An asymptotic relationship between coupling methods for stochastically modeled population processes

David F. Anderson[1] and Masanori Koyama[2]

August 1, 2014

Abstract

This paper is concerned with elucidating a relationship between two common coupling methods for the continuous time Markov chain models utilized in the cell biology literature. The couplings considered here are primarily used in a computational framework by providing reductions in variance for different Monte Carlo estimators, thereby allowing for significantly more accurate results for a fixed amount of computational time. Common applications of the couplings include the estimation of parametric sensitivities via finite difference methods and the estimation of expectations via multi-level Monte Carlo algorithms. While a number of coupling strategies have been proposed for the models considered here, and a number of articles have experimentally compared the different strategies, to date there has been no mathematical analysis describing the connections between them. Such analyses are critical in order to determine the best use for each. In the current paper, we show a connection between the common reaction path (CRP) method and the split coupling (SC) method, which is termed coupled finite differences (CFD) in the parametric sensitivities literature. In particular, we show that the two couplings are both limits of a third coupling strategy we call the "local-CRP" coupling, with the split coupling method arising as a key parameter goes to infinity, and the common reaction path coupling arising as the same parameter goes to zero. The analysis helps explain why the split coupling method often provides a lower variance than does the common reaction path method, a fact previously shown experimentally.

1 Introduction

Models of biochemical reaction networks with stochastic dynamics have become increasingly popular in the science literature over the previous fifteen years where they are often studied via computational methods and, in particular, Monte Carlo methods. These computational methods tend to be extremely expensive and time-consuming without the use of variance reduction techniques. One of the most common ways to achieve a large reduction of variance is to couple two relevant processes in order to increase their covariance. There are three main couplings found in the relevant literature: (i) the use of common random numbers (CRN), (ii) the common reaction path (CRP) coupling [17], and (iii) a *split coupling* (SC) method termed coupled finite differences in the setting of parametric sensitivities [2, 4]. It

[1]Department of Mathematics, University of Wisconsin, Madison, Wi. 53706, anderson@math.wisc.edu.

[2]Department of Systems Science, Graduate School of Informatics, Kyoto University, Integrated Systems Biology Lab, koyama.masanori@gmail.com. Corresponding Author.

[0]AMS 2000 subject classifications: Primary 60H35, 65C99; Secondary 92C40

has been observed in the literature that both the CRP and SC couplings are far superior to the CRN coupling in terms of variance reduction [2, 17, 18]. It has also been observed through examples that the SC method tends to perform much better than the CRP method, though some exceptions exist [2, 18]. To the best of the authors' knowledge there has to date been no analytical work on understanding the connections between these two couplings. In the present paper we prove that both the CRP and SC couplings arise naturally as different limits of a third family of couplings we term the *local-CRP* coupling. In particular, the CRP coupling arises as a limit in which the local-CRP coupling is as loosely coupled as possible, whereas the SC coupling arises from a limit of the local-CRP "recoupling" as often as possible. Such an analysis sheds light on why the split coupling often provides a lower variance than does the CRP coupling.

The outline for the remainder of the paper is as follows. In Section 2, we formally present the mathematical models considered in this paper, together with a brief description of the computational methods that serve as motivation for the present work. In Section 3, we present the different coupling strategies for the models presented in Section 2. In Section 4, we state and prove our main results. In Section 5, we provide numerical examples demonstrating our main results, and in Section 6 we conclude with some brief remarks.

2 Mathematical model and motivating computational methods

Motivated by models in biochemistry, we consider continuous time Markov chain models in $\mathbb{Z}^d$, in which the ith component of the process typically represents the number of molecules of "species" i present in the system. The transitions of the chain are specified by vectors, $\zeta_k \in \mathbb{Z}^d$, for $k \in \{1, \dots, R\}$ with $R < \infty$, determining the net change in the chain due to the occurrence of a single "reaction," and by the intensity functions $\lambda_k : \mathbb{Z}^d \to \mathbb{R}_{\geq 0}$, which determine the rate at which the different reactions are occurring.[1] Specifically, letting $N_k(t)$ be the number of times transition $k \in \{1, \dots, R\}$ has occurred by time $t \geq 0$, we will consider the continuous time Markov chain X satisfying the equation

$$X(t) = X(0) + \sum_{k=1}^{R} N_k(t)\zeta_k,$$

where N_k is a counting process with local intensity function λ_k. That is, $\{N_k\}$ are the counting processes for which the processes

$$N_k(t) - \int_0^t \lambda_k(X(s))ds$$

are local martingales. One useful representation for the counting processes $N_k(t)$ is via time-changed unit-rate Poisson processes [1, 6, 10, 15],

$$N_k(t) = Y_k\left(\int_0^t \lambda_k(X(s))ds\right),$$

yielding the stochastic equation

$$X(t) = X(0) + \sum_{k=1}^{R} Y_k\left(\int_0^t \lambda_k(X(s))ds\right)\zeta_k \tag{2.1}$$

[1] Intensity functions are termed "propensity" functions in the biochemistry literature.

where $\{Y_k\}_{k=1}^R$ is a collection of independent unit-rate Poisson processes. Note that X can also be specified by its infinitesimal generator,

$$(Af)(x) = \sum_{k=1}^{R} \lambda_k(x)(f(x+\zeta_k) - f(x)), \tag{2.2}$$

where f is any bounded function with compact support.

We denote Z as the process on $\mathbb{Z}^d$ with the same transition directions $\{\zeta_k\}$ as X, but with intensities $\widetilde{\lambda}_k : \mathbb{Z}^d \to \mathbb{R}_{\geq 0}$. That is, Z is the Markov process with infinitesimal generator

$$(Bf)(x) = \sum_{k=1}^{R} \widetilde{\lambda}_k(x)(f(x+\zeta_k) - f(x)), \tag{2.3}$$

and which satisfies the stochastic equation

$$Z(t) = Z(0) + \sum_{k=1}^{R} Y_k \left(\int_0^t \widetilde{\lambda}_k(Z(s))ds \right) \zeta_k, \tag{2.4}$$

where $\{Y_k\}_{k=1}^R$ is a collection of independent unit-rate Poisson processes. In the remainder of the paper, we consider different ways to couple X and Z and provide an asymptotic relationship between two of the couplings.

2.1 Motivating computational methods

We briefly present two computational methods that serve as the motivation for the analysis of the different coupling strategies: finite difference methods for parametric sensitivity analysis and multi-level Monte Carlo for the estimation of expectations.

2.1.1 Parametric sensitivity analysis

Suppose that $\{X^\theta\}$ is a parametric family of processes about θ on a state space E, and $f : E \to \mathbb{R}$ is some statistic of interest. For example, $f(X(t)) = X_i(t)$ may provide the abundance of species i at time $t \geq 0$. It is common to wish to evaluate

$$\frac{d}{d\theta}\mathbb{E}[f(X^\theta(t))] \approx \frac{\mathbb{E}[f(X^{\theta+h}(t))] - \mathbb{E}[f(X^\theta(t))]}{h} \tag{2.5}$$

as a measurement of the sensitivity of $\mathbb{E}[f(X^\theta(t))]$ with respect to θ. Such a strategy is usually called a *finite difference* method. We would like to empirically evaluate the right-hand side of (2.5) in as efficient a manner as possible. By coupling the processes $(X^{\theta+h}, X^\theta)$, we may evaluate

$$h^{-1}\mathbb{E}[f(X^{\theta+h}(t)) - f(X^\theta(t))],$$

with the magnitude of $\mathsf{Var}(f(X^{\theta+h}(t)) - f(X^\theta(t)))$ determining the quality of the coupling. In particular, we wish to minimize $\mathsf{Var}(f(X^{\theta+h}(t)) - f(X^\theta(t)))$ without greatly increasing the computational cost of producing realizations of the coupled processes $(X^{\theta+h}, X^\theta)$. We explicitly note that in the setting of the previous section, we have

$$\lambda_k(\cdot) = \eta_k(\theta, \cdot), \qquad \widetilde{\lambda}_k(\cdot) = \eta_k(\theta + h, \cdot),$$

where for each k, $\{\eta_k(\theta,\cdot):\mathbb{R}^d\to\mathbb{R}_{\geq 0}\}$ is a parametric family of functions about θ. In this case, we have

$$X = X^{\theta}, \quad \text{and} \quad Z = X^{\theta+h}.$$

As mentioned in Section 1, there has been a large amount of work in the literature on developing good coupling strategies for the estimation of parametric sensitivities via finite differences (2.5); see, for example, [2, 7, 17, 18]. To the best of the authors' knowledge there has been no mathematical analysis detailing the connections between the different couplings used, though see the discussion in Section 6 for details pertaining to a recent work by Arampatzis and Katsoulakis [8].

2.2 Multi-level Monte Carlo

In [11], Mike Giles introduced the multi-level Monte Carlo (MLMC) method for the approximation of expectations of diffusion processes. Specifically, if X is the diffusion process of interest and $\{Z_\ell\}$ are a family of approximations to X, with higher values of ℓ corresponding to better approximations, then we observe that for any function f of interest,

$$\mathbb{E}[f(X(t))] \approx \mathbb{E}[f(Z_L)] = \sum_{\ell=1}^{L} \mathbb{E}[f(Z_\ell(t)) - f(Z_{\ell-1}(t))] + \mathbb{E}[f(Z_0)],$$

where L is chosen large enough so that $|\mathbb{E}[f(X(t))] - \mathbb{E}[f(Z_L(t))]|$ is below some target accuracy. It is typical to choose Z_ℓ to be the process produced by Euler-Maruyama with a step size of $M^{-\ell}$ for some $M \in \{2,3,\ldots,7\}$. If each term $f(Z_\ell(t)) - f(Z_{\ell-1}(t))$ is tightly coupled, then the variance of each of the intermediate estimators will be low, thereby moving the computational cost to the lowest level, $\mathbb{E}[f(Z_0)]$, which can be estimated quickly via Euler-Maruyama with large time-steps.

