

Variance bounds for a class of biochemical reactions with bursts using a discrete expansion

Giovanni Pugliese Carratelli and Ioannis Lestas

Abstract—We consider the problem of quantifying the variance in the number of molecules of a species, in biochemical reactions with nonlinear reaction rates. We address this problem for a particular configuration where a species is formed with bursts, with a nonlinear rate that depends on another spontaneously formed species. By making use of an appropriately formulated expansion based on the Newton series, in conjunction with spectral properties of the master equation, we derive an analytical expression that provides a hard bound for the variance. We also show that this bound is exact when the propensities are linear. Furthermore, numerical simulations demonstrate that this is very close to the actual variance.

I. INTRODUCTION

Chemical Master Equations (CMEs) are frequently used to model biochemical systems. Nevertheless, when the reaction rates are nonlinear analytical expressions for the moments are in general not feasible. Approximate analysis can be carried out via simulations, such as trajectories generated via the *Gillespie Algorithm*, [1]. Using, however those for evaluating moments requires considerable computational effort. Alternative approaches that lead to an approximate analysis include the Finite State Projection algorithm [2] and the Linear Noise Approximations (LNAs), or *Van Kampen's Ω -Expansion* where the last two involve a linearisation of the reaction rates.

Recent studies have developed techniques for computing bounds for moments. In [3] bounds are computed by means of semidefinite programming. The idea of moment semidefinite programs was used in [4] and in [5], [6] optimization problems are formulated with positive semidefinite constraints associated with moment matrices for chemical reactions. Approximate moment dynamics with closure properties are derived in [7], [8] and [9] by selecting an appropriate nonlinear system to approximate the infinite set of moment Ordinary Differential Equations (ODEs) emerging from the CME.

In this paper we present an alternative approach whereby an appropriately formulated expansion based on the Newton series, is used to derive a hard bound for the variance for a class of biochemical reactions. In particular, the variance bound is derived for a species that is formed with bursts with a nonlinear rate that is a function of another spontaneously formed species. We also show that the bound derived is exact when the reaction rates are linear. Finally, we use numerical simulations to investigate how close the bound is to the actual

variance, and we find that, unlike LNAs, this is very close to the true variance even in regimes where the reaction rate is highly nonlinear. A side result associated with the covariance of the species under consideration is also presented.

It should be noted that the methodology used to derive the bound presented is of independent interest as it exploits appropriate spectral properties of the master equation. Generalizations of this approach to larger classes of biochemical reactions is part of ongoing work.

The paper is structured as follows. In Section II we introduce the notation and provide the problem formulation. The main results are given in Section III. A numerical evaluation of the results is provided in Section IV. Finally, conclusions are drawn in Section V. The proofs of the results presented in the text are provided in the appendix.

II. PROBLEM FORMULATION

A. Notation

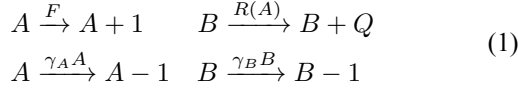
The following list introduces various symbols used within the paper

$(A)_k$	Falling factorial of $A \in \mathbb{Z}^{\geq}$ with $k \in \mathbb{Z}^{\geq}$: if $k \neq 0$ then $(A)_k = \prod_{n=0}^{k-1} A - n$, if $k = 0$ then $(A)_0 = 1$
$\Delta_k[\cdot]$	k step finite difference of the form $\Delta_k[f](A) \equiv f(A+k) - f(A)$
$\Delta_k^p[\cdot]$	k step finite difference operator applied p times
$\delta_{a,b}$	$\delta_{a,b} = 1$, if $a = b$, 0 otherwise
$\mathbb{E}[X]$	Expectation of random variable X
$\mathbb{E}_{\mathcal{P}}[\cdot]$	Expectation operator with respect to the Poisson distribution
$\mathbb{R}[A]$	Set of polynomials in A with real valued coefficients
$\mathbb{R}^>$	Set of positive real numbers $\{x \in \mathbb{R} : x > 0\}$
\mathbb{R}^{\geq}	Set of non negative real numbers $\{x \in \mathbb{R} : x \geq 0\}$
\mathbb{Z}^{\geq}	Set of non negative integers $\{0, 1, 2, 3, \dots\}$
$\mathcal{U}\{a, b\}$	Discrete uniform probability distribution with support $\{a, b\}$ with $a, b \in \mathbb{Z}^{\geq}$
$\mathbb{1}$	Unity operator: $\mathbb{1}[X] = X$
$A \sim \mathcal{P}(x)$	Random variable A has Poisson probability distribution with mean x
$cov(X, Y)$	Covariance of random variables X and Y
$var(X)$	Variance of random variable X
$\mathcal{B}(p, n)$	Binomial probability distribution, with parameters $0 < p \leq 1$ and $n \in \mathbb{Z}^{\geq}$
$\mathcal{G}(p)$	Geometric probability distribution, with parameter $0 < p \leq 1$

Giovanni Pugliese Carratelli and Ioannis Lestas are with the Department of Engineering, University of Cambridge, CB2 1PZ, United Kingdom; gp459, icl20@cam.ac.uk. This work was supported by ERC starting grant 679774.

