

## Research Article

# Perturbation and Truncation of Probability Generating Function Methods for Stiff Chemical Reactions

Soyeong Jeong, Pilwon Kim, and Chang Hyeong Lee

Department of Mathematical Sciences, School of Natural Science, Ulsan National Institute of Science and Technology (UNIST), Ulsan Metropolitan City 689-798, Republic of Korea

Correspondence should be addressed to Pilwon Kim; [pwkim@unist.ac.kr](mailto:pwkim@unist.ac.kr) and Chang Hyeong Lee; [chlee@unist.ac.kr](mailto:chlee@unist.ac.kr)

Received 15 September 2014; Revised 3 February 2015; Accepted 6 February 2015

Academic Editor: Lotfollah Najjar

Copyright © 2015 Soyeong Jeong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One can reformulate chemical master equations of the stochastic reaction network into a partial differential equation (PDE) of a probability generating function (PGF). In this paper, we present two improvements in such PGF-PDE approach, based on perturbation and double-truncation, respectively. The stiff system that involves fast and slow reactions together often requires high computational cost. By applying the perturbation method to PGF-PDEs, we expand the equation in terms of a small reaction rate which is often responsible for such stiffness of the system. Also by doubly truncating, we dump relatively small terms and reduce the computational load significantly at each time step. The terms corresponding to rare events are sieved out through truncations of Taylor expansion. It is shown through numerical examples of enzyme kinetics, transition model, and Brusselator model that the suggested method is accurate and efficient for approximation of the state probabilities.

## 1. Introduction

Stochastic chemical kinetics arises mostly in the modeling of chemical reactions in relatively small biochemical systems such as gene regulatory networks, cell systems, and protein folding models [1]. The governing equation, chemical master equation (CME), describes the probability distribution of the system under the Markov assumption as

$$\begin{aligned} \frac{\partial}{\partial t} p(\mathbf{n}, t) = & \sum_k a_k(\mathbf{n} - V_k) \cdot p(\mathbf{n} - V_k, t) \\ & - \sum_k a_k(\mathbf{n}) \cdot p(\mathbf{n}, t), \end{aligned} \quad (1)$$

where  $\mathbf{n}(t) = (n_1(t), n_2(t), \dots, n_s(t))$ , each  $n_i(t)$  denotes the molecular number of  $i$ th species at time  $t$ ,  $a_k$  is the propensity function or probability intensity for the  $k$ th reaction, and  $V_k$  denotes the  $k$ th column of the stoichiometric matrix  $V$  of which  $(i, j)$ th entry is the change in the number of molecules of the  $i$ th species by the occurrence of the  $j$ th reaction [2].

Since (1) is a high dimensional system in most real applications, it is very challenging to handle both analytically and computationally. One of the key methods to help simplify

the computation of the stochastic dynamics is done by using the stochastic simulation algorithm (SSA) [3]. One realization of the SSA shows the trajectory of the time-evolution of the states. Since it is basically a Monte-Carlo type algorithm, one has to perform many realizations to obtain important statistical quantities such as mean and variance. Moreover, it often requires a large number of realizations, especially in cases of the systems consisting of large numbers of molecules or very different time scale reactions. There have been many approaches to improve the SSA including Tau-leaping methods [4–6], reduction methods on the slow time scale [7–10], and moment closure methods [11–13].

In the recent papers [14, 15], the authors presented numerical schemes based on probability generating functions (PGF) for finding the probability distribution and moments. One can reformulate chemical master equations of the stochastic reaction network into a partial differential equation (PDE). Although such PDEs are mostly hard to deal with due to variable coefficients and lack of proper boundary conditions, several improvements in approximation of the moments based on Pade approximation and numerical superimposition have been made.

In this paper, we present two improvements in such PGF-PDE approach. The first method is an approximation method called “perturbed probability generating function (PPGF) method.” By applying the perturbation method to PDEs, we expand the equation in terms of a small reaction rate which is often responsible for the stiffness of the system. This enables us to handle the system more efficiently with less iteration and therefore capture same accuracy faster.

We also develop the double truncation method (DTM) that efficiently approximates the state probabilities. As suggested from the name, the method combines two truncation procedures for the series solution to PGF-PDEs.

Through consecutive truncations of the Taylor expansion of such PDEs in time, one can dump relatively rare events which correspond to excessive occurrence of reactions during a short time interval. This elimination is justified from the fact that the coefficient of each term in the PGF is the probability of a specific state. This enables us to ignore a great deal of small terms which do not affect the system significantly.

An outline of the paper is as follows. In Section 2, we review the properties of PDEs that PGF should satisfy. The perturbation of PGF is discussed in Section 3 with its numerical accuracy of through some examples. Section 4 presents the double truncation method for the PGF-PDEs. In Sections 3 and 4, we also illustrate numerical accuracy of the method in several examples.

