Remark 4.14. *Note that for any $v \in \Pi_2 \mathcal{S}$ and $z \in \mathbb{H}_v$, the mapping $t \mapsto f_\theta^N(t, v, z)$ is continuously differentiable with respect to t. Let $\partial f_\theta^N(t, v, z)/\partial t$ denote the derivative of this map. Since f is polynomially growing with respect to projection Π_2, the sequences of functions $\{f_\theta^N : N \in \mathbb{N}\}$, $\{\partial f_\theta^N/\partial t : N \in \mathbb{N}\}$ and $\{\mathbb{B}_\theta^N f_\theta^N : N \in \mathbb{N}\}$ are also polynomially growing with respect to projection $\Pi_{\hat{\mathcal{S}}}$.*

Proof. We prove part (A) by induction in i. Relation (4.46) certainly holds for $i = 0$. Suppose it holds for $(i-1)$ for some $i \in \mathbb{N}$. Then

$$\left(\delta_{i-1}^N, X_{S,\theta}^N(\delta_{i-1}^N), X_{F,\theta}^N(\delta_{i-1}^N)\right) \stackrel{d}{=} W_\theta^N(\sigma_{i-1}^N), \tag{4.48}$$

where the processes $X_{S,\theta}^N$ and $X_{F,\theta}^N$ are given by (4.27).

For any $v \in \Pi_2 \mathcal{S}$ and $z \in \mathbb{H}_v$ let $E_{i-1}(v, z)$ denote the event

$$E_{i-1}(v, z) = \{X_{S,\theta}^N(\delta_{i-1}^N) = v, X_{F,\theta}^N(\delta_{i-1}^N) = z\}. \tag{4.49}$$

Let η_i be the Γ_2-valued random variable that gives the jump direction of the process $X^N_{S,\theta}$ at time δ^N_i. For any $t > 0$, $k \in \Gamma_2$ and $e \in \mathbb{H}_v$ we can write

$$\lim_{h\to 0} \frac{\mathbb{P}\left(\delta^N_i \in (\delta^N_{i-1}+t, \delta^N_{i-1}+t+h), X^N_{S,\theta}(\delta^N_i) = v+\zeta^s_k, X^N_{F,\theta}(\delta^N_i) = e+\zeta^f_k \Big| E_{i-1}(v,z)\right)}{h}$$
$$= \lim_{h\to 0} \frac{\mathbb{P}\left(\delta^N_i - \delta^N_{i-1} \in (t,t+h), \eta_i = k, X^N_{F,\theta}(\delta^N_i -) = e \Big| E_{i-1}(v,z)\right)}{h}. \tag{4.50}$$

Let $\{\bar{Z}^N_\theta(t) : t \geq 0\}$ be an independent Markov process with initial state z and generator $N\mathbb{C}^v_\theta$. For each $k \in \Gamma_2$ let u_k be an independent $\mathrm{Unif}(0,1)$ random variable. Using the observation made in Remark (4.9), and the random time change representation (4.28) we can write

$$\begin{aligned}
&\mathbb{P}\left(\delta^N_i - \delta^N_{i-1} \in (t,t+h), \eta_i = k, X^N_{F,\theta}(\delta^N_i -) = e \big| E_{i-1}(v,z)\right)\\
&= \mathbb{P}\Bigg(\int_0^{t+h} \lambda_k(v+\bar{Z}^N_\theta(u),\theta)du \geq -\log u_k \geq \int_0^t \lambda_k(v+\bar{Z}^N_\theta(u),\theta)du,\\
&\qquad \int_0^t \lambda_j(v+\bar{Z}^N_\theta(u),\theta)du < -\log u_j \text{ for all } j \in \Gamma_2 - \{k\} \text{ and } \bar{Z}^N_\theta(t) = e\Bigg) + o(h)\\
&= \lambda_k(v+e,\theta)\mathbb{E}\left(\mathbb{1}_{\{\bar{Z}^N_\theta(t)=e\}} \exp\left(-\int_0^t \lambda_0(v+\bar{Z}^N_\theta(u),\theta)du\right)\right)h + o(h),
\end{aligned} \tag{4.51}$$

where $o(h)$ denotes any quantity which upon division by h, goes to 0 as $h \to 0$. To obtain (4.51) we integrated with respect to the joint density of $\{u_k : k \in \Gamma_2\}$. Note that due to (4.31) and (4.32) we get

$$\begin{aligned}
&\lambda_k(v+e,\theta)\mathbb{E}\left(\mathbb{1}_{\{\bar{Z}^N_\theta(t)=e\}} \exp\left(-\int_0^t \lambda_0(v+\bar{Z}^N_\theta(u),\theta)du\right)\right)\\
&= \lambda_k(v+e,\theta)\beta^N_\theta(t,v,z,e)\\
&= \rho^N_{k,\theta}(t,v,z)\exp\left(-\int_0^t \rho^N_{0,\theta}(s,v,z)ds\right)\Theta^N_{k,\theta}(t,v,z,e).
\end{aligned}$$

Hence relations (4.50) and (4.51) yield

$$\lim_{h\to 0} \frac{\mathbb{P}\left(\delta^N_i \in (\delta^N_{i-1}+t, \delta^N_{i-1}+t+h), X^N_{S,\theta}(\delta^N_i) = v+\zeta^s_k, X^N_{F,\theta}(\delta^N_i) = e+\zeta^f_k \Big| E_{i-1}(v,z)\right)}{h}$$
$$= \rho^N_{k,\theta}(t,v,z)\exp\left(-\int_0^t \rho^N_{0,\theta}(s,v,z)ds\right)\Theta^N_{k,\theta}(t,v,z,e).$$

From (4.44) it follows that for all $v \in \Pi_2\mathcal{S}$ and $z \in \mathbb{H}_v$

$$\lim_{h\to 0} \frac{\mathbb{P}\left(\sigma^N_i \in (\sigma^N_{i-1}+t, \sigma^N_{i-1}t+h), V^N_\theta(\sigma^N_i) = v+\zeta^s_k, Z^N_\theta(\sigma^N_i) = e+\zeta^f_k \Big| E_{i-1}(v,z)\right)}{h}$$
$$= \lim_{h\to 0} \frac{\mathbb{P}\left(\delta^N_i \in (\delta^N_{i-1}+t, \delta^N_{i-1}+t+h), X^N_{S,\theta}(\delta^N_i) = v+\zeta^s_k, X^N_{F,\theta}(\delta^N_i) = e+\zeta^f_k \Big| E_{i-1}(v,z)\right)}{h}.$$

This relation and (4.48) imply that

$$\left(\delta^N_i, X^N_{S,\theta}(\delta^N_i), X^N_{F,\theta}(\delta^N_i)\right) \stackrel{d}{=} \left(\sigma^N_i, V^N_\theta(\sigma^N_i), Z^N_\theta(\sigma^N_i)\right),$$

which completes the proof of part (A).

We now prove part (B). From Remark 4.14 and Lemma A.2 we can conclude that for any $t \geq 0$

$$\sup_{N \in \mathbb{N}} \mathbb{E}\left(f_\theta^N(W_\theta^N(t))\right) < \infty.$$

Moreover, one can rework the proof of part (C) of Lemma A.1 to show that

$$\sup_{N \in \mathbb{N}} \mathbb{E}\left(f(X^N_{\gamma_2,\theta}(t))\right) < \infty \text{ for any } t \geq 0.$$

Let $\{\mathcal{F}_t\}$ be the filtration generated by the process $\{X^N_{\gamma_2,\theta}(t) : t \geq 0\}$. Then we can write

$$\begin{aligned}
\mathbb{E}\left(f(X^N_{\gamma_2,\theta}(t))\right) &= \sum_{i=1}^{\infty} \mathbb{E}\left(\mathbb{1}_{\{\delta^N_{i-1} \leq t < \delta^N_i\}} f(X^N_{\gamma_2,\theta}(t))\right) \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left(\mathbb{1}_{\{\delta^N_{i-1} \leq t < \delta^N_i\}} f(X^N_{S,\theta}(t) + X^N_{F,\theta}(t))\right) \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left(\mathbb{1}_{\{\delta^N_{i-1} \leq t\}} \mathbb{E}\left(\mathbb{1}_{\{\delta^N_i - \delta^N_{i-1} > t - \delta^N_{i-1}\}} f(X^N_{S,\theta}(\delta^N_{i-1}) + X^N_{F,\theta}(t)) | \mathcal{F}_{\delta^N_{i-1}}\right)\right).
\end{aligned} \tag{4.52}$$

For any $v \in \Pi_2 \mathcal{S}$ and $z \in \mathbb{H}_v$, let $E_{i-1}(v,z)$ be the event given by (4.49). Suppose $\mathbb{H}_v = \{e_1, \dots, e_m\}$ and $\{\bar{Z}^N_\theta(t) : t \geq 0\}$ is an independent Markov process with initial state z and generator $N\mathcal{C}^v_\theta$. For each $k \in \Gamma_2$ let u_k be an independent Unif$(0,1)$ random variable. Using the observation made in Remark (4.9), and the random time change representation (4.28), for any $s < t$ we can write

$$\begin{aligned}
&\mathbb{E}\left(\mathbb{1}_{\{\delta^N_i - \delta^N_{i-1} > t - \delta^N_{i-1}\}} f(X^N_{S,\theta}(\delta^N_{i-1}) + X^N_{F,\theta}(t)) | E_{i-1}(v,z), \delta^N_{i-1} = s\right) \\
&= \mathbb{E}\left(\mathbb{1}_{\{\delta^N_i - \delta^N_{i-1} > t - s\}} f(v + \bar{Z}^N_\theta(t-s))\right) \\
&= \sum_{e \in \mathbb{H}_v} \mathbb{P}\left(\int_0^{t-s} \lambda_k(v + \bar{Z}^N_\theta(u), \theta) du < -\log u_k \text{ for all } k \in \Gamma_2,\ \bar{Z}^N_\theta(t-s) = e\right) f(v+e) \\
&= \sum_{e \in \mathbb{H}_v} \mathbb{E}\left(\mathbb{1}_{\{\bar{Z}^N_\theta(t-s) = e\}} \exp\left(-\int_0^{t-s} \lambda_0(v + \bar{Z}^N_\theta(u), \theta) du\right)\right) f(v+e).
\end{aligned}$$

The last inequality is obtained by integrating with respect to the joint density of $\{u_k : k \in \Gamma_2\}$. Due to (4.31) and (4.47) we obtain

$$\begin{aligned}
&\mathbb{E}\left(\mathbb{1}_{\{\delta^N_i - \delta^N_{i-1} > t - \delta^N_{i-1}\}} f(X^N_{S,\theta}(\delta^N_{i-1}) + X^N_{F,\theta}(t)) | E_{i-1}(v,z), \delta^N_{i-1} = s\right) \\
&= \sum_{e \in \mathbb{H}_v} f(v+e) \beta^N_\theta(t-s, v, z, e) \\
&= \exp\left(-\int_0^{t-s} \rho^N_{0,\theta}(u, v, z) du\right) f^N_\theta(t-s, v, z),
\end{aligned}$$

which shows that

$$\begin{aligned}
&\mathbb{E}\left(\mathbb{1}_{\{\delta^N_i - \delta^N_{i-1} > t - \delta^N_{i-1}\}} f(X^N_{S,\theta}(\delta^N_{i-1}) + X^N_{F,\theta}(t)) | \mathcal{F}_{\delta^N_{i-1}}\right) \\
&= \exp\left(-\int_0^{t-\delta^N_{i-1}} \rho^N_{0,\theta}(u, X^N_{S,\theta}(\delta^N_{i-1}), X^N_{F,\theta}(\delta^N_{i-1})) du\right) f^N_\theta(t - \delta^N_{i-1}, X^N_{S,\theta}(\delta^N_{i-1}), X^N_{F,\theta}(\delta^N_{i-1})).
\end{aligned}$$

Substituting this relation in (4.52) and using part (A) gives us

$$\begin{aligned}\mathbb{E}\left(f(X^N_{\gamma_2,\theta}(t))\right) &= \sum_{i=1}^{\infty}\mathbb{E}\left(\mathbb{1}_{\{\delta^N_{i-1}\leq t\}}\exp\left(-\int_0^{t-\delta^N_{i-1}}\rho^N_{0,\theta}(u,X^N_{S,\theta}(\delta^N_{i-1}),X^N_{F,\theta}(\delta^N_{i-1}))du\right)\right.\\ &\qquad\left.\times f^N_\theta(t-\delta^N_{i-1},X^N_{S,\theta}(\delta^N_{i-1}),X^N_{F,\theta}(\delta^N_{i-1}))\right)\\ &= \sum_{i=1}^{\infty}\mathbb{E}\left(\mathbb{1}_{\{\sigma^N_{i-1}\leq t\}}\exp\left(-\int_0^{t-\sigma^N_{i-1}}\rho^N_{0,\theta}(u,V^N_\theta(\sigma^N_{i-1}),Z^N_\theta(\sigma^N_{i-1}))du\right)\right.\\ &\qquad\left.\times f^N_\theta(t-\sigma^N_{i-1},V^N_\theta(\sigma^N_{i-1}),Z^N_\theta(\sigma^N_{i-1}))\right).\end{aligned}$$

However from (4.41) and (4.42) we can conclude that

$$\begin{aligned}&\sum_{i=1}^{\infty}\mathbb{E}\left(\mathbb{1}_{\{\sigma^N_{i-1}\leq t\}}\exp\left(-\int_0^{t-\sigma^N_{i-1}}\rho^N_{0,\theta}(u,V^N_\theta(\sigma^N_{i-1}),Z^N_\theta(\sigma^N_{i-1}))du\right)\right.\\ &\qquad\left.\times f^N_\theta(t-\sigma^N_{i-1},V^N_\theta(\sigma^N_{i-1}),Z^N_\theta(\sigma^N_{i-1}))\right)\\ &= \sum_{i=1}^{\infty}\mathbb{E}\left(\mathbb{1}_{\{\sigma^N_{i-1}\leq t<\sigma^N_i\}}f^N_\theta(t-\sigma^N_{i-1},V^N_\theta(\sigma^N_{i-1}),Z^N_\theta(\sigma^N_{i-1}))\right)\\ &= \sum_{i=1}^{\infty}\mathbb{E}\left(\mathbb{1}_{\{\sigma^N_{i-1}\leq t<\sigma^N_i\}}f^N_\theta(\tau^N_\theta(t),V^N_\theta(t),Z^N_\theta(t))\right)\\ &= \mathbb{E}\left(f^N_\theta(\tau^N_\theta(t),V^N_\theta(t),Z^N_\theta(t))\right)\\ &= \mathbb{E}\left(f^N_\theta(W^N_\theta(t)\right).\end{aligned}$$

This proves part (B) of the proposition. □

Part (B) of Assumption 3.1 says that a Markov process with generator $\mathbb{C}^v_\theta$ is ergodic and its unique stationary distribution is $\pi^v_\theta \in \mathcal{P}(\mathbb{H}_v)$. Since $\mathbb{H}_v$ is finite, we can view π^v_θ as a vector in $\mathbb{R}^n$ where $n = |\mathbb{H}_v|$. The differentiability of π^z_θ with respect to θ follows from arguments given in Section 4.1. Let $\{f^N : N \in \mathbb{N}\}$ be a sequence of real valued functions on $\mathbb{R}_+$ and let c be a constant. In the next lemma we will use the notation $f^N \to c$ to denote that the sequence of functions $\{(f^N - c) : N \in \mathbb{N}\}$ satisfies Condition 4.3.

