On the Validity of the Girsanov Transformation Method for Sensitivity Analysis of Stochastic Chemical Reaction Networks

Ting Wang*[1] and Muruhan Rathinam†[2]

[1]Department of Mathematical Sciences, University of Delaware
[2]Department of Mathematics and Statistics, University of Maryland Baltimore County

Abstract

We investigate the validity of the Girsanov Transformation (GT) method for parametric sensitivity analysis of stochastic models of chemical reaction networks. The validity depends on the likelihood ratio process being a martingale and the commutation of a certain derivative with expectation. We derive some exponential integrability conditions which imply both these requirements. We provide further conditions in terms of a reaction network that imply these exponential integrability conditions.

Keywords: Girsanov transformation, sensitivity analysis, chemical reaction networks, exponential integrability

1 Introduction

Parametric sensitivity analysis is an essential part of modeling and analysis of dynamical systems. In the context of stochastic dynamical systems the problem that is considered frequently is that of estimating the *sensitivity* defined by the partial derivative

$$\frac{\partial}{\partial c}\bigg|_{c=c^*} \mathbb{E}f(X(T,c))$$

where c is a parameter of interest, c^* is its nominal value, X is the stochastic process, f is a scalar function of the state and $T > 0$ is a fixed terminal time.

While we focus on the well-stirred stochastic model of chemical kinetics [6], we like to mention that similar models arise in applications that are concerned with populations (nonnegative integer vectors). Due to the high dimensionality of the state space of the Markov process X describing the chemical kinetics, Monte Carlo methods are usually the most viable. Monte Carlo methods of sensitivity analysis for stochastic chemical models can be classified into the *finite difference* methods (FD) [1, 13], the *Girsanov Transformation* (GT) method [10, 11], the *regularized pathwise derivative* (RPD) method [15] and what might be termed the *auxiliary path* (AP) type methods [7, 8]. When considering more general applications, one again finds roughly, a similar classification [2]. Among these methods, the FD methods are always biased and the RPD method is biased in the context of chemical kinetics and is not always applicable. The well known GT method is usually widely

*tingw@udel.edu
†muruhan@umbc.edu

applicable and is unbiased. The main shortcoming of the GT method is that it has been observed and that it often has large variance and hence less efficient [16]. For an asymptotic analysis of the variance of the GT and related methods, see [16] where a centralized GT method is shown to be more efficient under certain circumstances. The recently introduced auxiliary path type methods are unbiased as well, and provide for an alternative to the GT methods. Nevertheless, due to the ease of implementation, the GT method is of interest.

While the GT method is widely used, we are not aware of theoretical studies on the validity of GT method. In particular, in the area of stochastic reaction networks, no sufficient conditions have been provided to justify the method. Therefore, we believe a theoretical analysis of the method could provide a guideline about the applicability of GT to certain types of problems. Our goal in this manuscript is to provide some sufficient conditions that guarantee the validity of the GT method.

1.1 Stochastic chemical kinetics

We describe the stochastic model of well-stirred chemical reactions involving n molecular species undergoing m reaction channels [6]. In this model, the molecular copy number vector $X(t) \in \mathbb{Z}_+^n$ $(t \geq 0)$ is considered as a Markov process in continuous time. Occurrence of jth reaction leads to a change of $X(t)$ by $\nu_j \in \mathbb{Z}^n$ for $j = 1, \ldots, m$ where the ν_j are known as the *stoichiometric vectors*. We denote by $R_j(t)$ the counting process which counts the number of occurrences of reaction channel j during $(0, t]$, for $j = 1, \ldots, m$. The probabilistic rate of occurrence of reaction j, is given by the *intensity function* or *propensity function* $a_j(x)$ which is defined such that conditioned on $X(t) = x$, the probability that $X(t + h) = x + \nu_j$ is $a_j(x)h + o(h)$ as $h \to 0+$, and the probability that $X(t + h) = x$ is $1 - \sum_{j=1}^m a_j(x)h + o(h)$ as $h \to 0+$.

The propensity function, in addition to the state x, potentially depends on other factors such as the temperature and system volume and this dependence is captured by a set of parameters which are non-random and constant in time. In particular, in the stochastic form of *mass action* case, the propensity function is of the *product form*

$$a_j(x, c) = c_j b_j(x), \tag{1}$$

where $c_j > 0$ is a parameter independent of x and $b_j(x)$ is a (multivariate) polynomial in x [6]. While our final results in this paper assume the product form (1) (but not necessarily the mass action form), we shall keep our derivations as general as possible until the final steps.