B. Problem Formulation

We consider the following system



where A and B are two biochemical species. Random variable $A(t)$ denotes the number of molecules of species A at time t and similarly random variable $B(t)$ denotes the number of molecules of species B at time t . The random variable Q denotes the increase in the number of molecules of species B when a birth takes place. The parameter $F \in \mathbb{R}^>$ is a constant that denotes the rate at which the species A is produced, $R : \mathbb{Z}^{\geq} \rightarrow \mathbb{R}^{\geq}$ is the rate of production of the species B and constants $\gamma_A \in \mathbb{R}^>$ and $\gamma_B \in \mathbb{R}^>$ represent the death rate of each molecule of species A and B respectively.

For any $a, b \in \mathbb{Z}^{\geq}$, we denote by $\mathbb{P}(a, b, t)$ the probability $A(t) = a$, $B(t) = b$. We assume that the random variable Q is independent from the random variables $A(t), B(t)$ and denote by $\mathbb{P}_Q(q)$ the probability $Q = q$. The CME, a version of the Chapman Kolmogorov equation for Markov processes, for system (1) is

$$\begin{aligned} \frac{\partial \mathbb{P}(a, b, t)}{\partial t} &= \gamma_A [(a + 1)\mathbb{P}(a + 1, b, t) - a\mathbb{P}(a, b, t)] \\ &+ F [\mathbb{P}(a - 1, b, t) - \mathbb{P}(a, b, t)] \\ &+ \gamma_B [(b + 1)\mathbb{P}(a, b + 1, t) - b\mathbb{P}(a, b, t)] \\ &+ R(a) \left[\sum_{q=1}^b \mathbb{P}_Q(q)\mathbb{P}(a, b - q, t) - \mathbb{P}(a, b, t) \right] \end{aligned} \quad (2)$$

The problem we are trying to address is to find an expression of the mean and variance of B . This is a non-trivial problem due to the nonlinear rate $R(A)$.

In this paper we present a bound on the variance of B obtained by exploiting a discrete expansion of $R(A)$ based on the *Newton* series. The bound becomes exact when $R(A)$ is linear. Furthermore, we show via numerical examples that it is very close to the true variance even in regimes where $R(A)$ is highly nonlinear, in contrast to a LNA approximation for the variance.

It should be noted that filtering problems associated with (1) have also been studied in the literature. The optimal causal filter for estimating A when B is observed was derived in [10] and a close to optimal version was deployed *in vitro* in [11]. The average squared difference between A and B was quantified in [12].

The results that will be presented are associated with the equilibrium distribution of species A and B and we therefore make the following assumption.

Assumption 1: System (2) reaches a stationary distribution denoted by $\mathbb{P}(a, b)$ *i.e.* $\lim_{t \rightarrow +\infty} \mathbb{P}(a, b, t) = \mathbb{P}(a, b)$.

Note that from the CME (2) $\mathbb{P}(a, b)$ satisfies the following

$$\begin{aligned} \gamma_A [(a + 1)\mathbb{P}(a + 1, b) - a\mathbb{P}(a, b)] &+ F [\mathbb{P}(a - 1, b) - \mathbb{P}(a, b)] \\ &+ \gamma_B [(b + 1)\mathbb{P}(a, b + 1) - b\mathbb{P}(a, b)] \\ &+ R(a) \left[\sum_{q=1}^b \mathbb{P}_Q(q)\mathbb{P}(a, b - q) - \mathbb{P}(a, b) \right] = 0 \end{aligned} \quad (3)$$

For convenience in the notation we denote by A the random variable representing the number of molecules of species A at equilibrium, and its probability distribution is denoted by \mathbb{P}_A . Similarly, we denote by B the random variable representing the number of molecules of species B at equilibrium. Random variable Q can have any discrete distribution as long as it takes values in the set of positive integers excluding 0, and is independent of A and B . It should also be noted that $A \sim \mathcal{P}(F/\gamma_A)$, *i.e.* random variable A has a Poisson distribution with mean F/γ_A which is a known result that follows from the fact that F and γ_A are constant (see *e.g.* [13]).

III. MAIN RESULTS

Before we present our main results we state a condition upon $R(A)$.

Assumption 2: $R : \mathbb{Z}^{\geq} \rightarrow \mathbb{R}^{\geq}$ is a ratio of polynomials, *i.e.* $R(A) = \left\{ \frac{f}{g} \mid f, g \in \mathbb{R}[A], g \neq 0 \forall A \in \mathbb{Z}^{\geq} \right\}$.