Lemma 4.15. *Fix a $v \in \Pi_2\mathcal{S}$ and a $z \in \mathbb{H}_v$. Then we have the following.*

(A) For any $k \in \Gamma_2$

$$\rho^N_{k,\theta}(\cdot,v,z) \to \hat{\lambda}_k(v,\theta) \text{ and } \frac{\partial\rho^N_{k,\theta}(\cdot,v,z)}{\partial\theta} \to \frac{\partial\hat{\lambda}_k(v,\theta)}{\partial\theta},$$

where $\hat{\lambda}_k$ is defined by (3.4)*.*

(B) For any $k \in \Gamma_2$ and $e \in \mathbb{H}_v$

$$\Theta^N_{k,\theta}(\cdot,v,z,e) \to \frac{\lambda_k(v+e,\theta)\pi^v_\theta(e)}{\hat{\lambda}_k(v,\theta)} \text{ and } \frac{\partial\Theta^N_{k,\theta}(\cdot,v,z,e)}{\partial\theta} \to \frac{\partial}{\partial\theta}\left(\frac{\lambda_k(v+e,\theta)\pi^v_\theta(e)}{\hat{\lambda}_k(v,\theta)}\right).$$

(C) Fix a function $f : \mathcal{S} \to \mathbb{R}$. Let f^N_θ and f_θ be given by (4.47) *and* (3.6) *respectively. Then*

$$f^N_\theta(\cdot,v,z) \to f_\theta(v) \text{ and } \frac{\partial f^N_\theta(\cdot,v,z)}{\partial\theta} \to \frac{\partial f_\theta(v)}{\partial\theta}.$$

Proof. Assume that $\mathbb{H}_v = \{e_1, \dots, e_m\}$. For each $l = 1, \dots, m$, let $\hat{\beta}^N_{\theta,l} : \mathbb{R}_+ \to \mathbb{R}$ be given by

$$\hat{\beta}^N_{\theta,l}(t) = \beta^N_\theta(t, v, z, e_l) - \exp(-d_\theta(v)t)\pi^v_\theta(e_l),$$

where $d_\theta(v) = \sum_{e \in \mathbb{H}_v} \lambda_0(v + e, \theta)\pi^v_\theta(e)$. Observe that

$$\exp\left(-\int_0^t \rho^N_{0,\theta}(s, v, z)ds\right) = \sum_{e \in \mathbb{H}_v} \beta^N_\theta(t, v, z, e) = \sum_{l=1}^m \hat{\beta}^N_{\theta,l}(t) + \exp(-d_\theta(v)t).$$

From Corollary 4.6 we get that for any $T > 0$

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \left| \exp\left(-\int_0^t \rho^N_{0,\theta}(s, v, z)ds\right) - \exp(-d_\theta(v)t)\right| = 0 \tag{4.53}$$

$$\text{and} \quad \lim_{N\to\infty} \sup_{t\in[0,T]} \left| \frac{\partial}{\partial\theta} \exp\left(-\int_0^t \rho^N_{0,\theta}(s, v, z)ds\right) - \frac{\partial}{\partial\theta} \exp(-d_\theta(v)t)\right| = 0 \tag{4.54}$$

Using part (A) of Lemma 4.10 we can write

$$\rho^N_{k,\theta}(t, v, z) = \frac{\sum_{l=1}^m \lambda_k(v + e_l, \theta)\hat{\beta}^N_{\theta,l}(t)}{\sum_{l=1}^m \hat{\beta}^N_{\theta,l}(t)}.$$

From Proposition 4.4 we can see that each $\hat{\beta}^N_{\theta,l}$ satisfies Condition 4.3. This fact along with (4.53) and (4.54) proves part (A).

The proof of part (B) is immediate from the definition of $\Theta^N_{k,\theta}$ (see (4.32)), part (A), (4.53) and (4.54). Note that f^N_θ can be written as

$$f^N_\theta(t, v, z) = \frac{\sum_{l=1}^m f(v + e_l)\hat{\beta}^N_{\theta,l}(t)}{\sum_{l=1}^m \hat{\beta}^N_{\theta,l}(t)},$$

which enables us to prove part (C) in the same way as part (A). □

For the next proposition, recall the definition of the projection map $\Pi_{\hat{\mathcal{S}}}$ from (4.37) and the definition of $\hat{\mathbb{A}}_\theta$ from (3.3).

Proposition 4.16. *Fix $(t_0, v_0, z_0) \in \hat{\mathcal{S}}$ and let W^N_θ be the Markov process with generator $\mathbb{B}^N_\theta$ and initial state (t_0, v_0, z_0). Then the sequence of processes $\{W^N_\theta : N \in \mathbb{N}\}$ is tight in the space $D_{\hat{\mathcal{S}}}[0, \infty)$. Let W_θ be a limit point of this sequence and let $\hat{X}_\theta$ be the process with generator $\hat{\mathbb{A}}_\theta$ and initial state v_0. Then the process $\Pi_{\hat{\mathcal{S}}} W_\theta$ has the same distribution as the process $\hat{X}_\theta$.*

Remark 4.17. *Note that this proposition proves that $\Pi_{\hat{\mathcal{S}}} W^N_\theta \Rightarrow \hat{X}_\theta$ as $N \to \infty$.*

Proof. The tightness of the sequence of processes $\{W^N_\theta : N \in \mathbb{N}\}$ is argued in Lemma A.2. Let the process W_θ be a limit point of this sequence. For any function $g \in \mathcal{B}_c(\Pi_2\mathcal{S})$, define another function $f : \hat{\mathcal{S}} \to \mathbb{R}$ by

$$f(t, v, z) = g(v).$$

Then the function f is in the class $\mathcal{C}$ (see (4.38)) and the action of $\mathbb{B}^N_\theta$ (see (4.39)) on f is given by

$$\mathbb{B}^N_\theta f(t, v, z) = \sum_{k\in\Gamma_2} \rho^N_{k,\theta}(t, v, z) \left(g(v + \zeta^s_k) - g(v)\right).$$

This shows that the following is a martingale

$$\begin{aligned} m_g^N(t) &= f(W_\theta^N(t)) - \sum_{k\in\Gamma_2}\int_0^t \rho_{k,\theta}^N(W_\theta^N(s))\left(g(V_\theta^N(s)+\zeta_k^s)-g(V_\theta^N(s))\right)ds \\ &= g\left(\Pi_{\hat{\mathcal{S}}}W_\theta^N(t)\right) - \sum_{k\in\Gamma_2}\int_0^t \rho_{k,\theta}^N(W_\theta^N(s))\left(g\left(\Pi_{\hat{\mathcal{S}}}W_\theta^N(s)+\zeta_k^s\right)-g\left(\Pi_{\hat{\mathcal{S}}}W_\theta^N(s)\right)\right)ds. \end{aligned}$$

Since g is bounded, Lemma 4.15, the continuous mapping theorem and Lemma A.2 imply that as $N\to\infty$, we have $m_g^N \Rightarrow m_g$ where

$$m_g(t) = g\left(\Pi_{\hat{\mathcal{S}}}W_\theta(t)\right) - \sum_{k\in\Gamma_2}\int_0^t \hat{\lambda}_k(\Pi_{\hat{\mathcal{S}}}W_\theta(s),\theta)\left(g\left(\Pi_{\hat{\mathcal{S}}}W_\theta(s)+\zeta_k^s\right)-g\left(\Pi_{\hat{\mathcal{S}}}W_\theta(s)\right)\right)ds,$$

is also a martingale. This shows that $\{\Pi_{\hat{\mathcal{S}}}W_\theta(t): t\geq 0\}$ satisfies the martingale problem for operator $\hat{\mathbb{A}}_\theta$ (given by (3.3)). Moreover $\Pi_{\hat{\mathcal{S}}}W_\theta(0)=\hat{X}_\theta(0)=v_0$. Since the martingale problem for $\hat{\mathbb{A}}_\theta$ is well-posed, the process $\Pi_{\hat{\mathcal{S}}}W_\theta$ has the same distribution as the process $\hat{X}_\theta$ and this proves the proposition. □

4.3 Proof of Theorem 3.2

We now have all the tools to prove our main result. But first we need to define some quantities and provide some preliminary results. For any function $f:\hat{\mathcal{S}}\to\mathbb{R}$, $(t_0,v_0,z_0)\in\hat{\mathcal{S}}$ and $t\geq 0$ define

$$\Psi_{f,\theta}^N(t,t_0,v_0,z_0) = \mathbb{E}\left(f(W_\theta^N(t))\right), \tag{4.55}$$

where $\{W_\theta^N(t): t\geq 0\}$ is the process with generator $\mathbb{B}_\theta^N$ (see (4.39)) and initial state (t_0,v_0,z_0). Similarly for any function $g:\Pi_2\mathcal{S}\to\mathbb{R}$ define

$$\Psi_{g,\theta}(t,v_0) = \mathbb{E}\left(g(\hat{X}_\theta(t))\right), \tag{4.56}$$

where $\{\hat{X}_\theta(t): t\geq 0\}$ is the process with generator $\hat{\mathbb{A}}_\theta$ (see (3.3)) and initial state v_0. Now consider a function $f:\mathcal{S}\to\mathbb{R}$ which is polynomially growing with respect to projection Π_2. Corresponding to this function define $f_\theta^N:\hat{\mathcal{S}}\to\mathbb{R}$ by (4.47) and $f_\theta:\Pi_2\mathcal{S}\to\mathbb{R}$ by (3.6). Remark 4.14 and Lemma A.2 imply that for any $T>0$

$$\sup_{N\in\mathbb{N}}\sup_{t\in[0,T]}\mathbb{E}\left(|f_\theta^N(W_\theta^N(t))|\right)<\infty \quad\text{and}\quad \mathbb{E}\left(\int_0^T|\mathbb{B}_\theta^N f_\theta^N(W_\theta^N(t))|dt\right)<\infty. \tag{4.57}$$

If σ is a stopping time with respect to the filtration generated by W_θ^N, then due to part (E) of Lemma A.2 we have

$$\mathbb{E}\left(\int_0^{\sigma\wedge t}\mathbb{B}_\theta^N f(W_\theta^N(s))ds\right) = \mathbb{E}\left(\Psi_{f,\theta}^N(\sigma\wedge t,t_0,v_0,z_0)\right) - f(t_0,v_0,z_0). \tag{4.58}$$

Proposition 4.16 shows that the sequence of processes $\{W_\theta^N: N\in\mathbb{N}\}$ is tight and $\Pi_{\hat{\mathcal{S}}}W_\theta^N \Rightarrow \hat{X}_\theta$ as $N\to\infty$ (see Remark 4.17). This fact along with part (C) of Lemma 4.15 proves that for any $T>0$

$$\begin{aligned} &\lim_{N\to\infty}\sup_{t\in[\epsilon_N,T]}\left|\Psi_{f_\theta^N,\theta}^N(t,t_0,v_0,z_0)-\Psi_{f_\theta,\theta}(t,v_0)\right| = 0 \\ \text{and } &\lim_{N\to\infty}\int_0^T\left|\Psi_{f_\theta^N,\theta}^N(t,t_0,v_0,z_0)-\Psi_{f_\theta,\theta}(t,v_0)\right|dt = 0, \end{aligned} \tag{4.59}$$

where $\epsilon_N = 1/\sqrt{N}$.

