Spectral theory of stochastic gene expression: a Hilbert space framework

Bingjie Wu[1,2], Ramon Grima[3,*], Chen Jia[2,*]

[1] Center for Applied Mathematics, Tianjin University, Tianjin 300072, China.

[2] Applied and Computational Mathematics Division, Beijing Computational Science Research Center, Beijing 100193, China.

[3] School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom.

[*] Correspondence: ramon.grima@ed.ac.uk (R. Grima); chenjia@csrc.ac.cn (C. Jia)

Abstract

A survey of the literature reveals notable discrepancies among the purported exact results for the spectra of stochastic gene expression models. For self-repressing gene circuits, previous studies ([Phys. Rev. Lett. 99, 108103 (2007)], [Phys. Rev. E 83,062902 (2011)], [J. Chem. Phys. 160, 074105 (2024)], and [bioRxiv 2025.02.05.635946 (2025)]) have provided different exact solutions for the eigenvalues of the generator matrix. In this work, we propose a unified Hilbert space framework for the spectral theory of stochastic gene expression. Based on this framework, we analytically derive the spectra for models of constitutive, bursty, and autoregulated gene expression. The eigenvalues and eigenvectors obtained are then used to construct an exact spectral representation of the time-dependent distribution of gene product numbers. The spectral gap between the zero eigenvalue and the first nonzero eigenvalue, which reflects the relaxation rate of the system towards its steady state, is then compared with the prediction of the deterministic model, and we find that deterministic modeling fails to capture the relaxation rate when autoregulation is strong. In particular, our results demonstrate that for infinite-dimensional operators such as in stochastic gene expression models, many conclusions in linear algebra do not apply, and one must rely on the modern theory of functional analysis.

1 Introduction

Over the past few decades, significant progress has been made to derive a spectral theory of stochastic gene expression. Since the mRNA and protein numbers can take any nonnegative integer values, the state space of a stochastic gene expression model is usually infinite and the corresponding generator matrix of the Markovian dynamics is usually an infinite-dimensional operator. The spectral theory is crucial for the study of stochastic gene expression dynamics, mainly for the following two reasons. First, it is well known that the spectral gap between the zero eigenvalue and the first nonzero eigenvalue of the generator matrix characterizes the relaxation rate of the system towards its steady state [1]. The relaxation rate reflects the ability of gene expression levels to return to equilibrium after environmental changes or signal perturbations, serving as a key indicator of robustness and adaptation of the system. Second, the spectral decomposition method has been widely used to obtain the exact time-dependent solution for stochastic gene expression models [2–4], with different spectral components corresponding to different time scales of the system's dynamical evolution. Therefore, the spectral theory can reveal the detailed dynamical behavior of the system across different time scales.

Within the continuous-time Markovian framework, some studies model the protein dynamics explicitly and model the gene dynamics implicitly [5, 6]. For such models, the spectral theory has been well developed for both unregulated

and autoregulated genes when protein synthesis is constitutive [1], i.e. the molecules are produced one at a time. For unregulated genes, the eigenvalues are discrete and are given by

$$\lambda_n = -nd, \quad n = 0, 1, 2 \cdots ,$$

where d is the rate of protein degradation. By analyzing the properties of the first nonzero eigenvalue, it has been shown that positive autoregulation always slows down the relaxation kinetics to the steady state, while negative autoregulation always speeds it up [1]. In addition, the time-dependent solution for an autoregulatory feedback loop has been obtained when all eigenvalues of the Markovian model are known [7, 8].

Other studies have considered more realistic models where both the gene and protein dynamics are modeled explicitly [9–14]. For such models, the spectral theory has also been examined for both unregulated and autoregulated genes [3, 4, 15–17]. Since now the gene switching dynamics is taken into account, for unregulated genes, there are two sets of eigenvalues [4]:

$$\lambda_{1,n} = -nd, \quad \lambda_{2,n} = -(a + b + nd), \quad n = 0, 1, 2 \cdots ,$$

where d is the rate of protein degradation, and a and b are the switching rates between the active and inactive gene states. However, for autoregulatory gene circuits, results on the spectral theory do not agree with each other — previous studies have obtained different purported exact solutions of the eigenvalues. Ramos and coworkers [3, 15–17] claimed that the eigenvalues for self-repressing genes are all real numbers. However, our previous study [4] showed that the eigenvalues for the same system are the roots of a non-trivial continued fraction equation and are not always real as previously claimed. Thus far, there is a lack of a unified mathematical framework for the spectral theory of stochastic gene expression models that can be used to clarify which eigenvalues are correct and which are incorrect.

In the present work, we propose a rigorous Hilbert space framework for the spectral theory of stochastic gene expression. Based on this framework, we analytically derive the spectra for three classical models, including models of constitutive genes, bursty genes, and self-repressing genes (Fig. 1). For constitutive genes, we consider a simple model with only constitutive mRNA synthesis and mRNA degradation. For bursty genes, we consider a model with bursty protein synthesis and protein degradation, where the burst size of protein is assumed to have a geometric distribution, supported by experiments [18]. For self-repressing genes, we consider a more complicated model that includes protein synthesis, protein degradation, gene state switching, as well as autoregulation, whereby the protein produced from a gene represses its own transcription. Our results show that for self-repressing genes, the eigenvalues given in [4] are correct and can be complex numbers. This is contradictory to the real eigenvalues reported in [3, 15–17]. In particular, we demonstrate that for infinite-dimensional operators such as in stochastic gene expression models, many conclusions in linear algebra do not apply, and one must rely on the modern theory of functional analysis.

As an application of our analytical results, we use the eigenvalues and eigenvectors obtained to construct an exact spectral representation of the time-dependent distribution of gene product numbers for the three models. The magnitude of the real part of the first nonzero eigenvalue, which reflects the relaxation rate to the steady state, is also compared with the prediction of the deterministic model. We demonstrate that deterministic modeling accurately capture the relaxation rate for constitutive and bursty genes, while it fails to reproduce the relaxation rate for self-repressing genes.

2 Hilbert space framework

Stochastic gene expression dynamics is typically modeled as a Markov jump process (continuous-time Markov chain) with a discrete state space. Let V be the state space of the Markovian model and let $Q = (q_{xy})_{x,y \in V}$ be the

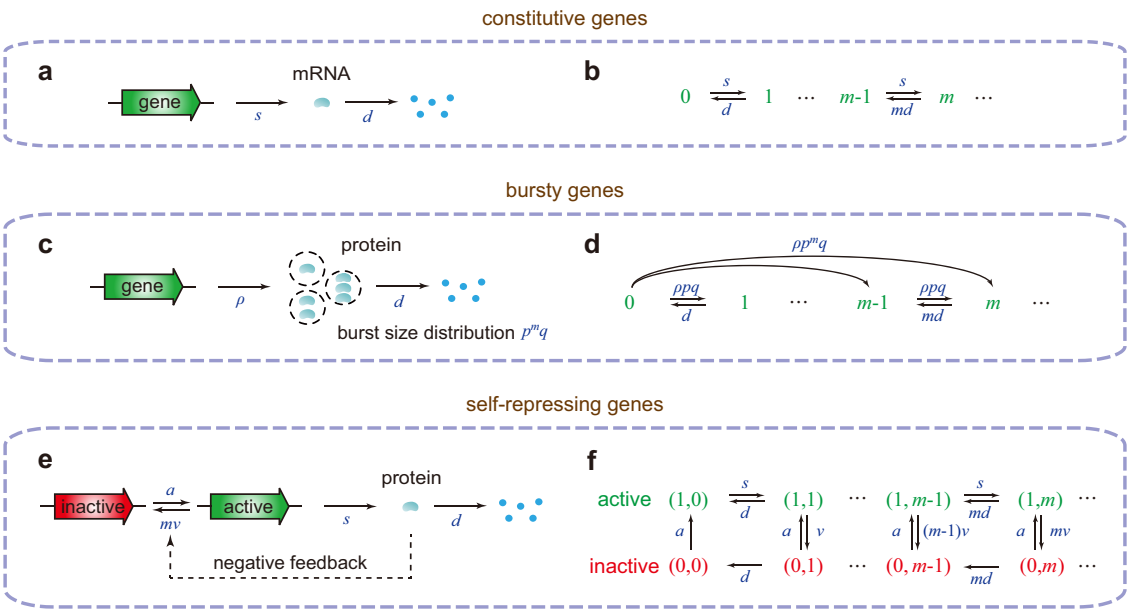

Figure 1: **Models**. **(a)** Model with constitutive mRNA synthesis and mRNA degradation. **(b)** Transition diagram of the Markovian dynamics for the model shown in (a). **(c)** Model with bursty protein synthesis and protein degradation. **(d)** Transition diagram of the Markovian dynamics for the model shown in (c). **(e)** Model of a self-repressing gene, where the protein expressed from a gene represses its own transcription. **(f)** Transition diagram of the Markovian dynamics for the model shown in (e). In (b) and (d), the microstate of the gene is only described by the gene product number $m = 0, 1, 2, \cdots$, while in (f) the microstate state of the gene is described by (i, m), where $i = 0, 1$ represents the gene state and $m = 0, 1, 2, \cdots$ represents the protein number.

associated generator matrix. The aim of the present paper is to find all the eigenvalues of Q for some important gene expression models. If the state space V is finite, then all the eigenvalues of the finite-dimensional Q can be determined straightforwardly using the standard linear algebraic method — they are simply all the roots of the characteristic equation $\det(\lambda I - Q) = 0$. However, the numbers of mRNA and protein in living cells can take any nonnegative integer values and hence the state space V is usually infinite. In such cases, the eigenvalue problem becomes fundamentally different from the finite-dimensional scenario, and continuing to use the conventional linear algebra method may result in non-physical outcomes.

To resolve this problem, we revisit modern functional analysis theory, where the eigenvalues and eigenvectors of an infinite-dimensional operator must be defined within a given Hilbert or Banach space. For Markov jump processes with infinite state spaces, a common choice for is the Hilbert space [1, 19]:

$$l^2(\pi) = \left\{ \phi = (\phi_x) \in \mathbb{C}^V : \sum_{x \in V} |\phi_x|^2 \pi_x < \infty \right\},$$

where $\pi = (\pi_x)_{x \in V}$ is the steady-state distribution of the Markovian model. Note that the generator matrix of a stochastic gene expression model is typically an unbounded operator on $l^2(\pi)$ (this can be readily observed in subsequent sections — the eigenvalues of stochastic gene expression models often have magnitudes that tend to infinity, whereas the eigenvalues of a bounded operator must be bounded by its operator norm [20]). Rigorously, the generator matrix can be viewed as the unbounded operator $Q : D(Q) \to l^2(\pi)$ on $l^2(\pi)$ defined by

$$(Q\phi)_x = \sum_{y \in V} q_{xy} \phi_y, \tag{1}$$

where

$$D(Q) = \left\{ \phi = (\phi_x) \in l^2(\pi) : (Q\psi)_x \text{ is well defined for all } x \in V \text{ and } \sum_{x \in V} |(Q\phi)_x|^2 \pi_x < \infty \right\}$$

is the domain of the definition of Q (for unbounded operators, the domain of definition is only a subspace of $l^2(\pi)$ rather than the entire Hilbert space).

We next recall the definitions of the spectrum and eigenvalues of the generator matrix Q [20].

Definition 1. The resolvent set of the generator matrix Q is the set of all $\lambda \in \mathbb{C}$ such that $Q - \lambda I$ is a one-to-one mapping of $D(Q)$ onto $l^2(\pi)$ whose inverse belongs to $\mathcal{B}(l^2(\pi))$, where I denotes the identity operator and $\mathcal{B}(l^2(\pi))$ denotes the space of all bounded linear operators on $l^2(\pi)$. In other words, $Q - \lambda I$ should have inverse $S \in \mathcal{B}(l^2(\pi))$ which satisfies

$$S(Q - \lambda I) \subset (Q - \lambda I)S = I.$$

The spectrum of Q is the complement of resolvent set of Q, i.e. the set of all $\lambda \in \mathbb{C}$ for which $Q - \lambda I$ does not have an inverse that is a bounded linear operator.

