Constrained Approximation of Effective Generators for Multiscale Stochastic Reaction Networks and Application to Conditioned Path Sampling

Cotter, Simon

2015

MIMS EPrint: **2015.81**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

Constrained Approximation of Effective Generators for Multiscale Stochastic Reaction Networks and Application to Conditioned Path Sampling

Simon L. Cotter

School of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom; e-mail: simon.cotter@manchester.ac.uk. *SC was funded by First Grant Award EP/L023989/1 from EPSRC.*

Abstract

Efficient analysis and simulation of multiscale systems of chemical kinetics is an ongoing area for research, and is the source of many theoretical and computational challenges. In this paper, we present a significant improvement to the constrained approach, which allows us to compute the effective generator of the slow variables, without the need for expensive stochastic simulations. This is done through finding the null space of the generator of the constrained system. For complex systems where this is not possible, the constrained approach can then be applied in turn to the constrained system in a nested manner, meaning that the problem can be broken down into solving many small eigenvalue problems. Moreover, this methodology does not rely on the quasi steady-state assumption, meaning that the effective dynamics that are approximated are highly accurate, and in the case of systems with only monomolecular reactions, are exact. We will demonstrate this with some numerics, and also use the effective generators to sample paths which are conditioned on their endpoints.

Keywords: Stochastic, multiscale, chemical kinetics, constrained dynamics

1. Introduction

Understanding of the biochemical reactions that govern cell function and regulation is key to a whole range of biomedical and biological applications and understanding mathematical modelling of gene regulatory networks has been an area of huge expansion over the last half century. Due to the low copy numbers of some chemical species within the cell, the random and sporadic nature of individual reactions can play a key part in the dynamics of the system, which cannot be well approximated by ODEs[7]. Methods for the simulation of such a system, such as Gillespie's stochastic simulation algorithm (SSA)[11] have been around for some decades. Versions which are more computationally efficient have also been developed in the intermediate years[10, 3].

Unfortunately, their application to certain systems can be computationally intractable. The algorithms simulate every single reaction individually. If the system is multiscale, i.e. there are some reactions (fast reactions) which are happening many times on a timescale for which others (slow reactions) are unlikely to happen at all, then in order for us to understand the occurrences of the slow reactions, an unfeasible number of fast reactions must be simulated. This is the motivation for numerical methods which allow us to approximate the dynamics of the slowly changing quantities in the system, without the need of simulating all of the fast reactions.

For systems which are assumed to be well-mixed, there are many different approaches and methods which have been developed. For example the τ-leap method[13] speeds up the simulation by timestepping by an increment within which several reactions may occur. This can lead to problems when the copy numbers of one or more of the species approaches zero, and a number of different methods for overcoming this have been presented[20, 1].

Several other methods are based on the quasi steady-state assumption (QSSA). This is the assumption that the fast variables converge in distribution in a time which is negligible in comparison with the rate of change of the slow variable. Through this assumption, a simple analysis of the fast subsystem yields an approximation of the dynamics of the slow variables. This fast subsystem can be analysed in several ways, either through analysis and approximation[2], or through direct simulation of the fast subsystem[22].

Another approach is to approximate the system by a continuous state-space stochastic differential equation (SDE), through the chemical Langevin equation (CLE)[12]. This system can then be simulated using numerical methods for SDEs. An alternative approach is to approximate only the slow variables by an SDE. The SDE parameters can be found using bursts of stochastic simulation of the system, initialised at a particular point on the slow state space[8], the so-called "equation-free" approach. This was further developed into the constrained multiscale algorithm (CMA)[5], which used a version of the SSA which also constrained the slow variables to a particular value. Using a similar approach to [2], the CMA can similarly be adapted so that approximations of the invariant distribution of this constrained system can be made without the need for expensive stochastic simulations[6]. However, depending on the system, as with the slow-scale SSA, these approximations may incur errors.

Analysis of mathematical models of gene regulatory networks (GRNs) is important for a number of reasons. It can give us further insight into how important biological processes within the cell, such as the circadian clock[21] or the cell cycle[16] work. In order for these models to be constructed, we need to observe how these systems work in the first place. Many of the observation techniques, such as the DNA microarray[18], are notoriously subject to a large amount of noise. Moreover, since the systems themselves are stochastic, the problem of identifying the structure of the network from this data is very difficult. As such, the inverse problem of characterising a GRN from observations is a big challenge facing our community[14].

One popular approach to dealing with inverse problems, is to use a Bayesian

[1] Define a dominating process to have transition rates given by the matrix $\mathcal{M} = \frac{1}{\rho}\mathcal{G} + I$.
[2] This process has uniformly distributed reaction events on the time interval $[t_0, t_1]$. The number r of such events is given by (1).
[3] Once $r = \hat{r}$ has been sampled, the type of each event must be decided, by sampling from the distribution (2), starting with the first event. An event which corresponds to rate $m_{i,i}$ indicates that no reaction event has occurred at this event.
[4] Once all event types have been sampled, we have formed a sample from the conditioned path space.

Table 1: *A summary of the methodology presented in [9], for sampling paths of Markov-modulated Poisson processes, conditioned on their endpoints.*

framework. The Bayesian approach allows us to combine prior knowledge about the system, complex models and the observations in a mathematically rigorous way[19]. In the context of GRNs, we only have noisy observations of the concentrations of species at a set of discrete times. As such, we have a lot of missing information. This missing data can be added to the state space of quantities that we wish to infer from the data that we do have. This complex probability distribution on both the true trajectories of the chemical concentrations, and on the network itself, can be sampled from using Markov chain Monte Carlo (MCMC) methods, in particular a Gibb's sampler[9]. Within this Gibb's sampler, we need a method for sampling a continuous path for the chemical concentrations given a guess at the reaction parameters, and our noisy measurements. Exact methods for sampling paths conditioned on their endpoints have been developed [9, 17].

The problems become even more difficult when, as is often the case, the systems in question are also multiscale. This means that these inverse problems require a degree of knowledge from a large number of areas of mathematics. Even though many of the approaches that are being developed are currently out of reach in terms of our current computational capacity, this capacity is continually improving. In this paper we aim to progress this methodology in a couple of areas.

1.1. Conditioned path sampling methods

We will briefly review the method presented in [9] for the exact sampling of conditioned paths in stochastic chemical networks. Suppose that we have a Markov jump process, possibly constructed from such a network, with a generator $\mathcal{G}$. We wish to sample a path, conditioned on $X(t_0) = x_0$ and $X(t_1) = x_1$. Such a path can be found by creating a dominating process (i.e. a process whose rate is greater than the fastest rate of any transitions of the original system) with a uniform rate.