Observe that the right side of (3.5) can be written as

$$\hat{S}_\theta(f_\theta, t) = \frac{\partial}{\partial\theta}\mathbb{E}\left(f_\theta(\hat{X}_\theta(t))\right) = \lim_{h\to 0}\frac{\mathbb{E}\left(f_{\theta+h}(\hat{X}_{\theta+h}(t))\right) - \mathbb{E}\left(f_\theta(\hat{X}_\theta(t))\right)}{h},$$

where $\hat{X}_\theta$ and $\hat{X}_{\theta+h}$ are processes with initial state $v_0 = \Pi_2 x_0$ and generators $\hat{\mathbb{A}}_\theta$ and $\hat{\mathbb{A}}_{\theta+h}$ respectively. This shows that we can write $\hat{S}_\theta(f_\theta, t)$ as

$$\begin{aligned}\hat{S}_\theta(f_\theta, t) &= \lim_{h\to 0}\frac{\mathbb{E}\left(f_{\theta+h}(\hat{X}_{\theta+h}(t))\right) - \mathbb{E}\left(f_\theta(\hat{X}_{\theta+h}(t))\right)}{h}\\ &\quad + \lim_{h\to 0}\frac{\mathbb{E}\left(f_\theta(\hat{X}_{\theta+h}(t))\right) - \mathbb{E}\left(f_\theta(\hat{X}_\theta(t))\right)}{h},\end{aligned} \tag{4.60}$$

provided that the two limits exist. If $\partial f_\theta/\partial\theta$ is the partial derivative of f_θ with respect to θ, then for any $v \in \Pi_2\mathcal{S}$

$$f_{\theta+h}(v) = f_\theta(v) + h\frac{\partial f_\theta}{\partial\theta}(v) + o(h).$$

This shows that the first limit in (4.60) is just

$$\lim_{h\to 0}\frac{\mathbb{E}\left(f_{\theta+h}(\hat{X}_{\theta+h}(t))\right) - \mathbb{E}\left(f_\theta(\hat{X}_{\theta+h}(t))\right)}{h} = \mathbb{E}\left(\frac{\partial f_\theta}{\partial\theta}(\hat{X}_\theta(t))\right). \tag{4.61}$$

Using coupling arguments, we proved in [17] that the second limit in (4.60) is given by

$$\begin{aligned}&\lim_{h\to 0}\frac{\mathbb{E}\left(f_\theta(\hat{X}_{\theta+h}(t))\right) - \mathbb{E}\left(f_\theta(\hat{X}_\theta(t))\right)}{h}\\ &= \sum_{k\in\Gamma_2}\mathbb{E}\left[\int_0^t \frac{\partial\hat{\lambda}_k(\hat{X}_\theta(s),\theta)}{\partial\theta}\left(f_\theta(\hat{X}_\theta(s)+\zeta_k^s) - f_\theta(\hat{X}_\theta(s))\right)ds\right]\\ &\quad + \sum_{k\in\Gamma_2}\mathbb{E}\left[\sum_{i=0,\sigma_i<t}^{\infty}\frac{\partial\hat{\lambda}_k(\hat{X}_\theta(\sigma_i),\theta)}{\partial\theta}R_{k,\theta}(\hat{X}_\theta(\sigma_i), f_\theta, t-\sigma_i\wedge t, k)\right].\end{aligned} \tag{4.62}$$

where $\zeta_k^s = \Pi_2\zeta_k$, σ_i is the i-th jump time[3] of the process $\hat{X}_\theta$ and

$$\begin{aligned}&R_{k,\theta}(x, f, t, k)\\ &= \int_0^t \left(\Psi_{f,\theta}(s, x+\zeta_k^s) - \Psi_{f,\theta}(s,x) - f(x+\zeta_k^s) + f(x)\right)\exp\left(-\hat{\lambda}_0(x,\theta)(t-s)\right)ds.\end{aligned} \tag{4.63}$$

From (4.61), (4.62), (4.60) and (3.5) we see that to prove Theorem 3.2 is suffices to show that

$$\begin{aligned}\lim_{N\to\infty}\frac{\partial}{\partial\theta}\mathbb{E}\left(f(X^N_{\gamma_2,\theta}(t))\right) &= \mathbb{E}\left(\frac{\partial f_\theta}{\partial\theta}(\hat{X}_\theta(t))\right)\\ &\quad + \sum_{k\in\Gamma_2}\mathbb{E}\left[\int_0^t \frac{\partial\hat{\lambda}_k(\hat{X}_\theta(s),\theta)}{\partial\theta}\left(f_\theta(\hat{X}_\theta(s)+\zeta_k^s) - f_\theta(\hat{X}_\theta(s))\right)ds\right]\\ &\quad + \sum_{k\in\Gamma_2}\mathbb{E}\left[\sum_{i=0,\sigma_i<t}^{\infty}\frac{\partial\hat{\lambda}_k(\hat{X}_\theta(\sigma_i),\theta)}{\partial\theta}R_{k,\theta}(\hat{X}_\theta(\sigma_i), f_\theta, t-\sigma_i\wedge t, k)\right].\end{aligned} \tag{4.64}$$

We now come to the proof of our main result, where we establish (4.64). The arguments used in the proof are motivated by the analysis in [17].

[3] We define $\sigma_0 = 0$ for convenience

Proof of Theorem 3.2. For the initial state x_0 let $v_0 = \Pi_2 x_0$ and $z_0 = (I - \Pi_2)x_0$. Let X_θ^N and $X_{\theta+h}^N$ be Markov processes with initial state x_0 and generators $\mathbb{A}_{\gamma_2,\theta}^N$ and $\mathbb{A}_{\gamma_2,\theta+h}^N$ respectively. Similarly let W_θ^N and $W_{\theta+h}^N$ be Markov processes with initial state $(0, v_0, z_0)$ and generators $\mathbb{B}_\theta^N$ and $\mathbb{B}_{\theta+h}^N$ respectively. From part (B) of Proposition 4.13 we know that

$$\mathbb{E}\left(f(X_\theta^N(t))\right) = \mathbb{E}\left(f_\theta^N(W_\theta^N(t))\right) \quad \text{and} \quad \mathbb{E}\left(f(X_{\theta+h}^N(t))\right) = \mathbb{E}\left(f_{\theta+h}^N(W_{\theta+h}^N(t))\right). \tag{4.65}$$

For any $(t, v, z) \in \hat{\mathcal{S}}$, $f_\theta^N(t, v, z)$ is a continuously differentiable function of θ. Hence we can write

$$f_{\theta+h}^N(t, v, z) = f_\theta^N(t, v, z) + h\frac{\partial f_\theta^N}{\partial \theta}(t, v, z) + o(h).$$

This expansion along with (4.65) gives us

$$\begin{aligned}
S_\theta^N(f, t) &= \frac{\partial}{\partial \theta}\mathbb{E}\left(f(X_{\gamma_2,\theta}^N(t))\right) \\
&= \lim_{h\to 0} \frac{\mathbb{E}\left(f(X_{\theta+h}^N(t))\right) - \mathbb{E}\left(f(X_\theta^N(t))\right)}{h} \\
&= \lim_{h\to 0} \frac{\mathbb{E}\left(f_{\theta+h}^N(W_{\theta+h}^N(t))\right) - \mathbb{E}\left(f_\theta^N(W_\theta^N(t))\right)}{h} \\
&= \lim_{h\to 0} \frac{\mathbb{E}\left(f_{\theta+h}^N(W_{\theta+h}^N(t))\right) - \mathbb{E}\left(f_\theta^N(W_{\theta+h}^N(t))\right)}{h} \\
&+ \lim_{h\to 0} \frac{\mathbb{E}\left(f_\theta^N(W_{\theta+h}^N(t))\right) - \mathbb{E}\left(f_\theta^N(W_\theta^N(t))\right)}{h} \\
&= S_\theta^{N,1}(f, t) + S_\theta^{N,2}(f, t),
\end{aligned} \tag{4.66}$$

where

$$S_\theta^{N,1}(f, t) = \mathbb{E}\left(\frac{\partial f_\theta^N}{\partial \theta}(W_\theta^N(t))\right) \tag{4.67}$$

$$\text{and} \quad S_\theta^{N,2}(f, t) = \lim_{h\to 0} \frac{\mathbb{E}\left(f_\theta^N(W_{\theta+h}^N(t))\right) - \mathbb{E}\left(f_\theta^N(W_\theta^N(t))\right)}{h}. \tag{4.68}$$

Proposition 4.16 shows that the sequence of processes $\{W_\theta^N : N \in \mathbb{N}\}$ is tight and if W_θ is a limit point then the process $\Pi_{\hat{\mathcal{S}}} W_\theta$ has the same distribution as the process $\hat{X}_\theta$. This fact along with part (C) of Lemma 4.15 shows that for any $t > 0$

$$\lim_{N\to\infty} S_\theta^{N,1}(f, t) = \mathbb{E}\left(\frac{\partial f_\theta}{\partial \theta}(\hat{X}_\theta(t))\right). \tag{4.69}$$

In order to compute the limit of $S_\theta^{N,2}(f, t)$ as $N \to \infty$, we will couple the processes W_θ^N and $W_{\theta+h}^N$ in a special way. We need to define certain quantities to describe the coupling. For any $(t_1, v_1, z_1), (t_2, v_2, z_2) \in \hat{\mathcal{S}}$ let

$$\begin{aligned}
\rho_{k,\theta,\min}^N(t_1, v_1, z_1, t_2, v_2, z_2, h) &= \rho_{k,\theta}^N(t_1, v_1, z_1) \wedge \rho_{k,\theta+h}^N(t_2, v_2, z_2), \\
r_{k,\theta}^{N,1}(t_1, v_1, z_1, t_2, v_2, z_2, h) &= \rho_{k,\theta}^N(t_1, v_1, z_1) - \rho_{k,\theta,\min}^N(t_1, v_1, z_1, t_2, v_2, z_2, h) \\
\text{and} \quad r_{k,\theta}^{N,2}(t_1, v_1, z_1, t_2, v_2, z_2, h) &= \rho_{k,\theta+h}^N(t_2, v_2, z_2) - \rho_{k,\theta,\min}^N(t_1, v_1, z_1, t_2, v_2, z_2, h).
\end{aligned}$$

We define the processes V_θ^N and $V_{\theta+h}^N$ by the following random time change representations

$$\begin{aligned} V_\theta^N(t) &= v_0 \qquad (4.70)\\ &+ \sum_{k\in\Gamma_2} Y_k\left(\int_0^t \rho^N_{k,\theta,\min}\left(\tau^N_\theta(s), V^N_\theta(s), Z^N_\theta(s), \tau^N_{\theta+h}(s), V^N_{\theta+h}(s), Z^N_{\theta+h}(s), h\right) ds\right)\zeta^s_k\\ &+ \sum_{k\in\Gamma_2} Y^{(1)}_k\left(\int_0^t r^{N,1}_{k,\theta}\left(\tau^N_\theta(s), V^N_\theta(s), Z^N_\theta(s), \tau^N_{\theta+h}(s), V^N_{\theta+h}(s), Z^N_{\theta+h}(s), h\right) ds\right)\zeta^s_k\\ V_{\theta+h}^N(t) &= v_0 \qquad (4.71)\\ &+ \sum_{k\in\Gamma_2} Y_k\left(\int_0^t \rho^N_{k,\theta,\min}\left(\tau^N_\theta(s), V^N_\theta(s), Z^N_\theta(s), \tau^N_{\theta+h}(s), V^N_{\theta+h}(s), Z^N_{\theta+h}(s), h\right) ds\right)\zeta^s_k\\ &+ \sum_{k\in\Gamma_2} Y^{(2)}_k\left(\int_0^t r^{N,2}_{k,\theta}\left(\tau^N_\theta(s), V^N_\theta(s), Z^N_\theta(s), \tau^N_{\theta+h}(s), V^N_{\theta+h}(s), Z^N_{\theta+h}(s), h\right) ds\right)\zeta^s_k, \end{aligned}$$

where $\{Y_k, Y^{(1)}_k, Y^{(2)}_k : k \in \Gamma_2\}$ is a family of independent unit rate Poisson processes. To V^N_θ ($V^N_{\theta+h}$) we associate processes τ^N_θ ($\tau^N_{\theta+h}$) and Z^N_θ ($Z^N_{\theta+h}$) as in Remark 4.12. The above representations couple the processes V^N_θ and $V^N_{\theta+h}$. For each $i \in \mathbb{N}$, let σ^1_i (σ^2_i) be the i-th jump time of the process V^N_θ ($V^N_{\theta+h}$) and let η^1_i (η^2_i) be the jump direction of the process V^N_θ ($V^N_{\theta+h}$) at time σ^1_i (σ^2_i). Define $\sigma^1_0 = \sigma^2_0 = 0$. Fix a sequence $\{u_i : i \in \mathbb{N}\}$ of independent Unif$(0,1)$ random numbers. We couple the processes Z^N_θ and $Z^N_{\theta+h}$, by letting $Z^N_\theta(\sigma^1_i) = \digamma^N_{\eta^1_i,\theta}(\sigma^1_i - \sigma^1_{i-1}, V^N_\theta(\sigma^1_{i-1}), Z^N_\theta(\sigma^1_{i-1}), u_i)$ and $Z^N_{\theta+h}(\sigma^2_i) = \digamma^N_{\eta^2_i,\theta+h}(\sigma^2_i - \sigma^2_{i-1}, V^N_{\theta+h}(\sigma^2_{i-1}), Z^N_{\theta+h}(\sigma^2_{i-1}), u_i)$ for each i, where the function $\digamma^N$ is defined by (4.35). Note that we are using the same u_i in the definition of $Z^N_\theta(\sigma^1_i)$ and $Z^N_{\theta+h}(\sigma^2_i)$. Define W^N_θ and $W^N_{\theta+h}$ by

$$W^N_\theta(t) = \left(\tau^N_\theta(t), V^N_\theta(t), Z^N_\theta(t)\right) \text{ and } W^N_{\theta+h}(t) = \left(\tau^N_{\theta+h}(t), V^N_{\theta+h}(t), Z^N_{\theta+h}(t)\right) \text{ for all } t \geq 0.$$

One can verify that the processes W^N_θ and $W^N_{\theta+h}$ have initial state $(0, v_0, z_0)$ and generators $\mathbb{B}^N_\theta$ and $\mathbb{B}^N_{\theta+h}$ respectively.