## 2. PGF-PDE Approach

The probability generating function is defined as

$$G(\mathbf{z}, t) = \sum_{\mathbf{n}=0}^{\infty} \mathbf{z}^{\mathbf{n}} p(\mathbf{n}, t), \quad (2)$$

where  $\mathbf{z} = (z_1, \dots, z_s)$ ,  $z_i \in [-1, 1]$ ,  $\mathbf{n} = (n_1, n_2, \dots, n_s)$ , and

$$\mathbf{z}^{\mathbf{n}} \equiv z_1^{n_1} z_2^{n_2} \cdots z_s^{n_s}. \quad (3)$$

After differentiation of (2) with respect to  $t$  and application of (1), one can derive a PDE

$$\frac{\partial G}{\partial t} = \sum_{|\mathbf{k}| \leq m} H_{\mathbf{k}}(\mathbf{z}, t) D^{\mathbf{k}} G. \quad (4)$$

Here  $\mathbf{k} = (k_1, \dots, k_s)$ ,  $|\mathbf{k}| = k_1 + \dots + k_s$ ,  $m$  is the highest order of the reactions in the chemical system, and  $D^{\mathbf{k}} G$  denotes any  $k$ th order partial derivatives of  $G$  as

$$D^{\mathbf{k}} G = \frac{\partial^{k_1}}{\partial z_1^{k_1}} \frac{\partial^{k_2}}{\partial z_2^{k_2}} \cdots \frac{\partial^{k_s}}{\partial z_s^{k_s}} G. \quad (5)$$

One can see that the PDE (4) is linear and  $H_{\mathbf{k}}$  is a polynomial function of degree at most order  $m$  which satisfies  $H_{\mathbf{k}}(\mathbf{z} = \mathbf{1}, t) = 0$  for  $|\mathbf{k}| = 0, 1, \dots, m$ .

The conditions of the PDE (4) can be found from the conditions on  $n$  [14]. The initial condition is

$$G(\mathbf{z}, t = 0) = \mathbf{z}^{\mathbf{n}_0}, \quad (6)$$

where  $\mathbf{n}_0$  is the initial condition of  $\mathbf{n}$ , and also it is true that

$$\begin{aligned} G(\mathbf{z} = \mathbf{0}, t) &= p(\mathbf{n} = \mathbf{0}, t), \\ G(\mathbf{z} = \mathbf{1}, t) &= \sum_{\mathbf{n}} p(\mathbf{n}, t) = 1. \end{aligned} \quad (7)$$

Using the solution  $G$  of the PDE (4), one can find the important information for stochastic reaction network such as marginal probability, mean, and the second moment [14]. If  $N_i(t)$  denotes the number of molecules of  $i$ th species at time  $t$ , the probability that  $N_i = \ell$  at time  $t$  is

$$p(N_i = \ell, t) = \frac{1}{\ell!} \left. \frac{\partial^{\ell} G(\mathbf{z}, t)}{\partial z_i^{\ell}} \right|_{z_i=0, z_j=1, j \neq i}, \quad (8)$$

and the mean and second moment are obtained by the first and second derivatives of  $G$

$$\begin{aligned} \frac{\partial}{\partial z_i} G(\mathbf{z} = \mathbf{1}, t) &= E[N_i(t)], \\ \frac{\partial^2}{\partial z_i \partial z_j} G(\mathbf{z} = \mathbf{1}, t) &= \begin{cases} E[N_i(t) N_j(t)], & \text{if } i \neq j, \\ (E[N_i^2(t)] - E[N_i(t)]^2), & \text{if } i = j, \end{cases} \end{aligned} \quad (9)$$

where  $E[X(t)]$  denotes the expectation of  $X(t)$ .

## 3. Perturbation of Probability Generating Function Method

**3.1. Perturbation of PGF.** For the PGF method recently proposed in [14],  $G(\mathbf{z}, t)$  is represented by a power expansion with respect to  $t$  as

$$G(\mathbf{z}, t) = \sum_{n=0}^{\infty} f_n(\mathbf{z}) t^n, \quad (10)$$

where  $f_n$ ,  $n = 0, 1, \dots$ , are coefficient functions of  $\mathbf{z}$ . Note that the initial condition determines the first coefficient as  $f_0(\mathbf{z}) = \mathbf{z}^{\mathbf{n}_0}$ .

Suppose that the reaction rates in the network do not change with time and therefore the coefficients of the PDE do not involve  $t$  explicitly. This implies  $H_{\mathbf{k}}(\mathbf{z}, t) = H_{\mathbf{k}}(\mathbf{z})$  in (4). Plugging (10) into the both sides of (4) and comparing the power series lead to a recursive relation between  $f_{n+1}(\mathbf{z})$  and  $f_n(\mathbf{z})$ ,  $n = 0, 1, \dots$ . One can derive coefficient functions  $f_1(\mathbf{z}), f_2(\mathbf{z}), \dots, f_J(\mathbf{z})$  from  $f_1(\mathbf{z}) = \mathbf{z}^{\mathbf{n}_0}$  and then construct an approximation  $\sum_{n=0}^J f_n(\mathbf{z}) t^n$  for sufficiently large  $J$ . Since the initial function  $f_0(\mathbf{z})$  and the coefficient of the PDE (2) are polynomials, all  $f_n(\mathbf{z})$  generated are polynomials as well.

Now, let us apply the perturbation method to PGF. Since many practical chemical networks that interest us involve reactions in a broad range of time scales, from seconds to microseconds, it is important to efficiently handle such stiffness in simulation of the systems. Suppose that  $c \ll 1$  is

a small reaction rate in the system; then one can expand the PGF as

$$G(\mathbf{z}, t) = g_0(\mathbf{z}, t) + c g_1(\mathbf{z}, t) + c^2 g_2(\mathbf{z}, t) + \cdots, \\ = \sum_{j=0}^{\infty} c^j g_j(\mathbf{z}, t), \quad (11)$$

where  $g_j(\mathbf{z}, t)$ ,  $j = 0, 1, 2, \dots$ , are functions to be determined. As in (10), one can set for each  $g_j$