It is possible to represent the processes for different parameter values c in the same sample space $(\Omega, \mathcal{F}, \mathbb{P})$ (hence the notation $X(t, c)$ and $R(t, c)$) via the *random time change representation* [5]

$$X(t, c) = x_0 + \sum_{j=1}^{m} \nu_j Y_j \left(\int_0^t a_j(X(s, c), c) ds \right), \tag{2}$$

where $Y_1, \ldots, Y_m$ are independent unit rate Poisson processes carried by $(\Omega, \mathcal{F}, \mathbb{P})$ and $x_0 \in \mathbb{Z}_+^n$ is the initial state. The reaction count processes R_j are then given by

$$R_j(t, c) = Y_j \left(\int_0^t a_j(X(s, c), c) ds \right) \quad j = 1, \ldots, m. \tag{3}$$

We also note the relationship

$$X(t, c) = x_0 + \sum_{j=1}^{m} \nu_j R_j(t, c). \tag{4}$$

We note that the processes $X(t,c)$ and $R(t,c)$ are *cadlag*. We shall also assume that $X(t,c)$ is non-explosive for each c, that is, $R_j(t,c) < \infty$ for each $t \geq 0$, each c and $j = 1, \ldots, m$. For sake of readability, throughout the paper we suppress the dependence of X and R on $\omega \in \Omega$ except when necessary.

Without loss of generality we shall focus on estimation of sensitivity with respect to one scalar parameter c and we assume that it corresponds to the first reaction channel, so that $c = c_1$ and thus $a_1(x,c) = c\, b_1(x)$ under the product form. Let $c^* > 0$ be a *nominal parameter value* and $T > 0$ be some terminal time. Given a function $f : \mathbb{Z}_+^n \to \mathbb{R}$, we are interested estimating the sensitivity defined by

$$\left.\frac{\partial}{\partial c}\right|_{c=c^*} \mathbb{E} f(X(T,c)).$$

Throughout this paper, we assume that the sensitivity exists. We refer the reader to [7] for some sufficient conditions that guarantee the existence of the sensitivity.

1.2 The Girsanov transformation method

One of the commonly used sensitivity estimation methods is the Girsanov transformation (GT) method which is also known as the likelihood ratio (LR) method in literature. We first describe the basics of this approach and then furnish details in the context of chemical kinetics. We consider a nominal parameter value c^* and an open interval $I_{c^*} = (c^* - \epsilon, c^* + \epsilon)$. As mentioned before, we assume that the processes $X(t,c)$ and $R(t,c)$ for $c \in I_{c^*}$ are all carried by a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let us denote by $\{\mathcal{F}_t\}_t$ the filtration generated by $X(t,c^*)$ and $R(t,c^*)$. (We remark that if we assume that the ν_j are all distinct then $\{\mathcal{F}_t\}_t$ will be generated by $X(t,c^*)$ alone).

The GT method involves defining for each $c \in I_{c^*}$ a probability measure $\mathbb{P}_c$ on $(\Omega, \mathcal{F})$ which satisfies the following condition.

Condition 1: For each $c \in I_{c^*}$, (i) $\mathbb{P}_c$ is absolutely continuous with respect to $\mathbb{P}$, (ii) $\mathbb{P}_{c^*} = \mathbb{P}$ and (iii) for every bounded function $f : \mathbb{Z}_+^n \to \mathbb{R}$

$$\mathbb{E} f(X(T,c)) = \mathbb{E}_c f(X(T,c^*)), \tag{5}$$

where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$ and $\mathbb{E}_c$ is the expectation with respect to $\mathbb{P}_c$.

Suppose such a family of probability measures $\mathbb{P}_c$ satisfying Condition 1 exist. Let us denote by $L(c,t)$ the Radon-Nykodim derivative

$$L(t,c) = \left.\frac{d\mathbb{P}_c}{d\mathbb{P}}\right|_{\mathcal{F}_t}. \tag{6}$$

Due to (5), the sensitivity can be written as

$$\left.\frac{\partial}{\partial c}\right|_{c=c^*} \mathbb{E} f(X(T,c)) = \left.\frac{\partial}{\partial c}\right|_{c=c^*} \mathbb{E}_c f(X(T,c^*)) = \left.\frac{\partial}{\partial c}\right|_{c=c^*} \mathbb{E}[f(X(T,c^*))L(T,c)].$$

Condition 2: Suppose that the derivative

$$Z(t,c^*) = \left.\frac{\partial}{\partial c}\right|_{c=c^*} L(t,c)$$

exists almost surely with respect to $\mathbb{P}_{c^*}$ and that the following commutation of derivative and expectation holds:

$$\mathbb{E}\left(\left.\frac{\partial}{\partial c}\right|_{c=c^*} f(X(T,c^*))L(T,c)\right) = \left.\frac{\partial}{\partial c}\right|_{c=c^*} \mathbb{E} f(X(T,c^*))L(T,c). \tag{7}$$

This leads to the formula

$$\left.\frac{\partial}{\partial c}\right|_{c=c^*} \mathbb{E}f(X(T,c)) = \mathbb{E}[f(X(T,c^*))Z(T,c^*)]. \tag{8}$$

Thus if Conditions 1 and 2 are satisfied, the required sensitivity equals the expected value of the random variable $f(X(T,c^*))Z(T,c^*)$ and hence can be estimated via iid sample estimation. Thus $\hat{s}_N$ given by

$$\hat{s}_N = \frac{1}{N}\sum_{i=1}^{N} f(X^{(i)}(T,c^*))Z^{(i)}(T,c^*),$$

where $(X^{(i)}(T,c^*), Z^{(i)}(T,c^*))$ for $i = 1,\ldots,N$ are iid pairs of samples, is the GT estimator for a sample size of N. We note that the simulation is carried out with respect to the probability measure $\mathbb{P}$.