Note that this assumption is mild within a biological perspective as reaction rates are typically rational functions of molecule numbers. In order to give our first result we define the quantities in (4) which are needed for the variance bound derived.

$$\sigma_0 = \mathbb{E}_{\mathcal{P}}[R(A)], \quad \sigma_1 = -\mathbb{E}_{\mathcal{P}}[R(A)] + \frac{\mathbb{E}_{\mathcal{P}}[AR(A)]}{\mathbb{E}_{\mathcal{P}}[R(A)]} \quad (4)$$

Note that in (4) $\mathbb{E}_{\mathcal{P}}[\cdot]$ denotes the expectation with respect to the Poisson distribution of the random variable A .

Lemma 1: Let $A \sim \mathcal{P}(F/\gamma_A)$. Then the quantities σ_0 and σ_1 from (4) exist and are finite.

We now give our main result which is an expression that provides a lower bound on the variance of B .

Proposition 1: Consider the system in (3) with $R(A)$ satisfying Assumption 2. The following inequality holds

$$\begin{aligned} \text{var}(B) &> \\ &\frac{(\gamma_A + \gamma_B)\sigma_0(\mathbb{E}[Q^2] + \mathbb{E}[Q]) + 2(\mathbb{E}[Q])^2\sigma_1^2\mathbb{E}_{\mathcal{P}}[A]}{2\gamma_B(\gamma_B + \gamma_A)} \end{aligned} \quad (5)$$

where σ_0 and σ_1 are given in (4).

Remark 1: If $R(A)$ is a linear function then, then equation (5) holds with equality. This is stated in Proposition 2.

Remark 2: Parameters σ_0 and σ_1 exist and are bounded as stated in Lemma 1. Note that $\mathbb{E}_{\mathcal{P}}[\cdot]$ is the expectation with respect to a Poisson distribution with a known mean and may be computed to an arbitrary high precision by taking a sufficient amount of terms in the evaluation of the expectation.

Below we show that the bound is exact if all transition rates are linear.

Proposition 2: Consider the system in (3). If $R(A) = R_c A$ with $R_c \in \mathbb{R}^>$ the bound for the variance in Proposition 1 holds with equality and is given by (6)

$$\text{var}(B) = \frac{R_c \mathbb{E}_{\mathcal{P}}[A] [R_c 2(\mathbb{E}[Q])^2 + (\gamma_A + \gamma_B)(\mathbb{E}[Q^2] + \mathbb{E}[Q])]}{2\gamma_B(\gamma_A + \gamma_B)} \quad (6)$$

Remark 3: The variance in the case of Proposition 2 can be computed analytically using LNA approaches, see *e.g.* [13] and [14].

Finally we give a side result on the covariance of A and B .

Proposition 3: Consider the systems described in (3). Then $\text{cov}(A, B) = \text{cov}(B, A)$ satisfies

$$\text{cov}(A, B) = \mathbb{E}[Q] \frac{\sigma_1 \mathbb{E}_{\mathcal{P}}[A]}{\gamma_A + \gamma_B} \quad (7)$$

where σ_1 is given in (4).

Remark 4: Note that Proposition 3 holds with equality also when $R(A)$ is nonlinear.

The results stated above have a number of useful properties relative to other more conventional approaches for quantifying the variance. As mentioned in Section I, Proposition 1 provides a bound for the variance when the transition rate $R(a)$ is nonlinear whereas existing methods like LNA and numerical simulations give only approximate values. As it will be seen in Section IV the variance calculation through the LNA is not as close to the variance of B as is the bound provided in Proposition 1. This is especially important in the case of highly nonlinear propensity functions where LNAs become less accurate. Furthermore, the numerical investigation in Section IV demonstrates that the bound maintains its accuracy despite the presence of bursts.

IV. EXAMPLES

In order to evaluate the conservativeness of the bound for the variance given in (5) we numerically investigate how close this is to the actual variance. The simulations have been performed using the software package GillesPy2 [15]. We select as function $R(A)$ a nonlinear function commonly found in biological systems, *i.e.* a Hill function $R(a) = \frac{(a/A_0)^{n_h}}{1+(a/A_0)^{n_h}}$ where $n_h \in \mathbb{R}^>$ and $A_0 \in \mathbb{R}^>$. We set as parameters $n_h = 9$ and $A_0 = 100$ and $F = 1$. This function clearly fulfills the conditions in Assumption 2.

Fig. 1 displays four cases corresponding to different burst distributions [16]. The figure highlights that the bound is not conservative regardless of the distribution of Q and it is very close to $\text{var}(B)$ for different values of $\mathbb{E}_{\mathcal{P}}[A]$ obtained by varying γ_A . The figures also show a comparison of the bound with the variance of B computed through the LNA of system (3). Note that the computation carried out via LNA performs poorly for certain regimes. Specifically, the performance is particularly poor when A is with high probability in the nonlinear regime of the Hill function, while it is more accurate when A is predominantly in the linear regime.