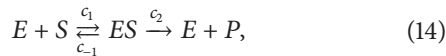
$$g_j(\mathbf{z}, t) = \sum_{k=0}^{\infty} f_{j,k}(\mathbf{z}) t^k. \quad (12)$$

Now, instead of (10), one can seek  $G(\mathbf{z}, t)$  in the following form:

$$G(\mathbf{z}, t) = \sum_{j=0}^{\infty} c^j \sum_{k=0}^{\infty} f_{j,k}(\mathbf{z}) t^k \\ = \sum_{j=0}^M c^j \sum_{k=0}^N f_{j,k}(\mathbf{z}) t^k + O(c^{M+1}, t^{N+1}) \quad (13)$$

for large  $M$  and  $N$ . By plugging (13) into the both sides of (4) and comparing the power series, one can find recurrence relations for  $f_{j,k}$ ,  $f_{j+1,k}$  and  $f_{j+1,k+1}$ , where  $j, k = 0, 1, 2, \dots$ . This decomposition enables us to handle stiffness caused by  $c$  properly, by adjusting the perturbation order  $M$  as will be shown in the numerical examples in the next sections.

**3.2. Example 1: Enzyme Kinetics.** As the first application, we consider a fundamental chemical kinetic model, the enzyme-substrate reaction system



where  $E$ ,  $S$ ,  $ES$ , and  $P$  denote enzyme, substrate, enzyme-substrate complex, and product, respectively, and  $c_1$ ,  $c_{-1}$ , and  $c_2$  are probability constants for reactions. If we denote the molecular numbers of  $E$ ,  $S$ ,  $ES$ , and  $P$  by  $n_1$ ,  $n_2$ ,  $n_3$ , and  $n_4$ , respectively, the governing equation of the stochastic enzyme-substrate system is obtained [16] as

$$\frac{\partial p(n_1, n_2, n_3, n_4, t)}{\partial t} \\ = c_1 (n_1 + 1) (n_2 + 1) \\ \cdot p(n_1 + 1, n_2 + 1, n_3 - 1, n_4, t) \\ + c_{-1} (n_3 + 1) p(n_1 - 1, n_2 - 1, n_3 + 1, n_4, t) \\ + c_2 (n_3 + 1) p(n_1 - 1, n_2, n_3 + 1, n_4 - 1, t) \\ - (c_1 n_1 n_2 + c_{-1} n_3 + c_2 n_3) p(n_1, n_2, n_3, n_4, t). \quad (15)$$

One can see that there are two conservation quantities in the system and thus we can remove two of variables, say  $n_3, n_4$ .

After the removal, we can derive a PDE of  $G(z_1, z_2, t)$  from the master equation [14]

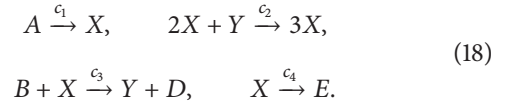
$$G_t = c_1 (1 - z_1 z_2) G_{12} \\ + ((c_{-1} + c_2) z_1 - c_{-1} z_1^2 z_2 - c_2 z_1^2) G_1 \\ + A (c_{-1} z_1 z_2 + c_2 z_1 - (c_{-1} + c_2)) G. \quad (16)$$

Here  $A$  is the conserved quantity  $n_1 + n_3$  and we use the notations

$$G_i = \frac{\partial G}{\partial z_i}, \quad G_{ij} = \frac{\partial^2 G}{\partial z_i \partial z_j}, \quad i, j = 1, 2, \dots \quad (17)$$

In the example, we use parameters  $c_1 = 0.01s^{-1}$ ,  $c_{-1} = 1s^{-1}$ , and  $c_2 = 1s^{-1}$  and let the perturbation be expanded in  $c_1$ . Figure 1 shows comparison of the exact solution of the master equation (15) and the approximate solution obtained from the PPGF method.

**3.3. Example 2: Brusselator Model.** We consider the Brusselator model:



Here  $X$ ,  $Y$  are molecular species and  $A$ ,  $B$ ,  $D$ , and  $E$  are constant species. The Brusselator is a generic model for a type of autocatalytic reaction. It is well known that the model reveals simple chemical oscillations. Refer to Figure 2(a) for a typical run of the model. If we denote the number of  $X$ ,  $Y$  by  $n_1$ ,  $n_2$ , respectively, we obtain the PDE for the probability generating function as follows:

$$G_t = c_1 A (z_1 - 1) G + \frac{c_2}{2} z_1^2 (z_1 - z_2) G_{112} \\ + c_3 B (z_2 - z_1) G_1 + c_4 (1 - z_1) G_1. \quad (19)$$

Here  $A = B = 1$  and the reaction rates  $c_1 = 1s^{-1}$ ,  $c_2 = 0.0001s^{-1}$ , and  $c_3 = c_4 = 0.1$  are used. We use PPGF with respect to the smallest parameter  $c_2$ . Figure 2 compares the corresponding numerical result to those from SSA.

## 4. Double Truncation Method

**4.1. Truncations in  $t$  and  $\mathbf{z}$ .** The truncation method was firstly introduced in [15] to deal with PGF-PDEs. Although the method provides a fast way to evaluate the mean and the variance, most of high order terms need to be thrown away for efficient computations at each time step. As a result, the statistical information that is carried in high order terms, such as the higher order moments and the probabilities of the states, is generally not available from the method.