The GT estimator is unbiased but often has large variance unless N is very large [16]. Nevertheless, due to its simplicity, GT has been widely applied for sensitivity analysis in numerous areas such as chemical kinetics and operations research.

While the GT method is widely used, we are not aware of theoretical studies on the validity of GT method. In particular, in the area of stochastic reaction networks, no sufficient conditions have been provided to justify the method. Therefore, we believe a theoretical analysis of the method could provide a guideline about the applicability of GT to certain types of problems. In this paper, we aim to provide sufficient conditions that ensure Conditions 1 and 2 stated above.

2 The validity of change of measure

2.1 Change of intensity

We explore some sufficient conditions that guarantee Condition 1 for the change of measure. Our exposition here is based on the change of intensity theory in Section VI 2 of [3]. We start with the probability space $(\Omega,\mathcal{F},\mathbb{P})$ on which the processes $X(t,c)$ and $R(t,c)$ are defined for $c \in I_{c^*} = (c^*-\epsilon, c^*+\epsilon)$. As before, we denote by $\mathcal{F}_t$ the filtration generated by $X(t,c^*)$ and $R(t,c^*)$. By definition, the counting process $R_j(t,c^*)$ has the $(\mathbb{P},\mathcal{F}_t)$ predictable intensity $a_j(X(t-,c^*),c^*)$. Now for any $c \in I_{c^*}$, we want to explicitly construct a probability measure $\mathbb{P}_c$ on $(\Omega,\mathcal{F})$ such that $R(t,c^*)$ admits the $(\mathbb{P}_c,\mathcal{F}_t)$ predictable intensity $a_j(X(t-,c^*),c)$. This is accomplished by defining the *likelihood ratio* process $L(t,c)$ which under the right conditions will serve as the Radon-Nykodim derivative.

We first define an auxiliary (m dimensional) process $\mu(t,c)$ under a certain assumption on the propensity functions as follows. Given an arbitrary $c \in I_{c^*}$, we assume that for all $x \in \mathbb{Z}_+^n$,

$$a_j(x,c^*) = 0 \text{ if and only if } a_j(x,c) = 0, \quad j = 1,2,\cdots,m. \tag{9}$$

We note that this is assumption holds in particular when the propensity functions are of the product form $a_j(x,c) = c_j b_j(x)$. Now, based on this assumption, the following process is well-defined (c^* is fixed). For each $c \in I_{c^*}$, we define

$$\mu_j(t,c) = \frac{a_j(X(t-,c^*),c)}{a_j(X(t-,c^*),c^*)}, \quad j = 1,\cdots,m. \tag{10}$$

In the case that $a_j(X(t-,c^*),c^*) = 0$, by assumption we have $a_j(X(t-,c^*),c) = 0$ as well, so we can simply define $\mu_j(t,c)$ to be any strictly positive constant. We note that $\mu_j(t,c)$ is $\mathcal{F}_t$-predictable by

its left continuity, and moreover for each $t \geq 0$, we have $0 < \mu_j(t,c) < \infty$ almost surely. We shall make the extra assumption that $\mu_j(t,c)$ is bounded almost surely for each $c \in I_{c^*}$. In the case of the product form of propensity functions with $c = c_1$, we note that the boundedness assumption holds since $\mu_1(t,c) = c/c^*$ and $\mu_j(t,c) = 1$ for $j = 2, \ldots, m$, that is the process $\mu(t,c)$ is a deterministic and constant in t.

Next, following [3], we explicitly define the likelihood ratio process $L(t,c)$ as follows

$$L(t,c) = \prod_{j=1}^{m} \left(\left(\prod_{n=1}^{R_j(t,c^*)} \mu_j(T_j^n, c) \right) \exp\left(\int_0^t (1 - \mu_j(s,c)) a_j(X(s,c^*), c^*)\, ds \right) \right), \tag{11}$$

where T_j^n is the n-th jump time of $R_j(t,c^*)$. By convention, we take the product $\prod_{n=1}^{R_j(t,c^*)}$ to be 1 if $R_j(t,c^*) = 0$. We remark that due to our non-explosivity assumption $R_j(t,c^*)$ is finite almost surely for each t and thus $L(t,c)$ is well defined and satisfies $0 < L(t,c) < \infty$ for each $t \geq 0$..