The left hand side of Fig. 2 shows the relative error, expressed as a percentage, between the variance of B and the

bound corresponding to different values of $\mathbb{E}_{\mathcal{P}}[A]$ (obtained by varying γ_A) and γ_B when Q has a geometric distribution, *i.e.* $Q \sim \mathcal{G}(0.5)$. The right hand side of Fig. 2 displays the relative error between the variance of B and the variance obtained through LNA. We find that the bound is still not conservative and performs consistently better than variance obtained through LNA for different values of γ_A and γ_B .

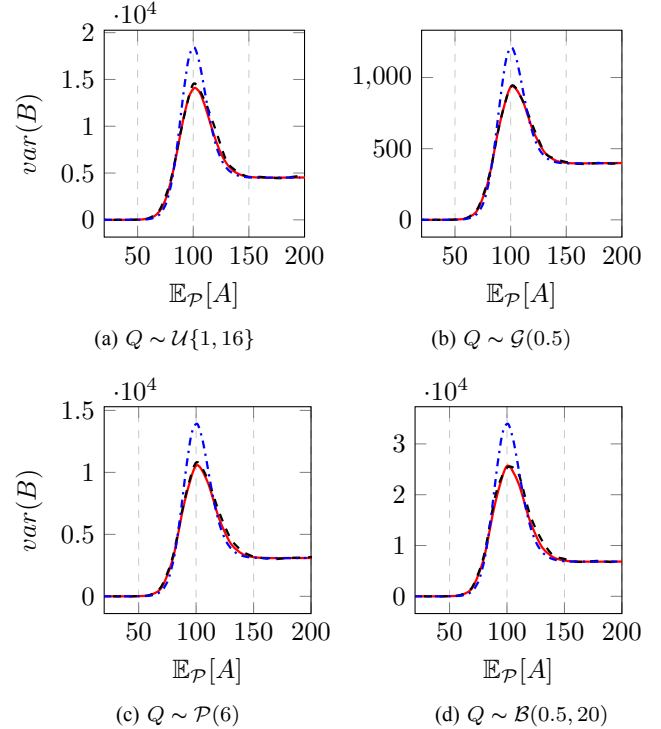


Fig. 1: Numerical simulation showing the bound obtained from (5) and the computed variance of B for varying $\mathbb{E}_{\mathcal{P}}[A] = F/\gamma_A$ with $F = 1$, $\gamma_B = 10^{-2}$ and for different distributions of Q . $R(a)$ is a Hill function with parameters $n_h = 9$ and $A_0 = 100$. (—) displays the bound computed from (5) and (---) displays the variance obtained from numerical simulations of (3). The variance computed through a LNA of system (3) is displayed as (-.-.-)

V. CONCLUSIONS

We have derived a hard bound for the variance in the molecule numbers of a species formed with bursts, with a nonlinear rate that depends on another spontaneously formed species. The bound follows from a discrete expansion based on the Newton series, and exploits spectral properties of the master equation. We have also shown that the bound holds with equality if the propensity functions are linear. The accuracy of the bound has been investigated with numerical simulations which demonstrate that this is very close to the actual variance also when the rate of formation of the species is highly nonlinear, and despite the presence of bursts. Future work will focus on extending the results to larger classes of CMEs.

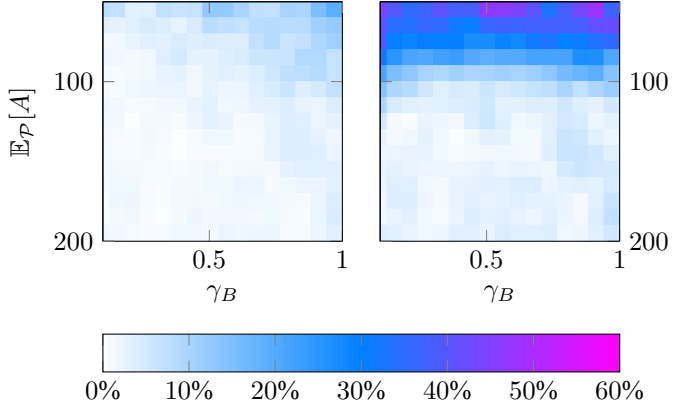


Fig. 2: The left hand side of the figure shows the relative error between the variance of B , obtained through a numerical simulation, and the bound in Proposition 1. The right hand side displays the relative error between the variance of B , and the LNA. $R(A)$ is set to be a Hill function with parameters $n_h = 9$ and $A_0 = 100$, and $F = 1$. The considered burst distribution is $Q \sim \mathcal{G}(0.5)$.

VI. APPENDIX

In the appendix we provide the proofs of the results presented in the main text.

A. Proof of Lemma 1

Proof: The proof follows by noting that σ_0 and σ_1 are expectations of rational functions with respect to a Poisson distribution. ■

B. Proof of Proposition 1

Proof:

We divide our proof in three parts. In the first part we compute $T'_a(1)$, where $T_a(z) = \sum_{b=0}^{\infty} z^b \mathbb{P}(a, b)$ is a probability generating function that is used for the calculation of the moments of B . In particular, we show that $T'_a(1)$ satisfies an equation that involves a difference operator that depends on the CME.