In this section, we develop an extension of the truncation method which enables us to carry high order terms in the computation. This is done by another truncation in  $\mathbf{z}$  at each step. Note that this is *not* just additional manipulation to

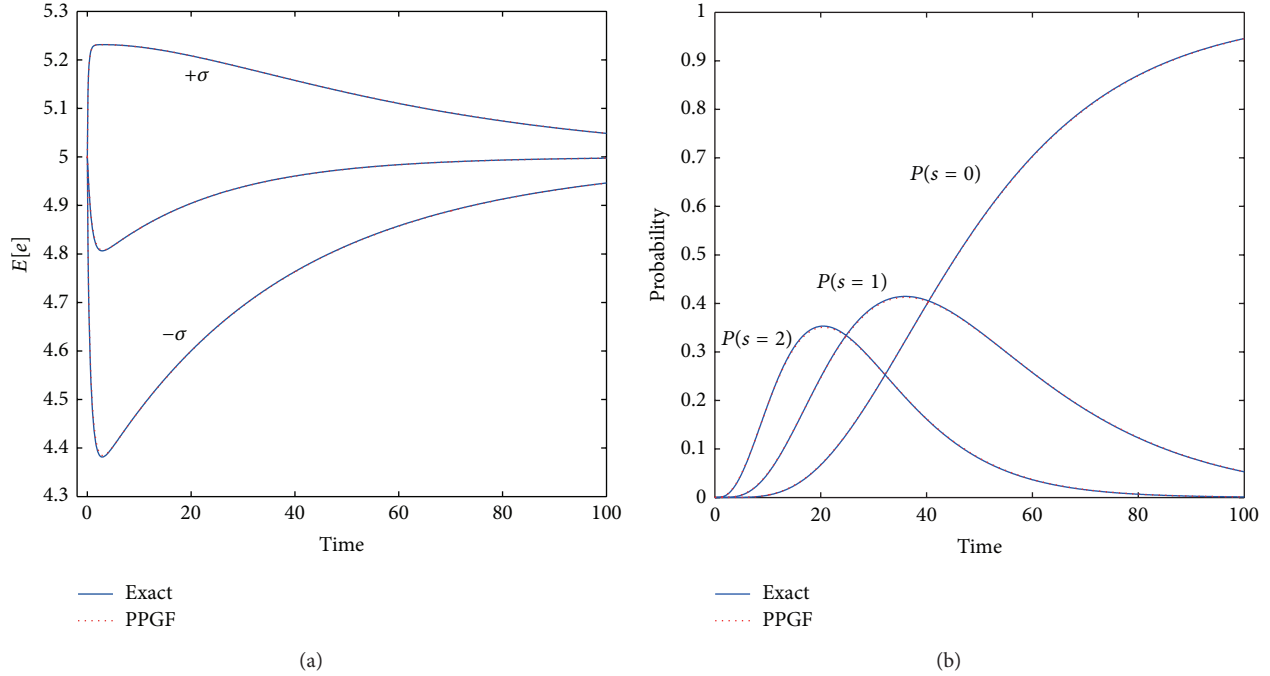


FIGURE 1: Comparison of mean, standard deviation and marginal probability for the enzyme-substrate model. In the three figures, we assume that  $n_1(0) = 5$ ,  $n_2(0) = 5$ , and  $A = 5$ . (a) shows the comparison of exact values and approximate values obtained from the PPGF method for mean and mean  $\pm$  standard deviations ( $\pm\sigma$ ) of  $n_1$  and (b) shows the comparison of the exact probability and the approximate probability obtained from the PPGF method. In the figure,  $P(s = i)$  denotes the probability that the number  $n_2$  of the substrate is  $i$ . The exact probability, mean, and variance are obtained by rewriting (15) into the linear ODE system  $dp/dt = Kp$  under the given initial condition and then solving it by MATLAB.

simply make further elimination after the first truncation. By introducing the second truncation, we can alleviate the restriction on the original truncation in  $t$  and allow higher order terms to remain, which is essential for computation of high order moments and probabilities.

In the second truncation, the terms with relatively small coefficients are dropped. It is justified from the fact that the coefficient of each term in the PGF is the probability of a specific state. We can therefore carry over high order terms, while keeping a reasonable number of terms only. This is important since evaluation of the high order moments and the probabilities is generally expensive, especially when the system involves stiff chemical reactions. Since DTM can preserve the high order terms without increasing the total number of the terms, it enables us to easily evaluate the probabilities for the states with many molecules.

*The First Truncation.* To overcome gradually increasing computational load in the series expansion (10), we restrict the series solution onto a short time interval. That is, on each interval  $[t_r, t_{r+1}]$ ,  $r = 0, 1, 2, \dots$ ,  $t_0 = 0 < t_1 < t_2 < \dots$ , we construct the separate Taylor expansion  $G^{[r]}(\mathbf{z}, t)$  at the center  $t = t_r$  of finite order  $N$ . We expect that side-by-side PGFs on consecutive intervals,  $G^{[r]}(\mathbf{z}, t)$  and  $G^{[r+1]}(\mathbf{z}, t)$ , are equal on the boundary  $t = t_{r+1}$  as

$$G^{[r]}(\mathbf{z}, t_{r+1}) \approx G^{[r+1]}(\mathbf{z}, t_{r+1}). \quad (20)$$