We define the rate to be greater than the fastest rate of the process with generator $\mathcal{G}$, so that

$$\rho > \max_i \mathcal{G}_{i,i}.$$

Then we define the transition operator of the dominant process by:

$$\mathcal{M} = \frac{1}{\rho}\mathcal{G} + I.$$

We can then derive the number of reaction events N_U of the dominating process in the time interval $[t_0, t_1]$ by:

$$\mathbb{P}(N_U = r) = \frac{\exp(-\rho t)(\rho t)^r / r! [\mathcal{M}^r]_{x_0,x_t}}{[\exp(\mathcal{G}t)]_{x_0,x_t}}. \tag{1}$$

A sample is taken from this distribution. The times $\{t_1^*, t_2^*, \dots t_r^*\}$ of all of the r reaction events can then be sampled uniformly from the interval $[t_0, t_1]$. The only thing that then remains is to ascertain which reaction has occurred at each reaction event. This can be found by computing, starting with $X(t_0) = x_0$, the probability distribution defined by:

$$\mathbb{P}(X(t_j^*) = x | X(t_{j-1}^* = x_{j-1}^*, X(t_1) = x_1) = \frac{[\mathcal{M}]_{x_{j-1}^*,x}[\mathcal{M}^{r-j}]_{x,x_1}}{[\mathcal{M}^{r-j+1}]_{x_{j-1}^*,x_1}}. \tag{2}$$

This method, summarised in Table 1 exactly samples from the desired distribution, but depending on the size and sparsity of the operator $\mathcal{G}$, it can also be very expensive. In the context of multiscale systems with a large number of possible states of the variables, the method quickly becomes computationally intractable. For numerical examples of this method, see Section 5.

1.2. Summary of Paper

In Section 2, we introduce a version of the Constrained Multiscale Algorithm (CMA), which allows us to approximate the effective generator of the slow processes within a multiscale system. In particular, we explore how stochastic simulations are not required in order to compute a highly accurate effective generator. In Section 3, we aim to compare the accuracy of the effective generators arrived at through the QSSA and CMA approaches. In Section 4, we describe how the constrained approach can be extended in a nested structure for systems for whose constrained subsystem is itself a large intractable multiscale system, By applying the methodology in turn to the constrained systems arising from the constrained approach, we can make the analysis of highly complex and high dimensional systems computationally tractable. In Section 5, we present some numerical results, including some examples of conditioned path sampling using effective generators approximated using the CMA. Finally, we will summarise our findings in Section 6.

[1] Calculate propensity functions $\alpha_i(t)$, $i = 1, 2, \ldots, M$.

[2] Next reaction time is given by

$$\tau = -\frac{\log(u)}{\alpha_0(\mathbf{X}(t))}, \quad \text{where} \quad \alpha_0(\mathbf{X}(t)) = \sum_{k=1}^{M} \alpha_k(\mathbf{X}(t)). \tag{3}$$

[3] Choose one $j \in \{1, \ldots, M\}$, with probability α_j/α_0, and perform reaction R_j.

[4] If $S \neq s$ due to reaction j occurring, then reset $S = s$ while not changing the value of $\mathbf{F}$.

[5] If $X_i < 0$ for any i, then revert to the state of the system before the reaction j occurred.

[6] Continue with step [**1**] with time $t = t + \tau$.

Table 2: *The* Constrained *Stochastic Simulation Algorithm (CSSA). Simulation starts with $S = s$ where s is a given value of the slow variable.*

2. The Constrained Multiscale Algorithm

The Constrained Multiscale Algorithm was originally designed as a multiscale method which allowed us to compute the effective drift and diffusion parameters of a diffusion approximation of the slow variables in a multiscale stochastic chemical network. The idea was simply to constrain the original dynamics to a particular value of the slow variable. This can be done through a simple alteration of the original SSA by Gillespie[11]. As shown in [5], the SSA is computed as normal, until one of the slow reactions occurs. After the reaction has occurred, the slow variable is then reset to its original value, in such a way that the fast variables are not affected. The constrained SSA is given in Table 2.

Let us illustrate this using an example which we shall be using also later in the paper.

$$\begin{aligned} R_1 &: \quad \emptyset \xrightarrow{k_1} X_1 \\ R_2 &: \quad X_2 \xrightarrow{k_2} \emptyset \\ R_3 &: \quad X_1 \xrightarrow{K} X_2 \\ R_4 &: \quad X_2 \xrightarrow{K} X_1. \end{aligned} \tag{4}$$

In certain parameter regimes, for example where $K \gg k_1 V + k_2$, this system is multiscale, with reactions R_3 and R_4 occurring many times on a time scale for which reactions R_1 and R_2 are unlikely to happen at all. The variable $S = X_1 + X_2$ is unaffected by these fast reactions, and as such is a good candidate for the slow variable which we wish to analyse. We have two choices for the fast variable, either $F = X_1$ or $F = X_2$. As detailed in [5], it is preferable to pick

fast variables, where possible, that are not involved in zeroth order reactions. Therefore, in this case, we choose $F = X_2$. Therefore, the constrained system can be written in the following way:

$$\begin{aligned} C_1 &: \quad X_1 + X_2 = S, \\ R_2 &: \quad X_2 \xrightarrow{k_2} X_1 \\ R_3 &: \quad X_1 \xrightarrow{K} X_2 \\ R_4 &: \quad X_2 \xrightarrow{K} X_1. \end{aligned} \tag{5}$$

Note that reaction R_1 has disappeared completely, since when we reset the slow variable for this reaction, we simply reset X_1 back to its previous value (as it is not our chosen fast variable) and as such there is no net effect of the reaction on either the fast or slow variables. Similarly, reaction R_2 has been altered. If this reaction occurs, the slow variable is reduced by one. We are not permitted to change the fast variable X_2 in order to reset the slow variable to its original value, and therefore we must increase X_1 by one, giving us a new stoichiometry for this reaction.

In the original CMA, statistics were taken regarding the frequency of the slow reactions, at each point of the slow domain, and were used to construct the effective drift and diffusion parameters of an effective diffusion[5, 4] process. However, this constrained approach can also be used to compute an effective generator for the original discrete slow process, as we will now demonstrate. The CMA can be very costly, due to the large computational burden of the stochastic simulations of the constrained system. In this section, we will also introduce a method for avoiding the need for these simulations, whilst also significantly improving accuracy.