Let γ^N_h be the stopping time given by

$$\gamma^N_h = \inf\{t \geq 0 : W^N_\theta(t) \neq W^N_{\theta+h}(t)\}. \qquad (4.72)$$

Then the coupling of processes W^N_θ and $W^N_{\theta+h}$ ensures that $\gamma^N_h \to \infty$ a.s. as $h \to 0$. Define

$$A^N_\theta = \lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[\int_0^{t\wedge\gamma^N_h}\left(\mathbb{B}^N_{\theta+h} f^N_\theta(W^N_{\theta+h}(s)) - \mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right) ds\right] \qquad (4.73)$$

$$\text{and} \quad B^N_\theta = \lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[\int_{t\wedge\gamma^N_h}^{t}\left(\mathbb{B}^N_{\theta+h} f^N_\theta(W^N_{\theta+h}(s)) - \mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right) ds\right]. \qquad (4.74)$$

Note that $f^N_\theta(0, v_0, z_0) = f(x_0)$. Using (4.58) we can write

$$\mathbb{E}\left(f^N_\theta(W^N_\theta(t))\right) = f(x_0) + \mathbb{E}\left(\int_0^t \mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s)) ds\right)$$

$$\text{and} \quad \mathbb{E}\left(f^N_\theta(W^N_{\theta+h}(t))\right) = f(x_0) + \mathbb{E}\left(\int_0^t \mathbb{B}^N_{\theta+h} f^N_\theta(W^N_{\theta+h}(s)) ds\right).$$

Therefore

$$
\begin{aligned}
S^{N,2}_\theta(f,t) &= \lim_{h\to 0}\frac{\mathbb{E}\left(f^N_\theta(W^N_{\theta+h}(t))\right)-\mathbb{E}\left(f^N_\theta(W^N_\theta(t))\right)}{h}\\
&= \lim_{h\to 0}\frac{1}{h}\left[\mathbb{E}\left(\int_0^t\left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(s))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right)ds\right)\right]\\
&= A^N_\theta + B^N_\theta. \qquad (4.75)
\end{aligned}
$$

Using Taylor's expansion, for any $f\in\mathcal{C}$ and $(t,v,z)\in\hat{\mathcal{S}}$ we get

$$
\begin{aligned}
&\mathbb{B}^N_{\theta+h}f(t,v,z)-\mathbb{B}^N_\theta f(t,v,z)\\
&= \sum_{k\in\Gamma_2}\rho^N_{k,\theta+h}(t,v,z)\sum_{e\in\mathbb{H}_v}\left(f(0,v+\zeta^s_k,e+\zeta^f_k)-f(t,v,z)\right)\Theta^N_{k,\theta+h}(t,v,z,e)\\
&- \sum_{k\in\Gamma_2}\rho^N_{k,\theta}(t,v,z)\sum_{e\in\mathbb{H}_v}\left(f(0,v+\zeta^s_k,e+\zeta^f_k)-f(t,v,z)\right)\Theta^N_{k,\theta}(t,v,z,e)\\
&= \sum_{k\in\Gamma_2}\sum_{e\in\mathbb{H}_v}f(0,v+\zeta^s_k,e+\zeta^f_k)\left(\rho^N_{k,\theta+h}(t,v,z)\Theta^N_{k,\theta+h}(t,v,z,e)-\rho^N_{k,\theta}(t,v,z)\Theta^N_{k,\theta}(t,v,z,e)\right)\\
&- \sum_{k\in\Gamma_2}f(t,v,z)\left(\rho^N_{k,\theta+h}(t,v,z)-\rho^N_{k,\theta}(t,v,z)\right)\\
&= \sum_{k\in\Gamma_2}\sum_{e\in\mathbb{H}_v}f(0,v+\zeta^s_k,e+\zeta^f_k)\left(\frac{\partial\rho^N_{k,\theta}(t,v,z)}{\partial\theta}\Theta^N_{k,\theta}(t,v,z,e)+\rho^N_{k,\theta}(t,v,z)\frac{\partial\Theta^N_{k,\theta}(t,v,z,e)}{\partial\theta}\right)h\\
&- \sum_{k\in\Gamma_2}f(t,v,z)\frac{\partial\rho^N_{k,\theta}(t,v,z)}{\partial\theta}h+o(h)\\
&= \sum_{k\in\Gamma_2}\frac{\partial\rho^N_{k,\theta}(t,v,z)}{\partial\theta}\left(\sum_{e\in\mathbb{H}_v}f(0,v+\zeta^s_k,e+\zeta^f_k)\Theta^N_{k,\theta}(t,v,z,e)-f(t,v,z)\right)h\\
&+ \sum_{k\in\Gamma_2}\rho^N_{k,\theta}(t,v,z)\sum_{e\in\mathbb{H}_v}f(0,v+\zeta^s_k,e+\zeta^f_k)\frac{\partial\Theta^N_{k,\theta}(t,v,z,e)}{\partial\theta}h+o(h). \qquad (4.76)
\end{aligned}
$$

Note that for any $t\in[0,\gamma^N_h)$ we have $W^N_{\theta+h}(t)=W^N_\theta(t)$. Relation (4.76) implies that

$$
\begin{aligned}
&\lim_{N\to\infty}A^N_\theta\\
&= \lim_{N\to\infty}\lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[\int_0^{t\wedge\gamma^N_h}\left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(s))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right)ds\right]\\
&= \lim_{N\to\infty}\sum_{k\in\Gamma_2}\mathbb{E}\left[\int_0^t\frac{\partial\rho^N_{k,\theta}(W^N_\theta(s))}{\partial\theta}\left(\sum_{e\in\mathbb{H}_v}f^N_\theta(0,\Pi_{\hat{\mathcal{S}}}W^N_\theta(s)+\zeta^s_k,e+\zeta^f_k)\Theta^N_{k,\theta}(W^N_\theta(s),e)\right.\right.\\
&\qquad\qquad\left.\left.-f^N_\theta(W^N_\theta(s))\right)ds\right]\\
&+ \lim_{N\to\infty}\sum_{k\in\Gamma_2}\mathbb{E}\left[\int_0^t\rho^N_{k,\theta}(W^N_\theta(s))\sum_{e\in\mathbb{H}_v}f^N_\theta(0,\Pi_{\hat{\mathcal{S}}}W^N_\theta(s)+\zeta^s_k,e+\zeta^f_k)\frac{\partial\Theta^N_{k,\theta}(W^N_\theta(s),e)}{\partial\theta}ds\right].
\end{aligned}
$$

Proposition 4.16 shows that the sequence of processes $\{W^N_\theta : N\in\mathbb{N}\}$ is tight and if W_θ is a limit point then the process $\Pi_{\hat{\mathcal{S}}}W_\theta$ has the same distribution as the process $\hat{X}_\theta$. This fact along with Lemma 4.15 implies that

$$
\lim_{N\to\infty}A^N_\theta=\sum_{k\in\Gamma_2}\mathbb{E}\left[\int_0^t\frac{\partial\hat{\lambda}_k(\hat{X}_\theta(s),\theta)}{\partial\theta}\left(f_\theta(\hat{X}_\theta(s)+\zeta^s_k)-f_\theta(\hat{X}_\theta(s))\right)ds\right]. \qquad (4.77)
$$

Our next goal is to compute $\lim_{N\to\infty} B_\theta^N$. Recall the definitions of $\Psi^N_{f,\theta}$ and $\Psi_{f,\theta}$ from (4.55) and (4.56) respectively. For $i=1,2$, let $(t_i,v_i,z_i)\in\hat{\mathcal{S}}$. Define an event

$$\begin{aligned}&E^N(t_1,v_1,z_1,t_2,v_2,z_2,s) \\ &=\{W^N_\theta(\gamma^N_h)=(t_1,v_1,z_1),W^N_{\theta+h}(\gamma^N_h)=(t_2,v_2,z_2)\text{ and }\gamma^N_h=s\}\end{aligned}\tag{4.78}$$

and let

$$\begin{aligned}&R^N_{\theta,h}(t_1,v_1,z_1,t_2,v_2,z_2,s,t) \\ &=\mathbb{E}\left[\int_{t\wedge\gamma^N_h}^t \left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(u))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(u))\right)du\middle|E^N(t_1,v_1,z_1,t_2,v_2,z_2,s)\right].\end{aligned}\tag{4.79}$$

Let $\epsilon_N=1/\sqrt{N}$. From (4.58) and the strong Markov property, we can deduce that for any $0<s<t$

$$\begin{aligned}&\lim_{N\to\infty}\lim_{h\to0}R^N_{\theta,h}(t_1,v_1,z_1,t_2,v_2,z_2,s,t)\\
&=\lim_{N\to\infty}\lim_{h\to0}\mathbb{E}\left[\int_{t\wedge\gamma^N_h}^t \left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(u))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(u))\right)du\middle|E^N(t_1,v_1,z_1,t_2,v_2,z_2,s)\right]\\
&=\lim_{N\to\infty}\lim_{h\to0}\mathbb{E}\left[\int_{s+\epsilon_N}^t \left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(u))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(u))\right)du\middle|E^N(t_1,v_1,z_1,t_2,v_2,z_2,s)\right]\\
&=\lim_{N\to\infty}\lim_{h\to0}\Big[\Psi^N_{f^N_\theta,\theta+h}(t-s,t_2,v_2,z_2)-\Psi^N_{f^N_\theta,\theta+h}(\epsilon_N,t_2,v_2,z_2)\\
&\qquad\qquad -\Psi^N_{f^N_\theta,\theta}(t-s,t_1,v_1,z_1)+\Psi^N_{f^N_\theta,\theta}(\epsilon_N,t_1,v_1,z_1)\Big]\\
&=\Psi_{f_\theta,\theta}(t-s,v_2)-\Psi_{f_\theta,\theta}(t-s,v_1)-f_\theta(v_2)+f_\theta(v_1),\end{aligned}\tag{4.80}$$

where the last equality holds due to (4.59).

Recall the random time change representations (4.70) and (4.71). For each $i\in\mathbb{N}$, let σ^N_i be the i-th jump time of the process C^N_θ defined by

$$C^N_\theta(t)=\sum_{k\in\Gamma_2}Y_k\left(\int_0^t\rho^N_{k,\theta,\min}\left(\tau^N_\theta(s),V^N_\theta(s),Z^N_\theta(s),\tau^N_{\theta+h}(s),V^N_{\theta+h}(s),Z^N_{\theta+h}(s),h\right)ds\right)\zeta^s_k.$$

Set $\sigma^N_0=0$ and note that $\gamma^N_h>\sigma^N_0$. For each $i\in\mathbb{N}$ define

$$\begin{aligned}&B^{N,1}_{\theta,i}=\lim_{h\to0}\frac{1}{h}\mathbb{E}\left[\mathbb{1}_{\{\sigma^N_i=\gamma^N_h\}}\int_{t\wedge\gamma^N_h}^t\left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(s))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right)ds\right]\\
\text{and}\quad &B^{N,2}_{\theta,i}=\lim_{h\to0}\frac{1}{h}\mathbb{E}\left[\mathbb{1}_{\{\sigma^N_{i-1}<\gamma^N_h<\sigma^N_i\}}\int_{t\wedge\gamma^N_h}^t\left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(s))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right)ds\right].\end{aligned}$$

Since $\mathbb{1}_{\{\sigma^N_{i-1}\le\gamma^N_h<\sigma^N_i\}}=\mathbb{1}_{\{\sigma^N_{i-1}=\gamma^N_h\}}+\mathbb{1}_{\{\sigma^N_{i-1}<\gamma^N_h<\sigma^N_i\}}$ we can write

$$\begin{aligned}B^N_\theta&=\lim_{h\to0}\frac{1}{h}\mathbb{E}\left[\int_{t\wedge\gamma^N_h}^t\left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(s))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right)ds\right]\\
&=\sum_{i=1}^\infty\lim_{h\to0}\frac{1}{h}\mathbb{E}\left[\mathbb{1}_{\{\sigma^N_{i-1}\le\gamma^N_h<\sigma^N_i\}}\int_{t\wedge\gamma^N_h}^t\left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(s))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right)ds\right]\\
&=\sum_{i=1}^\infty(B^{N,1}_{\theta,i}+B^{N,2}_{\theta,i}).\end{aligned}\tag{4.81}$$

We now show that the term $B^{N,1}_{\theta,i}$ converges to 0 as $N \to \infty$. Note that the event $\{\sigma^N_i = \gamma^N_h\}$ occurs if and only if the event $\{Z^N_\theta(\sigma^N_i-) \neq Z^N_{\theta+h}(\sigma^N_i-), V^N_\theta(\sigma^N_{i-1}) = V^N_{\theta+h}(\sigma^N_{i-1}), Z^N_\theta(\sigma^N_{i-1}) = Z^N_{\theta+h}(\sigma^N_{i-1})\}$ occurs. Let η^N_i be the Γ_2-valued random variable which gives the direction of the jump in C^N_θ at time σ^N_i. Pick a $\delta \geq 0$, $v \in \Pi_2\mathcal{S}$, $z \in \mathbb{H}_v$ and $k \in \Gamma_2$. Define an event

$$L_i(\delta, v, z, k) = \big\{\gamma^N_h \geq \sigma^N_i, (\sigma^N_i - \sigma^N_{i-1}) = \delta, \eta^N_i = k, V^N_\theta(\sigma^N_{i-1}) = V^N_{\theta+h}(\sigma^N_{i-1}) = v,$$
$$Z^N_\theta(\sigma^N_{i-1}) = Z^N_{\theta+h}(\sigma^N_{i-1}) = z\big\}\,.$$

Conditioned on this event, $Z^N_\theta(\sigma^N_i-) = \digamma^N_{k,\theta}(t, v, z, u_i)$ and $Z^N_{\theta+h}(\sigma^N_i-) = \digamma^N_{k,\theta+h}(t, v, z, u_i)$ where the function $\digamma^N_{k,\theta}$ is given by (4.35). For any distinct $z_1, z_2 \in \mathbb{H}_v$ define

$$G^N_\theta(z_1, z_2, \delta, v, z, k) = \lim_{h\to 0} \frac{\mathbb{P}\left(\sigma^N_i = \gamma^N_h, Z^N_\theta(\sigma^N_i-) = z_1 \text{ and } Z^N_{\theta+h}(\sigma^N_i-) = z_2 | L_i(\delta, v, z, k)\right)}{h}.$$