In [4], Anderson and Higham extended the multi-level Monte Carlo method to the setting of this paper by utilizing the split coupling detailed in Section 3. They further noted that an unbiased estimator can be produced for jump models by coupling the exact process X with the approximate process with the finest time discretizatoin

$$\mathbb{E}[f(X(t))] = \mathbb{E}[f(X(t)) - f(Z_L(t))] + \sum_{\ell=1}^{L} \mathbb{E}[f(Z_\ell(t)) - f(Z_{\ell-1}(t))] + \mathbb{E}[f(Z_0)],$$

where, again, it is the quality of the coupling at each level that determines the overall quality of the method.

We point out that in the diffusive case the most natural coupling is to re-use the driving Brownian path for each of the coupled processes. This is relatively easy to do via the Brownian bridge. However, as will be noted in the next section, there are multiple natural couplings to choose from in the context of jump processes with state dependent intensity functions, and different choices lead to computational methods with vastly different computational complexities and, hence, runtimes.

3 Different Couplings

We return to the notation introduced at the beginning of Section 2 and focus our discussion on ways to couple X and Z with intensities λ_k and $\widetilde{\lambda}_k$, respectively.

3.1 Split coupling

We will begin by introducing the split coupling (SC), which first appeared as an analytic tool in [16] and later appeared in the context of computational methods in [2, 3, 4, 5, 8]. Let $a \wedge b \stackrel{\text{def}}{=} \min\{a, b\}$, and let $\mathcal{U}$ and $\mathcal{V}$ be any càdlàg processes on $\mathbb{R}^d$. Then for each $k \in \{1, \ldots, R\}$ we define the operators r_{1k}, r_{2k}, and r_{3k} via

$$\begin{aligned} r_{1k}(\lambda_k, \widetilde{\lambda}_k, \mathcal{U}, \mathcal{V})(s) &\stackrel{\text{def}}{=} \lambda_k(\mathcal{U}(s)) \wedge \widetilde{\lambda}_k(\mathcal{V}(s)) \\ r_{2k}(\lambda_k, \widetilde{\lambda}_k, \mathcal{U}, \mathcal{V})(s) &\stackrel{\text{def}}{=} \lambda_k(\mathcal{U}(s)) - r_{1k}(\lambda_k, \widetilde{\lambda}_k, \mathcal{U}, \mathcal{V})(s) \\ r_{3k}(\lambda_k, \widetilde{\lambda}_k, \mathcal{U}, \mathcal{V})(s) &\stackrel{\text{def}}{=} \widetilde{\lambda}_k(\mathcal{V}(s)) - r_{1k}(\lambda_k, \widetilde{\lambda}_k, \mathcal{U}, \mathcal{V})(s). \end{aligned} \tag{3.1}$$

The split coupling of the processes X and Z is then given by

$$\begin{aligned} X_{\text{sc}}(t) = X(0) + \sum_{k=1}^{R} \Bigg\{ & Y_{1k}\left(\int_0^t r_{1k}(\lambda_k, \widetilde{\lambda}_k, X_{\text{sc}}, Z_{\text{sc}})(s)ds\right) \\ & + Y_{2k}\left(\int_0^t r_{2k}(\lambda_k, \widetilde{\lambda}_k, X_{\text{sc}}, Z_{\text{sc}})(s)ds\right)\Bigg\}\zeta_k \\ Z_{\text{sc}}(t) = Z(0) + \sum_{k=1}^{R} \Bigg\{ & Y_{1k}\left(\int_0^t r_{1k}(\lambda_k, \widetilde{\lambda}_k, X_{\text{sc}}, Z_{\text{sc}})(s)ds\right) \\ & + Y_{3k}\left(\int_0^t r_{3k}(\lambda_k, \widetilde{\lambda}_k, X_{\text{sc}}, Z_{\text{sc}})(s)ds\right)\Bigg\}\zeta_k, \end{aligned} \tag{3.2}$$

where $\{Y_{1k}\}_{k=1}^R \cup \{Y_{2k}\}_{k=1}^R \cup \{Y_{3k}\}_{k=1}^R$ are mutually independent unit-rate Poisson processes. Note that X_{sc} and Z_{sc} share the family of counting processes determined by the Poisson processes Y_{1k}. Further note that (X, Z) satisfying the stochastic equation (3.2) is simply a continuous time Markov chain on $\mathbb{Z}^d \times \mathbb{Z}^d$ with infinitesimal generator

$$\begin{aligned} (\mathcal{L}_{\text{sc}}g)(x, z) = & \sum_{k=1}^{R} \min\{\lambda_k(x), \widetilde{\lambda}_k(z)\}(g(x + \zeta_k, z + \zeta_k) - g(x, z)) \\ & + \sum_{k=1}^{R} (\lambda_k(x) - \min\{\lambda_k(x), \widetilde{\lambda}_k(z)\})(g(x + \zeta_k, z) - g(x, z)) \\ & + \sum_{k=1}^{R} (\widetilde{\lambda}_k(z) - \min\{\lambda_k(x), \widetilde{\lambda}_k(z)\})(g(x, z + \zeta_k) - g(x, z)), \end{aligned}$$

where $g : \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}$ is any bounded function with compact support.

3.2 Common random numbers

In the common random numbers (CRN) coupling, we simply simulate the embedded discrete time Markov chain for each process concurrently with the exponential holding time for each transition. The processes X and Z are then coupled by using (i) the same stream of random variables for the generation of the embedded discrete time chain, and (ii) the same stream of random variables for the exponential holding times.

More explicitly, let $\{U_i\}_{i=0}^{\infty}$ be a sequence of uniform random variables over the interval $[0,1]$, and let $\eta : \mathbb{R}^R_{\geq 0} \times [0,1] \to \{\zeta_1, \ldots, \zeta_R\}$ be defined via

$$\eta(c_1, ..., c_R, u) = \zeta_k \quad \text{if} \quad \frac{\sum_{i=1}^{k-1} c_i}{\sum_{i=1}^{R} c_i} \leq u < \frac{\sum_{i=1}^{k} c_i}{\sum_{i=1}^{R} c_i},$$

which is a categorical random variable parametrized by $c_1, ..., c_R$. Also, let us denote

$$\lambda_0(x) = \sum_{k=1}^{R} \lambda_k(x) \quad \text{and} \quad \widetilde{\lambda}_0(x) = \sum_{k=1}^{R} \widetilde{\lambda}_k(x). \tag{3.3}$$

Then for a common unit-rate poisson process Y, which will determine the exponential holding times, we consider the following system:

$$\begin{aligned}
R_X(t) &= Y\left(\int_0^t \lambda_0(X_{\text{crn}}(s))ds\right) \\
R_Z(t) &= Y\left(\int_0^t \widetilde{\lambda}_0(Z_{\text{crn}}(s))ds\right) \\
X_{\text{crn}}(t) &= X_{\text{crn}}(0) + \int_0^t \eta(\lambda_1(X_{\text{crn}}(s-)), \ldots, \lambda_R(X_{\text{crn}}(s-)), U_{R_X(s-)})dR_X(s) \\
Z_{\text{crn}}(t) &= Z_{\text{crn}}(0) + \int_0^t \eta(\widetilde{\lambda}_1(Z_{\text{crn}}(s-)), \ldots, \widetilde{\lambda}_R(Z_{\text{crn}}(s-)), U_{R_Z(s-)})dR_Z(s),
\end{aligned} \tag{3.4}$$

where we note that the processes shared not just the Poisson process Y, but also the sequence of uniform $[0,1]$ random variables $\{U_i\}_{i=0}^{\infty}$. The solution to this system exists and is unique by construction [6, 12, 13]. We note that while the representations are different, the marginal processes X_{crn} and X_{sc} have the same distribution, while the coupled processes $(X_{\text{crn}}, Z_{\text{crn}})$ and $(X_{\text{sc}}, Z_{\text{sc}})$ obviously do not.

3.3 Common reaction path coupling and the local common reaction path coupling

The common reaction path (CRP) coupling arises by simply noting that we may couple the processes (2.1) and (2.4) via the Poisson processes $\{Y_k\}$. That is, in the CRP coupling $(X_{\text{crp}}, Z_{\text{crp}})$ satisfies

$$\begin{aligned}
X_{\text{crp}}(t) &= X_{\text{crp}}(0) + \sum_{k=1}^{R} Y_k\left(\int_0^t \lambda_k(X_{\text{crp}}(s))ds\right)\zeta_k \\
Z_{\text{crp}}(t) &= Z_{\text{crp}}(0) + \sum_{k=1}^{R} Y_k\left(\int_0^t \widetilde{\lambda}_k(Z_{\text{crp}}(s))ds\right)\zeta_k,
\end{aligned} \tag{3.5}$$

where the Y_k are independent unit-rate Poisson processes.