In Part 2 we derive an expression for $T'_a(1)$ in terms of the eigenfunctions of this operator. This expression depends on the *Newton* series expansion of $R(A)$.

In Part 3 we make use the results in Part 1, 2 to derive an expression for the variance of B , and hence deduce a bound in terms of σ_0, σ_1 , as stated in Proposition 1.

Part 1: Consider (3) and let us make use of the generating function $T_a(z) = \sum_{b=0}^{\infty} z^b \mathbb{P}(a, b)$ and $M(z) = \sum_{q=0}^{\infty} z^q \mathbb{P}_Q(q)$. Under these transformations we compute the following equation.

$$\begin{aligned} \gamma_A[(a+1)T_{a+1}(z) - aT_a(z)] + F[T_{a-1}(z) - T_a(z)] \\ + \gamma_B(1-z)T'_a(z) + R(a)(M(z) - 1)T_a(z) = 0 \end{aligned} \quad (8)$$

Note that (8) provides an alternative way to express (3). By setting $z = 1$ we have $T_a(1) = \mathbb{P}_A(a)$, where A in the considered case has a Poisson distribution, see [13]. Secondly, we can compute moments of B by differentiating

(8) with respect to z and setting $z = 1$. In particular, we have

$$T'_a(1) = \sum_{b=0}^{\infty} b \mathbb{P}(a, b), \quad T''_a(1) = \sum_{b=0}^{\infty} b(b-1) \mathbb{P}(a, b) \quad (9)$$

Hence by differentiating (8) with respect z and then setting $z = 1$ we obtain the following expression involving the first moment

$$\begin{aligned} \gamma_A[(a+1)T'_{a+1}(1) - aT'_a(1)] + F[T'_{a-1}(1) - T'_a(1)] \\ - \gamma_B T'_a(1) + R(a)(M'(1)T_a(1)) = 0 \end{aligned} \quad (10)$$

where $M'(1) = \mathbb{E}[Q]$.

For the second moment we have

$$\begin{aligned} \gamma_A[(a+1)T''_{a+1}(1) - aT''_a(1)] + F[T''_{a-1}(1) - T''_a(1)] \\ - 2\gamma_B T'_a(1) + R(a)(M''(1)T_a(1) + 2M'(1)T'_a(1)) = 0 \end{aligned} \quad (11)$$

where $M''(1) = \mathbb{E}[Q^2] - \mathbb{E}[Q]$. By summing (10) over all a yields

$$\sum_{a=0}^{\infty} T'_a(1) = \frac{M'(1)}{\gamma_B} \sum_{a=0}^{\infty} R(a)T_a(1) \quad (12)$$

Note that the left hand side of (12) is the expectation of B with respect to the known distribution of A . From the definition of $T'_a(1)$ and by noting that $T_a(1) = \mathbb{P}_A(a)$, (12) can be written as

$$\mathbb{E}[B] = \frac{\mathbb{E}[Q]}{\gamma_B} \mathbb{E}_{\mathcal{P}}[R(A)] \quad (13)$$

For the second moment of B we proceed analogously as for the first moment by summing (11) over all a and obtain

$$\begin{aligned} \sum_{a=0}^{\infty} T''_a(1) = \\ \frac{1}{2\gamma_B} \left(M''(1) \sum_{a=0}^{\infty} R(a)T_a(1) + 2M'(1) \sum_{a=0}^{\infty} R(a)T'_a(1) \right) \end{aligned} \quad (14)$$

From (9) and (14) we obtain

$$\begin{aligned} \mathbb{E}[B^2] - \mathbb{E}[B] = \\ \frac{1}{2\gamma_B} (M''(1)\mathbb{E}_{\mathcal{P}}[R(A)] + 2M'(1)\mathbb{E}[BR(A)]) \end{aligned} \quad (15)$$

where for the last term on the right hand side of (14) we have applied the definition of $T'_a(1)$.

We now seek in the remainder of the proof to derive an expression for the last term in (15).

We now consider the following representation of $T'_a(1)$, where function $G(a)$ is to be determined:

$$T'_a(1) = \frac{T_a(1)G(a)}{\gamma_B} \quad (16)$$

We substitute (16) into (10) and obtain

$$\begin{aligned} \frac{\gamma_A}{\gamma_B} [(a+1)T_{a+1}(1)G(a+1) - aT_a(1)G(a)] \\ + \frac{F}{\gamma_B} [T_{a-1}(1)G(a-1) - T_a(1)G(a)] \\ - G(a)T_a(1) + M'(1)R(a)T_a(1) = 0 \end{aligned} \quad (17)$$

Note that from the definition of the Poisson distribution we have:

$$T_{a+1}(1) = \frac{F}{\gamma_A(a+1)}T_a(1), \quad T_{a-1}(1) = \frac{a\gamma_A}{F}T_a(1) \quad (18)$$