Definition 2. If there exists a nonzero complex-valued column vector $\phi = (\phi_x)_{x \in V} \in l^2(\pi)$ such that $Q\phi = \lambda\phi$, then we say that $\lambda \in \mathbb{C}$ is an eigenvalue of the generator matrix Q and ϕ is the associated eigenvector. In particular, the set of all eigenvalues $\lambda \in \mathbb{C}$ is called the point spectrum of Q.

A well known result in functional analysis is that for an unbounded operator, the set of eigenvalues, i.e. the point spectrum, must be included in the spectrum, but the spectrum is not just the set of eigenvalues — the spectrum can be decomposed into the point spectrum, continuous spectrum, and residual spectrum [20]. In the present work, we only focus on the point spectrum for stochastic gene expression models.

We emphasize that the nonzero column vector $\phi = (\phi_x)_{x \in V}$ in Definition 2 corresponds to the right eigenvector of the generator matrix Q, and the associated $\lambda \in \mathbb{C}$ corresponds to the right eigenvalue. Similarly, one can also define the left eigenvalues and eigenvectors of Q, where the Hilbert space should be chosen as

$$l^2(1/\pi) = \left\{ \psi = (\psi_x) \in \mathbb{C}^{\mathbb{V}} : \sum_{x \in V} |\psi_x|^2 / \pi_x < \infty \right\}.$$

Note that $l^2(1/\pi)$ can be viewed as the dual space of $l^2(\pi)$ [20] in the sense that every $\psi \in l^2(1/\pi)$ induces a bounded linear functional f_ψ on $l^2(\pi)$ defined by $f_\psi(\phi) = \sum_{x \in V} \phi_x \psi_x$.

Definition 3. If there exists a nonzero complex-valued row vector $\psi = (\psi_x)_{x \in V} \in l^2(1/\pi)$ such that $\psi Q = \lambda\psi$, then we say that $\lambda \in \mathbb{C}$ is a left eigenvalue of the generator matrix Q and ψ is the associated left eigenvector.

For stochastic gene expression models, the left eigenvalues (defined in $l^2(1/\pi)$) and the right eigenvalues (defined in $l^2(\pi)$) are generally identical. This equivalence is far from obvious will be discussed in detail in a forthcoming paper. In what follows, we mainly focus on the the left eigenvalues and eigenvectors since they are more amenable to analytical calculation. Note that the above framework is similar to that of quantum mechanics, where any physically meaningful observable must be a Hermitian operator on a Hilbert space and its eigenvalues correspond to all possible outcomes of a measurement of the given observable [21].

3 A simple model of constitutive mRNA production

Let G denote the gene of interest and let M denote the corresponding mRNA. We first focus on a simple gene expression model with constitutive mRNA synthesis (mRNA molecules are produced one at a time) and mRNA degradation. The expression dynamics for a constitutive gene is described by the reactions (Fig. 1(a))

$$G \xrightarrow{s} G + M, \quad M \xrightarrow{d} \varnothing, \tag{2}$$

where s is the rate of mRNA synthesis and d is the rate of mRNA degradation. The microstate of the gene is described by the mRNA number $m = 0, 1, 2, \cdots$. The stochastic evolution of the mRNA number in a single cell is characterized by the birth and death process shown in Fig. 1(b). Recall that it is a Markov jump process with generator matrix

$$Q = (q_{nm})_{n,m\geq 0} = \begin{pmatrix} -s & s & & & \\ d & -(d+s) & s & & \\ & 2d & -(2d+s) & \ddots & \\ & & \ddots & \ddots & \end{pmatrix},$$

where q_{mn} denotes the transition rate from microstate m to microstate n. In steady state, it is well known that the transcript number for the constitutive gene expression model shown in Fig. 1(a) has the Poisson distribution

$$\pi_m = \frac{(s/d)^m}{m!} e^{-s/d}. \tag{3}$$

A well-known result is that for this model, the eigenvalues of the generator matrix Q are given by [1, 17]

$$\lambda_n = -nd, \quad n = 0, 1, 2, \cdots.$$

In a recent paper [17], a linear algebraic method was used to compute the eigenvalues for stochastic gene expression dynamics. The logic of the linear algebraic method is as follows. Let Q be the generator matrix of a stochastic gene expression model; if one can find a square matrix Φ and a diagonal matrix Λ such that $\Phi Q = \Lambda\Phi$, then the diagonal elements of Λ are all the eigenvalues of Q. In fact, it can be proved that for any diagonal matrix $\Lambda = \mathrm{diag}(\lambda_0, \lambda_1, \cdots)$, one can always find a square matrix $\Phi = (\phi_{nm})_{n,m\geq 0}$ such that

$$\Lambda\Phi = \Phi Q. \tag{4}$$

Note that Eq. (4) can be written in component form as

$$\lambda_n \phi_{nm} = \sum_{l=0}^{\infty} \phi_{nl} q_{lm} = \phi_{n,m-1} s - \phi_{nm}(md + s) + \phi_{n,m+1}(m+1)d, \tag{5}$$

where $\phi_{n,-1} = 0$ by default. Directly deriving an explicit expression of ϕ_{nm} from the above recurrence relation is difficult and hence we use the generating function method. To proceed, we introduce the generating functions

$$f_n(z) = \sum_{m=0}^{\infty} \phi_{nm} z^m, \quad n = 0, 1, 2, \cdots.$$

Thus, Eq. (5) can be expressed as the following differential equation:

$$\lambda_n f_n(z) = s(z-1) f_n(z) - d(z-1) f_n'(z). \tag{6}$$

For convenience, we set $\phi_{n0} = 1$, which ensures that $f_n(z) = 1$. It then follows that Eq. (6) can be solved explicitly as

$$f_n(z) = e^{sz/d}(1-z)^{-\lambda_n/d}. \tag{7}$$

Hence the components of the matrix Φ can be computed in closed form as

$$\phi_{nm} = \frac{1}{m!} f_n^{(m)}(0) = \frac{1}{m!} \sum_{l=0}^{m} C_{m,l} (\lambda_n/d)_l (s/d)^{m-l}, \tag{8}$$

where $(x)_l = x(x+1)\cdots(x+l-1)$ denotes the Pochhammer symbol and $C_{m,l} = m!/l!(m-l)!$ denotes the combinatorial number. This implies that, regardless of how the diagonal matrix Λ is chosen, one can always find a matrix Φ such that Eq. (4) holds. Hence, even if Eq. (4) holds, we cannot conclude that the diagonal components of Λ are all eigenvalues of the generator matrix Q.

We emphasize that the linear algebraic method is indeed valid if Q is a finite-dimensional matrix; however, it may not work if Q is an infinite-dimensional matrix. In what follows, we use the Hilbert space framework to compute all eigenvalues for constitutive genes. Following the theoretical framework introduced in Section 2, we choose the Hilbert space

$$l^2(1/\pi) = \left\{ \psi = (\psi_m) \in \mathbb{C}^{\mathbb{N}} : \sum_{m=0}^{\infty} |\psi_m|^2/\pi_m < \infty \right\},$$

and our aim is to find all $\lambda \in \mathbb{C}$ and nonzero row vectors $\psi = (\psi_m) \in l^2(1/\pi)$ such that $\psi Q = \lambda \psi$. In fact, we can prove the following theorem.

Theorem 1. For constitutive genes, all the eigenvalues of the generator matrix Q are given by $\lambda_n = -nd$, $n \geq 0$, and the corresponding eigenvectors $\psi^{(n)} = (\psi_m^{(n)})$, $n \geq 0$ are given by

$$\psi_m^{(n)} = \frac{1}{m!} \sum_{l=0}^{m} C_{m,l} (\lambda_n/d)_l (s/d)^{m-l}. \tag{9}$$

Proof. Let $\psi = (\psi_m)$ be the eigenvector corresponding to the eigenvalue $\lambda \in \mathbb{C}$. To compute the explicit expression of the eigenvectors, we introduce the generating function

$$f(z) = \sum_{m=0}^{\infty} \psi_m z^m. \tag{10}$$

Similarly to Eq. (7), the generating function f can be computed exactly as

$$f(z) = e^{sz/d}(1-z)^{-\lambda/d}. \tag{11}$$

The eigenvector $\psi = (\psi_m)$ can then be computed by taking the derivatives of the generating function f, i.e.

$$\psi_m = \frac{1}{m!} \sum_{l=0}^{m} C_{m,l} (\lambda/d)_l (s/d)^{m-l}.$$

To find all eigenvalues of the generator matrix Q, we first focus on the case when $\lambda = -nd$ for some $n \geq 0$. In this case, the generating function $f(z) = e^{sz/d}(1-z)^n$ is holomorphic over the whole complex plane and thus the convergence radius of the power series given in Eq. (10) is ∞, which implies that (recall that the convergence radius of the power series $\sum_{m=0}^{\infty} \psi_m z^m$ is $1/\overline{\lim}_{m\to\infty} \sqrt[m]{|\psi_m|}$ [22])

$$\overline{\lim_{m\to\infty}} \sqrt[m]{|\psi_m|} = 0.$$

Hence we have $|\psi_m| \leq (1/2)^m$ when $m \gg 1$. On the other hand, when $\lambda = -nd$, the term $(\lambda/d)_l = (-n)_l$ will vanish when $l \geq n+1$. This suggests that for any $m \geq n$,

$$\begin{aligned}|\psi_m/\pi_m| &= \left| e^{s/d} \sum_{l=0}^{n} C_{m,l}(-n)_l (s/d)^{-l} \right| = e^{s/d} \left| \sum_{l=0}^{n} \frac{m!}{(m-l)!}(-1)^l C_{n,l}(d/s)^l \right| \\ &\leq e^{s/d} \sum_{l=0}^{n} C_{n,l} m^l (d/s)^l = e^{s/d}(md/s+1)^n.\end{aligned}$$

This further gives rise to

$$\sum_{m=0}^{\infty} |\psi_m|^2/\pi_m \leq \sum_{m=0}^{n-1} |\psi_m|^2/\pi_m + e^{s/d} \sum_{m=n}^{\infty} |\psi_m|(md/s+1)^n < \infty,$$

where we have used the fact that $|\psi_m| \leq (1/2)^m$ when $m \gg 1$. This clearly shows that $\psi \in l^2(1/\pi)$ and thus $\lambda = -nd$ is an eigenvalue of the generator matrix Q.

We next focus on the case when $\lambda \neq -nd$ for all $n \geq 0$. In this case, $z = 1$ is a singular point of the generating function $f(z) = e^{sz/d}(1-z)^{-\lambda/d}$ and thus the convergence radius of the power series given in Eq. (10) is 1, which implies that

$$\overline{\lim_{m\to\infty}} \sqrt[m]{|\psi_m|} = 1.$$

Hence we have $\overline{\lim}_{m\to\infty} |\psi_m| 2^m \geq 1$. It then follows from Eq. (3) that

$$\overline{\lim_{m\to\infty}} |\psi_m|^2/\pi_m \geq e^{s/d} \overline{\lim_{m\to\infty}} \frac{m!}{4^m (s/d)^m} = \infty,$$

where we have used the fact that $\lim_{m\to\infty} m!/C^m = \infty$ for any $C > 0$. This clearly shows that $\psi \notin l^2(1/\pi)$ and hence $\lambda \neq -nd$ is not an eigenvalue of the generator matrix Q. □

In summary, we have shown that for constitutive genes, all the eigenvalues of the generator matrix Q are given by $\lambda_n = -nd$, $n \geq 0$. In particular, our analysis highlights that for infinite-dimensional operators, many conclusions in linear algebra do not apply and we must rely on the theory of functional analysis. To validate this result, we illustrate the partial sum of the power series $\sum_{m=0}^{\infty} |\psi_m|^2/\pi_m$ for $\lambda = -nd$ and for $\lambda \neq -nd$ in Fig. 2. It is clear that for the former case, the power series converges (Fig. 2(a)) and hence $\lambda = -nd$ is a true eigenvalue; for the latter case, the power series diverges (Fig. 2(b)) and hence $\lambda \neq -nd$ is not an eigenvalue.