The constrained systems can often have a very small state space (which we will denote $\Gamma(s)$), since they are constrained to a single value of the slow variables. For example, for the constrained system (5), there are only $\left\lfloor \frac{S}{2} \right\rfloor$ possible states. Such a system can easily be fully analysed. For example, the invariant distribution can be found by characterising the one-dimensional null space of the generator matrix of the constrained process. For small to medium-sized systems, this is far more efficient than exhaustive Monte Carlo simulations. For other systems with larger constrained state spaces, stochastic simulation may still be the best option, although in Section 4 we show how the constrained approach can be applied iteratively until the constrained subsystem is easily analysed.

Suppose that we have a constrained system with N_F fast variables, $F_1, F_2, \ldots, F_{N_F}$. The generator for the constrained system with $S = s$ is given by $\mathcal{G}_F(s)$. Since the system is ergodic, there is a one-dimensional null space for this generator. This can be found by using standard methods for identifying eigenvectors, by searching for the eigenvector corresponding to the eigenvalue equal to zero. Krylov subspace methods allows us to find these eigenvectors with very fast convergence rates. Suppose we have found such a vector $\mathbf{v}$, such that

$$\mathcal{G}_F(s)\mathbf{v} = 0.$$

[1] For each value of the slow variable $S = s \in \Omega$, compute the generator of the constrained subsystem, $\mathcal{G}_s$.
[2] Find the zero eigenvector $\mathbf{v}$ of $\mathcal{G}_s$, and let $\mathbf{p}(s) = \frac{\mathbf{v}}{\sum v_i}$.
[3] Approximate the effective propensities at each point $s \in \Omega$ using (6).
[4] Construct an effective generator $\mathcal{G}$ of the slow processes of the system using these effective propensities.

Table 3: *The CMA approach to approximating the effective generator $\mathcal{G}$ of the slow variables, without the need for stochastic simulations.*

Then our approximation to the invariant distribution of this system is given by the discrete probability distribution represented by the vector

$$\mathbf{p}(s) = \frac{\mathbf{v}}{\sum v_i}.$$

Our aim is now to use this distribution to find the effective propensities of the slow reactions of the original system.

Suppose that we have M_S slow reactions in the original system. Each has an associated propensity function $\alpha_1(S,F), \alpha_2(S,F), \ldots, \alpha_{M_S}(S,F)$. We now simply want to find the expectation of each of these propensity functions with respect to the probability distribution $\mathbf{p}(s)$:

$$\mathbb{E}_{F\sim\mathbf{p}(s)}\alpha_i(S,F) = \sum_{f\in\Gamma(s)} p_f(s)\alpha_i(S,f). \tag{6}$$

Having computed this expectation for all of the slow propensities, over all required values of the slow variable, then an effective generator for the slow variable can be constructed.

3. Comparing the CMA and QSSA approaches

A very common approach to approximating the dynamics of slowly changing quantities in multiscale systems, is to invoke the quasi steady-state assumption (QSSA). The assumption is that the fast and slow variables are operating on sufficiently different time scales that it can be assumed that the fast subsystem enters equilibrium instantaneously following a change in the slow variables. This assumption means that if the fast subsystem's invariant distribution can be found (or approximated), then the effective propensities of the slow reactions can be computed. However, as demonstrated in [4], this assumption incurs an error, and for systems which do not have a large difference in time scales between the fast and slow variables, this error can be significant.

The CMA does not rely on the QSSA, which is a strong assumption that we can assume that no slow reactions occur on the timescale of relaxation of the fast variables. Therefore, the CMA is able to take into account the effect

that the slow variables has on the invariant distribution of the fast variables, conditioned on a value of the slow variables. In a true fast-slow system, this will yield the same results as the QSSA, but for most systems of interest, the constrained approach will have a significant increase in accuracy. A difference in time scales is still required for the algorithm to make any sense, but there are not large extra errors incurred when the time scale gap is smaller, (again see [4]). The assumptions for the CMA are weaker than the QSSA, namely that we assume that the dynamics of slow variable(s) can be approximated by a Markov-modulated Poisson process, independently of the value of the fast variables. This means that we have made the assumption that the current value of the fast variables has no effect on the transition rates of the slow variables once a slow reaction has occurred. This is subtly weaker than the QSSA, and importantly the effect of the slow reactions on the invariant distribution of the fast variables is accounted for.

If we follow the approach outlined in Table 3, we don't even need to conduct any stochastic simulations to approximate the effective dynamics, and the CMA becomes the preferred choice for estimation of effective dynamics.

3.1. A Linear Example

Let us illustrate this by returning to the example given by the linear system (4), first by using the QSSA. The QSSA tells us that the fast subsystem (made up of reactions R_3 and R_4) reaches probabilistic equilibrium on a timescale which is negligible in comparison with the timescale on which the slow reactions are occurring. Therefore we may treat this subsystem in isolation with fixed S:

$$X_1 \underset{k_4}{\overset{k_3}{\rightleftarrows}} X_2, \qquad S = X_1 + X_2.$$

This is a very simple autocatalytic reaction system, for which a great deal of analytical results are available. For instance, we can compute the invariant distribution for this system[15], which gives us that X_2 is a binomial random variable

$$X_2 \sim \mathcal{B}\left(\cdot, S, \frac{k_3}{k_3 + k_4}\right).$$

Therefore, we can compute the conditional expectation $\mathbb{E}(X_2|S) = \frac{k_3 S}{k_3+k_4}$ in this fast subsystem, and use this to approximate the effective rate of reaction R_2. Therefore, the effective slow system is given by the reactions:

$$\emptyset \xrightarrow{\hat{k}_1} S \xrightarrow{\hat{k}_2} \emptyset, \tag{7}$$

where

$$\hat{k}_1 = k_1, \qquad \hat{k}_2 = \frac{k_2 \mathbb{E}(X_2)}{S} = \frac{k_2 k_3}{k_3 + k_4}.$$