By rearranging (17) and using (18) leads to

$$T_a(1)[(\mathbb{S} - 1)G(a) + M'(1)R(a)] = 0 \quad (19)$$

where \mathbb{S} , defined in (20), is an operator acting upon $G(a)$

$$\mathbb{S} = \gamma_B^{-1}(\gamma_A a \Delta_{-1} + F \Delta_1) = \frac{\gamma_A}{\gamma_B}(a \Delta_{-1} + \mathbb{E}_{\mathcal{P}}[A] \Delta_1) \quad (20)$$

By making use of (19) we find that $G(a)$ satisfies

$$(\mathbb{1} - \mathbb{S})G(a) = M'(1)R(a) \quad (21)$$

We now seek to exploit spectral properties of \mathbb{S} in order to derive an expression for $G(a)$ and hence of $T'_a(1)$.

Part 2: In this part of the proof we derive an expression for $T'_a(1)$.

We first focus on the operator \mathbb{S} and its corresponding eigenfunctions. First let us note that $\Delta_1[(a)_k] = k(a)_{k-1}$ and $a\Delta_{-1}[(a)_k] = -k(a)_k$ and hence we can write $\mathbb{S}[(a)_k]$ as

$$\mathbb{S}[(a)_k] = -\frac{k\gamma_A}{\gamma_B}[(a)_k - \mathbb{E}_{\mathcal{P}}[A](a)_{k-1}] \quad (22)$$

We will derive an expression for $G(a)$ in terms of the eigenfunctions and eigenvalues of operator \mathbb{S} . We derive those below, following the approach in [17]. To facilitate the analysis of (22) we introduce operator \mathbb{F} , which for a monomial a^k , $a, k \in \mathbb{Z}^{\geq}$ it satisfies the property $\mathbb{F}[a^k] = (a)_k$. It is also defined to be linear, i.e. $\mathbb{F}[a_1^{k_1} + a_2^{k_2}] = (a_1)_{k_1} + (a_2)_{k_2}$, $\mathbb{F}[\lambda a^k] = \lambda(a)_k$, $\lambda \in \mathbb{R}$.

We can hence rewrite (22) as

$$\mathbb{S}[\mathbb{F}[a^k]] = -\frac{k\gamma_A}{\gamma_B}\mathbb{F}[a^{k-1}(a - \mathbb{E}_{\mathcal{P}}[A])] \quad (23)$$

which can also be written as

$$\mathbb{S}[\mathbb{F}[a^k]] = -\frac{\gamma_A}{\gamma_B}\mathbb{F}\left[\frac{d}{da}[a^k](a - \mathbb{E}_{\mathcal{P}}[A])\right] \quad (24)$$

We now apply operator \mathbb{S} , in combination with operator \mathbb{F} , to $(a - \mathbb{E}_{\mathcal{P}}[A])^k$ which yields

$$\begin{aligned} \mathbb{S}[\mathbb{F}[(a - \mathbb{E}_{\mathcal{P}}[A])^k]] &= \\ &= -\frac{\gamma_A}{\gamma_B}\mathbb{F}\left[\frac{d}{da}[(a - \mathbb{E}_{\mathcal{P}}[A])^k](a - \mathbb{E}_{\mathcal{P}}[A])\right] \end{aligned} \quad (25)$$

By observing that $\frac{d}{da}[(a - \mathbb{E}_{\mathcal{P}}[A])^k] = k(a - \mathbb{E}_{\mathcal{P}}[A])^{k-1}$ we have

$$\mathbb{S}[\mathbb{F}[(a - \mathbb{E}_{\mathcal{P}}[A])^k]] = -k\frac{\gamma_A}{\gamma_B}\mathbb{F}[(a - \mathbb{E}_{\mathcal{P}}[A])^k] \quad (26)$$

We now use the binomial expansion to rewrite the eigenfunctions and corresponding eigenvalues, which take the form

$$\lambda_n = -n\frac{\gamma_A}{\gamma_B} \quad (27)$$

$$\psi_n(a) = \sum_{k=0}^n \binom{n}{k} (-\mathbb{E}_{\mathcal{P}}[A])^k (a)_{n-k} \quad (28)$$

Using the expressions derived for the eigenfunctions and eigenvalues of \mathbb{S} we now proceed to derive an expression for $G(a)$. In particular, by noting the decomposition of $R(a)$ in terms of the eigenfunctions ψ_n given in Proposition 4 we make the following ansatz on $G(a)$

$$G(a) = M'(1) \sum_{n=0}^{\infty} \sigma_n \zeta(n) \psi_n(a) \quad (29)$$

where $\zeta(n)$ is a function to be determined and σ_n are as defined in Proposition 4. By substituting (29) in (21) we get

$$M'(1) \sum_{n=0}^{\infty} \sigma_n \left(1 + n\frac{\gamma_A}{\gamma_B}\right) \zeta(n) \psi_n(a) = M'(1)R(a) \quad (30)$$

Noting the series expansion of $R(a)$ in Proposition 4 with $r = \mathbb{E}_{\mathcal{P}}[A]$ we deduce that

$$\zeta(n) = \frac{\gamma_B}{n\gamma_A + \gamma_B} \quad (31)$$

Therefore from (16), (29), (31) we deduce the following expression for $T'_a(1)$

$$T'_a(1) = \frac{T_a(1)M'(1)}{\gamma_B} \sum_{n=0}^{\infty} \sigma_n \frac{\gamma_B}{n\gamma_A + \gamma_B} \psi_n(a) \quad (32)$$

Part 3: We now make use of the series expansion of $T'_a(1)$ in (32) to compute the moments of B .