Numerical experiments have shown that this coupling is significantly tighter than the CRN coupling, in that it produces a lower variance between the coupled processes, for many situations [2, 17, 18]. However, the variance between the processes often increases substantially as t grows. In fact, the variance of the relevant estimators oftentimes approaches that

of independent realizations of X and Z as t grows towards infinity [2, 18]. We postulate that the variance of the CRP coupling increases in this manner because of its inability to fix a "decoupling" once it occurs. To understand this heuristically, suppose that given $X_{\rm crp}(t_0)$ and $Z_{\rm crp}(t_0)$ for some $t_0 > 0$ we also have

$$\int_0^{t_0} \widetilde{\lambda}_k(Z_{\rm crp}(s))ds \ll \int_0^{t_0} \lambda_k(X_{\rm crp}(s))ds \tag{3.6}$$

for all k. Then the time and type of the next jump of $X_{\rm crp}$ is nearly uncorrelated from the time and type of the next jump of $Z_{\rm crp}$. This is true even if $X_{\rm crp}(t_0)$ and $Z_{\rm crp}(t_0)$ are very close or even equal. This problem does not occur with the split coupling since the next jump times of X and Z are always correlated via the counting processes with intensity

$$\lambda_k(X_{\rm sc}(s)) \wedge \widetilde{\lambda}_k(Z_{\rm sc}(s)).$$

The above discussion motivates us to consider the following modification to the CRP coupling. We discretize $[0, T]$ into multiple subintervals. For each such subinterval we generate the coupled processes using a new set of independent unit-rate Poisson processes and initial conditions given by the values of the processes at the terminal time of the previous subinterval. Note that if the processes $X_{\rm crp}$ and $Z_{\rm crp}$ are equal to each other at a transition between subintervals, then the processes will have recoupled. We will elaborate on this strategy. Let $\pi = \{0 = s_0 < s_1 \cdots < s_n = T\}$ be a partition of $[0, T]$. Also let $\{Y_{km} : k = 1, \ldots, R, m = 0, 1, 2, \ldots\}$ be a set of independent, unit-rate Poisson processes. Then we define the local-CRP coupling over $[0, T]$ with respect to π as the solution of

$$\begin{aligned} X^{\pi}_{\rm crp}(t) &= X(0) + \sum_{k=1}^{R}\sum_{m=0}^{\infty} Y_{km}\left(\int_{t\wedge s_m}^{t\wedge s_{m+1}} \lambda_k(X^{\pi}_{\rm crp}(s))ds\right)\zeta_k \\ Z^{\pi}_{\rm crp}(t) &= Z(0) + \sum_{k=1}^{R}\sum_{m=0}^{\infty} Y_{km}\left(\int_{t\wedge s_m}^{t\wedge s_{m+1}} \widetilde{\lambda}_k(Z^{\pi}_{\rm crp}(s))ds\right)\zeta_k. \end{aligned} \tag{3.7}$$

We remark that, irrespective of π, the marginal distribution of $X^{\pi}_{\rm crp}$ is the same as that of X, our process of interest, and the same goes for $Z^{\pi}_{\rm crp}$ and Z. Also, when π is a trivial partition with $n = 1$, the coupling (3.7) is precisely the CRP coupling of (3.5). In the next section, we will consider the limit of the the family of local-CRP couplings as $n \to \infty$ and prove that under reasonable conditions the coupled processes converge weakly to the processes coupled via the split coupling (3.2).

4 Limit of the local-CRP coupling

We begin this section by specifying some notation. First, when X and Z are stochastic processes built on the probability space $(\Omega, \mathcal{F}, P)$, we denote by $X(s, \omega)$ the process X evaluated at time s for a given choice $\omega \in \Omega$. Further, by $(X, Z)(s, \omega)$ we mean $(X(s, \omega), Z(s, \omega))$, a vector of random variables evaluated at time s. As is usual, we will often omit ω from the notation when no confusion is expected. Finally, when $\mathbf{t} = (t_1, \ldots, t_K)$ is a K dimensional vector of times points, we denote

$$X(\mathbf{t}) = [X(t_1), \ldots, X(t_K)].$$

Also, throughout the section, we assume that $X(0) = Z(0)$.

4.1 Weak convergence of finite dimensional distributions

We will first articulate what we mean by taking $n \to \infty$ in the context of the last section.

Definition. Let $\pi_n = \{0 = s_0 \le s_1 \le \cdots \le s_n = T\}$ be a partition of $[0, T]$. For $m \in \{0, \ldots, n-1\}$ let

$$\Delta_m \pi_n = s_{m+1} - s_m.$$

The mesh of π_n is defined as

$$\text{mesh}(\pi_n) \stackrel{\text{def}}{=} \max\{\Delta_m \pi_n : m \in \{0, \ldots, n-1\}\}.$$

Supposing that $\text{mesh}(\pi_n) \to 0$ as $n \to \infty$, the limit of interest to us is the weak limit of $(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})$ as $n \to \infty$. We begin with Proposition 4.1 showing the weak convergence of $X_{\text{crp}}^{\pi_n}$ to X_{sc} over finite coordinates as $n \to \infty$. In Subsection 4.2 we prove weak convergence at the process level.

Proposition 4.1. *Suppose that neither of the nominal processes X, Z are explosive and let $(X_{sc}(t), Z_{sc}(t))$ be coupled in the way of (3.2). Let*

$$\pi_n = \{0 = s_0 \le s_1 \le \cdots \le s_n = T\}$$

be a sequence of partitions such that mesh$(\pi_n) \to 0$, as $n \to \infty$, and for each n let $(X_{crp}^{\pi_n}(t), Z_{crp}^{\pi_n}(t))$ be coupled in the way of (3.7). Then for any $K \in \mathbb{Z}_{\ge 0}$ and $\mathbf{t} \in [0, T]^K$, and any bounded Lipshitz $f : (\mathbb{R}^d \times \mathbb{R}^d)^K \to \mathbb{R}$,

$$\mathbb{E}[f((X_{crp}^{\pi_n}, Z_{crp}^{\pi_n})(\mathbf{t}))] \to \mathbb{E}[f((X_{sc}, Z_{sc})(\mathbf{t}))], \quad \text{as } n \to \infty.$$

We will briefly outline the proof of 4.1. For a fixed n, let

$$\{Y_{ikm}^n; \quad i = 1, 2, 3, \quad k = 1, \ldots, R, \quad m = 0, 1, 2, \ldots\} \tag{4.1}$$

and

$$\{Y_{km}^n; \quad k = 1, \ldots, R, \quad m = 0, 1, 2, \ldots\} \tag{4.2}$$

be two sets of independent unit-rate Poisson processes. At this point, we do not make any assumption on the correlation between the processes in the set (4.1) and the processes in the set (4.2), except to note that they will not be independent. In fact, we will construct the Poisson processes of (4.1) as functions of the Poisson processes of (4.2). For now, simply consider the processes built using the Poisson processes of (4.1)

$$\begin{aligned}
X_{\text{sc}}^{\pi_n}(t) = X_{\text{sc}}(0) + \sum_{m=0}^{\infty} \sum_{k=1}^{R} \Bigg\{ & Y_{1km}^n \left(\int_{s_m \wedge t}^{s_{m+1} \wedge t} r_{1k}(\lambda_k, \widetilde{\lambda}_k, X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s) ds \right) \\
& + Y_{2km}^n \left(\int_{s_m \wedge t}^{s_{m+1} \wedge t} r_{2k}(\lambda_k, \widetilde{\lambda}_k, X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s) ds \right) \Bigg\} \zeta_k \\
Z_{\text{sc}}^{\pi_n}(t) = X_{\text{sc}}(0) + \sum_{m=0}^{\infty} \sum_{k=1}^{R} \Bigg\{ & Y_{1km}^n \left(\int_{s_m \wedge t}^{s_{m+1} \wedge t} r_{1k}(\lambda_k, \widetilde{\lambda}_k, X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s) ds \right) \\
& + Y_{3km}^n \left(\int_{s_m \wedge t}^{s_{m+1} \wedge t} r_{3k}(\lambda_k, \widetilde{\lambda}_k, X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s) ds \right) \Bigg\} \zeta_k,
\end{aligned} \tag{4.3}$$

along with

$$
\begin{aligned}
X^{\pi_n}_{\mathrm{crp}}(t) &= X_{\mathrm{crp}}(0) + \sum_{m=0}^{\infty}\sum_{k=1}^{R} Y^n_{km}\left(\int_{t\wedge s_m}^{t\wedge s_{m+1}} \lambda_k(X^{\pi_n}_{\mathrm{crp}}(s))ds\right)\zeta_k \\
Z^{\pi_n}_{\mathrm{crp}}(t) &= X_{\mathrm{crp}}(0) + \sum_{m=0}^{\infty}\sum_{k=1}^{R} Y^n_{km}\left(\int_{t\wedge s_m}^{t\wedge s_{m+1}} \widetilde{\lambda}_k(Z^{\pi_n}_{\mathrm{crp}}(s))ds\right)\zeta_k,
\end{aligned}
\tag{4.4}
$$

which are built with the Poisson processes (4.2). Note that $(X_{\mathrm{sc}}, Z_{\mathrm{sc}}) \overset{dist}{=} (X^{\pi_n}_{\mathrm{sc}}, Z^{\pi_n}_{\mathrm{sc}})$ irrespective of n. The construction we will employ will allow us to conclude that $(X^{\pi_n}_{\mathrm{sc}}, Z^{\pi_n}_{\mathrm{sc}})$ and $(X^{\pi_n}_{\mathrm{crp}}, Z^{\pi_n}_{\mathrm{crp}})$ satisfy

$$
\lim_{n\to\infty} P\left(\max_{i\in\{0,\dots,K\}} |(X^{\pi_n}_{\mathrm{sc}}(t_i), Z^{\pi_n}_{\mathrm{sc}}(t_i)) - (X^{\pi_n}_{\mathrm{crp}}(t_i), Z^{\pi_n}_{\mathrm{crp}}(t_i))| > \gamma\right) = 0 \tag{4.5}
$$

for any $\gamma > 0$. We can then appeal to a standard Portmanteau type argument to finish the proof of Proposition 4.1: let $\epsilon > 0$, and consider any bounded continuous map $f : (\mathbb{R}^d \times \mathbb{R}^d)^K \to \mathbb{R}$ with Lipshitz constant L. Then

$$
\begin{aligned}
&|\mathbb{E}f((X_{\mathrm{sc}}, Z_{\mathrm{sc}})(\mathbf{t}) - \mathbb{E}f((X^{\pi_n}_{\mathrm{crp}}, Z^{\pi_n}_{\mathrm{crp}})(\mathbf{t})| \\
&\quad = |\mathbb{E}f((X^{\pi_n}_{\mathrm{sc}}, Z^{\pi_n}_{\mathrm{sc}})(\mathbf{t}) - \mathbb{E}f((X^{\pi_n}_{\mathrm{crp}}, Z^{\pi_n}_{\mathrm{crp}})(\mathbf{t})| \\
&\quad \le L\mathbb{E}[|(X^{\pi_n}_{\mathrm{sc}}, Z^{\pi_n}_{\mathrm{sc}})(\mathbf{t}) - (X^{\pi_n}_{\mathrm{crp}}, Z^{\pi_n}_{\mathrm{crp}})(\mathbf{t})|] \\
&\quad \le LK\gamma + L\ P(\max_{i=0,\dots,K} |(X^{\pi_n}_{\mathrm{sc}}(t_i), Z^{\pi_n}_{\mathrm{sc}}(t_i)) - (X^{\pi_n}_{\mathrm{crp}}(t_i), Z^{\pi_n}_{\mathrm{crp}}(t_i))| > \gamma).
\end{aligned}
$$

We can first choose $\gamma < \epsilon/(2LK)$. With this γ fixed, we may choose n large enough so that the second piece can be bounded by $\epsilon/2$, and the claim is achieved.