An important application of our spectral results is to obtain the exact time-dependent distribution of gene product numbers. Let $p_m(t)$ denote the probability of having m mRNA molecules at time t and $\mathbf{p}(t) = (p_m(t))_{m\geq 0}$ denote the time-dependent mRNA distribution. The evolution of the Markovian dynamics is governed by the following chemical master equation (CME):

$$\dot{\mathbf{p}}(t) = \mathbf{p}(t)Q. \tag{12}$$

We next solve the CME using the spectral decomposition method [2]. Specifically, if the initial distribution $\mathbf{p}(0)$ of the system can be decomposed into a linear combination of eigenvectors, i.e.

$$\mathbf{p}(0) = \sum_{n=0}^{\infty} C_n \psi^{(n)}, \tag{13}$$

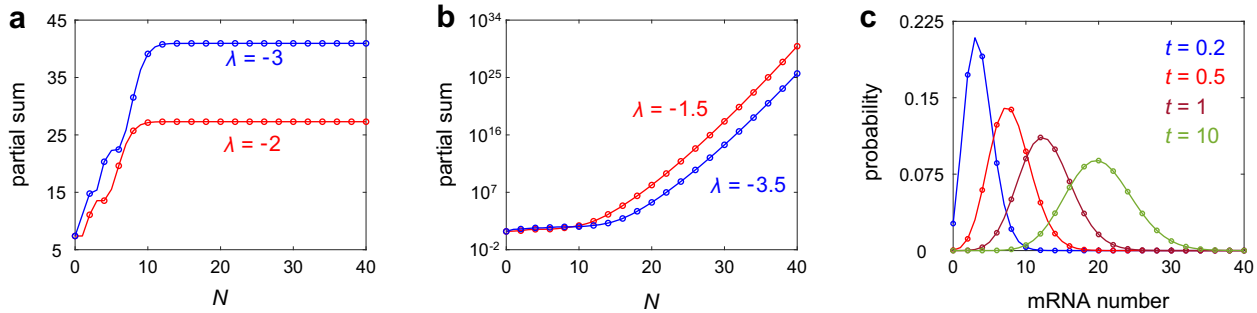

Figure 2: **Using the Hilbert space framework to determine the eigenvalues for a constitutive gene expression model**. Here the partial sum $\sum_{m=0}^{N} |\psi_m|^2/\pi_m$ of the power series $\sum_{m=0}^{\infty} |\psi_m|^2/\pi_m$ is plotted as a function of N under different choices of λ. **(a)** Partial sum $\sum_{m=0}^{N} |\psi_m|^2/\pi_m$ as a function of N for $\lambda = -2$ and $\lambda = -3$. Any $\lambda = -nd$ leads to the convergence of the power series. **(b)** Same as in (a) but for $\lambda = -1.5$ and $\lambda = -3.5$. Any $\lambda \neq -nd$ leads to the divergence of the power series. In (a), (b), the model parameters are chosen as $s = 2$ and $d = 1$. **(c)** Comparison between the exact and numerical mRNA distributions at four time points. The curves show the exact distributions given in Eqs. (14), (19), and (20), and the circles show the numerical ones obtained from FSP. The parameters are chosen to be $s = 20$ and $d = 1$. The initial mRNA number is chosen to be zero.

where $\psi^{(n)}$, $n \geq 0$ are the eigenvectors given in Eq. (9), then the time-dependent mRNA distribution admits the following spectral representation:

$$\mathbf{p}(t) = \sum_{n=0}^{\infty} C_n e^{-ndt} \psi^{(n)}. \tag{14}$$

The remaining question is how to determine the coefficients C_n in Eq. (13). To solve this, let

$$f_n(z) = \sum_{m=0}^{\infty} \psi_m^{(n)} z^m = e^{sz/d}(1-z)^n \tag{15}$$

be the generating function associated with the eigenvector $\psi^{(n)}$ (see Eq. (11)). It then follows from Eq. (13) that

$$\sum_{m=0}^{\infty} p_m(0) z^m = \sum_{n=0}^{\infty} C_n f_n(z). \tag{16}$$

Taking the ith derivative on both sides of Eq. (15) yields

$$f_n^{(i)}(z) = \begin{cases} \displaystyle\sum_{k=0}^{n} (-1)^k \frac{i!n!}{k!(i-k)!(n-k)!} (s/d)^{i-k} e^{sz/d} (1-z)^{n-k}, & i \geq n, \\ \displaystyle\sum_{k=0}^{i} (-1)^k \frac{i!n!}{k!(i-k)!(n-k)!} (s/d)^{i-k} e^{sz/d} (1-z)^{n-k}, & i < n. \end{cases}$$

Evaluating the above equation at $z = 1$, we obtain

$$f_n^{(i)}(1) = \begin{cases} (-1)^n \dfrac{i!}{(i-n)!} (s/d)^{i-n} e^{s/d}, & i \geq n, \\ 0, & i < n. \end{cases} \tag{17}$$

Taking the ith derivative of both sides of Eq. (16) and taking $z = 1$, we have

$$\sum_{m=i}^{\infty} p_m(0) \frac{m!}{(m-i)!} = \sum_{n=0}^{\infty} C_n f_n^{(i)}(1) = \sum_{n=0}^{i} C_n (-1)^n \frac{i!}{(i-n)!} (s/d)^{i-n} e^{s/d}. \tag{18}$$

Note that $f_0(1) = e^{s/d}$ and $f_n(1) = 0$ for any $n \geq 1$. It then follows from Eq. (16) that

$$C_0 = e^{-s/d} \sum_{m=0}^{\infty} p_m(0) = e^{-s/d}. \tag{19}$$

Furthermore, from Eq. (18), it is easy to obtain the following recurrence relation for C_i:

$$C_i = (-1)^i e^{-s/d} \sum_{m=i}^{\infty} p_m(0) \frac{m!}{i!(m-i)!} - \sum_{n=0}^{i-1} \frac{C_n}{(i-n)!} (-1)^{i-n} (s/d)^{i-n}, \quad i \geq 1. \tag{20}$$

Eqs. (14), (19), and (20) give the complete time-dependent solution of the CME.

Thus far, we have obtained the exact time-dependent mRNA distribution for constitutive genes using spectral decomposition. To test the exact solution, we compare it with the numerical one obtained from the finite state projection algorithm (FSP) [23]. For simplicity, we assume that the initial mRNA number is zero. This mimics the situation where the gene has been silenced by some repressor over a period of time such that all mRNA molecules have been removed via degradation. At time $t = 0$, the repressor is removed, and we study how gene expression recovers. The results are shown in Fig. 2(c). As expected, the two solutions agree perfectly at all times. We emphasize that distinct eigenvalues characterize different time scales of the system's dynamical evolution. Since the system is ergodic, the time-dependent distribution $\mathbf{p}(t)$ will converge to the steady-state distribution π in the long-term limit. Taking $t \to \infty$ in Eq. (14) yields

$$\pi = \mathbf{p}(t \to \infty) = C_0 \psi^0.$$

Hence the eigenvector $\psi^{(0)}$ corresponding to the zero eigenvalue $\lambda_0 = 0$ is precisely the steady-state distribution of the system (up to a constant). Moreover, it is well-known that the spectral gap between the zero eigenvalue and the first nonzero eigenvalue $\lambda_1 = -d$ characterizes the relaxation rate of the system towards its steady state [1]. A complete characterization of eigenvalues enables a more detailed investigation of the system's dynamical behavior on finer time scales.

4 A simple model of bursty protein production

Experiments have shown that many proteins are produced in a bursty manner [18], due to the rapid translation of protein from a single, short-lived mRNA molecule [24, 25]. Let G denote the gene of interest, let M denote the corresponding mRNA, and let P denote the corresponding protein. Consider the following classical two-stage model of stochastic gene expression:

$$G \xrightarrow{\rho} G + M, \quad M \xrightarrow{u} M + P, \quad M \xrightarrow{v} \varnothing, \quad P \xrightarrow{d} \varnothing, \tag{21}$$

where the first reaction describes transcription, the second reaction describes translation, and the last two reactions describe the degradation of mRNA and protein, respectively. When mRNA decays much faster compared to its protein counterpart ($\gamma = v/d \gg 1$) and when the number of protein molecules produced per mRNA lifetime ($B = u/v$) is strictly positive and bounded, it has been shown [24–26] that protein will be produced in a bursty manner. Specifically, in the limit of $\gamma \gg 1$ and $B = u/v$ is strictly positive and bounded, the two-stage model shown in Eq. (21) reduces to the following gene expression model with bursty protein synthesis and protein degradation (Fig. 1(c)):

$$G \xrightarrow{\rho p^k (1-p)} G + kP, \quad k \geq 1, \quad P \xrightarrow{d} \varnothing, \tag{22}$$

where the synthesis of protein occurs in bursts with frequency ρ and random size k sampled from a geometric distribution with parameter $p = u/(u+v) \in (0,1)$, and d is the rate of protein degradation. The microstate of the gene is described by the protein number $m = 0, 1, 2, \cdots$. The stochastic evolution of the protein number in a single cell is characterized by the Markov jump process shown in Fig. 1(d) with generator matrix

$$Q = \begin{pmatrix} -\rho p & \rho p(1-p) & \rho p^2(1-p) & \cdots \\ d & -(d+\rho p) & \rho p(1-p) & \ddots \\ & 2d & -(2d+\rho p) & \ddots \\ & & \ddots & \ddots \end{pmatrix}.$$

In steady state, it is well known that the protein number for the bursty gene expression model shown in Fig. 1(c) has the negative binomial distribution [5]

$$\pi_m = \frac{(\rho/d)_m}{m!} p^m (1-p)^{\rho/d}. \tag{23}$$

The following theorem shows that the eigenvalues for bursty genes are exactly the same as those for constitutive genes.