Again, we can compute the invariant distribution for this effective system[15], which in this instance is a Poisson distribution:

$$S \sim \mathcal{P}\left(\frac{k_1 V(k_3 + k_4)}{k_2 k_3}\right). \tag{8}$$

We can quantify the error we have made in using the quasi-steady state assumption by, for example, comparing this distribution with the true invariant distribution. Once again, using the results of [15], we can compute the invariant distribution of the system (4), which is a multivariate Poisson distribution:

$$[X_1, X_2] \sim \mathcal{P}(\bar{\lambda}_1, \bar{\lambda}_2),$$

where $\bar{\lambda}_1 = \frac{k_1 V(k_2+k_4)}{k_2 k_3}$, and $\bar{\lambda}_2 = \frac{k_1 V}{k_2}$. Trivially one can compute the marginal distribution on the slow variable S:

$$\begin{aligned}\mathbb{P}(S=s) &= \sum_{n=0}^{s} \frac{\bar{\lambda}_1^n}{n!} \frac{\bar{\lambda}_2^{s-n}}{(s-n)!} \exp(-(\bar{\lambda}_1+\bar{\lambda}_2), \\ &= \frac{(\bar{\lambda}_1+\bar{\lambda}_2)^s}{s!} \exp(-(\bar{\lambda}_1+\bar{\lambda}_2).\end{aligned}$$

Therefore S is also a Poisson variable with intensity $\lambda = \bar{\lambda}_1 + \bar{\lambda}_2 = \frac{k_1 V(k_2+k_3+k_4)}{k_2 k_3}$, which differs from the intensity approximated invariant density (8) by $\frac{k_1 V}{k_3}$. Note that k_3 is one of the fast rates, and $k_1 V$ is one of the slow rates, and therefore as the difference in timescales of the fast and slow reactions increases, this error decreases to zero, so that the QSSA gives us an asymptotically exact approximation of the slow dynamics.

For comparison, let us compute approximations of the effective slow rates by using the CMA. The CMA for this system tells us that we need to analyse the constrained system (5). The constrained system in this example only contains monomolecular reactions, and as such can be analysed using the results of [15]. The invariant distribution for this system is a binomial, such that

$$X_2 \sim \mathcal{B}\left(\cdot, S, \frac{k_3}{k_2+k_3+k_4}\right).$$

Using this, we can compute the effective propensity of reaction R_2,

$$\bar{\alpha}_2(S) = k_2 \mathbb{E}(X_2|S) = \frac{k_2 k_3 S}{k_2+k_3+k_4},$$

giving us the effective rate $\bar{k_2} = \frac{k_2 k_3}{k_2+k_3+k_4}$. The invariant distribution of (7) with this effective rate for $\bar{k_2}$ is once again a Poisson distribution with intensity

$$\lambda = \frac{k_1 V(k_2+k_3+k_4)}{k_2 k_3},$$

which is *identical* to the intensity of the true distribution on the slow variables. In other words, for this example, the CMA produces an approximation of the effective dynamics of the slow variables for this system, whose invariant distribution is identical to the marginal invariant distribution of the slow variables in the full system. The constrained approach corrects for the effect of the slow reactions on the invariant distribution of the fast variables. In this and other

examples of systems with monomolecular reactions, the constrained approach gives us a system whose invariant distribution is exactly equal to the marginal distribution on the slow variables for the full system. Another example is presented in Section 5.3, for which the constrained system is itself multiscale, and requires another iteration of the CMA to be applied.

For this example, we did not even need to compute the invariant distributions of the constrained systems numerically. In Section 5.2, we will come across a system for which it is necessary to numerically compute the invariant distribution of the constrained system.

The approaches described in Section 1.1 hit problems when the system for which you are trying to generate a conditioned path is multiscale. In a multiscale system, the rate ρ of the dominating process will be very large, and as such the number of reaction events will be large, even if the path we are trying to sample is short. Therefore M^r is likely to be a full matrix, and the number of calculations of (2) will be large. Moreover, the size of M is also likely to be large, since for each value $S = s$ of the slow variable, there are many states, one for each possible value of the fast variable. All of these factors make the problem of computing a conditioned path in such a scenario computationally intractable.

For example, let us consider the system (4). Naturally we cannot store the actual generator of this system, since the system is open and as such the generator is an infinite dimensional operator. However, the state space can be truncated carefully in such a way that the vast majority of the states with non-negligible invariant density are included, but an infinite number of highly unlikely states are presumed to have probability zero. Note that this means that we are effectively sampling paths satisfying $S(t_0) = s_1$, $S(t_1) = s_2$ conditioned on $S(t) \in \Omega \forall t$. However, even with careful truncation the number of states can be prohibitively large.

Suppose we consider system (4) with parameters given by

$$k_1 = k_2 = 1, \qquad K = 200, \qquad V = 100. \tag{9}$$

Suppose that we truncate the domain for this system to

$$\Omega = \{[X_1, X_2] | S = X_1, X_2 \in \{0, 1, \ldots, 200\}\,.$$

This truncated system has $201^2 = 40401$ different states, and therefore the generator $G \in \mathbb{R}^{40401 \times 40401}$. Although this matrix is sparse, the matrix exponential required in (1) is full, as is M^r for moderate $r \in \mathbb{N}$. A full matrix of this size stored at double precision would require over 13GB of memory. So even for this system, the most simple multiscale system that one could consider, the problem of sampling conditioned paths is computationally intractable.

In comparison, suppose that we use a multiscale method such as the CMA to approximate the effective rates of the slow reactions. Then, for the same Ω, we only have 401 possible states of the slow variable, a reduction of 99.25%. The effective generator $\mathcal{G} \in \mathbb{R}^{401 \times 401}$ would then only require 1.29MB to be stored as a full matrix in double precision. The dominating process for this system

must now have rate $\rho > 299.25$, instead of $\rho > 40300$, which is over 130 times bigger. This means far fewer calculations of (2). What is more, as we saw in Section 2, for some systems the CMA *exactly* computes the effective dynamics of the slow variables, with no errors.

Naturally, this approach only allows us to sample the paths of the slow variables. However, the fast variables, if required, can easily be sampled after the fact, using an adapted Gillespie approach which samples the fast variables given a trajectory of the slow variables.

4. The Nested CMA

There will be many systems for which the constrained subsystem is itself a highly complex and multiscale system. In this event, it will not be feasible to find the null space of a sensibly truncated generator for the constrained subsystem. Therefore, we need to consider how we might go about approximating this. Fortunately, we already have the tools to do this, since, we can iteratively apply the CMA methodology to this subsystem. This is analogous to the nested strategy proposed in the QSSA-based nested SSA[22].

This nested approach allows us to reduce much more complex systems in an accurate, computationally tractable way. The problem of finding the null space of the first constrained subsystem is divided into finding the null space of many small generators, through further constraining. An example of this nested approach will be demonstrated in Section 5.3.