We must still describe the specific construction alluded to above that will allow us to conclude (4.5). For each n, let

$$
\{Y^n_{km}, Y^{n,aug}_{ikm}, i = 1,2,3,\ k = 1,\dots,R,\ m = 0,1,2,\dots\}, \tag{4.6}
$$

be independent unit-rate Poisson processes. We generate $(X^{\pi_n}_{\mathrm{crp}}, Z^{\pi_n}_{\mathrm{crp}})$ up to time T using the processes Y^n_{km} according to (4.4). We now turn our attention to constructing the required independent unit-rate Poisson processes Y^n_{ikm}, and the coupled processes $(X^{\pi_n}_{\mathrm{sc}}, Z^{\pi_n}_{\mathrm{sc}})$ built using them according to (4.3).

Inductively arguing on m, suppose we have already generated $(X^{\pi_n}_{\mathrm{sc}}, Z^{\pi_n}_{\mathrm{sc}})$ given by (4.3) up to time $s_m \ge 0$. We further suppose that we have constructed the relevant Poisson processes $Y^n_{ik\tilde{m}}$ for all $\tilde{m} < m$. We must now describe how to construct Y^n_{ikm} for each valid pair (i,k). We define the following random times for each $i \in \{1,2,3\}$ and $k \in \{1,\dots,R\}$:

$$
\mathcal{T}_{ikm} \overset{\mathrm{def}}{=} r_{ik}(\lambda_k, \widetilde{\lambda}_k, X^{\pi_n}_{\mathrm{sc}}, Z^{\pi_n}_{\mathrm{sc}})(s_m)\cdot\Delta_m(\pi_n) \tag{4.7}
$$

and

$$
T^{\mathrm{crp}}_{km} \overset{\mathrm{def}}{=} \left(\int_{s_m}^{s_{m+1}} \lambda(X^{\pi_n}_{\mathrm{crp}}(s))ds\right) \vee \left(\int_{s_m}^{s_{m+1}} \widetilde{\lambda}(Z^{\pi_n}_{\mathrm{crp}}(s))ds\right),
$$

where, as usual, $a \vee b \stackrel{\text{def}}{=} \max\{a, b\}$, and we recall that $(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})$ has already been generated up to time T. For notational clarity we refrain from using n in the notation above for the random times. We now define Y^n_{1km} in the following manner

$$\begin{aligned} Y^n_{1km}(u) &= Y^n_{km}(u) && \text{for } \ u \le \mathcal{T}_{1km} \\ Y^n_{1km}(u) &= Y^n_{1km}(\mathcal{T}_{1km}) + Y^{n,aug}_{1km}(u - \mathcal{T}_{1km}) && \text{for } \ u > \mathcal{T}_{1km}. \end{aligned}$$

Having defined Y^n_{1km}, we turn to the construction of Y^n_{2km} and Y^n_{3km}. The construction is based on which of two of the following cases hold.

1. If $\lambda(Z^{\pi_n}_{\text{sc}}(s_m)) \le \lambda(X^{\pi_n}_{\text{sc}}(s_m))$, then let Y^n_{2km} satisfy

$$\begin{aligned} Y^n_{2km}(u) &= Y^n_{km}(u + \mathcal{T}_{1km}) - Y^n_{km}(\mathcal{T}_{1km}) && \text{for } \ u \le \mathcal{T}_{2km} \\ Y^n_{2km}(u) &= Y^n_{2km}(\mathcal{T}_{2km}) + Y^{n,aug}_{2km}(u - \mathcal{T}_{2km}) && \text{for } \ u > \mathcal{T}_{2km}, \end{aligned}$$

   and let $Y^n_{3km}(u) = Y^{n,aug}_{3km}(u)$ for all $u \ge 0$.

2. If $\lambda(Z^{\pi_n}_{\text{sc}}(s_m)) > \lambda(X^{\pi_n}_{\text{sc}}(s_m))$, then let Y^n_{3km} satisfy

$$\begin{aligned} Y^n_{3km}(u) &= Y^n_{km}(u + \mathcal{T}_{1km}) - Y^n_{km}(\mathcal{T}_{1km}) && \text{for } \ u \le \mathcal{T}_{3km} \\ Y^n_{3km}(u) &= Y^n_{3km}(\mathcal{T}_{3km}) + Y^{n,aug}_{3km}(u - \mathcal{T}_{3km}) && \text{for } \ u > \mathcal{T}_{3km}, \end{aligned}$$

   and let $Y^n_{2km}(u) = Y^{n,aug}_{2km}(u)$ for all $u \ge 0$.

Note that the strong Markov property guarantees that the processes $\{Y^n_{ikm}\}$ so constructed are independent, unit-rate Poisson processes. We then generate $(X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})$ between times s_m and s_{m+1} according to (4.3) with the processes $\{Y^n_{ikm}\}$. Note that in so doing, we have also created a coupling between $(X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})$ and $(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})$.

Note that for each i, k, and m, the value $\mathcal{T}_{ikm}$ as defined in (4.7) is an approximation to

$$T^{\text{sc}}_{ikm} \stackrel{\text{def}}{=} \int_{s_m}^{s_{m+1}} r_{ik}(\lambda_k, \widetilde{\lambda}_k, X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})(s)\, ds.$$

We would like to make a few observations about this approximation before proceeding further.

Lemma 4.2. *Fix n, and let $m \in \{0, 1, \dots\}$. If*

$$\sum_{k=1}^{R} \sum_{i=1}^{3} Y^n_{ikm}(\mathcal{T}_{ikm} \vee T^{sc}_{ikm}) = 1 \tag{4.8}$$

then there is a unique $j \in \{1, 2, 3\}$ and $\ell \in \{1, ..., R\}$ for which

$$Y^n_{j\ell m}(\mathcal{T}_{j\ell m} \wedge T^{sc}_{j\ell m}) = 1.$$

Note the difference between $\wedge$ and $\vee$ in the above statement.

Proof. For each (i,k), define

$$Q_{ik}(t) \stackrel{\text{def}}{=} Y^n_{ikm}\left(\int_{s_m}^{t+s_m} r_{ik}(\lambda_k, \widetilde{\lambda}_k, X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})(s)ds\right),$$

for $t \geq 0$. Note that (4.8) implies that $Y^n_{j\ell m}(\mathcal{T}_{j\ell m} \vee T^{\text{sc}}_{j\ell m}) = 1$ for some j and ℓ and $Y^n_{ikm}(\mathcal{T}_{ikm} \vee T^{\text{sc}}_{ikm}) = 0$ for all $(i,k) \neq (j,\ell)$. In particular, this implies $Q_{j\ell}$ is the first one among the set of counting processes $\{Q_{ik}\}$ to jump. (This follows since for all (i,k), $r_{ik}(\lambda_k, \widetilde{\lambda}_k, X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})(s)$ will not change from $r_{ik}(\lambda_k, \widetilde{\lambda}_k, X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})(s_m)$ until the first jump of $(X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})$ during $s > s_m$.) By the definitions of $\mathcal{T}_{j\ell m}$ and $T^{\text{sc}}_{j\ell m}$, it easily follows that $Y^n_{j\ell m}(\mathcal{T}_{j\ell m} \wedge T^{\text{sc}}_{j\ell m} = 1$. It is trivial that no other (i,k) pair can satisfy this relation. □

The following is an analogue to Lemma (4.2).

Lemma 4.3. *If* $(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(s_m) = (X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(s_m)$ *and*

$$\sum_k Y^n_{km}\left(\left(\sum_{i=1}^3 \mathcal{T}_{ikm}\right) \vee T^{crp}_{km}\right) = 1$$

then there is a unique j *for which*

$$Y^n_{jm}\left(\left(\sum_{i=1}^3 \mathcal{T}_{ijm}\right) \wedge T^{crp}_{jm}\right) = 1.$$

Further, the first jump time of the Poisson process Y^n_{jm} *occurs at some* t_0 *satisfying*

$$t_0 < \left(\lambda_j(X^{\pi_n}_{crp}(s_m)) \vee \widetilde{\lambda}_j(Z^{\pi_n}_{crp}(s_m))\right)\Delta_m.$$

Proof. Because the two processes are equal at time s_m, we have that

$$\sum_{i=1}^3 \mathcal{T}_{ikm} = \left(\lambda_{k_0}(X^{\pi_n}_{\text{crp}}(s_m)) \vee \widetilde{\lambda}_{k_0}(Z^{\pi_n}_{\text{crp}}(s_m))\right)\Delta_m.$$

As neither $Z^{\pi_n}_{\text{crp}}$ nor $X^{\pi_n}_{\text{crp}}$ changes until the first firing of Y_{jm}, the claim follows. □

Based on the last two observations, we have the following lemma which will be useful in proving Proposition 4.1.