Lemma 4.11 ensures that $G^N_\theta(z_1, z_2, \delta, v, z, k)$ exists and

$$G^N_\theta(z_1, z_2, \delta, v, z, k) \leq \sum_{e\in\mathbb{H}_v} \left| \frac{\partial \Theta^N_{k,\theta}(\delta, v, z, e)}{\partial\theta}\right|.$$

Assumptions 3.1 imply that the right hand side is a polynomially growing function with respect to projection $\Pi_{\hat{\mathcal{S}}}$ (see Definition 2.1). Given the events $L_i(\delta, v, z, k)$ and $\{Z^N_\theta(\sigma^N_i-) = z_1, Z^N_{\theta+h}(\sigma^N_i-) = z_2\}$ we have

$$\begin{aligned}&\left(\tau^N_\theta(\gamma^N_h), V^N_\theta(\gamma^N_h), Z^N_\theta(\gamma^N_h), \tau^N_{\theta+h}(\gamma^N_h), V^N_{\theta+h}(\gamma^N_h), Z^N_{\theta+h}(\gamma^N_h)\right)\\ &= (0, v+\zeta^s_k, z_1+\zeta^f_k, 0, v+\zeta^s_k, z_2+\zeta^f_k).\end{aligned}$$

Recall the definition of $R^N_{\theta,h}$ from (4.79). For any $\delta < s < t$ we can write

$$\begin{aligned}&\lim_{N\to\infty}\lim_{h\to 0}\\ &\frac{1}{h}\mathbb{E}\left[\mathbb{1}_{\{\sigma^N_i=\gamma^N_h\}}\int_{t\wedge\gamma^N_h}^t \left(\mathbb{B}^N_{\theta+h} f^N_\theta(W^N_{\theta+h}(s)) - \mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right) ds \middle| L_i(\delta, v, z, k), \sigma^N_{i-1} = s-\delta\right]\\ &= \lim_{N\to\infty}\lim_{h\to 0}\sum_{z_1\neq z_2\in\mathbb{H}_v} G^N_\theta(z_1, z_2, \delta, v, z, k) R^N_{\theta,h}(0, v+\zeta^s_k, z_1+\zeta^f_k, 0, v+\zeta^s_k, z_2+\zeta^f_k, s, t).\end{aligned} \tag{4.82}$$

Using (4.80) we see that

$$\lim_{N\to\infty}\lim_{h\to 0} R^N_{\theta,h}(0, v+\zeta^s_k, z_1+\zeta^f_k, 0, v+\zeta^s_k, z_2+\zeta^f_k, s, t) = 0. \tag{4.83}$$

This relation along with (4.82) implies that

$$\lim_{N\to\infty} B^{N,1}_{\theta,i} = 0. \tag{4.84}$$

Recall the random time change representations (4.70) and (4.71). On the event $\{\sigma^N_{i-1} < \gamma^N_h < \sigma^N_i\}$, the process V^N_θ (or $V^N_{\theta+h}$) jumps at time γ^N_h due to a jump in the Poisson process $Y^{(1)}_k$ (or $Y^{(2)}_k$) for some $k \in \Gamma_2$. Let η be the Γ_2-valued random variable which gives the direction of the jump in V^N_θ or $V^N_{\theta+h}$ at time γ^N_h. Define a random variable

$$\alpha^N_i = (\sigma^N_i - \sigma^N_{i-1}) \wedge (\gamma^N_h - \sigma^N_{i-1})$$

and an event

$$H_i(s,v,z) = \left\{\sigma^N_{i-1} = s, V^N_\theta(\sigma^N_{i-1}) = V^N_{\theta+h}(\sigma^N_{i-1}) = v, Z^N_\theta(\sigma^N_{i-1}) = Z^N_{\theta+h}(\sigma^N_{i-1}) = z\right\},$$

for $s \geq 0$, $v \in \Pi_2\mathcal{S}$ and $z \in \mathbb{H}_v$. The event $\{\sigma^N_{i-1} < \gamma^N_h < \sigma^N_i\}$ is equivalent to the event $\{\gamma^N_h > \sigma^N_{i-1}, \alpha^N_i = (\gamma^N_h - \sigma^N_{i-1})\}$. Given $\gamma^N_h > \sigma^N_{i-1}$ and $H_i(s,v,z)$, the density of the $\mathbb{R}_+$-valued random variable α^N_i on the event $\{\eta = k, \alpha^N_i = (\gamma^N_h - \sigma^N_{i-1})\}$ is given by

$$\begin{aligned}
&\lim_{\epsilon\to 0} \frac{\mathbb{P}\left(\alpha^N_i \in (t,t+\epsilon), \eta = k, \alpha^N_i = (\gamma^N_h - \sigma^N_{i-1}) \middle| H_i(s,v,z), \gamma^N_h > \sigma^N_{i-1}\right)}{\epsilon} \\
&= \left(\rho^N_{k,\theta}(t,v,z) + \rho^N_{k,\theta+h}(t,v,z) - 2\rho^N_{k,\theta}(t,v,z) \wedge \rho^N_{k,\theta+h}(t,v,z)\right) \\
&\times \exp\left[-\int_0^t \left(\rho^N_{0,\theta}(u,v,z) + \rho^N_{0,\theta+h}(u,v,z) - 2\rho^N_{0,\theta}(u,v,z) \wedge \rho^N_{0,\theta+h}(u,v,z)\right) du\right] \\
&= h\left|\frac{\partial \rho^N_{k,\theta}(t,v,z)}{\partial\theta}\right| \exp\left(-\int_0^t \rho^N_{0,\theta}(u,v,z)du\right) + o(h). \qquad (4.85)
\end{aligned}$$

On the event $H_i(s,v,z) \cap \{\gamma^N_h > \sigma^N_{i-1}, \eta = k, \alpha^N_i = (\gamma^N_h - \sigma^N_{i-1}) = \delta\}$ we have

$$\left(W^N_\theta(\gamma^N_h), W^N_{\theta+h}(\gamma^N_h)\right) = \begin{cases} (\delta, v, z, 0, v+\zeta^s_k, \xi_2 + \zeta^f_k) & \rho^N_{k,\theta+h}(\delta,v,z) > \rho^N_{k,\theta}(\delta,v,z) \\ (0, v+\zeta^s_k, \xi_1 + \zeta^f_k, \delta, v, z) & \rho^N_{k,\theta+h}(\delta,v,z) < \rho^N_{k,\theta}(\delta,v,z), \end{cases}$$

where $\xi_1 = F^N_{k,\theta}(\delta,v,z,u_i)$ and $\xi_2 = F^N_{k,\theta+h}(\delta,v,z,u_i)$ are $\mathbb{H}_v$-valued random variables with distributions $\Theta^N_{k,\theta}(\delta,v,z,\cdot)$ and $\Theta^N_{k,\theta+h}(\delta,v,z,\cdot)$ respectively. For small values of h, $\partial\rho^N_{k,\theta}(\delta,v,z)/\partial\theta > 0$ implies that $\rho^N_{k,\theta+h}(\delta,v,z) > \rho^N_{k,\theta}(\delta,v,z)$ and similarly $\partial\rho^N_{k,\theta}(\delta,v,z)/\partial\theta < 0$ implies that $\rho^N_{k,\theta+h}(\delta,v,z) < \rho^N_{k,\theta}(\delta,v,z)$. Using the density of α^N_i on the event $\{\eta = k, \alpha^N_i = (\gamma^N_h - \sigma^N_{i-1})\}$ (see (4.85)) we obtain

$$\begin{aligned}
&\lim_{N\to\infty}\lim_{h\to 0} \\
&\frac{1}{h}\mathbb{E}\left[\mathbb{1}_{\{\sigma^N_{i-1}<\gamma^N_h<\sigma^N_i\}} \int_{t\wedge\gamma^N_h}^t \left(\mathbb{B}^N_{\theta+h} f^N_\theta(W^N_{\theta+h}(u)) - \mathbb{B}^N_\theta f^N_\theta(W^N_\theta(u))\right) du \middle| H_i(s,v,z), \gamma^N_h > \sigma^N_{i-1}\right] \\
&= \lim_{N\to\infty}\lim_{h\to 0} \sum_{z_2\in\mathbb{H}_v}\sum_{k\in\Gamma_2} \int_0^{t-s} \left[\frac{\partial\rho^N_{k,\theta}(\delta,v,z)}{\partial\theta}\right]^+ \exp\left(-\int_0^\delta \rho^N_{0,\theta}(u,v,z)du\right) \\
&\qquad\qquad \times R^N_{\theta,h}(\delta,v,z,0,v+\zeta^s_k, z_2+\zeta^f_k, s+\delta, t)\Theta^N_{k,\theta+h}(\delta,v,z,z_2)d\delta \\
&+ \lim_{N\to\infty}\lim_{h\to 0} \sum_{z_1\in\mathbb{H}_v}\sum_{k\in\Gamma_2} \int_0^{t-s} \left[\frac{\partial\rho^N_{k,\theta}(\delta,v,z)}{\partial\theta}\right]^- \exp\left(-\int_0^\delta \rho^N_{0,\theta}(u,v,z)du\right) \\
&\qquad\qquad \times R^N_{\theta,h}(0,v+\zeta^s_k, z_1+\zeta^f_k, \delta, v, z, s+\delta, t)\Theta^N_{k,\theta+h}(\delta,v,z,z_1)d\delta. \qquad (4.86)
\end{aligned}$$

From (4.80) one can verify that

$$\begin{aligned}
&\lim_{N\to\infty}\lim_{h\to 0} R^N_{\theta,h}(\delta,v,z,0,v+\zeta^s_k, z_2+\zeta^f_k, s+\delta, t) \\
&= -\lim_{N\to\infty}\lim_{h\to 0} R^N_{\theta,h}(0,v+\zeta^s_k, z_1+\zeta^f_k, \delta, v, z, s+\delta, t) \\
&= \Psi_{f_\theta,\theta}(t-s-\delta, v+\zeta^s_k) - \Psi_{f_\theta,\theta}(t-s-\delta, v) - f_\theta(v+\zeta^s_k) + f_\theta(v). \qquad (4.87)
\end{aligned}$$

Using part (A) of Lemma 4.15, (4.87) and (4.86) we can conclude that

$$\begin{aligned}
&\lim_{N\to\infty}\lim_{h\to 0}\frac{1}{h}\\
&\mathbb{E}\left[\mathbb{1}_{\{\sigma^N_{i-1}<\gamma^N_h<\sigma^N_i\}}\int_{t\wedge\gamma^N_h}^{t}\left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(u))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(u))\right)du\middle| H_i(s,v,z),\gamma^N_h>\sigma^N_{i-1}\right]\\
&=\sum_{k\in\Gamma_2}\int_0^{t-s}\frac{\partial\hat{\lambda}_k(v,\theta)}{\partial\theta}\exp\left(-\hat{\lambda}_0(v,\theta)\delta\right)(\Psi_{f_\theta,\theta}(t-s-\delta,v+\zeta^s_k)-\Psi_{f_\theta,\theta}(t-s-\delta,v)\\
&\qquad\qquad -f_\theta(v+\zeta^s_k)+f_\theta(v))\,d\delta\\
&=\sum_{k\in\Gamma_2}\int_0^{t-s}\frac{\partial\hat{\lambda}_k(v,\theta)}{\partial\theta}\exp\left(-\hat{\lambda}_0(v,\theta)(t-s-u)\right)(\Psi_{f_\theta,\theta}(u,v+\zeta^s_k)-\Psi_{f_\theta,\theta}(u,v)\\
&\qquad\qquad -f_\theta(v+\zeta^s_k)+f_\theta(v))\,du, \qquad (4.88)
\end{aligned}$$

where $\hat{\lambda}_0(v,\theta)=\sum_{k\in\Gamma_2}\hat{\lambda}_k(v,\theta)$. Due to our coupling, as $h\to 0$, the process $W^N_{\theta+h}$ converges a.s. to the process W^N_θ and hence $\gamma^N_h\to\infty$ a.s. Proposition 4.16 and Remark 4.17 show that as $N\to\infty$ we have $V^N_\theta\Rightarrow\hat{X}_\theta$, where $\hat{X}_\theta$ is the limiting process in Theorem 3.2. This convergence and (4.88) yield the following

$$\begin{aligned}
&\lim_{N\to\infty}B^{N,2}_{\theta,i}\\
&=\lim_{N\to\infty}\lim_{h\to 0}\frac{1}{h}\mathbb{E}\left[\mathbb{1}_{\{\sigma^N_{i-1}<\gamma^N_h<\sigma^N_i\}}\int_{t\wedge\gamma^N_h}^{t}\left(\mathbb{B}^N_{\theta+h}f^N_\theta(W^N_{\theta+h}(s))-\mathbb{B}^N_\theta f^N_\theta(W^N_\theta(s))\right)ds\right]\\
&=\sum_{k\in\Gamma_2}\mathbb{E}\left[\frac{\partial\hat{\lambda}_k(\hat{X}_\theta(\sigma_{i-1}),\theta)}{\partial\theta}R_{k,\theta}(\hat{X}_\theta(\sigma_{i-1}),f_\theta,t-\sigma_{i-1}\wedge t,k)\right],
\end{aligned}$$

where σ_i is the i-th jump time of the process $\hat{X}_\theta$ (with $\sigma_0=0$) and the function $R_{k,\theta}$ is given by (4.63). Note that the quantity on the right hand side is 0 if $\sigma_{i-1}\geq t$. Using (4.81) and (4.84) we get

$$\lim_{N\to\infty}B^N_\theta=\sum_{k\in\Gamma_2}\mathbb{E}\left[\sum_{i=0,\sigma_i<t}^{\infty}\frac{\partial\hat{\lambda}_k(\hat{X}_\theta(\sigma_i),\theta)}{\partial\theta}R_{k,\theta}(\hat{X}_\theta(\sigma_i),f_\theta,t-\sigma_i\wedge t,k)\right].$$

This relation along with (4.66), (4.69), (4.75) and (4.77) gives us

$$\begin{aligned}
&\lim_{N\to\infty}S^N_\theta(f,t)\\
&=\mathbb{E}\left(\frac{\partial f_\theta}{\partial\theta}(\hat{X}_\theta(t))\right)+\sum_{k\in\Gamma_2}\mathbb{E}\left[\int_0^t\frac{\partial\hat{\lambda}_k(\hat{X}_\theta(s),\theta)}{\partial\theta}\left(f_\theta(\hat{X}_\theta(s)+\zeta^s_k)-f_\theta(\hat{X}_\theta(s))\right)ds\right]\\
&+\sum_{k\in\Gamma_2}\mathbb{E}\left[\sum_{i=0,\sigma_i<t}^{\infty}\frac{\partial\hat{\lambda}_k(\hat{X}_\theta(\sigma_i),\theta)}{\partial\theta}R_{k,\theta}(\hat{X}_\theta(\sigma_i),f_\theta,t-\sigma_i\wedge t,k)\right],
\end{aligned}$$

which is same as (4.64) and this completes the proof of the theorem. □

We end this section with a couple of important remarks regarding the proof of Theorem 3.2.