Theorem 2. For bursty genes, all the eigenvalues of the generator matrix Q are given by $\lambda_n = -nd$, $n \geq 0$, and the corresponding eigenvectors $\psi^{(n)} = (\psi_m^{(n)})$, $n \geq 0$ are given by

$$\psi_m^{(n)} = \frac{1}{m!} \sum_{l=0}^{m} C_{m,l} (\lambda_n/d)_l (\rho/d)_{m-l} p^{m-l}. \tag{24}$$

Proof. We next find all eigenvalues for bursty genes within the Hilbert space framework introduced above: if there exists a row vector $\psi = (\psi_m) \in l^2(1/\pi)$ such that $\psi Q = \lambda \psi$, then λ is as an eigenvalue of Q and ψ is the associated eigenvector. Note that the characteristic equation $\psi Q = \lambda \psi$ can be written in components as

$$(m+1)d\psi_{m+1} - (md + \rho p)\psi_m + \sum_{l=0}^{m-1} \rho p^{m-l}(1-p)\psi_l = \lambda \psi_m. \tag{25}$$

Using the generating function defined in Eq. (10), Eq. (25) can be converted into the differential equation

$$\frac{\rho p(z-1)}{1-pz} f(z) + d(1-z) f'(z) = \lambda f(z). \tag{26}$$

For convenience, we set $\psi_0 = 1$, which ensures that $f(z) = 1$. Then Eq. (26) can be solved explicitly as

$$f(z) = (1-pz)^{-\rho/d}(1-z)^{-\lambda/d}. \tag{27}$$

The eigenvector $\psi = (\psi_m)$ can then be computed by taking the derivatives of the generating function f, i.e.

$$\psi_m = \frac{1}{m!} \sum_{l=0}^{m} C_{m,l} (\lambda/d)_l (\rho/d)_{m-l} p^{m-l}.$$

To find all eigenvalues of the generator matrix Q, we first focus on the case when $\lambda = -nd$ for some $n \geq 0$. In this case, $z = 1/p$ is a singular point of the generating function $f(z) = (1-pz)^{-\rho/d}(1-z)^n$ and thus the convergence radius of the power series $f(z) = \sum_{m=0}^{\infty} \psi_m z^m$ is $1/p$, which implies that

$$\overline{\lim_{m\to\infty}} \sqrt[m]{|\psi_m|} = p.$$

Hence for any $\delta > 1$, there exists $C(\delta) > 0$ such that $|\psi_m| \leq C(\delta)(p\delta)^m$ for all $m \geq 0$. Moreover, it is easy to check that $(m-1)! < (\rho/d)_m$. It then follows from Eq. (23) that

$$\begin{aligned}\sum_{m=0}^{\infty} |\psi_m|^2/\pi_m &\leq C(\delta)^2(1-p)^{-\rho/d} \sum_{m=0}^{\infty} \frac{m!}{(\rho/d)_m}(p\delta^2)^m \\ &\leq C(\delta)^2(1-p)^{-\rho/d} \sum_{m=0}^{\infty} m(p\delta^2)^m < \infty,\end{aligned}$$

where we have chosen $\delta > 1$ such that $p\delta^2 < 1$. This clearly shows that $\psi \in l^2(1/\pi)$ and thus $\lambda = -nd$ is an eigenvalue.

We next focus on the case when $\lambda \neq -nd$ for all $n \geq 0$. In this case, both $z = 1$ and $z = 1/p > 1$ are singular points of $f(z) = (1-pz)^{-\rho/d}(1-z)^{-\lambda/d}$ and thus the convergence radius of the power series $f(z) = \sum_{m=0}^{\infty} \psi_m z^m$ is 1, which implies that

$$\overline{\lim_{m\to\infty}} \sqrt[m]{|\psi_m|} = 1.$$

Hence we have $\overline{\lim}_{m\to\infty} |\psi_m|\delta^m \geq 1$ for any $\delta > 1$. Since $(k)_m = (k+m-1)!/(k-1)!$, it is easy to check that

$$\frac{(\rho/d)_m}{m!} \leq \frac{(\lceil \rho/d \rceil)_m}{m!} = \frac{(m+1)(m+2)\cdots(m+\lceil \rho/d \rceil - 1)}{1\cdot 2\cdots(\lceil \rho/d \rceil - 1)} \leq (\lceil \rho/d \rceil + m - 1)^{\lceil \rho/d \rceil - 1},$$

where $\lceil x \rceil$ denotes the smallest integer that exceeds x. It then follows from Eq. (23) that

$$\begin{aligned}\overline{\lim_{m\to\infty}} |\psi_m|^2/\pi_m &\geq (1-p)^{-\rho/d} \overline{\lim_{m\to\infty}} \frac{m!}{(\rho/d)_m (p\delta^2)^m} \\ &\geq (1-p)^{-\rho/d} \overline{\lim_{m\to\infty}} \frac{1}{(\lceil \rho/d \rceil + m - 1)^{\lceil \rho/d \rceil - 1}(p\delta^2)^m} = \infty,\end{aligned}$$

where we have chosen $\delta > 1$ such that $p\delta^2 < 1$. This clearly shows that $\psi \notin l^2(1/\pi)$ and thus $\lambda \neq -nd$ is not an eigenvalue. □

The above analysis show that for bursty genes, all the eigenvalues of the generator matrix Q are the same as those for constitutive genes and are still given by $\lambda_n = -nd$, $n \geq 0$. To validate this, we illustrate the partial sum of the power series $\sum_{m=0}^{\infty} |\psi_m|^2/\pi_m$ for $\lambda = -nd$ and for $\lambda \neq -nd$ in Fig. 3. For the former case, the power series converges (Fig. 3(a)) and hence $\lambda = -nd$ is a true eigenvalue; for the latter case, the power series diverges (Fig. 3(b)) and hence $\lambda \neq -nd$ is not an eigenvalue.

We next use the spectral decomposition method to compute the exact time-dependent protein distribution for bursty genes. Let $p_m(t)$ denote the probability of having m protein molecules at time t and $\mathbf{p}(t) = (p_m(t))_{m\geq 0}$ denote the time-dependent protein distribution. Similarly, if the initial distribution $\mathbf{p}(0)$ of the system can be decomposed into a linear combination of eigenvectors, i.e.

$$\mathbf{p}(0) = \sum_{n=0}^{\infty} C_n \psi^{(n)}, \tag{28}$$

where $\psi^{(n)} = (\psi_m^{(n)})_{m\geq 0}$, $n \geq 0$ are the eigenvalues given in Eq. (24), then the time-dependent protein distribution admits the following spectral representation:

$$\mathbf{p}(t) = \sum_{n=0}^{\infty} C_n e^{-ndt} \psi^{(n)}. \tag{29}$$

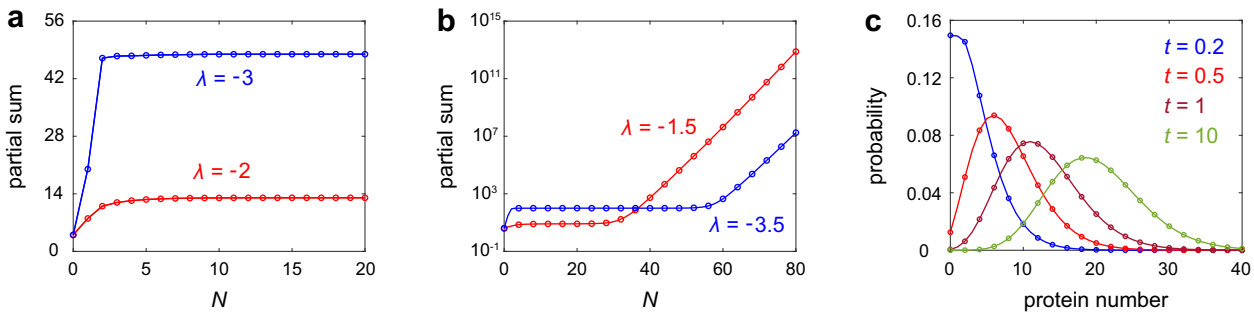

Figure 3: **Using the Hilbert space framework to determine the eigenvalues for a bursty gene expression model**. Here the partial sum $\sum_{m=0}^{N} |\psi_m|^2/\pi_m$ of the power series $\sum_{m=0}^{\infty} |\psi_m|^2/\pi_m$ is plotted as a function of N under different choices of λ. **(a)** Partial sum $\sum_{m=0}^{N} |\psi_m|^2/\pi_m$ as a function of N for $\lambda = -2$ and $\lambda = -3$. Any $\lambda = -nd$ leads to the convergence of the power series. **(b)** Same as in (a) but for $\lambda = -1.5$ and $\lambda = -3.5$. Any $\lambda \neq -nd$ leads to the divergence of the power series. In (a),(b), the model parameters are chosen as $\rho = 2, p = 0.5, d = 1$. **(c)** Comparison between the exact and numerical protein distributions at four time points. The curves show the exact distributions given in Eqs. (29) and (32), and the circles show the numerical ones obtained from FSP. The parameters are chosen to be $\rho = 20, p = 0.5, d = 1$. The initial protein number is chosen to be zero.

We next determine the coefficients C_n in Eq. (28). Let

$$f_n(z) = \sum_{m=0}^{\infty} \psi_m^{(n)} z^m = (1-pz)^{-\rho/d}(1-z)^n \tag{30}$$

be the generating function associated with the eigenvector $\psi^{(n)}$ (see Eq. (27)). Taking the ith derivative on both sides of Eq. (30) yields

$$f_n^{(i)}(z) = \begin{cases} \displaystyle\sum_{k=0}^{n} (-1)^k \frac{i!n!}{k!(i-k)!(n-k)!} p^{i-k} (\rho/d)_{i-k} (1-pz)^{-\rho/d-(i-k)} (1-z)^{n-k}, & i \geq n, \\ \displaystyle\sum_{k=0}^{i} (-1)^k \frac{i!n!}{k!(i-k)!(n-k)!} p^{i-k} (\rho/d)_{i-k} (1-pz)^{-\rho/d-(i-k)} (1-z)^{n-k}, & i < n. \end{cases}$$

Evaluating the above equation at $z = 1$, we obtain

$$f_n^{(i)}(1) = \begin{cases} (-1)^n \dfrac{i!}{(i-n)!} p^{i-n} (1-p)^{-\rho/d-(i-n)} (\rho/d)_{i-n}, & i \geq n, \\ 0, & i < n. \end{cases} \tag{31}$$

Taking the ith derivative of both sides of Eq. (16) and taking $z = 1$, we have

$$\sum_{m=i}^{\infty} p_m(0) \frac{m!}{(m-i)!} = \sum_{n=0}^{i} C_n (-1)^n \frac{i!}{(i-n)!} p^{i-n} (1-p)^{-\rho/d-(i-n)} (\rho/d)_{i-n}.$$

From this equation, it is easy to obtain the following recurrence relation for C_i:

$$\begin{aligned} C_0 &= (1-p)^{\rho/d} \sum_{m=0}^{\infty} p_m(0) = (1-p)^{\rho/d}, \\ C_i &= (-1)^i (1-p)^{\rho/d} \sum_{m=i}^{\infty} p_m(0) \frac{m!}{i!(m-i)!} - \sum_{n=0}^{i-1} \frac{C_n}{(i-n)!} (-1)^{i-n} \left(\frac{p}{1-p}\right)^{i-n} (\rho/d)_{i-n}, \quad i \geq 1. \end{aligned} \tag{32}$$

Eqs. (29) and (32) give the exact time-dependent protein distribution for bursty genes in terms of the eigenvalues and eigenvectors. To test our exact solution, we compare it with the numerical one obtained from FSP (Fig. 3(c)). Again, the two solutions coincide perfectly at all times.

5 Self-repressing genes

Finally, we focus on a more complex gene expression model, where the protein produced from a gene represses its own transcription, forming a negative autoregulatory gene circuit [9, 27]. Let G and G^* denote the inactive and active states of the gene, respectively, and let P denote the corresponding protein. The reaction scheme for the self-repressing gene is as follows (Fig. 1(e)):

$$\begin{aligned} G \xrightarrow{a} G^*, \quad G^* + P \xrightarrow{\nu} G + P, \\ G^* \xrightarrow{s} G^* + P, \quad P \xrightarrow{d} \varnothing. \end{aligned} \tag{33}$$

Here a is the switching rate from the inactive to the active gene state, ν is the binding rate of protein to the gene in its active state which characterizes the strength of negative feedback, s is the synthesis rate of protein when the gene is active, and d is the degradation rate of protein. We emphasize that this model is a simplified version of reality so that the calculations are analytical tractable and simple. Here we assume that there is no change in the protein number during gene activation and inactivation. However, in reality, the protein number decreases by one when a protein copy binds to a gene and increases by one when unbinding occurs [10]. In other words, the present model ignores the protein-gene binding fluctuations and hence it may not be accurate when the protein number is very small or when the feedback is very strong [13, 28].