More generally,

$$\begin{aligned} G^{[0]}(\mathbf{z}, t) &= \sum_{n=0}^N f_n^{[0]}(\mathbf{z}) t^n, & G^{[0]}(\mathbf{z}, 0) &= \mathbf{z}^n, \\ G^{[1]}(\mathbf{z}, t) &= \sum_{n=0}^N f_n^{[1]}(\mathbf{z}) (t - t_1)^n, \\ G^{[1]}(\mathbf{z}, t_1) &= F(G^{[0]}(\mathbf{z}, t_1)), \\ G^{[2]}(\mathbf{z}, t) &= \sum_{n=0}^N f_n^{[2]}(\mathbf{z}) (t - t_2)^n, \\ G^{[2]}(\mathbf{z}, t_2) &= F(G^{[1]}(\mathbf{z}, t_2)), \\ &\vdots \\ G^{[r]}(\mathbf{z}, t) &= \sum_{n=0}^N f_n^{[r]}(\mathbf{z}) (t - t_r)^n, \\ G^{[r]}(\mathbf{z}, t_r) &= F(G^{[r-1]}(\mathbf{z}, t_r)), \\ &\vdots \end{aligned} \quad (21)$$

where  $F$  is a “carry-over” functional. Our motivation of this successive superimpositions is to keep approximation order

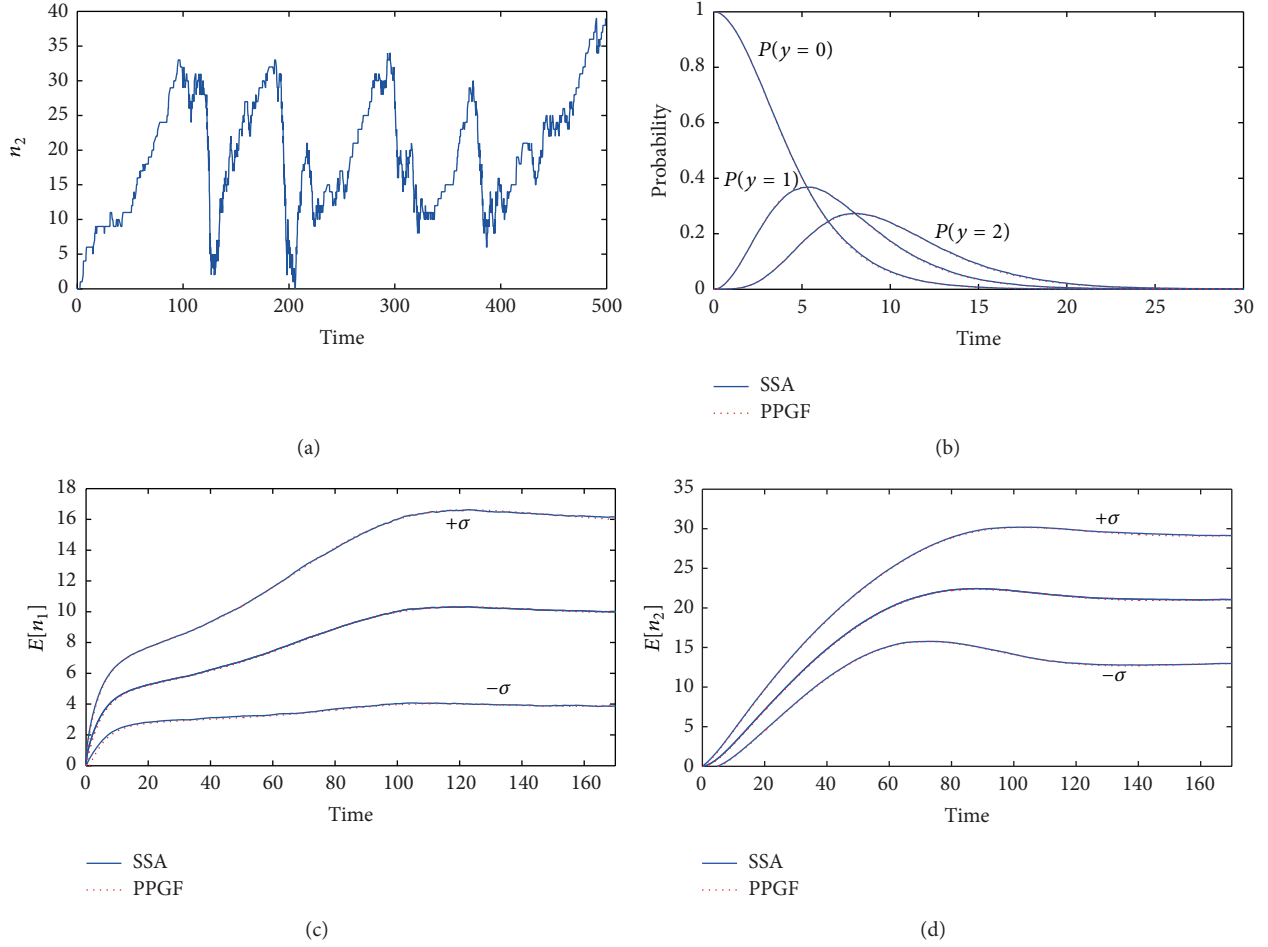


FIGURE 2: Comparison of mean and standard deviation for the Brusselator model. The initial condition  $n_1(0) = 0$  and  $n_2(0) = 0$  are assumed. (a) shows a typical run of  $Y$  in 500 seconds from a Monte Carlo simulation of the Brusselator model. In (b),  $P(y = i)$  denotes the probability that the number  $n_2$  of the substrate is  $i$ . (c) and (d) compare mean and mean  $\pm$  standard deviations ( $\pm\sigma$ ) of  $n_1$ ,  $n_2$ , and  $n_3$  obtained from the SSA and the PPGF. The results from the SSA are based on 50,000 realizations.