It can be shown that L defined above is the solution of the equation [3]

$$L(t,c) = 1 + \sum_{j=1}^{m} \int_{(0,t]} L(s-,c)(\mu_j(s,c) - 1) dM_j(s,c^*), \tag{12}$$

where

$$M_j(t,c^*) = R_j(t,c^*) - \int_0^t a_j(X(s,c^*),c^*) ds. \tag{13}$$

From the non-explosivity assumption, we see that for all $t \geq 0$ and $j = 1, \cdots, m$,

$$\int_0^t a_j(X(s,c^*),c^*)\, ds < \infty, \quad \mathbb{P} \quad \text{a.s.}, \tag{14}$$

and hence $M(t,c^*)$ is an m-dimensional local martingale [3]. We summarize some key results from [3] as a lemmas.

Lemma 1. (Bremaud [3], Section VI, Theorem T2) *Under the non-explosivity assumption (with respect to $\mathbb{P}$), for each $c \in I_{c^*}$, $L(t,c)$ is a $(\mathbb{P}, \mathcal{F}_t)$ nonnegative local martingale and hence a $(\mathbb{P}, \mathcal{F}_t)$ supermartingale.*

Lemma 2. (Bremaud [3], Section VI, Theorem T3) *Suppose that $\mathbb{E}L(T,c) = 1$. Then $L(t,c)$ is a $(\mathbb{P}, \mathcal{F}_t)$ martingale over $[0,T]$. Moreover, defining the probability measures $\mathbb{P}_c$ by the condition*

$$\frac{d\mathbb{P}_c}{d\mathbb{P}} = L(t,c),$$

it follows that over $[0,T]$, $R_j(t,c^*)$ has the $(\mathbb{P}_c, \mathcal{F}_t)$-intensity

$$a_j(X(t-,c^*),c) = \mu_j(t,c) a_j(X(t-,c^*),c^*).$$

Corollary 1. *Under the conditions of Lemma 2, for each bounded function $f : \mathbb{Z}_+^n \to \mathbb{R}$ and each $c \in I_{c^*}$*

$$\mathbb{E}f(X(T,c)) = \mathbb{E}_c f(X(T,c^*)). \tag{15}$$

In other words, the $\mathbb{P}_c$ law of $X(t,c^*)$ is the same as the $\mathbb{P}$ law of $X(t,c)$.

Proof. The law of a Markov process is uniquely determined by the Kolmogorov's forward equation and the Kolmogorov's forward equations for $p_1(x,t) = \mathbb{P}(X(t,c) = x)$ and $p_2(x,t) = \mathbb{P}_c(X(t,c^*) = x)$ are identical:

$$\frac{dp_i(x,t)}{dt} = \sum_{j=1}^{m}(p_i(x-\nu_j,t)a_j(x-\nu_j,c) - p_i(x,t)a_j(x,c)), \quad x \in \mathbb{Z}_+^n, \;\; i = 1,2.$$

□

In summary, the condition $\mathbb{E}(L(T,c)) = 1$ in Lemma 2 ensures the validity of the change of measure. In the next section we provide some sufficient conditions for it to hold.

2.2 Novikov type condition

In this section, we provide a Novikov type sufficient condition to ensure that $L(t,c)$ is a martingale over $[0,T]$ or equivalently $\mathbb{E}(L(T,c)) = 1$. Under the product form of propensities $a_j(x,c) = c_j b_j(x)$, and without loss of generality, taking $c = c_1$, the likelihood ratio $L(t,c)$ can be written as

$$L(t,c) = \left(\frac{c}{c^*}\right)^{R_1(t,c^*)} \exp\left(\int_0^t (c^* - c) b_1(X(s,c^*))\, ds\right). \tag{16}$$

We make the following useful observation. We have

$$L(t,c) \leq \left(\frac{c}{c^*}\right)^{R_1(t,c^*)} \tag{17}$$

for any $c \in I_{c^*}^+ = [c^*, c^* + \epsilon)$ and

$$L(t,c) \leq \exp\left(\int_0^t (c^* - c) b_1(X(s,c^*))\, ds\right) \tag{18}$$

for any $c \in I_{c^*}^- = (c^* - \epsilon, c^*]$. This simple observation turns out to be useful for our analysis.

Theorem 1. *Given $c \in I_{c^*}^+$, suppose that*

$$\mathbb{E}\left[\left(\frac{c}{c^*}\right)^{R_1(T,c^*)}\right] < \infty, \tag{19}$$

then $L(t,c)$ is a $(\mathbb{P}, \mathcal{F}_t)$ martingale over $[0,T]$.