Lemma 4.4. *Fix* n *and suppose that, for a given path of* $(X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(\omega)$, $(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(\omega)$ *coupled in the way we described above,*

$$H_{m,n}(\omega) \stackrel{def}{=} \sum_{k=1}^R \max\left\{\sum_{i=1}^3 Y^n_{ikm}(\mathcal{T}_{ikm} \vee T^{sc}_{ikm}), Y^n_{km}\left(\left(\sum_{i=1}^3 \mathcal{T}_{ikm}\right) \vee T^{crp}_{km}\right)\right\} \leq 1, \qquad (4.9)$$

for all m. *Then for all* $m = 0, \ldots, n$,

$$(X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(s_m, \omega) = (X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(s_m, \omega)$$

Proof. We will omit ω in the expressions. We have

$$(X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s_0) = (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(s_0)$$

by assumption. Arguing inductively, assume that

$$(X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s_m) = (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(s_m).$$

We will show that

$$(X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s_{m+1}) = (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(s_{m+1})$$

when (4.9) holds. If $H_{m,n} = 0$ for this m, then

$$(X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s_{m+1}) = (X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s_m) = (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(s_m) = (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(s_{m+1}),$$

and there is nothing to do. Therefore we consider the case in which $H_{m,n} = 1$. More specifically, suppose that for some k_0,

$$\max\left\{\sum_{i=1}^{3} Y_{ik_0m}^n(\mathcal{T}_{ik_0m} \vee T_{ik_0m}^{\text{sc}}), Y_{k_0m}^n\left(\left(\sum_{i=1}^{3}\mathcal{T}_{ik_0m}\right) \vee T_{k_0m}^{\text{crp}}\right)\right\} = 1.$$

This means that, by condition (4.9),

$$\max\left\{\sum_{i=1}^{3} Y_{ikm}^n(\mathcal{T}_{i\ell m} \vee T_{ikm}^{\text{sc}}), Y_{km}^n\left(\left(\sum_{i=1}^{3}\mathcal{T}_{ikm}\right) \vee T_{km}^{\text{crp}}\right)\right\} = 0$$

for all $k \neq k_0$. Combined with Lemmas 4.2 and 4.3, these conditions guarantee that each of the processes $X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n}, X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n}$ jump precisely one time in the time interval $[s_m, s_{m+1}]$, and the jump happens according to reaction channel k_0 (see [14] for more details). That is, we have

$$X_{\text{sc}}^{\pi_n}(s_{m+1}) = Z_{\text{sc}}^{\pi_n}(s_{m+1}) = X_{\text{crp}}^{\pi_n}(s_{m+1}) = Z_{\text{crp}}^{\pi_n}(s_{m+1}) = X_{\text{sc}}^{\pi_n}(s_m) + \zeta_{k_0},$$

and we are done. □

It is not too difficult to see that if λ_k and $\widetilde{\lambda}_k$ are uniformly bounded for all k, then we can make the condition in Lemma 4.4 hold with a probability greater than $1 - \epsilon$ for any $\epsilon > 0$ by setting $\text{mesh}(\pi_n)$ small enough. Of course, we do not have such a uniform bound on the intensity functions. Also, note that Lemma 4.4 does **not** imply that

$$(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(t) = (X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(t) \text{ for } t \in [s_m, s_{m+1}],$$

even if the conditions of the lemma are met, as the processes may (and most likely will) jump at slightly different times. However, we trivially note that under the conditions of Lemma 4.4,

$$(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(t) = (X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(t) \text{ for all } t \in [s_m, s_{m+1}] \tag{4.10}$$

if neither $(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})$ nor $(X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})$ jump at all in $[s_m, s_{m+1}]$.

We are now in a position to prove Proposition 4.1.

Proof of Proposition 4.1. We first recall that $\mathbf{t} = (t_1, \dots, t_K)$ for some $K \in \{1, 2, \dots\}$. Next, we define

$$K_0^n \stackrel{\text{def}}{=} \{m \in \{0, ..., n-1\} \; ; \; \{t_j\}_{j=1}^K \cap [s_m, s_{m+1}) \neq \emptyset\}.$$

Fix $\epsilon > 0$. As we remarked around (4.5), it suffices to show that, for large enough n,

$$P\left(\max_{i=0,\dots,K} |(X_{\text{sc}}^{\pi_n}(t_i), Z_{\text{sc}}^{\pi_n}(t_i)) - (X_{\text{crp}}^{\pi_n}(t_i), Z_{\text{crp}}^{\pi_n}(t_i))| > 0 \right) < \epsilon,$$

where we converted the γ in (4.5) to a zero as our processes take values in $\mathbb{Z}^d$.

We will resort to a localization argument and take advantage of the fact that X and Z are both nonexplosive. Let $M > 0$, and let $H_{m,n}$ be defined as in Lemma 4.4. Define

$$A_n(\mathbf{t}) \stackrel{\text{def}}{=} \{\omega : H_{m,n}(\omega) \leq 1 \text{ if } m \notin K_0^n \text{ and } H_{m,n}(\omega) = 0 \text{ if } m \in K_0^n\}, \tag{4.11}$$

and

$$B_{M,n} \stackrel{\text{def}}{=} \{\omega : \max\{\sup_{s \leq T} \lambda_k(X_{\text{sc}}^{\pi_n}(s)), \sup_{s \leq T} \widetilde{\lambda}_k(Z_{\text{sc}}^{\pi_n}(s)), \sup_{s \leq T} \lambda_k(X_{\text{crp}}^{\pi_n}(s)), \sup_{s \leq T} \widetilde{\lambda}_k(Z_{\text{crp}}^{\pi_n}(s)\} \leq M\}. \tag{4.12}$$

Note that by the non-explosivity of the processes, the supremums are achieved everywhere they appear above. By Lemma 4.4 and the arguments in and around (4.10), we have that

$$A_n(\mathbf{t}) \subset \{(X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(\mathbf{t}) = (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(\mathbf{t})\}.$$

Therefore

$$\begin{aligned} P((X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(\mathbf{t}) \neq (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(\mathbf{t})) &\leq P(A_n^C(\mathbf{t})) \\ &= P(A_n^C(\mathbf{t}) \cap B_{M,n}) + P(A_n^C(\mathbf{t}) \cap B_{M,n}^C). \end{aligned} \tag{4.13}$$

We handle the two pieces on the right hand side of (4.13) separately.

For the second term in (4.13), we first note that

$$\begin{aligned} B_{M,n}^C \subset \{\sup_{s \leq T} \lambda_k(X_{\text{sc}}^{\pi_n}(s)) > M\} \cup \{\sup_{s \leq T} \widetilde{\lambda}_k(Z_{\text{sc}}^{\pi_n}(s)) > M\} \cup \\ \{\sup_{s \leq T} \lambda_k(X_{\text{crp}}^{\pi_n}(s)) > M\} \cup \{\sup_{s \leq T} \widetilde{\lambda}_k(Z_{\text{crp}}^{\pi_n}(s)) > M\}. \end{aligned}$$

Now, recall that the marginal distributions of $X_{\text{crp}}^{\pi_n}$ and $X_{\text{sc}}^{\pi_n}$ are the same as the marginal distribution of X, and that the same goes for $Z_{\text{crp}}^{\pi_n}$ and $Z_{\text{sc}}^{\pi_n}$ compared with Z. Therefore, for all n we have

$$P(B_{M,n}^C) \leq 2 \times \left[P(\sup_{s \leq T}\{\lambda_k(X_s)\} > M) + P(\sup_{s \leq T}\{\widetilde{\lambda}_k(Z_s)\} > M) \right]. \tag{4.14}$$

By the monotone convergence theorem and the fact that the processes are all non explosive, the right hand side of (4.14) will tend to 0 as $M \to \infty$. Therefore, we can take M large enough so that the second piece of (4.13) is smaller than $\epsilon/2$. We fix this M, and turn attention to the first term on the right hand side of (4.13).

We consider the localized version of H. In particular, for our fixed $M > 0$ let

$$H^M_{m,n}(\omega) \stackrel{\text{def}}{=} \sum_{k=1}^{R} \max\left\{ \sum_{i=1}^{3} Y^n_{ikm}(M\Delta_m(\pi_n)), Y^n_{km}\left(3M\Delta_m(\pi_n)\right) \right\}.$$

Then it is clear that, for any $q > 0$,

$$\{\{H_{m,n} > q\} \cap B_{M,n}\} \subset \left\{\{H^M_{m,n} > q\} \cap B_{M,n}\right\} \subset \{H^M_{m,n} > q\}$$

and therefore

$$\begin{aligned} &P(A^C_n(\mathbf{t}) \cap B_{M,n}) \\ &\quad \leq P(H^M_{m,n} > 1 \text{ for some } m \notin K^n_0 \ \ \textbf{OR} \ \ H^M_{m,n} > 0 \text{ for some } m \in K^n_0) \\ &\quad \leq \sum_{m \notin K^n_0} P(H^M_{m,n} > 1) + \sum_{m \in K^n_0} P(H^M_{m,n} > 0). \end{aligned} \tag{4.15}$$

To handle these two pieces, we recall two basic facts pertaining to Poisson random variables. First, if we denote by $W(\Lambda) \sim \text{Poisson}(\Lambda)$ then

$$\begin{aligned} P(W(\Lambda) > 1) =& 1 - \exp(-\Lambda)(1 + \Lambda) \\ \leq& 1 - (1 - \Lambda)(1 + \Lambda) \\ =& \Lambda^2, \end{aligned}$$

where we used the inequality $\exp(-x) \geq 1 - x$. Second, and using the same inequality,

$$P(W(\Lambda) > 0) = 1 - \exp(-\Lambda) \leq \Lambda.$$

Now note that

$$P(\{H^M_{m,n} > q\}) \leq P(W(6RM\Delta_m(\pi_n)) > q).$$

Hence, if $\text{mesh}(\pi_n) = \delta_n$ then by the two facts above and (4.15), we have

$$\begin{aligned} P(A^C_n(\mathbf{t}) \cap B_M) &\leq \sum_{m \notin K^n_0} (6RM\Delta_m(\pi_n))^2 + \sum_{m \in K^n_0} (6RM\Delta_m(\pi_n)) \\ &\leq (6RM)^2 \delta_n \sum_{m \notin K^n_0} \frac{\Delta_m(\pi_n)}{\delta_n} \Delta_m(\pi_n) + 6RM|K^n_0|\delta_n \\ &\leq (6RM)^2 \delta_n T + 6RM|K^n_0|\delta_n, \end{aligned} \tag{4.16}$$

where in the third inequality we used that $\frac{\Delta_m(\pi_n)}{\delta_n} < 1$, which follows by the definition of mesh. We can now take n large enough so that (4.16) is less than $\epsilon/2$. Collecting the above, we may now conclude that for such n,

$$P(|(X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})(\mathbf{t}) - (X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})(\mathbf{t})| > 0) < \epsilon,$$

as required. □

The following is an immediate corollary to Proposition 4.1.