5. Numerical Results

In this section we will present some numerical results produced using the CMA approach.

5.1. A Simple Linear System

First we will consider the system (4), with parameters (9). As we demonstrated in Section 2, the CMA can be used to compute an effective generator for the slow variable $S = X_1 + X_2$, whose invariant distribution is exactly that of the slow variable in the full system without the multiscale reduction. Moreover, this can be achieved with no Monte Carlo simulations, since the constrained subsystem contains only monomolecular reactions, and as such its invariant distribution can be exactly computed[15].

At this juncture, we simply need to apply the method of Fearnhead and Sherlock[9] in order to be able sample paths conditioned on their endpoints. Suppose we wish to sample paths conditioned on $S(t_0 = 0) = 0$ and $S(t_1 = 10) = 200$. The invariant distribution of this system, as shown previously in this paper, is a Poisson distribution with mean $\lambda = \frac{k_1 V(k_2+k_3+k_4)}{k_2 k_3} = 200.5$. Therefore, we are attempting to sample paths which start in the tails of the invariant distribution, and end up close to the mean, in a timeframe for which an unconditioned path would easily be able to achieve the same feit.

Since the system is open, we are required to truncate the domain in order to be able to store and manipulate the effective generator. We truncate the domain to $\Omega = \{[X_1, X_2] | S = X_1 + X_2 \leq 400\}$. Therefore we aim to sample paths

$$\{S(t),\, t \in [0, 10] \,|\, S(0) = 0,\, S(10) = 200,\, S(t) \in \Omega\, \forall t \in [0, 10]\}.$$

As the number of possible states of the slow variable is relatively small, it was possible to compute and store full matrices for M^r as required in (1) and (2) for $r \in 1, 2, \ldots, 3420$. r has an upper bound of 3420 as the cumulative mass function for the probability distribution (1) is within machine precision of one at $r = 3420$. Storing all powers of the matrices is clearly not the most efficient way to implement this algorithm, but for this example was possible without any intensive computations, and with minimal numerical error. We will present a more efficient approach in the next section.

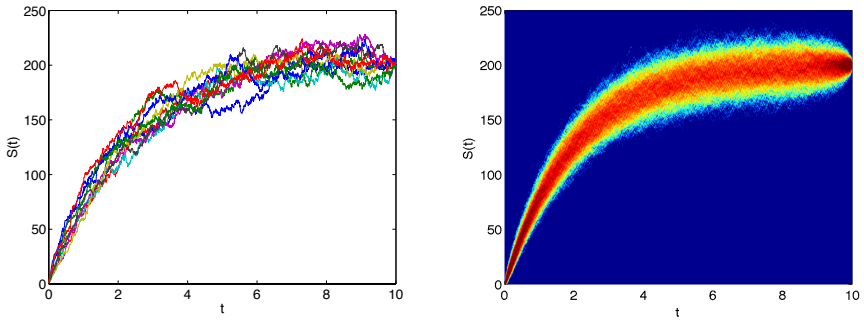


Figure 1: *(a) 10 trajectories of the slow variable $S = X_1 + X_2$ sampled conditioned on $S(0) = 0$, $S(10) = 200$, $S(t) \in \Omega \forall t \in [0, 10]$ for the system (4) with parameters (9), using the CMA approximation of the effective generator. (b) A heat map of the trajectories plotted in (a).*

Figure 1 (a) shows the results of 10 sampled paths using this approach, and (b) shows a heat map of 1000 trajectories. As expected, the trajectories start at $S(0) = 0$, but quickly enter probabilistic equilibrium in a Poisson distribution centered around $S = 200.5$. In the last 2 time units of the simulations, the effect of the conditioned endpoint begins to take effect, and soon all of the trajectories converge to $S = 200$ at time $t = 10$. Note that the length of the paths is much longer than the average relaxation time of the slow variables, and as such, the paths that we are sampling are not exhibiting rare behaviour. We will see an example of forcing paths to exhibit rare behaviour through endpoint conditioning in the next section.

We can be reasonably sure that the presented trajectories are samples from the space of conditioned paths, since they were formed using an effective generator whose invariant distribution would be exactly the same as the marginal

distribution on the slow variables of the full system, if it weren't for the necessary truncation of the domain. Moreover, since the invariant distributions of the constrained subsystems were solved analytically, the only numerical errors are those that are incurred when computing (1) and (2).

Previous papers have also shown the CMA to be highly accurate in more complex systems[5, 4], in the context of approximating the slow process by a diffusion. We will now use the CMA approach presented in this paper to generate an effective generator for the slow variable in a system which exhibits bistability.

5.2. A Bistable Example

Sampling of conditioned paths of this nature is an integral part of the approach of Bayesian inversion of biochemical data. A Gibb's sampler is used to alternately update the network structure and system parameters, and the missing data (i.e. the full trajectory), sampled for example using the method found in [9]. However, efficient methods to sample paths of multiscale systems may also be useful in other areas. For instance, it may allow us to sample paths which make rare excursions, or large deviations from mean behaviour.

Let us consider the following chemical system, which in certain parameter regimes exhibits bistable behaviour.

$$\begin{aligned} R_1, R_2 \quad &: \qquad X_2 \underset{k_2}{\overset{k_1}{\rightleftarrows}} X_1 + X_2, \\ R_3, R_4 \quad &: \qquad \emptyset \underset{k_4}{\overset{k_3}{\rightleftarrows}} X_1, \\ R_5, R_6 \quad &: \qquad X_1 + X_1 \underset{k_6}{\overset{k_5}{\rightleftarrows}} X_2, \end{aligned} \tag{10}$$

In particular, we consider parameter regimes where the occurrence of reactions R_5 and R_6 are on a relatively faster timescale than the other reactions. The following is just such a parameter regime:

$$\begin{aligned} k_1 = 123.0, \qquad k_2 = 1.0, \qquad k_3 = 66.0, \\ k_4 = 9.4, \qquad k_5 = 10.0, \qquad k_6 = 4000.0. \end{aligned} \tag{11}$$

The fast reactions in this example are reactions R_5 and R_6, and as such, $S = X_1 + 2X_2$ is a good choice of slow variable, since this quantity is invariant to these fast reactions. Figure 2 shows a plot of an approximation of the invariant distribution of the slow variable for this system. This approximation was found by constructing the full generator for the system, on a truncated domain, $\Omega = \{(x_1, x_2) \in \{0, 1, \ldots, 500\} \times \{0, 1, \ldots, 250\}$. This domain is sufficiently big that any increases lead to negligible changes in the computed invariant distribution on $S \in \{0, 1, \ldots, 200\}$, where the vast majority of the invariant probability mass is located, as we verified numerically. The zero eigenvector of this generator was then found, normalised, and then plotted. Since this system has 2nd order