Proof. By Lemma 1 $L(t,c)$ is a local martingale. Thus there exists a sequence (σ_n) of increasing stopping times with $\sigma_n \uparrow \infty$ such that $L(t \wedge \sigma_n, c)$ is a $(\mathbb{P}, \mathcal{F}_t)$ martingale for each n. Define the stopping times

$$\tau_n = \inf\{t \geq 0 \mid R_1(t,c^*) \geq n\}.$$

By the non-explosivity assumption, $\tau_n \uparrow \infty$. We define the stopped processes

$$L_n(t,c) = L(t \wedge \sigma_n \wedge \tau_n, c).$$

Now for each n, as a stopped martingale, $L_n(t,c)$ is a $\mathcal{F}_t$ martingale and hence $\mathbb{E}L_n(T,c) = 1$. By the estimates in (17),

$$L_n(T,c) \leq \left(\frac{c}{c^*}\right)^{R_1(T\wedge\sigma_n\wedge\tau_n,c^*)} \leq \left(\frac{c}{c^*}\right)^{R_1(T,c^*)}.$$

Hence, the integrability condition (19) implies that $\mathbb{E}L(T,c) = 1$ by the dominated convergence theorem and therefore $L(t,c)$ is a martingale over $[0,T]$. □

Similar result can be reproduced for the case $c \in I^-_{c^*}$ using the estimates (18).

Theorem 2. *Given $c \in I^-_{c^*}$, suppose that*

$$\mathbb{E}\left[\exp\left((c^* - c)\int_0^T b_1(X(s,c^*))\,ds\right)\right] < \infty, \tag{20}$$

then $L(t,c)$ is a $(\mathbb{P}, \mathcal{F}_t)$ martingale over $[0,T]$.

Proof. Define the stopping times τ_n by

$$\tau_n = \inf\left\{t \geq 0 \,|\, \int_0^t b_1(X(s,c^*))ds \geq n\right\}.$$

The rest of the proof is similar to that of Theorem 1. □

One can get rid of the time integral by verifying the following stronger condition.

Corollary 2. *If there exists $\epsilon > 0$ such that*

$$\sup_{s \leq T} \mathbb{E}\left[e^{\epsilon T b_1(X(s,c^*))}\right] < \infty, \tag{21}$$

then $L(t,c)$ is a $(\mathbb{P}, \mathcal{F}_t)$ martingale over $[0,T]$ for any $c \in I^-_{c^*} = (c^* - \epsilon, c^*)$.

Proof. Rearranging the right hand side of (20) and applying Jensen's inequality to the term inside the bracket, we obtain

$$\mathbb{E}\left[\exp\left(\frac{1}{T}\int_0^T (c^* - c)T b_1(X(s,c^*))\,ds\right)\right] \leq \frac{1}{T}\int_0^T \mathbb{E}\left[e^{(c^*-c)T b_1(X(s,c^*))}\right] ds. \tag{22}$$

Hence, it suffices to show that

$$\sup_{s \leq T} \mathbb{E}\left[e^{(\epsilon T b_1(X(s,c^*)))}\right]$$

is finite for ϵ satisfying $\epsilon > c^* - c$. □

3 Differentiation inside the integral

In this section, we provide a sufficient condition for the commutation (7) of Condition 2 via the use of Theorem 6 in appendix. Referring to Theorem 6, we take G to be

$$G(c) = f(X(T,c^*))L(T,c).$$

We shall assume the product form with $c = c_1$. Then it follows at once from (11) that

$$\frac{\partial}{\partial c}\ln L(T,c) = \frac{1}{c}R_1(T,c^*) - \int_0^T b_1(X(s,c^*))\,ds,$$

hence

$$\frac{\partial}{\partial c}L(T,c) = L(T,c)\left(\frac{1}{c}R_1(T,c^*) - \int_0^T b_1(X(s,c^*))\,ds\right) = \frac{1}{c}L(T,c)M_1(T,c^*).$$

Then a Lipschitz constant $K(\omega)$ (independent of c) for G on the interval I_{c^*} is

$$K = |f(X(T,c^*))M_1(T,c^*)| \sup_{c\in I_{c^*}} \frac{1}{c} L(T,c).$$

We first consider $c \in I_{c^*}^+$ (the right hand sensitivity) , in which case we have

$$L(T,c) \le \left(\frac{c}{c^*}\right)^{R_1(T,c^*)} \le \left(\frac{c^*+\epsilon}{c^*}\right)^{R_1(T,c^*)}.$$

Hence in order to justify the integrability of K, it suffices to show that

$$f(X(T,c^*))M_1(T,c^*)\left(\frac{c^*+\epsilon}{c^*}\right)^{R_1(T,c^*)}$$

is integrable. We also note that we may shrink the interval $I_{c^*} = (c^*-\epsilon, c^*+\epsilon)$ to be as small as we wish.

Theorem 3. *Assuming the product form* (1) *with* $c = c_1$*, suppose the following conditions are satisfied:*

- $\mathbb{E}[|f(X(T,c^*))|^3] < \infty$*;*
- *there exists* $\epsilon > 0$ *such that,*

$$\mathbb{E}\left[\left(\frac{c^*+\epsilon}{c^*}\right)^{R_1(T,c^*)}\right] < \infty. \tag{23}$$

Then

$$\lim_{c\to c^{*+}} \frac{\mathbb{E}f(X(T,c)) - \mathbb{E}f(X(T,c^*))}{c-c^*} = \mathbb{E}\left[f(X(T,c^*)) \lim_{c\to c^{*+}} \frac{L(T,c)-L(T,c^*)}{c-c^*}\right].$$

Proof. First we note that by Theorem 1, (23) implies the validity of the change of measure for $c \in I_{c^*}^+$.