The microstate of the gene of interest can be represented by the ordered pair (i, m), where i is the state of the gene with $i = 0, 1$ corresponding to the inactive and active states, respectively, and m is the number of protein. The stochastic gene expression dynamics is described by the Markov jump process illustrated in Fig. 1(f) with an infinite-dimensional generator matrix. In the Letter [15], the authors computed the exact spectrum for a self-repressing gene. The authors claimed that the generator matrix Q of the Markovian dynamics had a set of eigenvalues (see Eq. (26) in [15]):

$$\lambda_{1,n} = -n(d+\nu) - a - \frac{s\nu}{d+\nu}, \quad n = 0, 1, 2, \cdots, \tag{34}$$

In their subsequent work [3, 16], they gave another set of eigenvalues for the same model (see Eq. (9) in [3] and Eq. (5) in [16]):

$$\lambda_{2,n} = -nd, \quad n = 0, 1, 2, \cdots, \tag{35}$$

In particular, the two sets of eigenvalues are all real numbers. A "proof" of this result was provided in a recent paper using the linear algebraic method [17] — if one can find a square matrix Φ and a diagonal matrix Λ such that $\Phi Q = \Lambda \Phi$, then the diagonal elements of Λ are all eigenvalues of the generator matrix Q. In Appendix A, we have proved that for any diagonal matrix Λ (its diagonal elements do not need to be all eigenvalues), one can always find a matrix Φ such that $\Phi Q = \Lambda \Phi$. This again shows that the linear algebraic method cannot capture the spectrum of stochastic gene expression if the corresponding generator matrix is infinite-dimensional.

Here we will show that $\lambda_{1,n}$ and $\lambda_{2,n}$ as given by Eqs. (34) and (35) are not the true eigenvalues of Q. To see this, we next apply the Hilbert space framework introduced above to find the true eigenvalues for self-repressing genes: if there exists a row vector $\psi = (\psi_{i,m}) \in l^2(1/\pi)$ such that $\psi Q = \lambda \psi$, then λ is as an eigenvalue of Q and ψ is the associated eigenvector, where

$$l^2(1/\pi) = \left\{ \psi = (\psi_{i,m}) : \sum_{i=0}^{1} \sum_{m=0}^{\infty} |\psi_{i,m}|^2 / \pi_{i,m} < \infty \right\}.$$

a

eigenvalues	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
correct (HF)	0	-1.36	-2.97	-4.05+0.81**i**	-4.05-0.81**i**	-5.10+1.13**i**	-5.10-1.13**i**
correct (GM)	0	-1.36	-2.97	-4.05+0.81**i**	-4.05-0.81**i**	-5.10+1.13**i**	-5.10-1.13**i**
incorrect	0	-1	-2	-3	-4	-5	-5.82

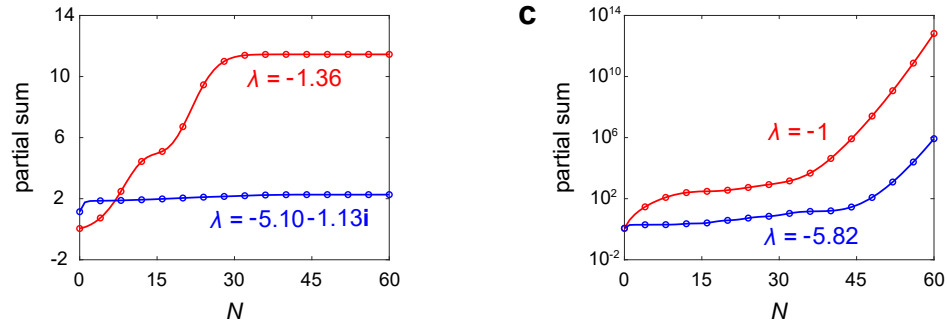

Figure 4: **Spectrum for a self-repressing gene**. **(a)** Comparison of the true eigenvalues with the "eigenvalues" given in Eqs. (34) and (35). The true eigenvalues are obtained by solving the continued fraction equation Eq. (38). Note that the true eigenvalues agree with the approximate eigenvalues of the truncated generator matrix (truncated at $N = 200$ with N being the maximum protein number). The model parameters are chosen as $s = 20, d = 1, a = 4, \nu = 0.1$. **(b)** Partial sum $\sum_{i=0}^{1}\sum_{m=0}^{N}|\psi_{i,m}|^2/\pi_{i,m}$ as a function of N for $\lambda = -1.36$ and $\lambda = -5.10 - 1.13\mathbf{i}$. **(c)** Same as (b) but for $\lambda = -1$ and $\lambda = -5.82$. In (b),(c), the values of λ are chosen to be the first and sixth nonzero true and false eigenvalues shown in (c), respectively. It is clear that the power series converges for the true eigenvalues and diverges for the false ones.

Here $\pi = (\pi_{i,m})$ is the steady-state distribution of the self-repressing gene expression model which has been obtained exactly in [9] and is given by

$$\begin{aligned}\pi_{1,m} &= \frac{d}{s}\frac{(\alpha)_{m+1}}{(\beta)_{m+1}}\frac{w^{m+1}}{m!}\frac{{}_1F_1(\alpha+m+1;\beta+m+1;-wz_0)}{{}_1F_1(\alpha;\beta;w(1-z_0))},\\ \pi_{0,m} &= \frac{(\alpha)_m}{(\beta)_m}\frac{w^m}{m!}\frac{{}_1F_1(\alpha+m;\beta+m;-wz_0)}{{}_1F_1(\alpha;\beta;w(1-z_0))} - \pi_{1,m},\end{aligned} \tag{36}$$

where

$${}_1F_1(\alpha;\beta;z) = \sum_{n=0}^{\infty}\frac{(\alpha)_n}{(\beta)_n}\frac{z^n}{n!}$$

is Kummer's confluent hypergeometric function [29, Sec. 13.2] and

$$\alpha = \frac{a}{d}, \quad \beta = \frac{a}{\nu+d} + \frac{s\nu}{(\nu+d)^2}, \quad w = \frac{s}{\nu+d}, \quad z_0 = \frac{d}{\nu+d}. \tag{37}$$

The following theorem gives the eigenvalues for self-repressing genes.

Theorem 3. For self-repressing genes, there is a zero eigenvalue $\lambda_0 = 0$; all the nonzero eigenvalues $\lambda_n, n \geq 1$ of the generator matrix Q are the roots of the following nontrivial continued fraction equation:

$$q = \frac{R_0P_1}{Q_1-}\frac{R_1P_2}{Q_2-}\frac{R_2P_3}{Q_3-}\cdots, \tag{38}$$

where $P_n = \theta + (n-1)\epsilon$, $Q_n = n[-(n-1+\gamma) - \delta + \epsilon] + q$, and $R_n = -(n+1)(n+\gamma)$ with

$$\begin{aligned}q &= -\frac{(\lambda+d)(a+\lambda)}{d(\nu+d)} - \frac{[\nu(a+\lambda)+d\nu]s}{d(\nu+d)^2}, \quad \theta = -\frac{(a+d+\lambda)\nu s}{d(\nu+d)^2},\\ \gamma &= \frac{a+\lambda}{\nu+d} + \frac{s\nu}{(\nu+d)^2}, \quad \delta = 2 + \frac{\lambda}{d}, \quad \epsilon = -\frac{\nu d s}{d(\nu+d)^2}.\end{aligned}$$

Here we have used the standard shorthand notation for continued fractions [29, Sec. 1.12], i.e.

$$\frac{d_1}{e_1\pm}\frac{d_2}{e_2\pm}\frac{d_3}{e_3\pm}\cdots = \frac{d_1}{e_1 \pm \frac{d_2}{e_2\pm\frac{d_3}{e_3\pm\cdots}}}.$$

Proof. The rigorous proof based on the Hilbert space framework can be found in Appendix B. An informal derivation of the eigenvalues can be found in [4]. □

Fig. 4(a) compares the "eigenvalues" given in Eqs. (34) and (35) and the true ones which are the roots of the continued fraction equation Eq. (38). We also compare the true eigenvalues with the eigenvalues of the truncated generator matrix (truncated at $N = 200$ with N being the maximum protein number), which can be viewed as an accurate approximation of the true eigenvalues when the truncation size N is large. The perfect agreement between the two once again verifies the correctness of our results. It can be seen that the eigenvalues are not always real as previously claimed in [3, 15–17]. This agrees with a well-known result that a negative feedback loop can produce oscillatory behaviors which correspond to complex eigenvalues [30, 31]. For self-repressing genes, the time-dependent distribution of protein numbers has already been constructed using the spectral decomposition method in [4], and hence we omit it here.

It has been previously argued [17] that the generator matrix of a self-repressing gene is self-adjoint and hence must have real eigenvalues. A well-established result in probability theory states that the generator of a Markov jump process is self-adjoint only when the system satisfies detailed balance [32, 33], i.e. the product of transition rates along any cycle is exactly equal to that along its reversed cycle, which is known as the Kolmogorov cyclic condition [34, 35]. It is easy to verify that the Markovian model shown in Fig. 1(f) does not satisfy detailed balance. Hence, its generator cannot be self-adjoint, which implies that the eigenvalues are not restricted to be real.

To further validate our exact result, we compute the eigenvector corresponding to each eigenvalue numerically and then check whether the corresponding eigenvector $\psi = (\psi_{i,m})$ is in the Hilbert space $l^2(1/\pi)$. Fig. 4(b),(c) illustrate the partial sum of the power series $\sum_{i=0}^{1}\sum_{m=0}^{\infty}|\psi_{i,m}|^2/\pi_{i,m}$ with the eigenvalues being chosen as either in Eqs. (34) and (35) or as in Eq. (38). It is clear that the power series converges for the eigenvalues given in Eq. (38) (Fig. 4(b)), while the power series diverges for the eigenvalues given in Eqs. (34) and (35) (Fig. 4(c)). This once again demonstrates the power of the Hilbert space framework in studying the spectral structure of gene expression models.

6 Comparison with the prediction of the deterministic model

In general, stochastic gene expression dynamics is exponentially ergodic [19], which means that the time-dependent probability $p_m(t)$ of having m gene product molecules will converge to the steady-state probability π_n at exponential speed: for any $m \geq 0$, there exists a constant $K_m > 0$ such that

$$|p_m(t) - \pi_m| \leq K_m e^{-\gamma t}, \quad \forall\, t \geq 0.$$

Here the optimal (maximal) parameter γ characterizes the relaxation rate of the system towards its steady state. In the mathematics literature, it is also called exponentially ergodic convergence rate [19]. From the spectral representation of the time-dependent distribution given in Eqs. (14) and (29), it is easy to see that the relaxation rate γ is exactly $|\mathrm{Re}(\lambda_1)|$, the spectral gap between the zero eigenvalue and first nonzero eigenvalue. On the other hand, the relaxation rate of gene expression dynamics can also be analyzed within the conventional deterministic framework, where it is defined as the exponential convergence rate to the fixed point. This raises an interesting question: what is the relationship between the relaxation rates for the stochastic and deterministic models?

For constitutive genes (Fig. 1(a)), according to the reaction scheme given in Eq. (2) and the law of mass action, the corresponding deterministic model is given by

$$\dot{x} = s - dx, \tag{39}$$

where x denotes the mean mRNA number. This is a simple linear differential equation and can be solved explicitly as

$$x(t) = (x(0) - s/d)e^{-dt} + s/d.$$

Clearly, $x^* = s/d$ is the unique globally stable equilibrium, and the deterministic system converges to this equilibrium exponentially at rate d. Note that the first nonzero eigenvalue of the stochastic model is $\lambda_1 = -d$, and hence the relaxation rate for the stochastic model is also $|\text{Re}(\lambda_1)| = d$. Hence for constitutive genes, the deterministic and stochastic models yield the same prediction for the relaxation rate.

For bursty genes (Fig. 1(c)), according to the reaction scheme given in Eq. (22), the corresponding deterministic dynamics is given by

$$\dot{x} = \sum_{k=1}^{\infty} \rho p^k (1-p)k - dx = \rho p - dx,$$

where x denotes the mean protein number. Note that the deterministic model for bursty genes has almost the same form as that for constitutive genes, and hence the relaxation rate for the deterministic model is also d. The prediction of the deterministic model also provides an physical intuition of why $\lambda = -nd$ commonly show up as eigenvalues for stochastic gene expression models.