Corollary 4.5. *Let $s = \{s_0 < s_1 < s_2 < \cdots < s_{m_1}\}$ and $t = \{t_0 < t_1 < t_2 < \cdots < t_{m_2}\}$. Let $f_i : \mathbb{R}^d \to \mathbb{R}$, $i = 0, \ldots, m_1$, and $g_j : \mathbb{R}^d \to \mathbb{R}$, $j = 0, \ldots, m_2$, be bounded and continuous functions on $\mathbb{R}^d$, and assume the conditions set forth in Proposition 4.1. Then*

$$\mathbb{E}\left[\prod_{i=0}^{m_1} f_i((X_{crp}^{\pi_n}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{crp}^{\pi_n}(t_j)))\right] \to \mathbb{E}\left[\prod_{i=0}^{m_1} f_i((X_{sc}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{sc}(t_j)))\right], \text{ as } n \to \infty.$$

Of course, we hope that Proposition 4.1 together with Corollary 4.5 imply the weak convergence of $(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})$ to $(X_{\text{sc}}, Z_{\text{sc}})$ at the process level. Since it is natural to view $(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n}) \in \mathbb{R}^{2d}$, we would ideally like to show that $(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n}) \implies (X_{\text{sc}}, Z_{\text{sc}})$ weakly as stochastic processes on $\mathbb{R}^{2d}$. For such convergence to hold we require the laws of $\{(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})\}$ to be relatively compact (i.e. every sequence has a convergent subsequence) Unfortunately, and perhaps surprisingly, this is not the case as we now show.

The following result is Theorem 7.2 on page 128 of [10]. Following the notation in [10], when E is a metric space we let $D_E[0, \infty)$ be the set of all càdlàg functions from $[0, \infty)$ to E.

Theorem 4.6. *Let (E, r) be a complete and separable metric space, and let $\{X_n\}$ be a family of processes with sample paths in $D_E[0, \infty)$ endowed with the Skorohod metric. Then $\{X_n\}$ is relatively compact if and only if the following two conditions hold:*

1. *For each $\eta > 0$ and rational $t \geq 0$, there is exists a compact set $\Gamma_{\eta,t} \subset E$ such that*

$$\inf_n P(X_n(t) \in \Gamma_{\eta,t}) \geq 1 - \eta.$$

2. *For every $\eta > 0$ and $T > 0$, there exists $\delta > 0$ such that*

$$\sup_n P(w'(X_n, \delta, T) \geq \eta) < \eta$$

where

$$w'(X, \delta, T) \stackrel{def}{=} \inf_{\pi} \max_i \sup_{a,b \in [t_i, t_{i+1})} |X(a) - X(b)|$$

where π ranges over all partitions of $[0, T]$ satisfying $t_{i+1} - t_i > \delta$ for all $i \geq 0$.

Unfortunately the conditions of Theorem 4.6 do not hold in general for our set of processes $\{(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})\}$ over the skorohod space $D_{\mathbb{R}^{2d}}[0, \infty)$. To see this, we note the following two facts:

1. For jump processes whose jump sizes are bounded below, for example by integer values in our present setting, for small enough $\eta > 0$ we have

$$\{w'((X, Z), \delta, T) < \eta\} = \{w'((X, Z), \delta, T) = 0\},$$

2. The event $w'((X, Z), \delta, T) = 0$ can be achieved if and only if the minimum time between jumps of (X, Z) is greater than δ.

To understand the second statement, simply note that if the minimum time between jumps is less than δ, then for any partition π satisfying $t_{i+1} - t_i > \delta$ for all i, the process must change by at least the smallest jump size ($\min_k |\zeta_k|$ in our case) in *some* interval of the partition. Conversely, if the minimum holding time of the process is greater than δ, then we achieve a value of 0 for w' by choosing π so that the jump times correspond with a subset of the partition times t_i.

The following example explicitly shows that Theorem 4.6 does not hold for our choice of $\{(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})\}$ with $E = \mathbb{R}^{2d}$. Essentially the same argument would work for any model considered in this paper.

Example 1. *Consider the chemical reaction network*

$$A \to 2A,$$

which models increases in A as a counting process with a linear intensity (i.e. a linear birth process). We consider the corresponding coupled processes (X_{sc}, Z_{sc}) and $(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})$ with

$$\lambda_1(x) = \theta x, \qquad \widetilde{\lambda}_1(x) = (\theta + h)x,$$

and initial condition

$$X_{sc}(0) = Z_{sc}(0) = X^{\pi_n}_{crp}(0) = Z^{\pi_n}_{crp}(0) > 0.$$

For any $\delta > 0$, the probability that the processes X_{sc} and Z_{sc} jump simultaneously in the time period $[0, \delta]$ and that their simultaneous jump is the first jump for both processes is

$$\alpha_\delta \stackrel{def}{=} \frac{\theta}{\theta + h}\left(1 - e^{-(\theta+h)X(0)\delta}\right) > 0.$$

By the arguments we made in the proof above, for any $\epsilon > 0$ there exists some M_ϵ such that if $n > M_\epsilon$, then with probability greater than $\alpha_\delta - \epsilon$, both $X^{\pi_n}_{crp}$ and $Z^{\pi_n}_{crp}$ will also make a first jump in $[0, \delta]$. However, with a probability of one, $X^{\pi_n}_{crp}$ and $Z^{\pi_n}_{crp}$ jump at different times. Hence, when they jump in the time interval $[0, \delta)$, we have

$$\sup_{a,b\in[0,\delta)} |(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(a) - (X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(b)| \geq 1.$$

This in particular means that for any $0 < \eta < 1$,

$$\sup_n P(w'((X^{\pi_n}_{crp}, Z^{\pi_n}_{crp}), \delta, T) \geq \eta) \geq \alpha_\delta,$$

and the laws of $\{(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})\}$ fail to be relatively compact.

4.2 Weak Convergence in the product Skorohod topology

Example 1 demonstrates that the measures induced by $(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})$ on $D_{\mathbb{R}^{2d}}[0, \infty)$ are not relatively compact. Hence, the processes $(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})$ do not converge weakly to $(X_{\text{sc}}, Z_{\text{sc}})$ in $D_{\mathbb{R}^{2d}}[0, \infty)$. However, in this section we demonstrate that there is convergence in

$$\mathcal{D} := D_{\mathbb{R}^d}[0, \infty) \times D_{\mathbb{R}^d}[0, \infty)$$

endowed with the product Skorohod topology.

As is usual, the main work that remains to be done is in showing that $\{(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})\}$ is relatively compact in the appropriate topological space.

Proposition 4.7. *Let $\mathcal{D} \stackrel{def}{=} D_{\mathbb{R}^d}[0,\infty) \times D_{\mathbb{R}^d}[0,\infty)$, with the product Skorohod topology. The family of processes $\{(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})\}$ is relatively compact in $\mathcal{D}$.*

Proof. By Theorem 2.2 on page 104 of [10], it is enough to show that for any $\epsilon > 0$, there exists a compact set $C^\epsilon \in \mathcal{D}$ such that

$$\inf_n P((X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}}) \in C^\epsilon) > 1 - \epsilon.$$

To show this, we consider the marginal processes, which we recall satisfy $X \sim X^{\pi_n}_{\text{crp}}$ and $Z \sim Z^{\pi_n}_{\text{crp}}$ for each $n \geq 1$. Note that if $A^\epsilon, B^\epsilon \subset D_{\mathbb{R}^d}[0,\infty)$ are compact, then the inequalities

$$\begin{aligned} P(X \in A^\epsilon) = P(X^{\pi_n}_{\text{crp}} \in A^\epsilon) > 1 - \frac{\epsilon}{2} \\ P(Z \in B^\epsilon) = P(Z^{\pi_n}_{\text{crp}} \in B^\epsilon) > 1 - \frac{\epsilon}{2} \end{aligned} \tag{4.17}$$

imply the inequality

$$P((X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}}) \in A^\epsilon \times B^\epsilon) > 1 - \epsilon,$$

with $A^\epsilon \times B^\epsilon$ compact in $\mathcal{D}$. Hence, it is sufficient to simply prove the pair of inequalities (4.17) for the marginal processes, which live in $D_{\mathbb{R}^d}[0,\infty)$. However, inequality (4.17) holds so long as the marginal processes are tight (in $D_{\mathbb{R}^d}[0,\infty)$), and so Theorem 4.6 may be used. Therefore it suffices to show that X and Z both separately satisfy the conditions in Theorem 4.6, which we do now.