Now we need to verify the limit using Theorem 6. Using the inequality $3abc \le a^3+b^3+c^3$, we can separate the terms and provide the following sufficient conditions,

$$\begin{aligned} &\mathbb{E}[|f(X(T,c^*))|^3] < \infty, \\ &\mathbb{E}[|M_1(T,c^*)|^3] < \infty, \\ &\mathbb{E}\left[\left(\frac{c^*+\epsilon}{c^*}\right)^{3R_1(T,c^*)}\right] < \infty. \end{aligned} \tag{24}$$

It is sufficient to show that the third condition implies the second condition. Since the quadratic variation of the local martingale $M_1(t,c^*)$ is $R_1(t,c^*)$, by the Burkholder-Davis-Gundy (BDG) inequality [12],

$$\mathbb{E}(|M_1(T,c^*)|^4) \le C\mathbb{E}[R_1(T,c^*)^2].$$

for some constant C. It is obvious that the right hand side is integrable given the second condition. Since $|M_1(T,c^*)|^4$ is integrable the result follows. □

Similarly, for the left hand side sensitivity, we have

$$L(T,c) \le \exp\left(\int_0^T (c^* - c) b_1(X(s,c^*))\, ds\right) \le \exp\left(\int_0^T \epsilon b_1(X(s,c^*))\, ds\right)$$

for $c \in I_{c^*}^-$. Hence, the Lipschitz constant is proportional to

$$|f(X(T,c^*))M_1(T,c^*)| \exp\left(\int_0^T \epsilon b_1(X(s,c^*))\, ds\right).$$

It boils down to verifying the following three integrability conditions,

$$\begin{aligned} &\mathbb{E}[|f(X(T,c^*))|^3] < \infty, \\ &\mathbb{E}[|M_1(T,c^*)|^3] < \infty, \\ &\mathbb{E}\left[\exp\left(\int_0^T 3\epsilon b_1(X(s,c^*))\, ds\right)\right] < \infty. \end{aligned} \tag{25}$$

We have the following result concerning the left hand side sensitivity.

Theorem 4. *Assuming the product form* (1) *with* $c = c_1$*, suppose further that*

- $\mathbb{E}[|f(X(T,c^*))|^3] < \infty$*;*
- *there exists* $\epsilon > 0$ *such that,*

$$\mathbb{E}\left[\exp\left(\int_0^T \epsilon b_1(X(s,c^*))\, ds\right)\right] < \infty. \tag{26}$$

Then

$$\lim_{c \to c^{*-}} \frac{\mathbb{E}f(X(T,c)) - \mathbb{E}f(X(T,c^*))}{c - c^*} = \mathbb{E}\left[f(X(T,c^*)) \lim_{c \to c^{*-}} \frac{L(T,c) - L(T,c^*)}{c - c^*}\right].$$

Proof. We only need to show the second integrability condition in (25) holds. Note that $M_1(t,c^*)$ in the second term is a local martingale, we apply the (BDG) inequality such that

$$\mathbb{E}\{|M_1(T,c^*)|^4\} \le C\mathbb{E}[R_1(T,c^*)^2]$$

for some constant C. Hence, it is sufficient to verify that $\mathbb{E}[R_1(T,c^*)^2] < \infty$. Applying the BDG inequality again to $\mathbb{E}[M_1(T,c^*)^2]$, there exists some constant $\bar{C}$ such that

$$\mathbb{E}[M_1(T,c^*)^2] \le \bar{C}\mathbb{E}[R_1(T,c^*)] = \bar{C}\mathbb{E}\left[\int_0^T a_1(X(s,c^*),c^*)\, ds\right] < \infty.$$

Owing to the simple inequality $(a+b)^2 \le 2(a^2+b^2)$,

$$\mathbb{E}[R_1(T,c^*)^2] \le 2\mathbb{E}\left[\left(\int_0^T a_1(X(s,c^*),c^*)\, ds\right)^2\right] + 2\mathbb{E}[M_1(T,c^*)^2] < \infty.$$

□

Remark 1. *We note that the conditions of Theorem 3 guarantee the existence of the right hand derivative (sensitivity) and that the GT method would provide an unbiased estimator of it. Likewise for Theorem 4. However, the conditions of these theorems include the restrictive exponential integrability conditions* (23) *and* (26)*. We do know from [7] that the existence of sensitivity can be guaranteed under milder conditions. Thus, if we assume the existence of sensitivity at* c^**, then verification of either the conditions of Theorem 3 or those of Theorem 4 can guarantee the validity of the GT method. This will be our focus in the next section.*

4 Sufficient conditions in terms of the network

The conditions of of Theorems 3 and 4 are not directly stated in terms of a chemical reaction network. A chemical reaction network is characterized by the stoichiometric matrix ν (whose columns are the vectors ν_j) and the propensity functions $a_j(x,c)$ which we assume to be of the product form (1) and we take the parameter of interest to be $c = c_1$ at a nominal value $c^* > 0$.