Remark 4.18. *In the proof of Theorem 3.2 we had assumed that the set* $(\Gamma_1\cup\Gamma_2)^c$ *is empty (recall Remark 4.8). We now explain how the proof changes when this is not the case. Note that if* k *is in the set* $(\Gamma_1\cup\Gamma_2)^c$*, then* $\beta_k+\gamma_2<0$*, which shows that reaction* k

is "slow" for the time-scale γ_2*. For any* $k \in (\Gamma_1 \cup \Gamma_2)^c$*, let* $\hat{\lambda}_k(\cdot,\theta)$ *be identically* 0 *and let* $\rho^N_{k,\theta}$ *be defined by* (4.30) *with an additional multiplicative factor of* $N^{\beta_k+\gamma_2}$*. By replacing* Γ_2 *with* $\Gamma_2' := \Gamma_2 \cup (\Gamma_1 \cup \Gamma_2)^c$*, Theorem 3.2 can be proved in the same way as above.*

Remark 4.19. *In proving Theorem 3.2, we assumed that the set* $\mathbb{H}_v$ *is finite for any* $v \in \Pi_2\mathcal{S}$ *(see part (A) of Assumption 2.4). This means that if the state of the "natural" dynamics is* v *then the "fast" dynamics is constrained within a compact set* $\mathbb{H}_v$*. This assumption can be relaxed at the expense of making the proof more technical. The only place where finiteness of* $\mathbb{H}_v$ *is crucial is in the proof of Proposition 4.4. As explained in Remark 4.7, this proposition can be extended for Markov chains with countable state spaces. Assuming the existence of a suitable Lyapunov function for the fast dynamics, the proof of Theorem 3.2 goes through with minor modifications.*

5 An Illustrative Example

In this section we present a simple example to illustrate how our main result, Theorem 3.2, can be useful for the estimation of parameter sensitivity for multiscale networks. Consider a chemical reaction network with three species S_1, S_2 and S_3, and three reactions given by

$$S_1 \xrightarrow{c_1} S_2, \quad S_2 \xrightarrow{c_2} S_1 \quad \text{and} \quad S_2 \xrightarrow{c_3} S_3.$$

The rate constant of the i-th reaction is c_i, for $i = 1,2,3$. Such a network is used to model the cellular heat-shock response in [7], where S_1, S_2 and S_3 correspond to the $\sigma_{32}-$DnaK complex, the σ_{32} heat shock regulator and the σ_{32}-RNAP complex, respectively. In this example, the first and second reactions are much faster than the third reaction. We assume that the rate constants are given by

$$c_1 = 1, \quad c_2 = 2 \quad \text{and} \quad c_3 = 5 \times 10^{-4}. \tag{5.1}$$

We choose our sensitive parameter to be $\theta = c_1 = 1$ and the large *normalization* parameter to be $N_0 = 10^4$. The three reactions along with their *scaling* factors (β_k's), *propensity* functions (λ_k's) and their *stoichiometric* vectors (ζ_k's) are presented in Table 1.

Table 1: Example of Heat Shock Response Model

No.	Reaction	Scaling Factor	Propensity Function	Stoichiometric Vector
1	$S_1 \longrightarrow S_2$	$\beta_1 = 0$	$\lambda_1(x_1,x_2,x_3) = \theta x_1$	$\zeta_1 = (-1,1,0)$
2	$S_2 \longrightarrow S_1$	$\beta_2 = 0$	$\lambda_2(x_1,x_2,x_3) = 2x_2$	$\zeta_2 = (1,-1,0)$
3	$S_2 \longrightarrow S_3$	$\beta_3 = -1$	$\lambda_3(x_1,x_2,x_3) = 5x_2$	$\zeta_3 = (0,-1,1)$

Let $\left\{X^{N_0}_\theta(t) = (X^{N_0}_{\theta,1}(t), X^{N_0}_{\theta,2}(t), X^{N_0}_{\theta,3}(t)) : t \geq 0\right\}$ be the stochastic process representing the dynamics of this multiscale reaction network. Hence for any time $t \geq 0$ and $i = 1,2,3$, $X^{N_0}_{\theta,i}(t)$ denotes the number of molecules of S_i. Suppose that the initial state of the system is $X^{N_0}_\theta(0) = (v_0,0,0)$ for $v_0 = 20$. Note that the sum of the three species numbers is preserved by all the reactions. Hence the state space for the process $X^{N_0}_\theta$ is

$$\mathcal{S} = \left\{(x_1,x_2,x_3) \in \mathbb{N}^d_0 : \ x_1 + x_2 + x_3 = v_0\right\}.$$

Clearly for this multiscale network, the *first* time-scale is $\gamma_1 = 0$ (see Section 2.1) and the corresponding set of "natural" reactions is $\Gamma_1 = \{1,2\}$. Similarly the *second* time-scale is $\gamma_2 = 1$ (see Section 2.2) and the corresponding set of "natural" reactions is

$\Gamma_2 = \{3\}$. If the time-scale of reference is γ_2 then the dynamics is given by the Markov process $X^N_{\gamma_2,\theta}$ with generator $\mathbb{A}^N_{\gamma_2,\theta}$ (see (3.1)) with $N = N_0$. As described in Section 2.2, under certain conditions we can construct a projection Π_2 for which the process $\Pi_2 X^N_{\gamma_2,\theta}$ has a well-behaved limit as $N \to \infty$. In this example, this projection is given by

$$\Pi_2(x_1, x_2, x_3) = \left(\frac{x_1 + x_2}{2}, \frac{x_1 + x_2}{2}, x_3\right).$$

Note that $\Pi_2\zeta_k = (0,0,0)$ for each $k \in \Gamma_1$ and $\Pi_2\zeta_3 = (-1/2, -1/2, 1)$. For any $v = (v_1, v_1, v_2) \in \Pi_2\mathcal{S}$, define the space $\mathbb{H}_v$ (see (2.10)) by

$$\mathbb{H}_v = \{(x - v_1, v_1 - x, 0) : \ x = 0, 1, \dots, 2v_1\}$$

and let $\mathbb{C}^v_\theta$ be the generator given by (3.2). A Markov process with state space $\mathbb{H}_v$ and generator $\mathbb{C}^v_\theta$ is ergodic. The unique stationary distribution has the form of a *binomial* distribution

$$\pi^\theta_v(x, y) = \frac{2v_1!}{(x + v_1)!(y + v_1)!}\left(\frac{\theta}{2+\theta}\right)^{y+v_1}\left(\frac{2}{2+\theta}\right)^{x+v_1} \quad \text{for } (x, y, 0) \in \mathbb{H}_v.$$

Define $\hat{\lambda}_3 : \Pi_2\mathcal{S} \to \mathbb{R}_+$ by

$$\hat{\lambda}_3(v_1, v_1, v_2) = \sum_{(x,y,0)\in\mathbb{H}_v} 5(y + v_1)\pi^\theta_v(x, y) = \left(\frac{10v_1\theta}{2+\theta}\right).$$

Let $\{\hat{X}_\theta(t) = (\hat{X}_{\theta,1}(t), \hat{X}_{\theta,2}(t), \hat{X}_{\theta,3}(t)) : t \geq 0\}$ be the $\Pi_2\mathcal{S}$-valued process with the following random time change representation

$$\hat{X}_\theta(t) = \begin{bmatrix} v_0/2 \\ v_0/2 \\ 0 \end{bmatrix} + Y\left(\left(\frac{10\theta}{2+\theta}\right)\int_0^t \hat{X}_{\theta,1}(s)ds\right)\begin{bmatrix} \text{-1/2} \\ \text{-1/2} \\ 1 \end{bmatrix},$$

where Y is a unit rate Poisson process. The due to Proposition 2.5 we have $\Pi_2 X^N_{\gamma_2,\theta} \Rightarrow \hat{X}_\theta$ as $N \to \infty$.

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be the function given by

$$f(x_1, x_2, x_3) = x_3,$$

and suppose we want to estimate

$$S^{N_0}_{\gamma_2,\theta}(f, t) = \frac{\partial}{\partial\theta}\mathbb{E}\left(f(X^{N_0}_{\gamma_2,\theta}(t))\right) = \frac{\partial}{\partial\theta}\mathbb{E}\left(X^{N_0}_{\theta,3}(t)\right).$$

Note that $f(x) = f(\Pi_2 x)$ for all $x \in \mathcal{S}$, and hence the function f_θ (given by (3.6)) coincides with the function f on the set $\Pi_2\mathcal{S}$. Therefore from Theorem 3.2 we obtain

$$S^{N_0}_{\gamma_2,\theta}(f, t) \approx \hat{S}_\theta(f, t) := \frac{\partial}{\partial\theta}\mathbb{E}\left(f(\hat{X}_\theta(t))\right) = \frac{\partial}{\partial\theta}\mathbb{E}\left(\hat{X}_{\theta,3}(t)\right), \tag{5.2}$$

for large values of N_0. We now demonstrate the usefulness of (5.2) in estimating $S^{N_0}_{\gamma_2,\theta}(f, t)$. We will numerically show that $S^{N_0}_{\gamma_2,\theta}(f, t)$ and $\hat{S}_\theta(f, t)$ are “close" to each other and the estimation of $\hat{S}_\theta(f, t)$ is far less computationally demanding than the estimation of $S^{N_0}_{\gamma_2,\theta}(f, t)$.

To estimate parameter sensitivities we will use the *coupled finite difference* (CFD) scheme developed in [1]. In this method, the sensitivity value $S^{N_0}_{\gamma_2,\theta}(f, t)$ is estimated by a *finite-difference* of the form

$$\frac{1}{h}\mathbb{E}\left(f\left(X^{N_0}_{\gamma_2,\theta+h}(t)\right) - f\left(X^{N_0}_{\gamma_2,\theta}(t)\right)\right)$$

for a *small* h, and the processes $X^{N_0}_{\gamma_2,\theta+h}$ and $X^{N_0}_{\gamma_2,\theta}$ are coupled together in a special way to reduce the variance of the associated estimator. Replacing derivative by a finite-difference introduces a *bias* in the sensitivity estimate, but we will ignore this issue here. Using CFD, we estimate $S^{N_0}_{\gamma_2,\theta}(f,t)$ and $\hat{S}_\theta(f,t)$, with $h = 0.01$, $t = 1$, $N_0 = 10^4$, $\theta = 1$ and $v_0 = 20$. The results are reported in Table 2. The sensitivity values are written in the form $s \pm l$, which means that the 95% confidence interval of the estimated value is $[s-l, s+l]$. For each estimation we use the minimum number of samples that is needed to ensure that $l \leq 0.05|s|$, where $|\cdot|$ is the absolute value function. In the table, we also indicate the CPU time[4] (in seconds) that was needed for the estimation. The CPU time can be taken as a measure of the computational effort that was required to estimate the sensitivity value. Note that Table 2 shows that relation (5.2) holds but the time needed

Table 2: Estimation of sensitivity value for $f(x_1,x_2,x_3) = x_3$

	Sensitivity Value	Number of Samples	CPU time (s)
$S^{N_0}_{\gamma_2,\theta}(f,t)$	4.2138 ± 0.2107	34932	1663.34
$\hat{S}_\theta(f,t)$	4.2017 ± 0.2100	35056	0.2333

to estimate $\hat{S}_\theta(f,t)$ is approximately 7000 times less than the time needed to estimate $S^{N_0}_{\gamma_2,\theta}(f,t)$. Note that the true sensitivity value in this case is 4.1982 (see Section B in the Appendix).

Now suppose we want to estimate $S^{N_0}_{\gamma_2,\theta}(f,t)$ for $f : \mathbb{R}^3 \to \mathbb{R}$ given by

$$f(x_1,x_2,x_3) = x_1.$$

In this case, $f_\theta : \Pi_2\mathcal{S} \to \mathbb{R}$ can be computed as

$$f_\theta(v) = \sum_{(x,y)\in\mathbb{H}_v} (x+v_1)\pi^\theta_v(x,y) = \left(\frac{4v_1}{2+\theta}\right) \text{ for any } v = (v_1,v_1,v_2) \in \Pi_2\mathcal{S}.$$

Hence Theorem 3.2 implies that

$$\begin{aligned} S^{N_0}_{\gamma_2,\theta}(f,t) \approx \hat{S}_\theta(f_\theta,t) &= \frac{\partial}{\partial\theta}\mathbb{E}\left(f_\theta(\hat{X}_\theta(t))\right) \\ &= \frac{\partial}{\partial\theta}\left(\frac{4\mathbb{E}\left(\hat{X}_{\theta,1}(t)\right)}{2+\theta}\right) \\ &= \left(\frac{4}{2+\theta}\right)\left[\frac{\partial}{\partial\theta}\mathbb{E}\left(\hat{X}_{\theta,1}(t)\right) - \frac{\mathbb{E}\left(\hat{X}_{\theta,1}(t)\right)}{2+\theta}\right]. \end{aligned}$$

As before we estimate $S^{N_0}_{\gamma_2,\theta}(f,t)$ and $\hat{S}_\theta(f_\theta,t)$ using CFD, with $h = 0.01$, $t = 1$, $N_0 = 10^4$, $\theta = 1$ and $v_0 = 20$. The results are reported in Table 3. As before, Table 3 shows that

Table 3: Estimation of sensitivity value for $f(x_1,x_2,x_3) = x_1$

	Sensitivity Value	Number of Samples	CPU time (s)
$S^{N_0}_{\gamma_2,\theta}(f,t)$	-3.3946 ± 0.1697	43745	2181.5
$\hat{S}_\theta(f_\theta,t)$	-3.6369 ± 0.1818	20827	0.1396

$S^{N_0}_{\gamma_2,\theta}(f,t) \approx \hat{S}_\theta(f_\theta,t)$ but the estimation of $S^{N_0}_{\gamma_2,\theta}(f,t)$ is around 15000 times slower than

[4] All the computations in this paper were performed using C++ programs on an Apple machine with a 2.2 GHz Intel i7 processor.

the estimation of $\hat{S}_\theta(f_\theta, t)$. In this case, the true sensitivity value is -3.6386 (see Section B in the Appendix).