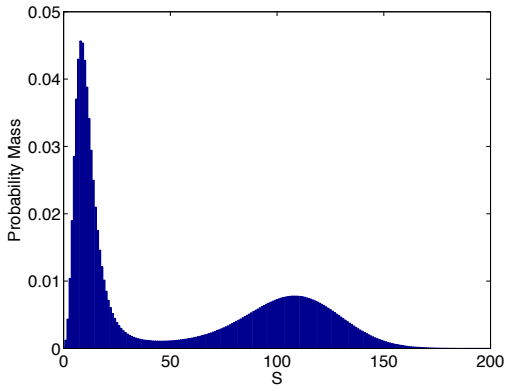


Figure 2: *Approximation of the invariant distribution on the slow variable* $S = X_1 + 2X_2$ *of system* (10) *with parameters* (12)*, demonstrating the bistable nature of the system. Approximation was computed by finding the null space of the full generator of the system on the truncated domain* $\{0, 1, \ldots, 500\} \times \{0, 1, \ldots, 250\}$.

reactions, its invariant density cannot currently be written in closed form, and as such, we could use this approximation on the truncated domain in order to quantify the accuracy of the CMA approach. This plot demonstrates the bistable nature of this system, which can take a long time to switch between the two favourable regions.

First, we will use the CMA to approximate the effective generator of the slow variable. We will then find the invariant distribution arising from that generator, and compare it with the distribution shown in Figure 2.

There are two choices for the fast variable, but as explained in detail in [5], $F = X_2$ is the best choice, since there is a zeroth order reaction involving X_1. This leads to the following constrained system:

$$\begin{aligned} C_1 : &\qquad X_1 + 2X_2 = S, \\ R_5, R_6 : &\qquad X_1 + X_1 \underset{k_6}{\overset{k_5}{\rightleftarrows}} X_2. \end{aligned} \tag{12}$$

This system is an interesting example, since X_2 is not affected by any of the slow reactions. This means that the constrained version of this system is made up only of the fast reactions, and therefore the CMA and QSSA-based methods are in complete agreement. However, the methodology we outlined in Table 3 allows us to approximate the effective generator arising from these approaches without either the need for expensive stochastic simulations, or errors incurred through various approximations of the invariant density of the constrained (or equivalently for this system, fast) subsystem.

Following this methodology, an effective generator $\mathcal{G}$ can be computed. The null space of this generator gives us an approximation of the invariant distribution. We can quantify the error we have incurred in our approximation by

comparing this density with the marginal density that we computed and plotted in Figure 2.

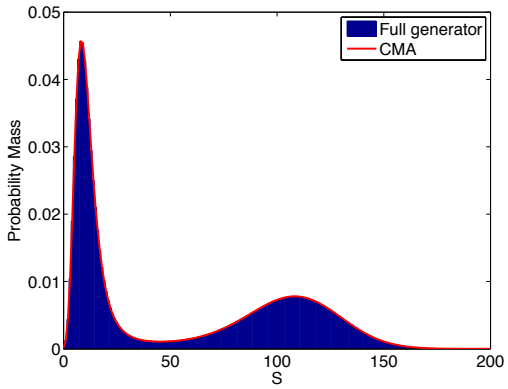


Figure 3: *Plots to show approximation of the invariant distribution of the slow variable* $S = X_1 + 2X_2$ *of system* (10) *with parameters* (12)*, through computing the null space of the truncated generator of the full system (blue), and of the effective generator computed using the CMA.*

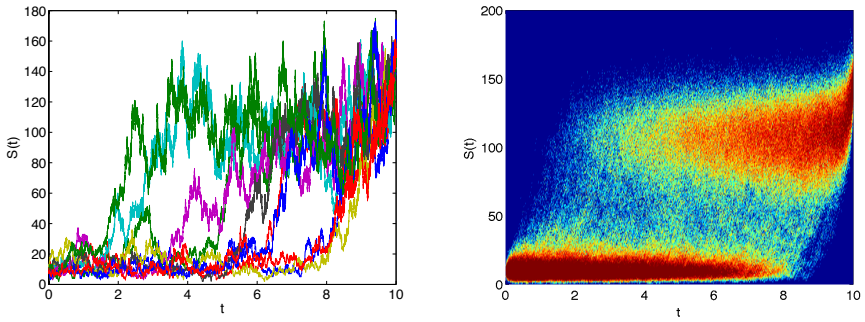
The two distributions, as can be seen in Figure 3, are indistinguishable by eye, and the relative l^2-error, given by

$$\frac{\|\mathbf{p}_{\text{CMA}} - \mathbf{p}_{\text{approx}}\|_{l^2}}{\mathbf{p}_{\text{approx}}\|_{l^2}},$$

was equal to 3.215×10^{-3}. The size of this discrepancy is very small, and what is more since we were comparing to another approximation (since this was all that we were able to do), it is not clear where this error was incurred, or which method is more accurate. However, the difference is small enough to indicate that the effective generator that we have computed using the CMA is a highly accurate representation of the dynamics of the slow variables within this system. Therefore, it is entirely reasonable to use this approximation of the effective generator in order to attempt to sample conditioned paths of the slow variable.

Given an approximation of the effective generator of the slow variables, computed using the CMA, we can now employ the methodology of [9], as summarised in Section 1.1, to sample paths conditioned on their endpoints. This time, a full eigenvalue decomposition of the matrix $\mathcal{M} = \frac{1}{\rho}\mathcal{G} + I$ was computed, so that matrices V and D could be found with V unitary and D diagonal, with $\mathcal{M} = V^{-1}DV$. Then rows of $\mathcal{M}^r = V^{-1}D^rV$ can be efficiently and accurately computed, as required in (1) and (2).