We shall focus on the case that $f : \mathbb{Z}_+^n \to \mathbb{R}$ is of polynomial growth. This may be stated by the condition that there exists $C > 0$ and $r \in \mathbb{Z}_+$ such that

$$|f(x)| \leq C(1 + \|x\|^r) \quad \forall x \in \mathbb{Z}_+^n.$$

In this case, there exist multiple results in the literature that guarantee the condition that $|f(X(T, c^*)|^3$ is integrable [14, 9, 4].

On the other hand, the exponential integrability conditions (23) or (26) are harder to satisfy. When $b_1(x)$ is linear, the condition (21) is implied by the conditions for the *uniform light-tailedness* property presented in [9], and since (21) implies (26), this provides a sufficient condition for the validity of the GT method.

In this section, in Theorem 5, we present a condition that implies (23) and hence provides another sufficient condition for the validity of the GT method.

Given a fixed initial state $x_0 \in \mathbb{Z}_+^n$ let $S_{x_0} \subset \mathbb{Z}_+^n$ denote the set of all states that can be reached by the process starting at x_0. Thus S_{x_0} is the effective state space of the process $X(t,c)$ and it may be finite or infinite.

We present a useful lemma that plays an important role in our result.

Lemma 3. *Let $J \subset \{1, \ldots, m\}$ be a subset of reaction channels and suppose there exists $K > 0$ such that the propensity $a_j(x,c)$ of any reaction $j \in J$ satisfies*

$$a_j(x,c) \leq K \quad \forall x \in S_{x_0}.$$

Then $\mathbb{E}(e^{\epsilon \sum_{j\in J} R_j(t,c)}) < \infty$ for every $\epsilon > 0$ and $t \geq 0$.

Proof. By the random time change representation, for $j \in J$

$$R_j(t,c) = Y_j\left(\int_0^t a_j(X(s,c),c)ds\right) \leq Y_j(Kt).$$

Thus $\sum_{j\in J} R_j(t,c) \leq \sum_{j\in J} Y_j(Kt)$ and hence

$$\mathbb{E}(e^{\epsilon \sum_{j\in J} R_j(t,c)}) \leq \prod_{j\in J} \mathbb{E}(e^{\epsilon Y_j(Kt)}).$$

The result follows from the fact that $\mathbb{E}(e^{\epsilon Y_j(Kt)}) < \infty$ for every $\epsilon > 0$. □

It is clear from this lemma that if reaction 1 has bounded propensity on S_{x_0} then the required bound (23) follows at once.

Even if reaction 1 does not satisfy a propensity bound on S_{x_0}, it may be possible to bound $R_1(t)$ above in terms of some other reactions which have bounded propensities. As a motivating example, let us consider the chemical kinetics example with two species and four reactions. Let

$$\nu_1 = (-1,1)^T, \;\; \nu_2 = (-1,-1)^T, \;\; \nu_3 = (1,2)^T, \;\; \nu_4 = (1,0)^T.$$

We make explicit use of the fact that species population process $X(t)$ remains nonnegative. The process at any time $t \geq 0$ satisfies (4)

$$X(t) = x_0 + \nu R(t)$$

and since $X(t) \geq 0$ we have that $x_0 + \nu R(t) \geq 0$. We readily see that

$$R_1(t) \leq R_1(t) + R_2(t) \leq x_{0,1} + R_3(t) + R_4(t).$$

Suppose further that the reactions 3 and 4 have bounded propensities on S_{x_0}. That is, there exists $K > 0$ such that

$$a_j(x, c^*) \leq K \quad j = 1, 2, \ \ \forall x \in S_{x_0}.$$

Now Lemma 3 readily implies that for all $\epsilon > 0$ and $t \geq 0$

$$\mathbb{E}(e^{\epsilon R_1(t)}) < \infty.$$

This example suggests the possibility that the relation $x_0 + \nu R(t) \geq 0$ may imply that $R_1(t)$ is bounded above in terms of a positive affine combination of some other reactions which have bounded propensities on S_{x_0}. In general, this determination could be made as follows.