This example clearly illustrates that our main result, Theorem 3.2, can be used to obtain enormous savings in the computational effort that is required for the estimation of parameter sensitivities for multiscale networks.

A Appendix.

Let S_1 and S_2 be open subsets of $\mathbb{R}^n_+$ and $\mathbb{R}^m$ respectively. Let $\mathbb{A} \subset \mathcal{B}(S_1 \times S_2) \times \mathcal{B}(S_1 \times S_2)$ be an operator whose domain $\mathcal{D}(\mathbb{A})$ includes all functions $f : S_1 \times S_2 \to \mathbb{R}$ of the form

$$f(x, y) = g(x), \tag{A.1}$$

where g is some function in $\mathcal{B}_c(S_1)$. Let $U \subset S_1 \times S_2$ be an open set and let X be a stochastic process with initial distribution $\nu \in \mathcal{P}(S_1 \times S_2)$ and sample paths in $D_{S_1 \times S_2}[0, \infty)$. Define a stopping time with respect to the filtration generated by the process X as

$$\tau = \inf\{t \geq 0 : X(t) \notin U \text{ or } X(t-) \notin U\}. \tag{A.2}$$

Then X is a solution of the *stopped martingale problem* (see Section 6, Chapter 4 in [9]) for $(\mathbb{A}, \nu, U)$ if $X(\cdot) = X(\cdot \wedge \tau)$ a.s. and

$$f(X(t)) - \int_0^{t\wedge\tau} \mathbb{A} f(X(s)) ds$$

is a martingale for each $f \in \mathcal{D}(\mathbb{A})$.

Let $\Pi : S_1 \times S_2 \to S_1$ be the projection map defined by $\Pi(x, y) = x$. Suppose that for any $g \in \mathcal{B}_c(S_1)$ and f given by (A.1) we have

$$\mathbb{A} f(x, y) = \sum_{k=1}^{K} \lambda_k(x, y) \left(g(x + \zeta_k) - g(x) \right), \tag{A.3}$$

where $\zeta_1, \dots, \zeta_K$ are certain vectors in $\mathbb{R}^n$ and $\lambda_1, \dots, \lambda_K$ are positive functions on $S_1 \times S_2$ satisfying the following : if $\lambda_k(x, y) > 0$ for some $(x, y) \in S_1 \times S_2$ then $(x + \zeta_k) \in S_1$. Furthermore we assume that the function

$$\sum_{k=1, \langle \bar{1}_d, \zeta_k \rangle > 0}^{K} \lambda_k(x, y) \tag{A.4}$$

is linearly growing with respect to projection Π (see Definition 2.1) .

Lemma A.1. *Fix a $w_0 = (x_0, y_0) \in S_1 \times S_2$ and let $\delta_{w_0} \in \mathcal{P}(S_1 \times S_2)$ be the distribution that puts all the mass at w_0. For any $M \in \mathbb{N}$, let U_M be the open set*

$$U_M = \{(x, y) \in S_1 \times S_2 : \|x\| < M\}.$$

Assume that the stopped martingale problem for $(\mathbb{A}, \delta_{w_0}, U_M)$ has a unique solution W_M for each M. Let τ_M be the stopping time defined by (A.2) with U replaced by U_M and the process X replaced by W_M. Then we have the following.

(A) For any $T > 0$, $\lim_{M\to\infty} \mathbb{P}(\tau_M < T) = 0$.

(B) There exists a unique solution W for the (unstopped) martingale problem for $(\mathbb{A}, \delta_{w_0})$. Moreover for any positive integer p and $T > 0$ we have

$$\sup_{t \in [0,T]} \mathbb{E}(\|\Pi W(t)\|^p) < \infty.$$

(C) If a function $f : S_1 \times S_2 \to \mathbb{R}$ is polynomially growing with respect to projection Π, then for any $T \geq 0$

$$\sup_{t \in [0,T]} \mathbb{E}\left(|f(W(t))|\right) < \infty.$$

(D) The martingale problem for $\mathbb{A}$ is well-posed.

Proof. Suppose that $W_M(t) = (X_M(t), Y_M(t))$ for any $t \geq 0$, where X_M and Y_M are processes with state spaces S_1 and S_2 respectively. Let $q = \max\{\langle \bar{1}_d, \zeta_k \rangle : k = 1, \dots, K\}$. For a large M and a positive integer p, define $g \in \mathcal{B}_c(S_1)$ by

$$g(x) = \|x\|^p \mathbb{1}_{\{\|x\| \leq M+q\}}(x).$$

Assume that $\|x_0\|^p < M$ and note that the definition of g implies that for any $t \leq \tau_M$, we have $g(X_M(t)) = \|X_M(t)\|^p$ and $g(X_M(t) + \zeta_k) = \|X_M(t) + \zeta_k\|^p$ for each $k = 1, \dots, K$. Let $f : S_1 \times S_2 \to \mathbb{R}$ be the function given by $f(x,y) = g(x)$. Then $f \in \mathcal{D}(\mathbb{A})$ and using (A.3) we obtain

$$\begin{aligned}
&f(W_M(t \wedge \tau_M)) - \int_0^{t \wedge \tau_M} \mathbb{A} f(W_M(s)) ds \\
&= \|X_M(t \wedge \tau_M))\|^p - \int_0^{t \wedge \tau_M} \sum_{k=1}^K \lambda_k(X_M(s), Y_M(s)) \left(\|X_M(s) + \zeta_k\|^p - \|X_M(s))\|^p\right) ds
\end{aligned}$$

is a martingale starting at $\|x_0\|^p$. Taking expectations we get

$$\begin{aligned}
&\mathbb{E}\left(\|X_M(t \wedge \tau_M))\|^p\right) \\
&= \|x_0\|^p + \mathbb{E}\left(\int_0^{t \wedge \tau_M} \sum_{k=1}^K \lambda_k(X_M(s), Y_M(s)) \left(\|X_M(s) + \zeta_k\|^p - \|X_M(s))\|^p\right) ds\right).
\end{aligned}$$

Our assumption on the functions $\lambda_1, \dots, \lambda_K$ implies that when $\lambda_k(X_M(s), Y_M(s)) > 0$, then $(X_M(s) + \zeta_k) \in S_1 \subset \mathbb{R}^d_+$ and hence $\|X_M(s) + \zeta_k\| = \langle \bar{1}_d, X_M(s) \rangle + \langle \bar{1}_d, \zeta_k \rangle$. This gives us

$$\begin{aligned}
&\mathbb{E}\left(\|X_M(t \wedge \tau_M))\|^p\right) \\
&= \|x_0\|^p + \mathbb{E}\left(\int_0^{t \wedge \tau_M} \sum_{k=1}^K \lambda_k(X_M(s), Y_M(s)) \left((\langle \bar{1}_d, X_M(s) \rangle + \langle \bar{1}_d, \zeta_k \rangle)^p - \langle \bar{1}_d, X_M(s) \rangle^p\right) ds\right) \\
&\leq \|x_0\|^p + 2^p q^p \mathbb{E}\left(\int_0^t \sum_{k \in P} \lambda_k(X_M(s \wedge \tau_M), Y_M(s \wedge \tau_M)) \left(\|X_M(s \wedge \tau_M)\|^{p-1} + 1\right) ds\right),
\end{aligned}$$

where $P = \{k = 1, \dots, K : \langle \bar{1}_d, \zeta_k \rangle > 0\}$. Since the function given by (A.4) is linearly growing with respect to projection Π, we can find a positive constant C (independent of M) such that

$$\mathbb{E}\left(\|X_M(t \wedge \tau_M)\|^p)\right) \leq \|x_0\|^p + Ct + C \int_0^t \mathbb{E}\left(\|X_M(s \wedge \tau_M)\|^p\right) ds.$$

Gronwall's inequality implies that

$$\mathbb{E}\left(\|X_M(t\wedge\tau_M)\|^p\right) \leq \left(\|x_0\|^p + Ct\right)e^{Ct}. \tag{A.5}$$

Using Markov's inequality we obtain

$$\lim_{M\to\infty}\mathbb{P}\left(\tau_M < t\right) = \lim_{M\to\infty}\mathbb{P}\left(\|X_M(t\wedge\tau_M)\|^p \geq M^p\right) \leq \lim_{M\to\infty}\frac{\mathbb{E}\left(\|X_M(t\wedge\tau_M)\|^p\right)}{M^p} = 0.$$

The last limit is 0 due to (A.5). This proves part (A) of the lemma. From Theorem 6.3 in Chapter 4 of [9] we can conclude that the martingale problem for $(\mathbb{A}, \delta_{w_0})$ has a unique solution W. In fact for any $M \in \mathbb{N}$, the process $W_M(\cdot\wedge\tau_M)$ has the same distribution as the process $W(\cdot\wedge\tau_M)$. Therefore using (A.5) we get

$$\mathbb{E}\left(\|\Pi W(t\wedge\tau_M)\|^p\right) = \mathbb{E}\left(\|\Pi W_M(t\wedge\tau_M)\|^p\right) \leq \left(\|x_0\|^p + Ct\right)e^{Ct}. \tag{A.6}$$

Since τ_M is monotonically increasing with M, we must have that $\tau_M \to \infty$ a.s. as $M \to \infty$. Letting $M\to\infty$ in (A.6) and using Fatou's lemma we obtain

$$\mathbb{E}\left(\|\Pi W(t)\|^p\right) \leq \lim_{M\to\infty}\mathbb{E}\left(\|\Pi W(t\wedge\tau_M)\|^p\right) \leq \left(\|x_0\|^p + Ct\right)e^{Ct}.$$

Taking supremum over $t\in[0,T]$ proves part (B) of the lemma. The proof of part (C) is immediate from part (B). Since part (B) of this lemma holds for any w_0, the martingale problem for $\mathbb{A}$ is well-posed and this proves part (D). □

Using the above lemma we now prove the main result of this section.

Lemma A.2. *Recall the definition of operator $\mathbb{B}_\theta^N$ from* (4.39)*.*

(A) The martingale corresponding to $\mathbb{B}_\theta^N$ is well-posed.

(B) Let W_θ^N be the $\hat{\mathcal{S}}$-valued Markov process with generator $\mathbb{B}_\theta^N$ and initial state $(t_0, v_0, z_0) \in \hat{\mathcal{S}}$. For any $M\in\mathbb{N}$, define a stopping time by

$$\sigma_M^N = \inf\{t\geq 0 : \|\Pi_{\hat{\mathcal{S}}}W_\theta^N(t)\| > M\}. \tag{A.7}$$

Then for any $T>0$

$$\lim_{M\to\infty}\sup_{N\in\mathbb{N}}\mathbb{P}\left(\sigma_M^N < T\right) = 0. \tag{A.8}$$

(C) For any positive integer p and any $T>0$

$$\sup_{t\in[0,T]}\sup_{N\in\mathbb{N}}\mathbb{E}\left(\|\Pi_{\hat{\mathcal{S}}}W_\theta^N(t)\|^p\right) < \infty. \tag{A.9}$$

(D) Let $f:\mathcal{S}\to\mathbb{R}$ be a function which is polynomially growing with respect to projection Π_2, and define f_θ^N by (4.47)*. Then for any positive integer p and $T>0$*

$$\sup_{N\in\mathbb{N}}\sup_{t\in[0,T]}\mathbb{E}\left(\left|f_\theta^N(W_\theta^N(t))\right|^p\right) < \infty \quad \text{and} \quad \sup_{N\in\mathbb{N}}\mathbb{E}\left(\int_0^T\left|\mathbb{B}_\theta^N f_\theta^N(W_\theta^N(t))\right|^p dt\right) < \infty. \tag{A.10}$$

(E) Let f and f_θ^N be as in part (D). For any $T\geq 0$ and any stopping time σ we have

$$\mathbb{E}\left(f_\theta^N(W_\theta^N(T\wedge\sigma))\right) = f_\theta^N(t_0, v_0, z_0) + \mathbb{E}\left(\int_0^{T\wedge\sigma}\mathbb{B}_\theta^N f_\theta^N(W_\theta^N(t))dt\right). \tag{A.11}$$

(F) The sequence of processes $\{W^N_\theta : N \in \mathbb{N}\}$ is tight in the space $D_{\hat{\mathcal{S}}}[0,\infty)$.