$N$  low in favor of computational efficiency while carrying over essential information like first several moments to a next function over the following time interval. Note that truncation of high order temporal terms can be justified in that, in PGF computation,  $N$  represents how many reaction would happen for that interval  $[t_r, t_{r+1}]$ ,  $r = 0, 1, 2, \dots$ . Elimination of such terms has an effect of excluding excessive occurrence of reactions during such short interval.

One of the issues in the first truncation is how to determine carry-over functional  $F$ . Although the best approximation in this framework would be obtained with the identity functional, it is not a practical choice since the number of terms involved in the computations keeps increasing rapidly. In [15],  $F$  was suggested as a projection onto  $P_3(\mathbf{z})$ , a set of polynomials of degree three or less, which preserves the mean and the variance. However, such functional does not preserve the information related to high order derivatives of PGF and therefore approximation of quantities such as the state probabilities is not possible. Here we propose a different type of carry-over functional that preserves high

order information. This functional indeed works as another truncation on what remains after the first truncation.

*The Second Truncation.* At time  $t = t_{r+1}$ , the separate PGF  $G^{[r]}$  becomes

$$G^{[r]}(\mathbf{z}, t_{r+1}) = \sum_{\mathbf{n}=0}^{\infty} \mathbf{z}^{\mathbf{n}} p^{[r]}(\mathbf{n}, t_{r+1}), \quad (22)$$

where  $p^{[r]}(\mathbf{n}, t_{r+1})$  means the probability for the state being  $\mathbf{n}$  at time  $t_{r+1}$ . Now we define the carry-over functional  $F$  such that it maps (22) into

$$\sum_{\mathbf{n}=0}^{\infty} \mathbf{z}^{\mathbf{n}} I_{\epsilon} \left( p^{[r]}(\mathbf{n}, t_{r+1}) \right), \quad (23)$$

where  $H_{\epsilon}$  is a function defined as

$$I_{\epsilon}(x) = \begin{cases} x, & \text{if } x > \epsilon, \\ 0, & \text{if } x \leq \epsilon. \end{cases} \quad (24)$$

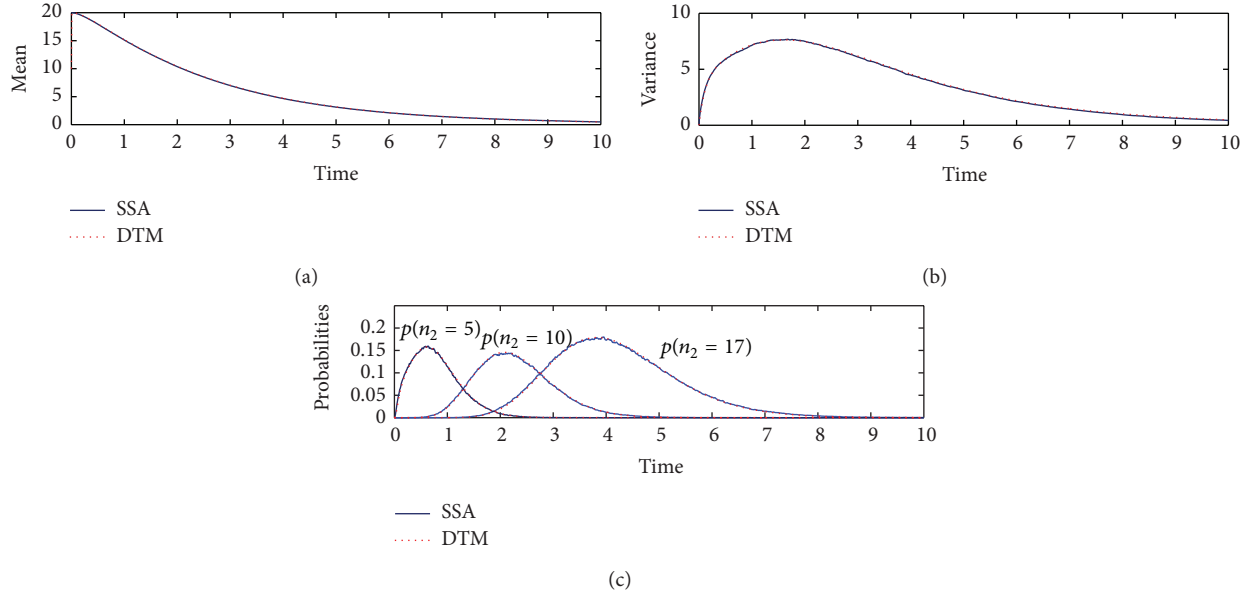


FIGURE 3: The initial condition  $n = [10, 20, 20, 0]$  and parameters  $c_1 = 0.1$ ,  $c_{-1} = 1$ , and  $c_2 = 0.5$  are assumed. In (c), each curve  $p(n_2 = i)$  denotes the time-dependent probability solution that  $n_2 = i$ ,  $i = 5, 10, 17$ , respectively. The terms which have a coefficient less than  $10^{-7}$  are dropped out.

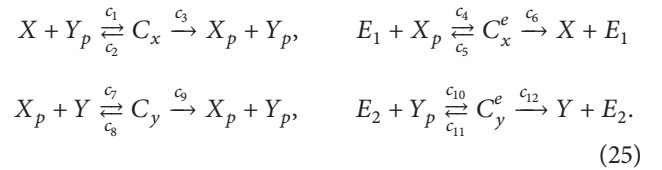
Application of such  $F$  implies that we sieve the terms out of which coefficients are less than  $\epsilon$ . One can see from (2) that those terms correspond to rare events of which coefficients probability is less than  $\epsilon$ . Considering that the iterative process in the PGF approach is a linear operation with respect to  $\mathbf{z}^n$ , manipulation on a term only affects entailed terms, each of which represents possible consequence of the corresponding state. Since the probabilities of all the entailed events and their sum as well cannot exceed the probability of the eliminated event, the effect of elimination does not propagate or grow in probability.