We evaluate first the unknown term in the right hand side of (15). In particular, by noting that $\mathbb{E}[BR(A)] = \sum_{a=0}^{\infty} R(a)T'_a(1)$ and from the expression of $T'_a(1)$ in (32) we have

$$\begin{aligned} \mathbb{E}[BR(A)] &= \\ &= \frac{M'(1)}{\gamma_B} \sum_{a=0}^{\infty} R(a) \sum_{n=0}^{\infty} \sigma_n \frac{\gamma_B}{n\gamma_A + \gamma_B} \psi_n(a) T_a(1) \\ &= \frac{\mathbb{E}[Q]}{\gamma_B} \mathbb{E}_{\mathcal{P}} \left[\sum_{n'=0}^{\infty} \sum_{n=0}^{\infty} \sigma_{n'} \sigma_n \frac{\gamma_B}{n\gamma_A + \gamma_B} \psi_{n'}(A) \psi_n(A) \right] \end{aligned} \quad (33)$$

where $M'(1) = \mathbb{E}[Q]$ and in the second step we have used the results of Proposition 4. We now evaluate the expectation in (33) as follows

$$\begin{aligned} \mathbb{E}_{\mathcal{P}} \left[\sum_{n'=0}^{\infty} \sum_{n=0}^{\infty} \sigma_{n'} \sigma_n \frac{\gamma_B}{n\gamma_A + \gamma_B} \psi_{n'}(A) \psi_n(A) \right] &= \\ &= \sum_{n=0}^{\infty} \sigma_n^2 \frac{\gamma_B n! (\mathbb{E}_{\mathcal{P}}[A])^n}{n\gamma_A + \gamma_B} \end{aligned} \quad (34)$$

where the equality holds given the results of Proposition 5. Hence we can now write the following expression for $\mathbb{E}[BR(A)]$

$$\mathbb{E}[BR(A)] = \frac{M'(1)}{\gamma_B} \sum_{n=0}^{\infty} \sigma_n^2 \frac{\gamma_B n! (\mathbb{E}_{\mathcal{P}}[A])^n}{n\gamma_A + \gamma_B} \quad (35)$$

We finally express $\text{var}(B)$ as in (36) by making use of (13), (15), and (35)

$$\begin{aligned} \text{var}(B) = & \frac{1}{2\gamma_B} \left[(\mathbb{E}[Q^2] - \mathbb{E}[Q])\mathbb{E}_{\mathcal{P}}[R(A)] \right. \\ & \left. + \frac{2(\mathbb{E}[Q])^2}{\gamma_B} \sum_{n=0}^{\infty} \sigma_n^2 \frac{\gamma_B n! (\mathbb{E}_{\mathcal{P}}[A])^n}{n\gamma_A + \gamma_B} \right] \\ & + \frac{\mathbb{E}_{\mathcal{P}}[R(A)]\mathbb{E}[Q]}{\gamma_B} - \left(\frac{\mathbb{E}_{\mathcal{P}}[R(A)]\mathbb{E}[Q]}{\gamma_B} \right)^2 \end{aligned} \quad (36)$$

It should be noted that from Proposition 5 we have $\mathbb{E}_{\mathcal{P}}[\psi_n(A)] = \delta_{n,0}$ and hence by computing the expectation with respect to the Poisson distribution on both sides of (40) we have that $\mathbb{E}_{\mathcal{P}}[R(A)] = \sigma_0$. By substituting σ_0 in (36) and through some manipulation we obtain

$$\begin{aligned} \text{var}(B) = & \frac{1}{2\gamma_B} \left[(\mathbb{E}[Q^2] - \mathbb{E}[Q])\sigma_0 \right. \\ & \left. + \frac{2(\mathbb{E}[Q])^2}{\gamma_B} \sum_{n=1}^{\infty} \frac{\gamma_B \sigma_n^2 n! (\mathbb{E}_{\mathcal{P}}[A])^n}{n\gamma_A + \gamma_B} \right] + \frac{\mathbb{E}[Q]}{\gamma_B} \sigma_0 \end{aligned} \quad (37)$$

Since all terms in (37) are positive by using only the first term in the summation appearing in (37) we obtain

$$\text{var}(B) > \frac{(\gamma_A + \gamma_B)\sigma_0(\mathbb{E}[Q^2] + \mathbb{E}[Q]) + 2(\mathbb{E}[Q])^2\sigma_1^2\mathbb{E}_{\mathcal{P}}[A]}{2\gamma_B(\gamma_B + \gamma_A)} \quad (38)$$

C. Expansion of $R(A)$

The proposition below is based on ideas in [12]. The proof is omitted due to page constraints.