Let $\mathcal{B} \subset \{1, \ldots, m\}$ denote the indices j of the reactions that have bounded propensities on S_{x_0} and suppose $1 \notin \mathcal{B}$ (otherwise the result follows immediately by Lemma 3). The fact that $X(t) \geq 0$ can be expressed by

$$x_0 + \sum_{j \notin \mathcal{B}} \nu_j R_j(t) + \sum_{j \in \mathcal{B}} \nu_j R_j(t) \geq 0.$$

For $j \in \mathcal{B}$ let $\mu_j \in \mathbb{R}^n$ be defined by $(\mu_j)_i = \max\{0, \nu_{i,j}\}$. Then $X(t) \geq 0$ implies

$$x_0 + \sum_{j \notin \mathcal{B}} \nu_j R_j(t) + \sum_{j \in \mathcal{B}} \mu_j R_j(t) \geq 0.$$

Letting

$$y = x_0 + \sum_{j \in \mathcal{B}} \mu_j R_j(t)$$

and noting that $y \geq x_0$, motivates the linear programming (feasibility) problem:

$$y + \sum_{j \notin \mathcal{B}} \nu_j \xi_j \geq 0, \ \ \xi \geq 0, \tag{27}$$

where $\xi \in \mathbb{R}^n$. The feasible region for ξ is given by a convex polytope $\mathcal{R}_y$ which may be unbounded. If $\mathcal{R}_y$ is bounded in the ξ_1 direction, then one can obtain an upper bound for $R_1(t)$ as an affine combination of $R_j(t)$ for $j \in \mathcal{B}$. We note that whether $\mathcal{R}_y$ is bounded in the ξ_1 direction or not depends only on ν_j for $j \notin \mathcal{B}$ and not on y.

Then Lemma 3 can be used to obtain (23). We summarize this discussion as a theorem.

Theorem 5. *Given a non-explosive chemical reaction network with product form propensity functions, suppose that either reaction* 1 *has bounded propensity on* S_{x_0} *or the feasible region of the linear program* (27) *is bounded in the first coordinate* ξ_1*. Then* (23) *holds for every* $\epsilon > 0$*.*

We remark that if S_{x_0} is finite, then the validity of the GT method follows trivially.

5 Appendix

5.1 Differentiating Inside an Integral

Theorem 6. (Asmussen & Glynn) [2]) *Suppose $G(c,\omega)$ is a random variable for each c in some interval of the real line. Let c_{ref} be a specific value of c. Suppose the following hold:*

1. *For a set of ω with probability one, $G(c,\omega)$ is differentiable with respect to c at $c = c_{ref}$.*
2. *There exists an interval (c_l, c_u) containing c_{ref} (independent of ω) on which $G(c,\omega)$ is Lipschitz (in c) for a set of ω with probability one, with constant K which may depend on ω. That is, for any c_1, c_2 in the interval (c_l, c_u), the following holds:*

$$|G(c_1,\omega) - G(c_2,\omega)| \leq K(\omega)|c_1 - c_2|.$$

3. *$\mathbb{E}(K)$ is finite.*
4. *$\mathbb{E}(|G(c,\omega)|)$ is finite for all c in (c_l, c_u).*

Then the following holds:

$$\left.\frac{d}{dc}\right|_{c=c_{ref}} \mathbb{E}(G(c)) = \mathbb{E}\left(\left.\frac{d}{dc}\right|_{c=c_{ref}} G(c)\right).$$

References

[1] D. F. Anderson. An efficient finite difference method for parameter sensitivities of continuous time Markov chains. *SIAM J. Numer. Anal.*, 50(5):2237–2258, 2012.

[2] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.

[3] Pierre Brémaud. Point processes and queues: martingale dynamics. 1981.

[4] Stefan Engblom. On the stability of stochastic jump kinetics. *Applied Mathematics*, 5(19):3217–3239, 2014.

[5] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, Inc., New York, second edition, 2005.

[6] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[7] A. Gupta and M. Khammash. Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks. *SIAM J. Sci. Comput.*, 35(6):A2598–A2620, 2013.

[8] A. Gupta and M. Khammash. An efficient and unbiased method for sensitivity analysis of stochastic reaction networks. *J. R. Soc. Interface*, (101):20140979, 2014.

[9] Ankit Gupta, Corentin Briat, and Mustafa Khammash. A scalable computational framework for establishing long-term behavior of stochastic reaction networks. *PLoS computational biology*, 10(6):e1003669, 2014.

[10] S. Plyasunov and A. P. Arkin. Efficient stochastic sensitivity analysis of discrete event systems. *J. Comput. Phys.*, 221(2):724–738, 2007.

[11] Sergey Plyasunov and Adam P Arkin. Efficient stochastic sensitivity analysis of discrete event systems. *Journal of Computational Physics*, 221(2):724–738, 2007.

[12] Philip E Protter. *Stochastic differential equations*. Springer, 2005.

[13] M. Rathinam, P. W. Sheppard, and M. Khammash. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *J. Chem. Phys.*, 132:034103, 2010.

[14] Muruhan Rathinam. Moment growth bounds on continuous time markov processes on non-negative integer lattices. *Quart. Appl. Math.*, 53(2):347–364, 2015.

[15] P. W. Sheppard, M. Rathinam, and M. Khammash. A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems. *J. Chem. Phys.*, 136:034115, 2012.

[16] Ting Wang and Muruhan Rathinam. Efficiency of the Girsanov transformation approach for parametric sensitivity analysis of stochastic chemical kinetics. *JUQ*, 4(1):1288–1322, 2016.