Note that this truncation process is clearly consistent in that the scheme reverts to the first truncation as  $\epsilon \rightarrow 0$ . However, one should be careful when applying the second truncation to approximation of the moments, since the moments do not solely depend on the state probabilities. For example, suppose that a rare state eventually gives a birth to large number of a species in an open system. If the mean is what is sought, one cannot simply omit such state for the reason that the corresponding probability is small, since the elimination may substantially change the resulted mean or variance in approximation. Hence, for the moment approximation of the open systems, it is better to use the above functional  $F$  together with other carry-over functional as suggested in [15].

**4.2. Example 1: Enzyme Kinetics.** Once again, we use the enzyme kinetic model to test the performance of DTM. Figure 3 shows comparison of approximate solutions obtained from DTM and those from SSA. Note that, in the perturbation method in Section 3.2, evaluation of the probabilities was available only for the states that correspond to small number of that species,  $n_2 = 0, 1, 2$ . This is

because the probability computation (8) requires as many times of differentiation as the number of the species at the state. However, one can see that DTM successfully resolves this problem and produces the probabilities of the states corresponding as many as  $n_2 = 10, 15$ .

**4.3. Example 2:  $G_2/M$  Transition Model.** We consider the  $G_2/M$  transition network in the eukaryotic cell cycle as follows [14, 17]:



We denote the number of molecules of  $X_p, Y_p, X, Y, E_1, E_2, C_x, C_x^e, C_y$ , and  $C_y^e$  by  $n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9$ , and  $n_{10}$ , respectively.

We obtain a PDE of the moment generating function  $G(\mathbf{z}, t)$  [14]

$$\begin{aligned} G_t = & c_1 G_{23} (z_7 - z_2 z_3) + c_2 G_7 (z_2 z_3 - z_7) \\ & + c_3 G_7 (z_1 z_2 - z_7) + c_4 G_{15} (z_8 - z_1 z_5) \\ & + c_5 G_8 (z_1 z_5 - z_8) + c_6 G_8 (z_3 z_5 - z_8) \\ & + c_7 G_{14} (z_9 - z_1 z_4) + c_8 G_9 (z_1 z_4 - z_9) \\ & + c_9 G_9 (z_1 z_2 - z_9) + c_{10} G_{26} (z_{10} - z_2 z_6) \\ & + c_{11} G_{10} (z_2 z_6 - z_{10}) + c_{12} G_{10} (z_4 z_6 - z_{10}). \end{aligned} \quad (26)$$

In Figure 4, we compare the simulation results obtained from the SSA and the DTM.