Finally, we focus on self-repressing genes (Fig. 1(e)). According to the reaction scheme given in Eq. (33), the corresponding deterministic dynamics is given by

$$\begin{cases} \dot{g} = a(1-g) - \nu g x, \\ \dot{x} = sg - dx, \end{cases} \tag{40}$$

where g denotes the mean number of genes in the active state and x denotes the mean protein number. Although the above system may not have an closed-form solution, its relaxation rate is still analytically tractable. It is easy to see that the two-dimensional system Eq. (40) has an unique globally stable equilibrium (g^*, x^*) with

$$\begin{cases} g^* = \dfrac{-ad + \sqrt{a^2d^2 + 4ad\nu s}}{2\nu s}, \\ x^* = \dfrac{-ad + \sqrt{a^2d^2 + 4ad\nu s}}{2\nu d}. \end{cases}$$

In the long-term limit, the deterministic system converges to this equilibrium at exponential speed. We next compute the convergence rate analytically. To this end, note that at the equilibrium (g^*, x^*), the Jacobian matrix of the deterministic model is given by

$$J = \begin{pmatrix} -a - \nu x^* & -\nu g^* \\ s & -d \end{pmatrix}.$$

The exponential convergence rate to the equilibrium is determined by the eigenvalues of the matrix J. It is easy to see that the eigenvalue γ of J satisfies

$$\gamma^2 + (a + d + \nu x^*)\gamma + (ad + \nu d x^* + \nu s g^*) = 0.$$

This is a quadratic equation and hence has the following two roots:

$$\gamma_1 = \frac{-(a+d+\nu x^*)+\sqrt{\triangle}}{2}, \quad \gamma_2 = \frac{-(a+d+\nu x^*)-\sqrt{\triangle}}{2}, \tag{41}$$

where $\triangle = (a+d+\nu x^*)^2 - 4(ad+\nu d x^* + \nu s g^*)$. When $\Delta \geq 0$, both γ_1 and γ_2 are real-valued with $|\gamma_1| \leq |\gamma_2|$; when $\Delta < 0$, both γ_1 and γ_2 are complex-valued and have the same real part. The relaxation rate for the deterministic model, i.e. the exponential convergence rate to the equilibrium, is then given by $|\mathrm{Re}(\gamma_1)|$, i.e. the absolute value of the real part of γ_1. From Eq. (41), it is easy to see that

$$\lim_{\nu\to\infty} |\mathrm{Re}(\gamma_1)| = 2d, \quad \lim_{\nu\to 0} |\mathrm{Re}(\gamma_1)| = \frac{a+d+|a-d|}{2}.$$

This indicates that the relaxation rate for the deterministic model is equal to $2d$ in the limit of strong feedback ($\nu \to \infty$), and is equal to $(a+d+|a-d|)/2$ in the limit of weak feedback ($\nu \to 0$).

Fig. 5 compares the relaxation rates of the deterministic and stochastic models as the feedback strength ν and the gene activation rate a vary. We have observed an interesting critical phenomenon. When $a > 2d$, the relaxation rates for the two models are nearly identical, and both increase with ν. In this regime, the deterministic model successfully captures the relaxation rate of its stochastic counterpart. In contrast, when $a < 2d$, the relaxation rates for the two models remain close only for small ν, but diverge markedly as ν becomes large. These results indicate that under strong feedback, deterministic modeling fails to reproduce the relaxation rate of the system, particularly when the gene activation rate is relatively low.

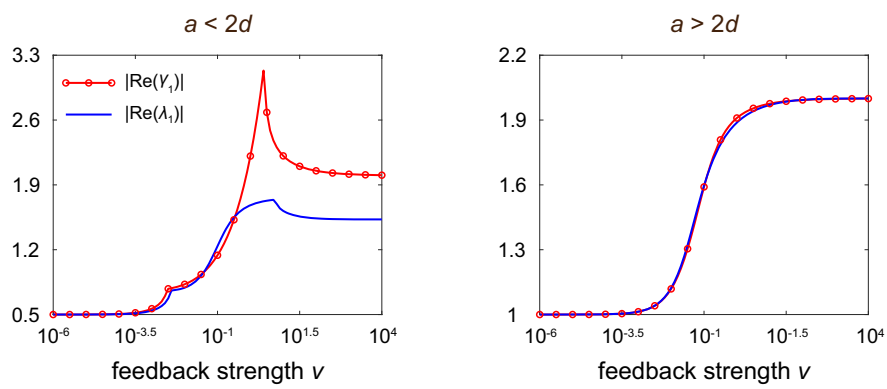

Figure 5: **Comparison between the relaxation rate $|\mathrm{Re}(\lambda_1)|$ for the deterministic model and the relaxation rate $|\mathrm{Re}(\gamma_1)|$ for the stochastic model as a and ν vary**. The gene activation rate a is chosen to be $a = 0.5$ for the left panel and $a = 2.5$ for the right panel. The remaining parameters are chosen as $s = 20$ and $d = 1$.

7 Conclusions

In this work, we propose a rigorous and general mathematical framework to investigate the spectral structure of three classical stochastic gene expression models, including the models for consecutive genes, bursty genes, and self-repressing genes. Our results demonstrate that for these models, the conventional linear algebraic method may not apply because the corresponding generator matrices are infinite-dimensional operators. Instead, for a rigorous and correct treatment we must rely on the modern theory of functional analysis — the eigenvalues and eigenvectors of the generator matrix for the Markovian model must be defined within a unified Hilbert space framework. The Hilbert space is chosen to be the typical l^2 space and we require the the eigenvectors to be in that space. This is different

from the conventional linear algebraic theory of finite-dimensional matrices, where we do not make any restrictions on the eigenvectors. In particular, we show that the exact eigenvalues obtained are consistent with the approximate eigenvalues obtained by truncating the generator matrix to a finite dimension — the agreement between the two verifies the correctness of our results. Note that the Hilbert space framework is essential for ensuring consistency between the infinite-dimensional spectral results and the finite-dimensional truncation results. In particular, we clarify that for autoregulatory feedback loops, the exact eigenvalues given in [4] are correct and can be complex numbers. This is contradictory to the real eigenvalues reported in [3, 15–17].

The eigenvalues and eigenvectors are then used to construct an exact spectral representation of the time-dependent distribution of gene product numbers, where distinct spectral modes capture the system's dynamical behavior across different time scales. The spectral gap between the zero eigenvalue and the first nonzero eigenvalue, which reflects the relaxation rate of the system towards its steady state, is then compared with the prediction of the deterministic model. Interestingly, we find that deterministic modeling accurately capture the relaxation rate for constitutive and bursty genes, while it fails to reproduce the relaxation rate for self-repressing genes, especially when self-repression is strong. We anticipate that the Hilbert space framework introduced in the present work can reveal the spectral structure of more complex gene expression models, such as models with multiple gene states [36, 37] and models with complex feedback regulation [30, 38].

Appendices

A. Inapplicability of the linear algebraic method for self-repressing genes

Consider the stochastic dynamics of a self-repressing gene, whose reaction scheme is given by Eq. (33). The stochastic gene expression dynamics is described by the Markov jump process shown in Fig. 1(f) with generator matrix

$$Q = \begin{pmatrix} Q_0 & Q_{01} & & \\ Q_{10} & Q_1 & Q_{12} & \\ & Q_{21} & Q_2 & \ddots \\ & & \ddots & \ddots \end{pmatrix},$$

where

$$Q_m = \begin{pmatrix} -(a+md) & a \\ m\nu & -[s+m(\nu+d)] \end{pmatrix}, \quad Q_{m+1,m} = \begin{pmatrix} (m+1)d & 0 \\ 0 & (m+1)d \end{pmatrix}, \quad Q_{m,m+1} = \begin{pmatrix} 0 & 0 \\ 0 & s \end{pmatrix}.$$

The spectral structure for self-repressing genes has been studied in [3, 4, 15–17]; however, the results obtained in these papers are contradictory. In our previous paper [4], we provided a complete spectrum characterization for the stochastic model of an autoregulatory feedback loop based on a complex analysis method; however, the derivation is informal because we have not proved that the eigenvectors obtained belong to the Hilbert space $l^2(1/\pi)$. Specifically, it is clear that $\lambda_0 = 0$ is an eigenvalue of the generator matrix Q. Furthermore, all nonzero eigenvalues are the roots of the following nontrivial continued fraction equation:

$$q = \frac{R_0P_1}{Q_1-}\frac{R_1P_2}{Q_2-}\frac{R_2P_3}{Q_3-}\cdots, \tag{42}$$

where

$$R_n = -(n+1)(n+\gamma), \quad P_n = \theta + (n-1)\epsilon, \quad Q_n = n[-(n-1+\gamma) - \delta + \epsilon] + q,$$

with

$$q = -\frac{(\lambda+d)(a+\lambda)}{d(\nu+d)} - \frac{(a+d+\lambda)\nu s}{d(\nu+d)^2}, \quad \theta = -\frac{(a+d+\lambda)\nu s}{d(\nu+d)^2},$$
$$\gamma = \frac{a+\lambda}{\nu+d} + \frac{s\nu}{(\nu+d)^2}, \quad \delta = 2 + \frac{\lambda}{d}, \quad \epsilon = -\frac{\nu s}{(\nu+d)^2}. \tag{43}$$

However, Ramos and coauthors have obtained different eigenvalues for the same model. In [3, 15], the authors claimed that a self-repressing gene circuit has two families of eigenvalues, i.e.

$$\lambda_{1,n} = -nd, \quad \lambda_{2,n} = -n(d+\nu) - a - \frac{s\nu}{d+\nu}, \quad n \geq 0. \tag{44}$$

A proof of this result was given in [17] using a linear algebraic method. In particular, the eigenvalues given in [3, 15] are all real. We now show that the linear algebraic method proposed in [17] is questionable. In [17], the author investigated the eigenvalues of the matrix

$$U_s = \mathrm{diag}(L, L, \cdots) Q^T \mathrm{diag}(L, L, \cdots)^{-1}, \tag{45}$$

with

$$L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

A matrix V was then constructed such that

$$VU_s = \Lambda V, \tag{46}$$

for some diagonal matrix Λ. It was then claimed that the diagonal elements of Λ are all eigenvalues of U_s and hence are also all eigenvalues of the generator matrix Q (note that U_s and Q^T are the same up to a similar transformation and they have the same eigenvalues). Note that U_s and Λ can be represented in block form as

$$U_s = \begin{pmatrix} LQ_0^T L^{-1} & LQ_{10}^T L^{-1} & & \\ LQ_{01}^T L^{-1} & LQ_1^T L^{-1} & LQ_{21}^T L^{-1} & \\ & LQ_{12}^T L^{-1} & LQ_2^T L^{-1} & \ddots \\ & & \ddots & \ddots \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \Lambda_0 & & & \\ & \Lambda_1 & & \\ & & \Lambda_2 & \\ & & & \ddots \end{pmatrix},$$

where Λ_n, $n \geq 0$ are 2×2 matrices. Moreover, the matrix V can be represented in block form as

$$V = \begin{pmatrix} V_{00} & V_{01} & V_{02} & \cdots \\ V_{10} & V_{11} & V_{12} & \cdots \\ V_{20} & V_{21} & V_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where V_{mn}, $m, n \geq 0$ are also 2×2 matrices. From the condition $VU_s = \Lambda V$, we obtain the equation

$$V_{m,n-1}U_{n-1,n} + V_{mn}U_{nn} + V_{m,n+1}U_{n+1,n} = \Lambda_m V_{mn}, \quad \forall m, n \geq 0, \tag{47}$$

where

$$U_{n-1,n} = LQ_{n,n-1}^T L^{-1}, \quad U_{nn} = LQ_n^T L^{-1}, \quad U_{n+1,n} = LQ_{n,n+1}^T L^{-1}.$$