Figure 4 presents results using this approach. An effective generator for the system (10) was computed for the domain $X_1 + 2X_2 = S \in \Omega = \{0, 1, \ldots, 500\}$, and then fed into the conditioned path sampling algorithm. Figure 4 (a) shows



(a) (b)

Figure 4: *(a) 10 trajectories of the slow variable* $S = X_1 + 2X_2$ *sampled conditioned on* $S(0) = 10$, $S(10) = 150$, $S(t) \in \Omega = \{0, 1, \ldots, 500\} \forall t \in [0, 10]$ *for the system* (10) *with parameters* (12)*, using the CMA approximation of the effective generator. (b) A heat map of a set of 1000 trajectories.*

10 samples of conditioned paths. Notice that as the transition time between the two favourable regions is relatively short compared with the length of the simulation, the time of the transition varies greatly between the different trajectories. This indicates that we are producing trajectories with a fair reflection of what happens in a transition between these regions. Figure 4 (b) shows a heat map of 1000 sampled paths. As time progresses, more of the trajectories make the transition from the lower stable region to the higher stable region, finally all converging to $S(t_1) = 150$.

5.3. An Example of the Nested CMA Approach

In this section, we will illustrate how the nested approach outlined in Section 4 can be applied. We will consider an example, that as before, we know what the invariant distribution of the slow variables should be. This gives us a way of quantifying any errors that we incur by applying the nested CMA approach.

$$\begin{aligned} R_1 &: \quad \emptyset \xrightarrow{k_1} X_1 \\ R_2 &: \quad X_3 \xrightarrow{k_2} \emptyset \\ R_3 &: \quad X_1 \xrightarrow{\kappa} X_2 \\ R_4 &: \quad X_2 \xrightarrow{\kappa} X_1. \\ R_5 &: \quad X_2 \xrightarrow{\gamma} X_3 \\ R_6 &: \quad X_3 \xrightarrow{\gamma} X_2. \end{aligned} \tag{13}$$

We will consider the following parameter system:

$$k_1 = k_2 = 1, \qquad \kappa = 2000, \qquad \gamma = 200, \qquad V = 100. \tag{14}$$

In this regime, there are multiple different time scales on which the reactions are occurring. This is demonstrated in Figure 5, where there is a clear gap in the frequency of reactions R_1 and R_2 (the slowest), R_5 and R_6 (fast reactions) and R_3 and R_4 (fastest reactions).

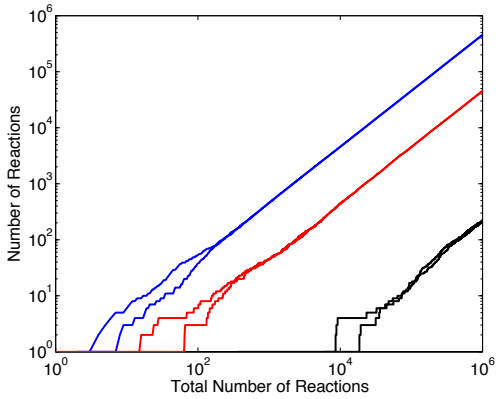


Figure 5: *Relative occurrences of the reactions R_1-R_6, for the system* (13) *with parameters* (14). *The most frequent reactions are reactions R_3 and R_4, reactions R_4 and R_6 are the next most frequent, with reactions R_1 and R_2 being the least frequent.*

Suppose that we wish to use the CMA approach to reduce the dimension of this problem to a one dimensional system, with $S_1 = X_1 + X_2 + X_3$ being the slow variable. We wish to approximate the effective generator for the resultant reduced system.

Firstly, we apply the CMA as we have done previously. There are 3 choices for the fast reactions, each involving two out of X_1, X_2 and X_3. Since X_1 is the product of a zeroth order reactions, it is preferable not to include this as one of the fast variables, and so we pick $\mathbf{F}_1 = [X_2, X_3]$. We then construct the constrained subsystem for this choice of slow and fast variables:

$$\begin{aligned}
C_1 &: \quad X_1 + X_2 + X_3 = S, \\
R_2 &: \quad X_3 \xrightarrow{k_2} X_1 \\
R_3 &: \quad X_1 \xrightarrow{\kappa} X_2 \\
R_4 &: \quad X_2 \xrightarrow{\kappa} X_1. \\
R_5 &: \quad X_2 \xrightarrow{\gamma} X_3 \\
R_6 &: \quad X_3 \xrightarrow{\gamma} X_2.
\end{aligned} \tag{15}$$

Note that R_1 is removed, since it does not change the fast variables. R_2 is the only other reaction which has changes. We have reduced the dimension of the system (due to the constraint $X_1 + X_2 + X_3 = \sigma$ for some $\sigma \in \mathbb{N}$), but we are still left with a multiscale system, which in theory could be computationally intractable for us to find the invariant distribution for, through funding the null

space of its generator. Therefore, we can apply another iteration of the CMA to this constrained system.

Reactions R_3 and R_4 are the fastest reactions in the system, and therefore we pick our next slow variable that we wish to constrain to be $S_2 = X_1 + X_2$, which is invariant with respect to these reactions. Due to the previous constraint $S_1 = X_1 + X_2 + X_3$, we are only required to define one fast variable for this system. All three choices $F_2 = X_1, X_2, X_3$, are essentially equivalent, and so we pick $F_2 = X_3$. These choices lead us to this further constrained system:

$$\begin{aligned}
C_1 \quad &: \qquad X_1 + X_2 + X_3 = S_1, \\
C_2 \quad &: \qquad X_1 + X_2 = S_2, \\
R_2 \quad &: \qquad \alpha_2(\mathbf{X}) = \begin{cases} k_2 X_3, & \text{if } X_2 > 0, \\ 0 & \text{otherwise.} \end{cases} \\
&\qquad\quad \nu_2 = [1, -1, 0]^T \\
R_3 \quad &: \qquad X_1 \xrightarrow{\kappa} X_2 \\
R_4 \quad &: \qquad X_2 \xrightarrow{\kappa} X_1.
\end{aligned} \tag{16}$$

Notice that we now have two separate constraints, and as such reactions R_5 and R_6 now have zero stoichiometric vectors. Moreover, these constraints lead us to a somewhat unphysical reaction for R_2. The reactant for this reaction is X_3, but only X_2 and X_3 are affected by this reaction. When reaction R_2 happens, we lose one X_3, and gain X_1. Therefore, both constraints have been violated. In order to reset these constraints, without changing the fast variable $F = X_3$, we arrive at the stoichiometry presented in (16). Note that we add the condition that this reaction can only happen if $X_2 > 0$, as we cannot have negative numbers of this species.