Proposition 4: Consider $R(A) : \mathbb{Z}^{\geq} \rightarrow \mathbb{R}^{\geq}$, that satisfies Assumption 2. Let $A \in \mathbb{Z}^{\geq}$ and consider the function

$$\psi_n(A) = \sum_{k=0}^n \binom{n}{k} (-r)^k (A)_{n-k} \quad (39)$$

where $r \in \mathbb{R}^{\geq}$. Then the following equality holds:

$$R(A) = \sum_{n=0}^{\infty} \sigma_n \psi_n(A) \quad (40)$$

where

$$\sigma_n = \sum_{k=n}^{\infty} \binom{k}{n} \frac{\Delta_1^k [R](A)}{k!} \Big|_{A=0} r^{k-n} \quad (41)$$

D. Proof of Proposition 2

Proof: The proof follows by substituting the values of σ_0 and σ_1 in (5) when $R(a) = R_c a$ and by comparing with that obtained using LNA (see [13] and [14]). ■

E. Properties of ψ_n

The proof of the following proposition is analogous to one in [12] and it is omitted due to page constraints.

Proposition 5: Consider $\psi_n(A)$ as in Proposition 4 with $r = \mathbb{E}_{\mathcal{P}}[A]$, and A is a random variable with Poisson distribution. Then the following equalities hold:

$$\mathbb{E}_{\mathcal{P}}[\psi_n(A)] = \delta_{n,0}, \quad \mathbb{E}_{\mathcal{P}}[\psi_n(A)\psi_{n'}(A)] = n!(\mathbb{E}_{\mathcal{P}}[A])^n \delta_{n',n} \quad (42)$$

REFERENCES

- [1] D. T. Gillespie, "Stochastic simulation of chemical kinetics," *Annual Review of Physical Chemistry*, vol. 58, no. 1, pp. 35–55, may 2007.
- [2] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *The Journal of Chemical Physics*, vol. 124, no. 4, p. 044104, jan 2006.
- [3] J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona, "Bounding the stationary distributions of the chemical master equation via mathematical programming," *The Journal of Chemical Physics*, vol. 151, no. 3, p. 034109, jul 2019.
- [4] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*. Imperial College Press, oct 2009.
- [5] K. R. Ghusinga, C. A. Vargas-Garcia, A. Lamperski, and A. Singh, "Exact lower and upper bounds on stationary moments in stochastic biochemical systems," *Physical Biology*, vol. 14, no. 4, p. 04LT01, jun 2017.
- [6] G. R. Dowdy and P. I. Barton, "Bounds on stochastic chemical kinetic systems at steady state," *The Journal of Chemical Physics*, vol. 148, no. 8, p. 084106, feb 2018.
- [7] J. P. Hespanha and A. Singh, "Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems," *International Journal of Robust and Nonlinear Control*, vol. 15, no. 15, pp. 669–689, 2005.
- [8] A. Singh and J. P. Hespanha, "Approximate moment dynamics for chemically reacting systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 414–418, feb 2011.
- [9] P. Smadbeck and Y. N. Kaznessis, "A closure scheme for chemical master equations," *Proceedings of the National Academy of Sciences*, vol. 110, no. 35, pp. 14261–14265, aug 2013.
- [10] D. L. Snyder, *Random Point Processes*. John Wiley & Sons Inc, 1975.
- [11] C. Zechner, G. Seelig, M. Rullan, and M. Khammash, "Molecular circuits for dynamic noise filtering," *Proceedings of the National Academy of Sciences*, vol. 113, no. 17, pp. 4729–4734, apr 2016.
- [12] M. Hinczewski and D. Thirumalai, "Cellular signaling networks function as generalized wiener-kolmogorov filters to suppress noise," *Physical Review X*, vol. 4, no. 4, oct 2014.
- [13] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier LTD, Oxford, 2007.
- [14] I. Lestas, J. Paulsson, N. E. Ross, and G. Vinnicombe, "Noise in gene regulatory networks," *IEEE Transactions on Automatic Control*, vol. 53, no. Special Issue, pp. 189–200, jan 2008.
- [15] J. H. Abel, B. Drawert, A. Hellander, and L. R. Petzold, "GillesPy: A python package for stochastic model building and simulation," *IEEE Life Sciences Letters*, vol. 2, no. 3, pp. 35–38, sep 2016.
- [16] C. H. L. Beentjes, R. Perez-Carrasco, and R. Grima, "Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics," *Physical Review E*, vol. 101, no. 3, mar 2020.
- [17] S. Roman, *The umbral calculus*. New York: Academic Press, 1984.