Note that V_{mn}, L, and Λ_n are 2×2 matrices. Set

$$V_{mn}L = \begin{pmatrix} v_{(0,m),(0,n)} & v_{(0,m),(1,n)} \\ v_{(1,m),(0,n)} & v_{(1,m),(1,n)} \end{pmatrix}, \quad \Lambda_n = \begin{pmatrix} \lambda_{0,n} & \\ & \lambda_{1,n} \end{pmatrix}.$$

Then Eq. (47) can be written in component form as

$$\begin{aligned} &md v_{(i,n),(0,m-1)} - (a+md)v_{(i,n),(0,m)} + a v_{(i,n),(1,m)} = \lambda_{i,n} v_{(i,n),(0,m)}, \\ &md v_{(i,n),(1,m-1)} + m\nu v_{(i,n),(0,m)} - [s + m(\nu+d)]v_{(i,n),(1,m)} + s v_{(i,n),(1,m+1)} = \lambda_{i,n} v_{(i,n),(1,m)}, \end{aligned} \tag{48}$$

where $v_{(i,n),(0,-1)} = v_{(i,n),(1,-1)} = 0$ by default. For convenience, we set $v_{(i,n),(0,0)} = 1$. It is clear that

$$v_{(i,n),(1,0)} = (1 + \lambda_{i,n}/a) v_{(i,n),(0,0)}.$$

Solving Eq. (48) recursively yields that for any $n, m \geq 0$ and $i = 0, 1$,

$$\begin{aligned} v_{(i,n),(1,m+1)} &= \frac{[s + m(\nu+d) + \lambda_{i,n}]v_{(i,n),(1,m)} - n\nu v_{(i,n),(0,m)} - md v_{(i,n),(1,m-1)}}{s}, \\ v_{(i,n),(0,m+1)} &= \frac{a v_{(i,n),(1,m+1)} + (m+1)d v_{(i,n),(0,m)}}{a + (m+1)d + \lambda_{i,n}}. \end{aligned}$$

This implies that the matrix $V_{mn}L$ can be determined recursively. Then V_{mn} can be constructed as

$$V_{mn} = \begin{pmatrix} v_{(0,m),(0,n)} & v_{(0,m),(1,n)} \\ v_{(1,m),(0,n)} & v_{(1,m),(1,n)} \end{pmatrix} L^{-1}.$$

The above computations show that, regardless of how the diagonal matrix Λ is chosen, one can always find a matrix V such that $VU_s = \Lambda V$. Hence even if $VU_s = \Lambda V$ holds, we cannot conclude that the diagonal elements of Λ are all eigenvalues of the generator matrix Q.

B. Derivation of eigenvalues for self-repressing genes using the Hilbert space framework

We next apply the Hilbert space framework introduced above to find the eigenvalues for self-repressing genes: if there exists a row vector $\psi = (\psi_{i,m}) \in l^2(1/\pi)$ such that $\psi Q = \lambda\psi$, then λ is as an eigenvalue of Q and ψ is the associated eigenvector. It is clear that $\lambda_0 = 0$ is an eigenvalue of the generator matrix Q. We now prove that the solutions of Eq. (42) give all nonzero eigenvalues of Q. Note that the characteristic equation $\psi Q = \lambda\psi$ can be written in components as

$$\begin{aligned} &(m+1)d\psi_{0,m+1} + m\nu\psi_{1,m} - (a+md)\psi_{0,m} = \lambda\psi_{0,m}, \\ &s\psi_{1,m-1} + (m+1)d\psi_{1,m+1} + a\psi_{0,m} - [s + m(\nu+d)]\psi_{1,m} = \lambda\psi_{1,m}, \end{aligned} \tag{49}$$

where $\psi_{1,-1} = 0$ by default. To proceed, we define two generating functions

$$f_i(z) = \sum_{m=0}^{\infty} \psi_{i,m} z^m, \quad i = 0, 1, \tag{50}$$

and let $f(z) = f_0(z) + f_1(z)$ be the total generating function. In [4], by using a complex analysis method, we have shown that the total generating function f can be computed in closed form as

$$f(z) = (z-1)Hc\left(q, \theta, \gamma, \delta, \epsilon; \frac{\nu+d}{\nu}(z - z_0)\right), \tag{51}$$

and the generating functions f_0 and f_1 can be computed exactly as

$$f_1 = \frac{1}{s(z-1)}[\lambda f - d(1-z)\partial_z f], \quad f_0 = f - f_1, \tag{52}$$

where $Hc(z)$ is the local confluent Heun function [39, 40], and $q, \theta, \gamma, \delta, \epsilon$ are the parameters given in Eq. (43). Hence, from Eq. (50), $\psi_{0,m}$ and $\psi_{1,m}$ can be recovered from the generating functions as

$$\psi_{i,m} = \frac{1}{m!} f_i^{(m)}(z)\Big|_{z=0}, \quad i = 0, 1. \tag{53}$$

To simplify, in what follows, we abbreviate $Hc(q, \theta, \gamma, \delta, \epsilon; z)$ as $Hc(z)$. Before giving the rigorous proof, we recall that the local confluent Heun function has the following important properties [39, 40]:

(i) If λ is a solution of Eq. (42), then $Hc(z)$ is holomorphic over whole complex plane. In this case, the local confluent Heun function is called a confluent Heun function. Let $Hc(z) = \sum_{l=0}^{\infty} c_l z^l$ be the power series expansion of the confluent Heun function. Then the convergence radius of $Hc(z) = \sum_{l=0}^{\infty} c_l z^l$ is ∞. Moreover, there exists $\kappa > 0$ such that $|c_l| \leq \kappa^l / l!$ for any $l \geq 0$ (see Appendix C for the proof of this fact).

(ii) If λ is not a solution of Eq. (42), then $Hc(z)$ has only one singularity at $z = 1$ and the convergence radius of the power series $Hc(z) = \sum_{l=0}^{\infty} c_l x^l$ is 1.

We first focus on the case where λ is a solution to Eq. (42). In this case, we have

$$\begin{aligned}
\left|\partial_z^m Hc\left(-\frac{\nu+d}{\nu} z_0\right)\right| &= \left|\sum_{l=m}^{\infty} c_l \left(\frac{\nu+d}{\nu}\right)^l \frac{l!}{(l-m)!}(-z_0)^{l-m}\right| \\
&\leq \left(\frac{\nu+d}{\nu}\kappa\right)^m \sum_{l=m}^{\infty} \frac{1}{(l-m)!}\left(\frac{\nu+d}{\nu}\kappa z_0\right)^{l-m} \\
&= \left(\frac{\nu+d}{\nu}\kappa\right)^m \exp\left(\frac{\nu+d}{\nu}\kappa z_0\right).
\end{aligned}$$

Hence, it is easy to see that there exist two positive constants M and η such that

$$\begin{aligned}
\frac{1}{m!}\left|f^{(m)}(0)\right| &= \frac{1}{m!}\left|m\partial_z^{m-1} Hc\left(-\frac{\nu+d}{\nu} z_0\right) - \partial_z^m Hc\left(-\frac{\nu+d}{\nu} z_0\right)\right| \\
&\leq M\frac{\eta^m}{(m-1)!}.
\end{aligned}$$

Similarly, it follows from Eqs. (52) and (53) that

$$\begin{aligned}
|\psi_{0,m}| &= \frac{1}{m!}\left|f_0^{(m)}(0)\right| \leq M\frac{\eta^m}{(m-1)!}, \\
|\psi_{1,m}| &= \frac{1}{m!}\left|f_1^{(m)}(0)\right| \leq M\frac{\eta^m}{(m-1)!}.
\end{aligned} \tag{54}$$

On the other hand, recall the following property of the confluent hypergeometric function [29]:

$${}_1F_1(\alpha+m; \beta+m; -wz_0) = e^{-wz_0}(1+o(1)), \quad m \to \infty. \tag{55}$$

Hence, it follows from Eqs. (36) and (55) that

$$\begin{aligned}
\pi_{1,m} &= \frac{d}{\nu+d} e^{\alpha-\beta}\frac{\beta^{\beta-1}}{\alpha^{\alpha-1}} e^{-wz_0}\frac{m^{\alpha-\beta}w^m}{m!}(1+o(1)), \\
\pi_{0,m} &= \frac{\nu}{\nu+d} e^{\alpha-\beta}\frac{\beta^{\beta-1}}{\alpha^{\alpha-1}} e^{-wz_0}\frac{m^{\alpha-\beta}w^m}{m!}(1+o(1)).
\end{aligned} \tag{56}$$

This indicates that $\pi_{i,m} \geq \sigma m^{\alpha-\beta} w^m / m!$ for some constant $\sigma > 0$. It then follows from Eq. (54) that

$$\sum_{i=0}^{1}\sum_{m=0}^{\infty} |\psi_{i,m}|^2/\pi_{i,m} \leq \frac{2M^2}{\sigma}\sum_{m=0}^{\infty}\frac{m^{1-\alpha+\beta}(\eta^2/w)^m}{(m-1)!} < \infty.$$

This shows that $\psi = (\psi_{i,m}) \in l^2(1/\pi)$ and thus the roots of the continued fraction equation given in Eq. (42) are indeed the eigenvalues of the generator matrix Q.

We next focus on the case where λ is not a solution of Eq. (42). From Eq. (51), the convergence radius of the power series

$$f(z) = \sum_{m=0}^{\infty} (\psi_{0,m} + \psi_{1,m}) z^m$$

is the same as that of $Hc((\nu + d)(z - z_0)/\nu) = \sum_{l=0}^{\infty} b_l z^l$. Since $Hc(z)$ has only one singularity at $z = 1$, it is clear that $Hc((\nu + d)(z - z_0)/\nu)$ has only one singularity at $z = z_0 + \nu/(\nu + d) = 1$. Thus the convergence radius of $f(z)$ must be 1. Then we immediately obtain

$$\overline{\lim_{m\to\infty}} \sqrt[m]{|\psi_{i,m}|} \geq 1, \quad i = 0, 1.$$

Hence we have $\overline{\lim}_{m\to\infty} |\psi_{i,m}|\sigma^m \geq 1$ for any $\sigma > 1$. It follows from Eq. (56) that $\pi_{i,m} \leq M' m^{\alpha-\beta} w^m/m!$ for some constant $M' > 0$. This indicates that

$$\overline{\lim_{m\to\infty}} |\psi_{i,m}|^2/\pi_{i,m} \geq \frac{1}{M'} \overline{\lim_{m\to\infty}} \frac{m!}{m^{\alpha-\beta}\sigma^{2m} w^m} = \infty.$$

This shows that $\psi = (\psi_{i,m}) \notin l^2(1/\pi)$ and thus any complex number that is not the solution of Eq. (42) is also not an eigenvalue of the generator matrix Q.

C. Proof of a property for the confluent Heun function

Recall that the local confluent Heun function is defined as the solution of the following Heun differential equation that is holomorphic at $z = 0$ and is equal to 1 there [39]:

$$h''(z) + \left(\frac{\gamma}{z} + \frac{\delta}{z-1} + \epsilon\right) h'(z) + \frac{\theta z - q}{z(z-1)} h(z) = 0.$$

Let $Hc(z) = \sum_{l=0}^{\infty} c_l z^l$ be the power series expansion of the local confluent Heun function. Direct computations show that

$$c_0 = 1, \quad c_1 = -q/\gamma, \quad R_n c_{n+1} - Q_n c_n + P_n c_{n-1} = 0, \quad n \geq 1, \tag{57}$$

where

$$R_n = -(n+1)(n+\gamma), \quad P_n = \theta + (n-1)\epsilon, \quad Q_n = n[-(n-1+\gamma) - \delta + \epsilon] + q. \tag{58}$$

It follows from [39] that $\lim_{l\to\infty} c_{l+1}/c_l = 0$ if and only if $\gamma, \delta, \epsilon, \theta, q$ satisfy the following continued fraction equation:

$$q = \frac{R_0 P_1}{Q_1 -} \frac{R_1 P_2}{Q_2 -} \frac{R_2 P_3}{Q_3 -} \cdots . \tag{59}$$

In fact, it can be proved that there exists $\kappa > 0$ such that $|c_l| \leq \kappa^l/l!$ for any $l \geq 0$.