Since X is a nonexplosive pure jump process, it clearly passes the first condition of Theorem 4.6. Also, recall that X is constructed with $R \in \mathbb{Z}_{>0}$ Poisson processes, one for each jump direction. Then for any $T > 0$ and $M > 0$,

$$\begin{aligned} P(w'(X,\delta,T) > 0) &\leq P\left(w'(X,\delta,T) > 0, \sup_{k=1,..,R, s<T} \lambda_k(X(s)) \leq M\right) \\ &\quad + P\left(\sup_{k=1,..,R,s<T} \lambda_k(X(s)) > M\right) \\ &\leq P\left(w'(Y(MR\cdot),\delta,T) > 0\right) + P\left(\sup_{k=1,..,R,s<T} \lambda_k(X(s)) > M\right) \end{aligned} \tag{4.18}$$

where $Y(MR\cdot)$ is a Poisson process with rate MR. Since X is non explosive, we may take M large enough to control the second piece, and for this M we can choose δ small enough to control the first piece. That is, $\lim_{\delta\to 0} P(w'(X,\delta,T) > 0) = 0$. This tells us that X also passes the second condition of Theorem 4.6. The same procedure works for Z. Thus, $\{(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})\}$ is relatively compact in $\mathcal{D}$ with the product topology. □

With this proposition at our hand, we can prove the main result of our paper.

Theorem 4.8. *Suppose X and Z are both non-explosive, càdlàg process as given above. Let $D_{\mathbb{R}^d}[0,\infty)$ be the Skorohod Space as defined in [10]. Consider the product topology on*

$$\mathcal{D} := D_{\mathbb{R}^d}[0,\infty) \times D_{\mathbb{R}^d}[0,\infty).$$

Also, let $\pi_n = \{s^n_j\}$ be a sequence of partitions of $[0,\infty)$ such that

$$mesh(\pi_n) = \max_{j<\infty}(s^n_j - s^n_{j-1}) \to 0, \quad \text{as } n \to \infty.$$

Then for all $f : \mathcal{D} \to \mathbb{R}$ that are bounded and continuous,

$$\mathbb{E}[f(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})] \to \mathbb{E}[f(X_{sc}, Z_{sc})], \quad \text{as } n \to \infty.$$

That is, $(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp}) \to (X_{sc}, Z_{sc})$, as $n \to \infty$, weakly in the product Skorohod topology.

We would like to emphasize that the test function f considered above maps a path in $\mathcal{D}$ to $\mathbb{R}$. The test functions for Proposition 4.1, on the other hand, are evaluated at discrete time points.

Now we put everything together to prove Theorem 4.8.

Proof of Theorem 4.8. By Proposition 4.7 it is sufficient to show that every convergent (in distribution) subsequence of $(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})$ converges in distribution to $(X_{\text{sc}}, Z_{\text{sc}})$. By Corollary 4.5, it is sufficient to show that if

$$\mathbb{E}\left[\prod_{i=0}^{m_1} f_i(X_{\text{sc}}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{\text{sc}}(t_j))\right] = \mathbb{E}\left[\prod_{i=0}^{m_1} f_i(X^*(s_i)) \prod_{j=0}^{m_2} g_j(Z^*(t_j))\right] \tag{4.19}$$

for all $\{s_i\}, \{t_j\} \subset [0, \infty)$, and $f_i, g_i \in \overline{C}(\mathbb{R}^d)$ (bounded and continuous functions), then $\mathbb{E}[h(X_{sc}, Z_{sc})] = \mathbb{E}[h(X^*, Z^*)]$ for any bounded and continuous function $h : \mathcal{D} \to \mathbb{R}$. A standard monotone class argument (for example, see page 132 in [10]) shows that 4.19 is more than enough to guarantee that $\mathbb{E}[h(X_{\text{sc}}, Z_{\text{sc}})] = \mathbb{E}[h(X^*, Z^*)]$ for all h continuous with respect to $D_{\mathbb{R}^{2d}}[0, \infty)$. From the definition of the Skorohod metric, it is straightforward to show that the topology of $D_{\mathbb{R}^{2d}}[0, \infty)$ is finer than that of $\mathcal{D}$. This in particular means that the continuous functions with respect to $\mathcal{D}$ are a subset of those of $D_{\mathbb{R}^{2d}}[0, \infty)$. Thus, we may conclude that $\mathbb{E}[h(X_{\text{sc}}, Z_{\text{sc}})] = \mathbb{E}[h(X^*, Z^*)]$ if h is continuous with respect to $\mathcal{D}$, and the result is shown. □

While the results presented so far pertain to the specific couplings found in the numerical analysis literature, a slightly more general theorem can be achieved by following an identical line of reasoning.

Theorem 4.9. *For $i \in \{1, 2, 3\}$ and $k \in \{1, \ldots, R\}$, let $r_{ik} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be a non-negative measurable function. Suppose that $\{\pi_n\}$ is a sequence of partitions of $[0, \infty)$ for which mesh$(\pi_n) \to 0$, as $n \to \infty$. Define (X_{sc}, Z_{sc}) and $(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})$ via*

$$X_{sc}(t) = X(0) + \sum_{k=1}^{R} \left\{ Y_{1k}\left(\int_0^t r_{1k}(X_{sc}, Z_{sc})(s)ds\right) + Y_{2k}\left(\int_0^t r_{2k}(X_{sc}, Z_{sc})(s)ds\right) \right\} \zeta_k$$

$$Z_{sc}(t) = Z(0) + \sum_{k=1}^{R} \left\{ Y_{1k}\left(\int_0^t r_{1k}(X_{sc}, Z_{sc})(s)ds\right) + Y_{3k}\left(\int_0^t r_{3k}(X_{sc}, Z_{sc})(s)ds\right) \right\} \zeta_k,$$

and

$$X^{\pi_n}_{crp}(t) = X(0) + \sum_{m=0}^{\infty} \sum_{k=1}^{R} Y^n_{km}\left(\int_{t \wedge s_m}^{t \wedge s_{m+1}} \{r_{1k}(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(s) + r_{2k}(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(s)\}ds\right) \zeta_k$$

$$Z^{\pi_n}_{crp}(t) = Z(0) + \sum_{m=0}^{\infty} \sum_{k=1}^{R} Y^n_{km}\left(\int_{t \wedge s_m}^{t \wedge s_{m+1}} \{r_{1k}(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(s) + r_{3k}(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(s)\}ds\right) \zeta_k,$$

where all notation is as before. Finally, we suppose that all processes are non-explosive. Then, $(X^{\pi_n}_{crp}, Z^{\pi_n}_{crp}) \to (X_{sc}, Z_{sc})$, as $n \to \infty$, weakly in the product Skorohod topology.

Proposition 4.8 is therefore a special case of Theorem 4.9 in which each r_{ik} depends on λ_k and $\tilde{\lambda}_k$ in a specific way.

5 Numerical examples

In this section, we provide two numerical examples demonstrating the convergence of the local-CRP coupling to that of the split coupling. Based upon our motivation in terms of variance reduction, we focus upon the convergence of the variance between the coupled processes.

Example 2. *We begin by considering a basic model of gene transcription and translation, where the model tracks the counts for the numbers of genes (G), mRNA molecules (M), and proteins (P) in the system. We suppose that the system can undergo the following possible reactions,*

$$G \to G + M \tag{R1}$$
$$M \to M + P \tag{R2}$$
$$M \to \emptyset \tag{R3}$$
$$P \to \emptyset, \tag{R4}$$

where, for example, reaction (R1) implies a net change to the system of one extra mRNA molecule. Since no reaction changes the number of genes present in the system, we may take that to be a fixed quantity. Hence, there are two dynamic components, and the stochastic model for this system is

$$\begin{aligned} X(t) = X(0) &+ Y_1\left(\int_0^t \lambda_1(X(s))ds\right)\begin{bmatrix}1\\0\end{bmatrix} + Y_2\left(\int_0^t \lambda_2(X(s))ds\right)\begin{bmatrix}0\\1\end{bmatrix} \\ &+ Y_3\left(\int_0^t \lambda_3(X(s))ds\right)\begin{bmatrix}-1\\0\end{bmatrix} + Y_4\left(\int_0^t \lambda_4(X(s))ds\right)\begin{bmatrix}0\\-1\end{bmatrix}, \end{aligned}$$

where X_1 counts the numbers of mRNA molecules, and X_2 counts the numbers of proteins. We now let X be the process with intensity functions

$$\lambda_1(x) = 2, \quad \lambda_2(x) = 10x_1, \quad \lambda_3(x) = (1/4 + 1/80)x_1, \quad \lambda_4(x) = x_2,$$

and let Z be the process with intensity functions

$$\lambda_1(x) = 2, \quad \lambda_2(x) = 10x_1, \quad \lambda_3(x) = (1/4 - 1/80)x_1, \quad \lambda_4(x) = x_2.$$

These are reasonable choices, for example, if we were attempting to estimate the sensitivity of some statistic with respect to the rate parameter for the third intensity function evaluated at $1/4$.

Let π_n be a partition of $[0, 30]$ into n equally sized intervals. In Figure 1, we plot numerical estimates of $Var(X_{sc}(t) - Z_{sc}(t))$, $Var(X_{crp}(t) - Z_{crp}(t))$, and $Var(X^{\pi_n}_{crp}(t) - Z^{\pi_n}_{crp}(t))$, for $n \in \{2, 6, 30, 300\}$, over the time period $[0, 30]$. The estimates were achieved via Monte Carlo methods with 10,000 sample paths. We observe the uniform convergence of $Var(X^{\pi_n}_{crp}(\cdot) - Z^{\pi_n}_{crp}(\cdot))$ to $Var(X_{sc}(\cdot) - Z_{sc}(\cdot))$ as $mesh(\pi_n) \to 0$. We also observe a sharp drop in the variance of $X^{\pi_n}_{crp}(\cdot) - Z^{\pi_n}_{crp}(\cdot)$ at the "resetting" of the Poisson processes, which occur at the end of each interval of the discretization π_n.