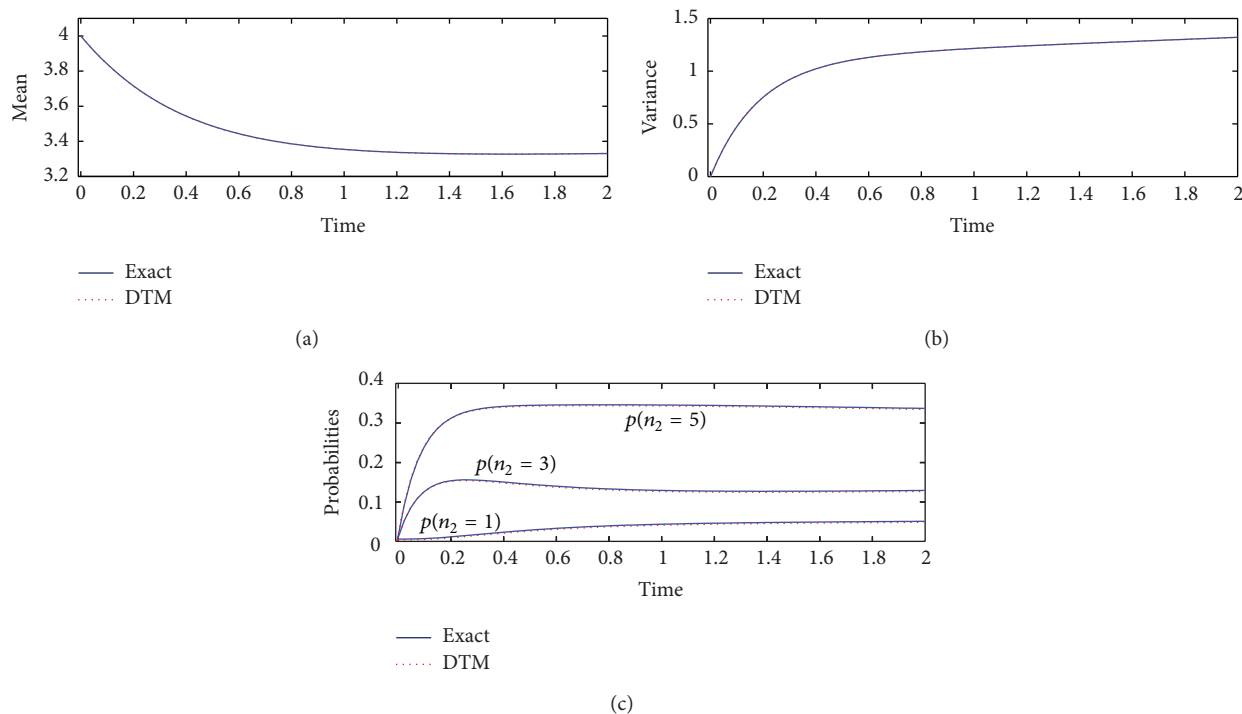


FIGURE 4: The initial condition  $n = [5, 4, 2, 3, 2, 3, 2, 0, 0, 0]$  and parameters  $c_1 = c_4 = c_7 = c_{10} = 0.2$ ,  $c_2 = c_5 = c_8 = c_{11} = 1$ , and  $c_3 = c_6 = c_9 = c_{12} = 0.1$  are assumed. In (c), each curve  $p(n_2 = i)$  denotes the time-dependent probability solution that  $n_2 = i$ ,  $i = 1, 3, 5$ , respectively. The terms which have a coefficient less than  $10^{-8}$  are dropped out.

## 5. Conclusion

In this work, we presented the double truncation method (DTM) that improves the probability generating function (PGF) method for computation of the solutions of stochastic reaction networks. Advantage of the PGF method is that, rather than struggling with the chemical master equation of extremely high order, one can handle a partial differential equation (PDE) of low order. However, it still requires intense computation for high order approximation, especially for the systems with stiff parameters. The method proposed in this paper applies the perturbation method to the corresponding PGF for reaction networks. That is, we expand the PGF in the parameter that is responsible for stiffness of the system. Then the PGF-PDE leads to a recurrence relation of the coefficient functions with respect to both the time and the corresponding parameter, which we can easily solve symbolically. It turns out that the suggested method requires a lower approximation order than the original PGF method, while it produces as good approximation as such.

While mean and variance of species are easily traceable in single truncation suggested in [15], efficient evaluation of the state probability was still a challenging task. In Section 4, we showed that the probability evaluation though PGF-PDEs can be achieved by doubly truncating the Taylor series at each time step. Through consecutive truncation processes, we dump relatively rare events, most of which are generated from excessive occurrence of reactions during a short time

interval. This also implies elimination of a great deal of terms and significant reduction in computation.

It turns out in several numerical examples that the suggested method produces good approximation for the state probabilities. We expect that the DTM will be used for a fast and accurate approximation for computation of real and complex stochastic chemical reaction networks.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

Pilwon Kim acknowledges that this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2011-0023486). Chang Hyeong Lee was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1A2054976).

## References

- [1] C. V. Rao, D. M. Wolf, and A. P. Arkin, "Control, exploitation and tolerance of intracellular noise," *Nature*, vol. 420, no. 6912, pp. 231–237, 2002.

- [2] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A: Statistical Mechanics and its Applications*, vol. 188, no. 1–3, pp. 404–425, 1992.
- [3] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.
- [4] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *Journal of Chemical Physics*, vol. 115, no. 4, pp. 1716–1733, 2001.
- [5] Y. Cao, D. T. Gillespie, and L. R. Petzold, "Efficient step size selection for the tauleaping simulation method," *The Journal of Chemical Physics*, vol. 124, no. 4, Article ID 044109, 2006.
- [6] D. F. Anderson, "Incorporating postleap checks in tau-leaping," *Journal of Chemical Physics*, vol. 128, no. 5, Article ID 054103, 2008.
- [7] Y. Cao, D. T. Gillespie, and L. R. Petzold, "The slow-scale stochastic simulation algorithm," *Journal of Chemical Physics*, vol. 122, no. 1, Article ID 014116, 2005.
- [8] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *Journal of Chemical Physics*, vol. 124, no. 4, Article ID 044104, 2006.
- [9] C. H. Lee and R. Lui, "A reduction method for multiple time scale stochastic reaction networks," *Journal of Mathematical Chemistry*, vol. 46, no. 4, pp. 1292–1321, 2009.
- [10] C. H. Lee and R. Lui, "A reduction method for multiple time scale stochastic reaction networks with non-unique equilibrium probability," *Journal of Mathematical Chemistry*, vol. 47, no. 2, pp. 750–770, 2010.
- [11] C. H. Lee, K.-H. Kim, and P. Kim, "A moment closure method for stochastic reaction networks," *Journal of Chemical Physics*, vol. 130, no. 13, Article ID 134107, 2009.
- [12] P. Milner, C. S. Gillespie, and D. J. Wilkinson, "Moment closure approximations for stochastic kinetic models with rational rate laws," *Mathematical Biosciences*, vol. 231, no. 2, pp. 99–104, 2011.
- [13] C. H. Lee, "A moment closure method for stochastic chemical reaction networks with general kinetics," *MATCH. Communications in Mathematical and in Computer Chemistry*, vol. 70, no. 3, pp. 785–800, 2013.
- [14] P. Kim and C. H. Lee, "A probability generating function method for stochastic reaction networks," *Journal of Chemical Physics*, vol. 136, no. 23, Article ID 234108, 2012.
- [15] P. Kim and C. H. Lee, "Fast probability generating function method for stochastic chemical reaction networks," *MATCH: Communications in Mathematical and in Computer Chemistry*, vol. 71, no. 1, pp. 57–69, 2014.
- [16] D. A. McQuarrie, "Stochastic approach to chemical kinetics," *Journal of Applied Probability*, vol. 4, pp. 413–478, 1967.
- [17] A. Kumar and K. Josić, "Reduced models of networks of coupled enzymatic reactions," *Journal of Theoretical Biology*, vol. 278, pp. 87–106, 2011.