We next give a rigorous proof of this fact. Set $u_l = c_{l+1}/c_l$. It is clear that $\lim_{l\to\infty} u_l = 0$. Moreover, it follows from Eq. (57) that

$$u_0 = -q/\gamma, \quad R_n u_n u_{n-1} - Q_n u_{n-1} + P_n = 0, \quad n \geq 1.$$

This implies that

$$u_{n-1} = \frac{P_n}{Q_n - R_n u_n}, \quad n \geq 1. \tag{60}$$

Iterating Eq. (60) m times yields

$$u_n = \frac{P_{n+1}}{Q_{n+1}-}\frac{R_{n+1}P_{n+2}}{Q_{n+2}-}\cdots\frac{R_{n+m-1}P_{n+m}}{Q_{n+m}-}\frac{R_{n+m}P_{n+m+1}}{Q_{n+m+1}-R_{n+m+1}u_{n+m+1}}. \tag{61}$$

To estimate u_n, we consider its truncated version of the above equation:

$$u_{nm} = \frac{P_{n+1}}{Q_{n+1}-}\frac{R_{n+1}P_{n+2}}{Q_{n+2}-}\cdots\frac{R_{n+m-1}P_{n+m}}{Q_{n+m}-}\frac{R_{n+m}P_{n+m+1}}{Q_{n+m+1}}. \tag{62}$$

Combining Eqs. (59), (61) and (62), it is easy to see that

$$\lim_{m\to\infty} u_{nm} = u_n.$$

Let N be a sufficiently large positive integer such that

$$(|\gamma| + |\delta| + |\epsilon| + |\theta| + |q| + 10)^3 \le N. \tag{63}$$

We claim that for any $n \ge N$ and $m \ge 0$,

$$u_{nm} = -\frac{\epsilon}{n+1} + \frac{v_{nm}}{(n+1)^{3/2}}, \tag{64}$$

where $|v_{nm}| \le 1$. We next prove the claim by induction.

First, the above claim holds for $m = 0$. Combining Eqs. (58), (62), and (64), we have

$$\begin{aligned}|v_{n0}| &= \left|u_{n0}(n+1)^{3/2} + \epsilon(n+1)^{1/2}\right| = \left|\frac{n\epsilon+\theta}{(n+1)[-(n+\gamma)-\delta+\epsilon]+q}(n+1)^{3/2} + \epsilon(n+1)^{1/2}\right| \\ &= (n+1)^{1/2}\frac{|[\theta+\epsilon(\epsilon-\gamma-\delta)](n+1)+\epsilon q|}{|(n+1)(-n\gamma-\delta+\epsilon)+q|} \le 1,\end{aligned}$$

where the last step follows from Eq. (63). Second, assume that the claim holds for $m = k-1$ $(k \ge 1)$. We then prove that it also holds for $m = k$. It follows from Eq. (62) that

$$u_{nk} = \frac{P_{n+1}}{Q_{n+1} - R_{n+1}u_{n+1,k-1}}.$$

Similarly, we immediately obtain

$$\begin{aligned}|v_{nk}| &= \left|\frac{n\epsilon+\theta}{(n+1)[-(n+\gamma)-\delta+\epsilon]+q+(n+2)(n+\gamma+1)u_{n+1,k-1}}(n+1)^{3/2} + \epsilon(n+1)^{1/2}\right| \\ &= (n+1)^{1/2}\frac{|[\theta-\epsilon(\gamma+\delta)](n+1)+\epsilon v_{n+1,k-1}(n+\gamma+1)(n+2)^{-1/2}+\epsilon q-\epsilon^2\gamma|}{|(n+1)(-n-\gamma-\delta)+v_{n+1,k-1}(n+\gamma+1)(n+2)^{-1/2}+q-\epsilon\gamma|} \le 1,\end{aligned}$$

where the last step also follows from Eq. (63). Thus far, we have finished the proof of the above claim by induction.

Since $|v_{nm}| \le 1$, there exists $\kappa > 0$ such that $|u_{nm}| \le \kappa/(n+1)$. This gives rise to

$$|u_n| = \lim_{m\to\infty}|u_{nm}| \le \frac{\kappa}{n+1}, \quad n \ge N.$$

Since $u_l = c_{l+1}/c_l$, it is clear that

$$|c_l/c_N| = |u_N|\cdots|u_{l-1}| \le \frac{(\kappa)^{l-N}N!}{l!}, \quad l \ge N.$$

This implies that there exists a sufficiently large positive number, still denoted by κ, such that $|c_l| \le \kappa^l/l!$ for any $l \ge 0$.

Acknowledgements

We are grateful to Dr. Nikola Popovic (University of Edinburgh) for valuable comments and suggestions. R. G. acknowledges support from the Leverhulme Trust (RPG-2020-327). C. J. acknowledges support from National Natural Science Foundation of China with grant No. 12271020.

Data availability

The manuscript has no associated data.

Ethical statement

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

All authors have given approval for publication.

References

[1] Jia, C., Qian, H., Chen, M. & Zhang, M. Q. Relaxation rates of gene expression kinetics reveal the feedback signs of autoregulatory gene networks. *J. Chem. Phys.* **148** (2018).

[2] Mugler, A., Walczak, A. M. & Wiggins, C. H. Spectral solutions to stochastic models of gene expression with bursts and regulation. *Phys. Rev. E* **80**, 041921 (2009).

[3] Ramos, A. F., Innocentini, G. C. & Hornos, J. E. M. Exact time-dependent solutions for a self-regulating gene. *Phys. Rev. E* **83**, 062902 (2011).

[4] Wu, B., Holehouse, J., Grima, R. & Jia, C. Solving the time-dependent protein distributions for autoregulated bursty gene expression using spectral decomposition. *J. Chem. Phys.* **160** (2024).

[5] Paulsson, J. & Ehrenberg, M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Phys. Rev. Lett.* **84**, 5447 (2000).

[6] Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).

[7] Smith, S. & Shahrezaei, V. General transient solution of the one-step master equation in one dimension. *Phys. Rev. E* **91**, 062119 (2015).

[8] Jia, C. & Grima, R. Dynamical phase diagram of an auto-regulating gene in fast switching conditions. *J. Chem. Phys.* **152** (2020).

[9] Hornos, J. E. *et al.* Self-regulating gene: an exact solution. *Phys. Rev. E* **72**, 051907 (2005).

[10] Grima, R., Schmidt, D. R. & Newman, T. J. Steady-state fluctuations of a genetic feedback loop: An exact solution. *J. Chem. Phys.* **137** (2012).

[11] Kumar, N., Platini, T. & Kulkarni, R. V. Exact distributions for stochastic gene expression models with bursting and feedback. *Phys. Rev. Lett.* **113**, 268105 (2014).

[12] Liu, P., Yuan, Z., Wang, H. & Zhou, T. Decomposition and tunability of expression noise in the presence of coupled feedbacks. *Chaos* **26**, 043108 (2016).

[13] Jia, C. & Grima, R. Small protein number effects in stochastic models of autoregulated bursty gene expression. *J. Chem. Phys.* **152** (2020).

[14] Liu, S., Wu, B. & Jia, C. Exact switching time distributions for autoregulated gene expression models with mRNA and protein descriptions. *SIAM J. Appl. Math.* **85**, 1998–2025 (2025).

[15] Ramos, A. F. & Hornos, J. E. Symmetry and stochastic gene regulation. *Phys. Rev. Lett.* **99**, 108103 (2007).

[16] Ramos, A. F. & Reinitz, J. Physical implications of so (2, 1) symmetry in exact solutions for a self-repressing gene. *J. Chem. Phys.* **151** (2019).

[17] Ramos, A. F. Comment on "Solving the time-dependent protein distributions for autoregulated bursty gene expression using spectral decomposition" [J. Chem. Phys. 160, 074105 (2024)]. *bioRxiv 10.1101/2025.02.05.635946* (2025).

[18] Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).

[19] Chen, M.-F. *Eigenvalues, Inequalities, and Ergodic Theory* (Springer, 2006).

[20] Rudin, W. *Functional analysis* (McGraw-Hill, New York, 1991), 2nd edn.

[21] Teschl, G. *Mathematical methods in quantum mechanics*, vol. 157 (American Mathematical Soc., 2014).

[22] Rudin, W. *Principles of mathematical analysis* (McGraw Hill, 1953).

[23] Munsky, B. & Khammash, M. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* **124**, 044104 (2006).

[24] Bokes, P., King, J. R., Wood, A. T. & Loose, M. Multiscale stochastic modelling of gene expression. *J. Math. Biol.* **65**, 493–520 (2012).

[25] Jia, C. Simplification of Markov chains with infinite state space and the mathematical theory of random gene expression bursts. *Phys. Rev. E* **96**, 032402 (2017).

[26] Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. USA* **105**, 17256–17261 (2008).

[27] Sheng, Y., Lin, G., Jiao, F. & Jia, C. Geometric theory of distribution shapes for autoregulatory gene circuits. *SIAM J. Appl. Math.* **85**, 636–661 (2025).

[28] Holehouse, J., Cao, Z. & Grima, R. Stochastic modeling of autoregulatory genetic feedback loops: a review and comparative study. *Biophys. J.* **118**, 1517–1525 (2020).

[29] Olver, F. W. J. *et al.* NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.17 of 2017-12-22 (2017).

[30] Thomas, P., Popovic, N. & Grima, R. Phenotypic switching in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **111**, 6994–6999 (2014).

[31] Jia, C., Qian, H. & Zhang, M. Q. Exact power spectrum in a minimal hybrid model of stochastic gene expression oscillations. *SIAM J. Appl. Math.* **84**, 1204–1226 (2024).

[32] Jiang, D.-Q., Qian, M. & Qian, M.-P. *Mathematical theory of nonequilibrium steady states: on the frontier of probability and dynamical systems* (Springer, Berlin Heidelberg, 2004).

[33] Zhang, X.-J., Qian, H. & Qian, M. Stochastic theory of nonequilibrium steady states and its applications. Part I. *Phys. Rep.* **510**, 1–86 (2012).

[34] Kolmogoroff, A. Zur theorie der Markoffschen ketten. *Math. Ann.* **112**, 155–160 (1936).

[35] Jiang, Y., Wu, B. & Jia, C. Large deviations and fluctuation theorems for cycle currents defined in the loop-erased and spanning tree manners: a comparative study. *Phys. Rev. Res.* **5**, 013207 (2023).

[36] Zhou, T. & Zhang, J. Analytical results for a multistate gene model. *SIAM J. Appl. Math.* **72**, 789–818 (2012).

[37] Jia, C. & Li, Y. Analytical time-dependent distributions for gene expression models with complex promoter switching mechanisms. *SIAM J. Appl. Math.* **83**, 1572–1602 (2023).

[38] Fang, Z., Gupta, A., Kumar, S. & Khammash, M. Advanced methods for gene network identification and noise decomposition from single-cell data. *Nat. Commun.* **15**, 4911 (2024).

[39] Ronveaux, A. *Heun's differential equations* (The Clarendon Press Oxford University Press, New York, 1995).

[40] Motygin, O. V. On evaluation of the confluent Heun functions. In *2018 Days on Diffraction (DD)*, 223–229 (IEEE, 2018).