This is a closed system, with a very limited number of different states. Therefore, it is computationally cheap to construct its generator, and to find that generator's null space. Our aim with this system, is to find the invariant distribution of the fast variable given particular values for the constraints C_1 and C_2. This distribution will then allow us to compute the expectation of the reaction R_4 within the constrained system (5), which is the only reaction in which is dependent on the results of the second constrained system (since $X_3 = S_1 - S_2$). Once the invariant distribution has been found, this can be used to find the effective propensity of reaction R_5 given a values of $S_1 = X_1 + X_2 + X_3$ and $S_2 = X_1 + X_2$. In turn, the constrained system (15) can then be solved to find the invariant distribution on X_3 given a value of S_1. Finally, this leads us to the construction of an effective generator for the slow variable S_1.

The MATLAB code that was written to implement this process is provided in the electronic supplementary material*****. This system was chosen as we are able to, using the results in [15], find the exact invariant distribution of the slow variable S_1. In this instance, it is a Poisson distribution with mean parameter

$$\lambda = k_1 V \left(\frac{2(\kappa + 1)}{\kappa} + \frac{1 + \gamma}{\gamma} \right) = 301.05.$$

The invariant distribution of the approximated effective generator of S_1 was identical to this distribution up to machine precision.

In comparison, if we had taken a nested QSSA-based approach, such as the nested SSA, we would have converged to a Poisson distribution with mean $\lambda = 300$, which gives a relative error of 4.285×10^{-2}. This demonstrates the improvement that can be made by taking a constrained approach to the characterisation of conditional distributions of fast variables, as opposed to the QSSA approach. What is more, this can be achieved without the need for any expensive stochastic simulations.

6. Conclusions

In this paper, we presented a significant improvement and extension to the original constrained multiscale algorithm (CMA). Through constructing and finding the null space of the generator of the constrained process, we can find its invariant distribution without the need for expensive stochastic simulations. The CMA in this format can also be used not just to approximate the parameters of an approximate diffusion, but to approximate the rates in an effective generator for the slow variables.

Through iterative nesting, the CMA can be applied to much more complex systems, as it can be applied repeatedly if the resulting constrained system is itself multiscale. This makes it a viable approach for a bigger family of (possibly biologically relevant) systems. This nested approach breaks up the original task of solving an eigenvalue problem for one large matrix per row of the effective generator, down into many eigenvalue solves for significantly smaller generators for smaller dimensional problems, making the overall problem computationally tractable.

It was shown that for two examples which contained only monomolecular reactions, that the effective generator produced by the CMA had a null space which was exactly equal (up to machine precision) to the true invariant distribution of the slow variable for those systems. This was in contrast to the generators computed using the QSSA, which exhibited significant errors, which will be bigger the smaller the gap in timescales between the different reactions is. This demonstrates the clear advantage of the constrained approach over the QSSA-based approaches. The second of these systems required the use of the nesting structure.

A more complex bistable system was also analysed using the CMA, and the invariant distribution of the computed effective generator was shown to be very close to the best approximation that we could make of the invariant distribution of the slow variables, using the null space of the original generator with as little truncation as we could sensibly manage with our computational resources.

We showed how these effective generators can be used in the sampling of paths conditioned on their endpoints. Such an approach could be employed as a method to sample missing data within a Gibb's sampler when attempting to find the structure of a network that was observed[9]. This approach could also

be used simply to simulate trajectories of the slow variables, in the same vein as [2] or [22]. In this instance, it would only be necessary to compute the column of the effective generator corresponding to the current value of the slow variables. We also intend to explore how similar ideas could be used in the context of multiscale SDEs, as an alternative method for homogenisation.

Acknowledgements: The author would like to thank Kostas Zygalakis for useful conversations regarding this work. This work was funded by First Grant Award EP/L023989/1 from EPSRC.

[1] A. Auger, P. Chatelain, and P. Koumoutsakos. R-leaping: Accelerating the stochastic simulation algorithm by reaction leaps. *The Journal of chemical physics*, 125(8):084103, 2006.

[2] Y. Cao, D. Gillespie, and L. Petzold. The slow-scale stochastic simulation algorithm. *The Journal of chemical physics*, 122(1):014116, 2005.

[3] Y. Cao, H. Li, and L. Petzold. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *The journal of chemical physics*, 121(9):4059–4067, 2004.

[4] S. Cotter and R. Erban. Error analysis of diffusion approximation methods for multiscale systems in reaction kinetics. *SIAM journal on Scientific Computing*, submitted.

[5] S. Cotter, K. Zygalakis, I. Kevrekidis, and R. Erban. A constrained approach to multiscale stochastic simulation of chemically reacting systems. *The Journal of chemical physics*, 135(9):094102, 2011.

[6] M. Cucuringu and R. Erban. Adm-cle approach for detecting slow variables in continuous time markov chains and dynamic data. *arXiv preprint arXiv:1504.01786*, 2015.

[7] R. Erban, S. Chapman, I. Kevrekidis, and T. Vejchodsky. Analysis of a stochastic chemical system close to a SNIPER bifurcation of its mean-field model. *SIAM Journal on Applied Mathematics*, 70(3):984–1016, 2009.

[8] R. Erban, I. Kevrekidis, D. Adalsteinsson, and T. Elston. Gene regulatory networks: A coarse-grained, equation-free approach to multiscale computation. *Journal of Chemical Physics*, 124:084106, 2006.

[9] P. Fearnhead and C. Sherlock. An exact Gibbs sampler for the Markov-modulated Poisson process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(5):767–784, 2006.

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

[11] D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

[12] D. Gillespie. The chemical langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.

[13] D. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[14] A. Golightly and D. Wilkinson. Bayesian inference for markov jump processes with informative observations. *Statistical Applications in Genetics and Molecular Biology*, 2014.

[15] T. Jahnke and W. Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of mathematical biology*, 54(1):1–26, 2007.

[16] S. Kar, W. Baumann, M. Paul, and J. Tyson. Exploring the roles of noise in the eukaryotic cell cycle. *Proceedings of the National Academy of Sciences*, 106(16):6471–6476, 2009.

[17] V. Rao and Y.W. Teh. Fast MCMC sampling for Markov jump processes and extensions. *The Journal of Machine Learning Research*, 14(1):3295–3320, 2013.

[18] M. Schena, D. Shalon, R. Davis, and P. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235):467–470, 1995.

[19] A. Stuart. Inverse problems: a bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[20] T. Tian and K. Burrage. Binomial leap methods for simulating stochastic chemical kinetics. *The Journal of chemical physics*, 121(21):10356–10364, 2004.

[21] J. Vilar, Hao Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992, 2002.

[22] E Weinan, D. Liu, and E. Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *The Journal of chemical physics*, 123(19):194107, 2005.