Proof. Note that on the set

$$U_M = \{(t,v,z) \in \hat{\mathcal{S}} : \|v\| < M\},$$

the functions $\{\rho^N_{k,\theta} : k \in \Gamma_2\}$ are bounded. If we define each $\rho^N_{k,\theta}$ to be 0 outside the set U_M, then the resulting operator $\mathbb{B}^N_{M,\theta}$ can be seen as a bounded perturbation of the translation operator

$$\mathbb{T}f(t,v,z) = \frac{\partial f(t,v,z)}{\partial t},$$

which certainly has a well-posed martingale problem. From Theorem 4.10.3 in [9] we can conclude that the martingale problem for $\mathbb{B}^N_{M,\theta}$ is well-posed. This implies that for any initial state $w_0 \in \hat{\mathcal{S}}$, the stopped martingale problem for $(\mathbb{B}^N_\theta, \delta_{w_0}, U_M)$ is well-posed. Assumption 3.1 implies that the function

$$\sum_{k \in \Gamma_2, \langle \bar{1}_d, \zeta^s_k \rangle > 0} \rho^N_{k,\theta}(t,v,z) \tag{A.12}$$

is linearly growing with respect to projection $\Pi_{\hat{\mathcal{S}}}$ (given by (4.37)). Therefore part (B) of Lemma A.1 shows that there is a unique solution for the martingale problem for $(\mathbb{B}^N_\theta, \delta_{w_0})$. Hence the martingale problem for $\mathbb{B}^N_\theta$ is well-posed and this proves part (A). Let W^N_θ be the $\hat{\mathcal{S}}$-valued Markov process with generator $\mathbb{B}^N_\theta$ and initial state (t_0, v_0, z_0). We can rework the proof of Lemma A.1 to prove parts (B) and (C).

Let $f : \mathcal{S} \to \mathbb{R}$ be a function which is polynomially growing with respect to projection Π_2 and define f^N_θ by (4.47). Remark 4.14 implies that the sequences of functions $\{f^N_\theta : N \in \mathbb{N}\}$ and $\{\mathbb{B}^N_\theta f^N_\theta : N \in \mathbb{N}\}$ are polynomially growing with respect to projection $\Pi_{\hat{\mathcal{S}}}$. Therefore part (D) is an easy consequence of part (C).

Corresponding to the function f define a function $f_M : \mathcal{S} \to \mathbb{R}$ by

$$f_M(x) = f(x) \mathbb{1}_{\{\|x\| \leq M\}}(x).$$

Let $f^N_{M,\theta}$ be the function given by (4.47), with f replaced by f_M. Since f_M is in $\mathcal{B}_c(\mathcal{S})$, the function $f^N_{M,\theta}$ is in class $\mathcal{C}$ (see (4.38)). Using Dynkin's theorem (see Lemma 19.21 in [20]) we get

$$\mathbb{E}\left(f^N_{M,\theta}(W^N_\theta(T \wedge \sigma))\right) = f^N_{M,\theta}(t_0, v_0, z_0) + \mathbb{E}\left(\int_0^{T\wedge\sigma} \mathbb{B}^N_\theta f^N_{M,\theta}(W^N_\theta(t))dt\right).$$

Taking the limit $M \to \infty$, and using part (D) along with the dominated convergence theorem proves part (E).

To show that the sequence $\{W^N_\theta : N \in \mathbb{N}\}$ is tight, we first have to prove the compact containment criterion (see Chapter 3 in [9]). This means that for any $T, \epsilon > 0$ we exhibit a compact set $K_{\epsilon,T} \subset \hat{\mathcal{S}}$ such that

$$\inf_{N \in \mathbb{N}} \mathbb{P}\left(W^N_\theta(t) \in K_{\epsilon,T} \text{ for all } t \in [0,T]\right) \geq 1 - \epsilon. \tag{A.13}$$

Let σ^N_M be the stopping time given by (A.7). For any $t \geq 0$, we can write $W^N_\theta(t) = (\tau^N_\theta(t), V^N_\theta(t), Z^N_\theta(t))$ (see (4.40)). Fix an $\epsilon > 0$ and $T > 0$. Part (B) shows that we can find a $M > 0$ large enough so that

$$\sup_{N \in \mathbb{N}} \mathbb{P}(\sigma^N_M \leq T) < \epsilon. \tag{A.14}$$

Note that for any $t \geq 0$, if $V_\theta^N(t) = v$ then $\tau_\theta^N(t) \in [0, t+t_0]$ and $Z_\theta^N(t) \in \mathbb{H}_v$ where $\mathbb{H}_v$ is a finite set. This shows that for any $t < \sigma_M^N$ we have $W_\theta^N(t) \in K_{\epsilon,T}$ where $K_{\epsilon,T}$ is the compact set given by

$$K_{\epsilon,T} = \left\{ (t, v, z) \in \hat{\mathcal{S}} : t \in [0, T+t_0], \|v\| \leq M \text{ and } z \in \mathbb{H}_v \right\}.$$

Hence

$$\mathbb{P}\left(W_\theta^N(t) \in K_{\epsilon,T} \text{ for all } t \in [0,T] \right) \geq \mathbb{P}(\sigma_M^N > T) = 1 - \mathbb{P}(\sigma_M^N \leq T).$$

Taking supremum over N and using (A.14) proves (A.13).

Now that we have shown the compact containment condition, Theorem 3.9.1 in [9] allows us to verify the tightness of $\{W_\theta^N : N \in \mathbb{N}\}$ by proving that for any $f \in \mathcal{C}$, the sequence of processes $\{f(W_\theta^N(\cdot)) : N \in \mathbb{N}\}$ is tight in the space $D_{\mathbb{R}}[0,\infty)$. Note that

$$f(W_\theta^N(t)) - \int_0^t \mathbb{B}_\theta^N f(W_\theta^N(s)) ds$$

is a martingale and part (D) of the lemma shows that

$$\mathbb{E}\left(\int_0^t \left| \mathbb{B}_\theta^N f_\theta^N(W_\theta^N(s)) \right|^2 ds \right) < \infty$$

for any $t \geq 0$. The tightness of the sequence $\{f(W_\theta^N(\cdot)) : N \in \mathbb{N}\}$ is immediate from Theorem 3.9.4 in [9]. This completes the proof of part (E) of the lemma. □

B Appendix.

Recall the multiscale reaction network considered in Section 5. This network consists of 3 species S_1, S_2 and S_3, and the following three reactions

$$S_1 \xrightarrow{c_1} S_2, \quad S_2 \xrightarrow{c_2} S_1 \quad \text{and} \quad S_2 \xrightarrow{c_3} S_3.$$

The values of the rate constants c_1, c_2 and c_3 are given in (5.1). For any time $t \geq 0$ and $i = 1,2,3$, let $X_{\theta,i}^{N_0}(t)$ denote the number of molecules of S_i, where $N_0 = 10^4$ is the normalization parameter and $\theta = c_1$ is the sensitive parameter. We assume that the initial state of the system is $(X_{\theta,1}^{N_0}(t), X_{\theta,2}^{N_0}(0), X_{\theta,3}^{N_0}(0)) = (v_0, 0, 0)$.

For each $i = 1,2,3$ and time $t \geq 0$, define the *first-moment* of species i, and its sensitivity with respect to θ as

$$m_{\theta,i}(t) = \mathbb{E}\left(X_{\theta,i}^{N_0}(t) \right) \qquad \text{and} \qquad M_{\theta,i}(t) = \frac{\partial}{\partial \theta} \mathbb{E}\left(X_{\theta,i}^{N_0}(t) \right).$$

Let $m_\theta(t) = (m_{\theta,1}(t), m_{\theta,2}(t), m_{\theta,3}(t))$ be the vector of first-moments and let $M_\theta(t) = (M_{\theta,1}(t), M_{\theta,2}(t), M_{\theta,3}(t))$ denote the vector of its θ-sensitivities. Since the network only consists of *unimolecular* reactions, we can explicitly compute $m_\theta(t)$ and $M_\theta(t)$ by solving the following system of ordinary differential equations:

$$\begin{aligned} \frac{dm_\theta(t)}{dt} &= A_\theta m_\theta(t) \\ \frac{dM_\theta(t)}{dt} &= A_\theta M_\theta(t) + B_\theta m_\theta(t), \end{aligned}$$

with the initial conditions $m_\theta(0) = (v_0, 0, 0)$ and $M_\theta(0) = (0,0,0)$. Here A_θ and B_θ are 3×3 matrices given by

$$A_\theta = \begin{bmatrix} -\theta & c_2 & 0 \\ \theta & -(c_2+c_3) & 0 \\ 0 & c_3 & 0 \end{bmatrix} \qquad \text{and} \qquad B_\theta = \frac{\partial A_\theta}{\partial \theta} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

On solving this system of equations until time $t = N_0 = 10^4$ we get

$$M_{\theta,3}(N_0) = \frac{\partial}{\partial\theta}\mathbb{E}\left(X^{N_0}_{\theta,3}(N_0)\right) = 4.1982 \quad \text{and } M_{\theta,1}(N_0) = \frac{\partial}{\partial\theta}\mathbb{E}\left(X^{N_0}_{\theta,1}(N_0)\right) = -3.6386,$$

which shows the correctness of the sensitivity values given in Tables 2 and 3 respectively.

References

[1] D. Anderson. An efficient finite difference method for parameter sensitivities of continuous time markov chains. *SIAM: Journal on Numerical Analysis*, 2012. MR-3022217

[2] D. Anderson and T. Kurtz. Continuous time Markov chain models for chemical reaction networks. In H. Koeppl, G. Setti, M. di Bernardo, and D. Densmore, editors, *Design and Analysis of Biomolecular Circuits*. Springer-Verlag, 2011.

[3] A. P. Arkin, C. V. Rao, and D. M. Wolf. Control, exploitation and tolerance of intracellular noise. *Nature*, 420:231–237, 2002.

[4] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.*, 16(4):1925–1961, 2006. MR-2288709

[5] Y. Cao, D. Gillespie, and L. Petzold. The slow-scale stochastic simulation algorithm. *Journal of Chemical Physics*, 122(1):1–18, 2005.

[6] W. E, D. Liu, and E. Vanden-Eijnden. Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. *J. Comput. Phys.*, 221(1):158–180, Jan. 2007. MR-2290567

[7] H. El-Samad, H. Kurata, J. C. Doyle, C. A. Gross, and M. Khammash. Surviving heat shock: Control strategies for robustness and performance. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8):2736–2741, 2005.

[8] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[9] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence. MR-0838085

[10] F. O. Farid. Spectral properties of perturbed linear operators and their application to infinite matrices. *Proc. Amer. Math. Soc.*, 112(4):1013–1022, 1991. MR-1057943

[11] X.-j. Feng, S. Hooshangi, D. Chen, R. Li, Genyuan; Weiss, and H. Rabitz. Optimizing genetic circuits by global sensitivity analysis. *Biophysical journal*, 87(4):2195 – 2202, 2004.

[12] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

[13] D. Gillespie. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58:35–55, 2007.

[14] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[15] J. Goutsias. Classical versus stochastic kinetics modeling of biochemical reaction systems. *Biophysical Journal*, 92(7):2350 – 2365, 2007.

[16] R. Gunawan, Y. Cao, and F. Doyle. Sensitivity analysis of discrete stochastic systems. *Biophysical Journal*, 88(4):2530–2540, 2005.

[17] A. Gupta and M. Khammash. Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks. *SIAM : Journal on Scientific Computing*, 35(6):2598 – 2620, 2013. MR-3129760

[18] E. Haseltine and J. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *Journal of Chemical Physics*, 117(15), 2002.

[19] E. L. Haseltine, D. B. Patience, and J. B. Rawlings. On the stochastic simulation of particulate systems. *SIAM : Journal on Scientific Computing*, 60(10):2627 – 2641, 2013.

[20] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. MR-1876169

[21] H.-W. Kang. A multiscale approximation in a heat shock response model of e. coli. *BMC Systems Biology*, 6(143), 2012.

[22] H.-W. Kang and T. G. Kurtz. Separation of time-scales and model reduction for stochastic reaction networks. *Ann. Appl. Probab.*, 23(2):529–583, 2013. MR-3059268

[23] A. M. Kierzek, J. Zaim, and P. Zielenkiewicz. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *Journal of Biological Chemistry*, 276(11):8165–8172, 2001.

[24] M. D. Levin, C. J. Morton-Firth, W. N. Abouhamad, R. B. Bourret, and D. Bray. Origins of individual swimming behavior in bacteria. *Biophysical Journal*, 74(1):175 – 181, 1998.

[25] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci., Biochemistry*, 94:814–819, 1997.

[26] H. H. McAdams and A. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in genetics : TIG*, 15(2):65–69, Feb. 1999.

[27] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993. MR-1234295

[28] M. Morelli, R. Allen, S. Tanase-Nicola, and P. ten Wolde. Eliminating fast reactions in stochastic simulations of biochemical networks: a bistable genetic switch. *Journal of Chemical Physics*, 128(4), 2008.

[29] C. Rao and A. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm. *Journal of Chemical Physics*, 118(11):4999–5010, 2003.

[30] M. Rathinam, P. W. Sheppard, and M. Khammash. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *Journal of Chemical Physics*, 132, 2010.

[31] B. Sericola. Occupation times in Markov processes. *Comm. Statist. Stochastic Models*, 16(5):479–510, 2000. MR-1796344

[32] A. P. Seyranian and A. A. Mailybaev. *Multiparameter stability theory with mechanical applications*, volume 13 of *Series on Stability, Vibration and Control of Systems. Series A: Textbooks, Monographs and Treatises*. World Scientific Publishing Co. Inc., River Edge, NJ, 2003. MR-2056325

[33] P. W. Sheppard, M. Rathinam, and M. Khammash. A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems. *Journal of Chemical Physics*, 136, 2012.

[34] J. Stelling, E. D. Gilles, and F. J. Doyle. Robustness properties of circadian clock architectures. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13210–13215, 2004.

[35] P. Warren and R. Allen. Steady-state parameter sensitivity in stochastic modeling via trajectory reweighting. *Journal of Chemical Physics*, 136(10), 2012.

[36] E. Weinan, D. Liu, and E. Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *Journal of Chemical Physics*, 123(19):1–8, 2005.

Acknowledgments. This work has been supported by ETH and the Human Frontier Science Program Grant RGP0061/2011.