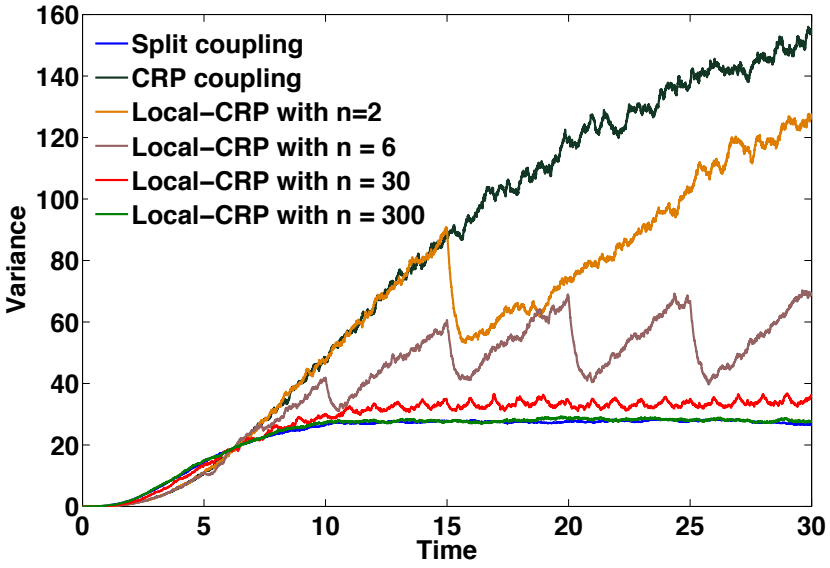

Figure 1: Numerical approximations (via Monte Carlo with 10,000 sample paths) for the variance of the difference between the processes X and Z of Example 2 for the split coupling (blue), CRP coupling (black), and various local-CRP couplings. Convergence of the variance of the local-CRP coupling to the variance of the split coupling is clear.

Example 3. *Consider a simple quadratic birth and death model*

$$\emptyset \to 2A \tag{r1}$$
$$2A \to \emptyset \tag{r2}$$

with initial count $X(0)$ given by a Poisson random variable with parameter 15. We can model the dynamics of this system with the stochastic equations

$$X(t) = X(0) + 2Y_1\left(\int_0^t \lambda_1(X(s))ds\right) - 2Y_2\left(\int_0^t \lambda_2(X(s))ds\right),$$

where

$$\lambda_1(x) = 400, \quad \text{and} \quad \lambda_2(x) = kx(x-1),$$

and where k is a parameter of the model. We consider the model X with $k = 0.1 + 1/25$ and the model Z with $k = 0.1 - 1/25$. Further, we let the initial conditions of X and Z be independent Poisson random variables with a parameter of 15 (that is, the initial conditions of X and Z are independent from each other). Let π_n be a partition of $[0,1]$ into n equally sized intervals. In Figure 3, we plot numerical estimates of $Var(X_{sc}(t) - Z_{sc}(t))$, $Var(X_{crp}(t) - Z_{crp}(t))$, and $Var(X^{\pi_n}_{crp}(t) - Z^{\pi_n}_{crp}(t))$, for $n \in \{2, 4, 8, 100\}$, over the time period $[0,1]$. The estimates were achieved via Monte Carlo methods with 5,000 sample paths. We again observe the sharp drop in variance at the "resetting" times of the processes.

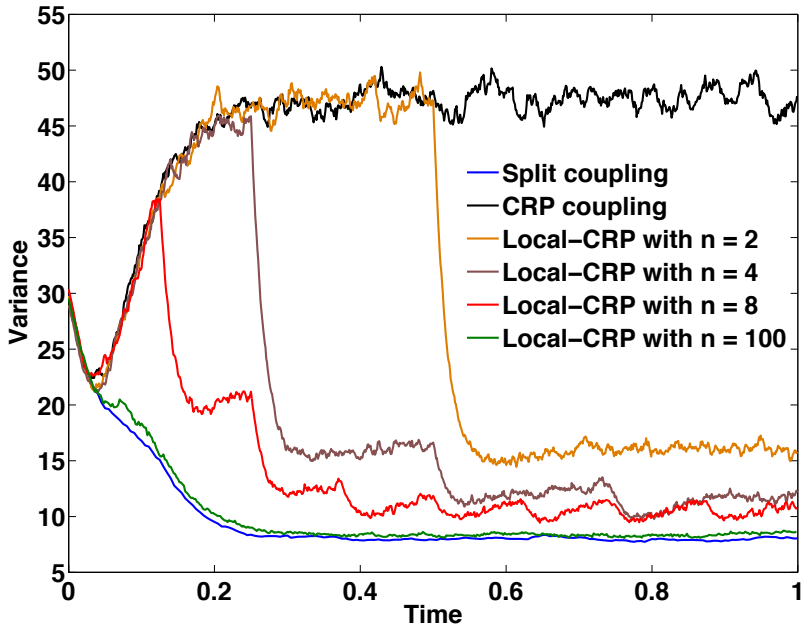

Figure 2: Numerical approximations (via Monte Carlo with 5,000 sample paths) for the variance of the difference between the processes X and Z of Example 3 for the split coupling (blue), CRP coupling (black), and various local-CRP couplings. Convergence of the variance of the local-CRP coupling to the variance of the split coupling is clear.

6 Discussion

The stochastic models finding widespread use in the cell biology literature are typically immensely complicated, and computational methods often provide the only effective way to probe the dynamics. As Persi Diaconis recently noted [9], this presents mathematicians with an opportunity to make contributions by explicitly studying the different simulation and computational algorithms themselves. Such analyses will not only shed light on which methods to use in different contexts, but will inevitably lead to a deeper understanding of the underlying processes, and hence to better computational methods.

In this work we have clarified the connection between two couplings commonly found in the computational cell biology literature and, in particular, showed that the split coupling can be regarded as a natural limit of a localized version of the CRP coupling. There are other interesting ways to understand the split coupling. For example, Arampatzis and Katsoulakis [8] recently studied a group of couplings that is included in the family of general split couplings considered in Theorem 4.9. They note that for each test function f there is an optimal choice for the function $r_{1k}(\lambda_k, \widetilde{\lambda}_k, \mathcal{U}, \mathcal{V})(s)$ in (3.1) that minimizes the variance of the finite difference $\mathbb{E}[f(X_t) - f(Z_t)]$ in the setting of (3.2). When the test function is $f(x) = x$, the correct choice of r_{1k} is the one given in (3.1), which yields the split coupling.

Acknowledgments. We thank Thomas Kurtz for several illuminating discussions and for reading an early version of this work. We thank two anonymous referees for careful

readings that substantially improved the clarity of Section 4.2. Anderson was supported by NSF grants DMS-1009275 and DMS-1318832 and Army Research Office grant W911NF-14-1-0401. Koyama was supported by NSF grants DMS-1009275, DMS-1318832, and DMS-0805793.

References

[1] David F. Anderson, *A modified next reaction method for simulating chemical systems with time dependent propensities and delays*, J. Chem. Phys. **127** (2007), no. 21, 214107.

[2] _____, *An efficient finite difference method for parameter sensitivities of continuous time Markov chains*, SIAM Journal on Numerical Analysis **50** (2012), no. 5, 2237 – 2258.

[3] David F. Anderson, Arnab Ganguly, and Thomas G. Kurtz, *Error analysis of tau-leap simulation methods*, Annals of Applied Probability **21** (2011), no. 6, 2226 – 2262.

[4] David F. Anderson and Desmond J. Higham, *Multi-level Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics*, SIAM: Multiscale Modeling and Simulation **10** (2012), no. 1, 146 – 179.

[5] David F. Anderson, Desmond J. Higham, and Yu Sun, *Complexity analysis of multilevel Monte Carlo tau-leaping*, Submitted. Arxiv: 1310.2676, 2014.

[6] David F. Anderson and Thomas G. Kurtz, *Continuous time markov chain models for chemical reaction networks*, Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology (H. Koeppl et al., ed.), Springer, 2011, pp. 3–42.

[7] David F. Anderson and Elizabeth Skubak Wolf, *A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks*, J. Chem. Phys. **137** (2012), no. 22, 224112.

[8] Georgios Arampatzis and Markos A. Katsoulakis, *Goal-oriented sensitivity analysis for lattice kinetic Monte Carlo simulations*, J. Chem. Phys. **140** (2014), 124108.

[9] Persi Diaconis, *Some things we've learned (about markov chain monte carlo)*, Bernoulli **19** (2013), no. 4, 1294 – 1305.

[10] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes: Characterization and convergence*, John Wiley & Sons, New York, 1986.

[11] Mike B. Giles, *Multilevel Monte Carlo path simulation*, Operations Research **56** (2008), 607–617.

[12] Daniel T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys. **22** (1976), 403–434.

[13] _____, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. **81** (1977), no. 25, 2340–2361.

[14] Masanori Koyama, *Analysis of stochastically modeled biochemical processes with applications to numerical methods*, Ph.D. thesis, University of Wisconsin at Madison, December, 2013.

[15] Thomas G. Kurtz, *Approximation of population processes*, CBMS-NSF Reg. Conf. Series in Appl. Math.: 36, SIAM, 1981.

[16] _____, *Representation and approximation of counting processes*, Lecture Notes in Control and Information Science **42** (1982), 177–191.

[17] Muruhan Rathinam, Patrick W. Sheppard, and Mustafa Khammash, *Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks*, Journal of Chemical Physics **132** (2010), 034103.

[18] Rishi Srivastava, David F. Anderson, and James B. Rawlings, *Comparison of finite difference based methods to obtain sensitivities of stochastic chemical kinetic models*, Journal of Chemical Physics **138** (2013), no. 7, 074110.