



## Fluctuating-rate model with multiple gene states

Jingwei Li<sup>1</sup> · Hao Ge<sup>2</sup>  · Yunxin Zhang<sup>3</sup>

Received: 20 February 2019 / Revised: 28 August 2020 / Published online: 30 September 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

### Abstract

Multiple phenotypic states of single cells often co-exist in the presence of positive feedbacks. Stochastic gene-state switchings and low copy numbers of proteins in single cells cause considerable fluctuations. The chemical master equation (CME) is a powerful tool that describes the dynamics of single cells, but it may be overly complicated. Among many simplified models, a fluctuating-rate (FR) model has been proposed recently to approximate the full CME model in the realistic intermediate region of gene-state switchings. However, only the scenario with two gene states has been carefully analysed. In this paper, we generalise the FR model to the case with multiple gene states, in which the mathematical derivation becomes more complicated. The leading order of fluctuations around each phenotypic state, as well as the transition rates between phenotypic states, in the intermediate gene-state switching region is characterized by the rate function of the stationary distribution of the FR model in the Freidlin–Wentzell-type large deviation principle (LDP). Under certain reasonable assumptions, we show that the derivative of the rate function is equal to the unique nontrivial solution of a dominant generalised eigenvalue problem, leading to a new numerical algorithm for obtaining the LDP rate function directly. Furthermore, we prove the Lyapunov property of the rate function for the corresponding deterministic mean-field dynamics. Finally, through a tristable example, we show that the local fluctuations (the asymptotic variance of the stationary distribution at each phenotypic state) in the intermediate and rapid regions of gene-state switchings are different. Finally, a tri-stable example is constructed to illustrate the validity of our theory.

**Keywords** Fluctuating-rate model · Nonequilibrium landscape function · Transition rates · Dominant generalised eigenvalue

**Mathematics Subject Classification** 60F10 · 47A75 · 15B48

### Abbreviations

CME                      Chemical master equation  
FR                         Fluctuating-rate

---

Extended author information available on the last page of the article

DMFD	Deterministic mean-field dynamics
LDP	Large deviation principle
NLF	Nonequilibrium landscape function
GE	Generalised eigenvalue
DGE	Dominant generalised eigenvalue

## 1 Introduction

Genes switch among different states due to the regulation of transcription factors and synthesise proteins at a state-dependent rate. This paper considers self-regulatory genes with positive feedback regulations, in which the transcription factors are synthesised by the regulated gene itself and reversely help the genes switch to a state with a relatively large synthesis rate. This may lead to a copy-number distribution with multiple modals, which, from a biological point of view, correspond to multiple phenotypic states of a living cell continuously exchanging materials and energy with its surroundings (Choi et al. 2008; Gupta et al. 2011; Ozbudak et al. 2004). Transitions among phenotypic states induced by intrinsic stochasticity are advantageous for cells to survive in fluctuating environments (Kussell and Leibler 2005; Acar et al. 2008).

Early mathematical works use the reaction rate equation to model the expression of a gene with two (Babloyantz and Sanglier 1972) or multiple (Santillán 2008) states, which neglects the randomness. The chemical-master-equation (CME) model, on the contrary, describes a random dynamics inside a single cell (Delbrück 1940; Gillespie 1977), and has been applied to study lots of gene-regulatory networks (Samad et al. 2005; Berg 1978; Thattai and van Oudenaarden 2001; Paulsson 2005; Jia et al. 2018; Newby and Chapman 2014; Newby 2015). Although the CME captures both types of randomness, i.e. stochastic gene-state switchings and low copy numbers of chemical species inside a single cell (Li and Xie 2011; Taniguchi et al. 2010; Eldar and Elowitz 2010), its mathematical computation is usually complex, and the exact solution is only available in simple cases (Hornos et al. 2005; Ramos et al. 2011).

Therefore, many algorithms have been proposed to numerically solve the probability distributions of CME accurately for problems of interests, without Monte Carlo simulations which are usually computationally expensive. For example, the Finite State Projection (FSP) algorithm (Munsky and Khammashb 2006) and its improved versions (Peleš et al. 2006; MacNamara et al. 2008; Kazeev et al. 2014; Hegland et al. 2008) utilize efficient projections in the vector-based state space of CME. Recently, the FSP is further developed to efficiently estimate the stationary distribution and the parameter sensitivities of the CME (Gupta et al. 2017; Dürrenberger et al. 2019). Other examples include the on-the-fly variant of the uniformisation technique, which improves the original algorithm at the cost of a small approximation error (Mateescu et al. 2010), the method of conditional moments (MCM), which employs a discrete stochastic description for low-copy number species and a moment-based description for medium/high-copy number species (Hasenauer et al. 2014), and so on.

Alternatively, various simplified mechanistic models rather than just numerical algorithms have also been proposed to approximate the CME and investigate the mechanism of single-cell dynamics. Their mathematical foundation is the limit behavior

of general CME under different scales of reaction rate, species abundance, and time (Crudu et al. 2012; Kang and Kurtz 2013). Specifically, for gene-regulatory networks, simplified models are applicable under different parameter regions based on the relationship between the gene-state switchings and the birth–death kinetics of proteins. Most previous works assumed that gene-state switchings are extremely slow (Karmakar and Bose 2004; Artyomov et al. 2007; Qian et al. 2009; To and Maheshri 2010; Feng et al. 2011; Ochab-Marcinek and Tabaka 2010) or extremely rapid (Ge and Qian 2009; Wang et al. 2010; Zhou et al. 2012; Qian 2014; Lu et al. 2014; Hufton et al. 2019a, b) to avoid the mathematical difficulty in subsequent analyses. However, at least in bacteria, the single-cell gene-state switchings are neither extremely slow nor extremely rapid (Li and Xie 2011; Taniguchi et al. 2010; Choi et al. 2008; Gupta et al. 2011; Ozbudak et al. 2004). The relative stability of phenotypic states and the transition rates among them in such an intermediate region are far from quantitatively understood.

Recently, Ge et al. (2015) have proposed a so-called fluctuating-rate (FR) model for the more realistic intermediate region, which neglects the randomness caused by the low copy number of proteins but retains the randomness caused by gene-state switchings. The FR model is much more accessible for mathematical analyses than the full CME model because its mathematical prototype, i.e. the piecewise deterministic Markov processes (PDMP), has been well studied (Davis 1984, 1993). PDMP is a Markov process, whose randomness is only given by the jumps among different deterministic dynamics. PDMP has appeared in several previous studies of the gene-regulatory networks (Kepler and Elston 2001; Newby 2012; Hufton et al. 2016, 2018). These works are all restricted to the case of specific number (two or three) of gene states. In addition, Lin and Doering (2016) studied two-state cases theoretically, but only provided the numerical results for multiple-state ones.

Actually, it has already been proved that, after taking the limit of large active synthesis rates of proteins, the full CME model approaches the FR model, and if the switching rates between discrete gene states further tend to infinity, the FR model finally converges to the deterministic mean-field dynamics (DMFD) (Crudu et al. 2012; Faggionato et al. 2010). Studying the relative stability of phenotypic states and the transition rates among them should go one step further beyond this law of large number. That is the large deviation principle (LDP), especially the Freidlin–Wentzell-type one (Freidlin and Wentzell 2014). Although the general dynamic LDP theory of PDMP models has already been derived (Kifer 2009; Faggionato et al. 2009, 2010), explicit expressions of the LDP rate function of steady-state distribution as well as the proof for its Lyapunov property for the DMFD are still lacking for gene-regulatory FR models with multiple gene states. The case with only two gene states has been solved via the WKB expansion in Ge et al. (2015). The connection between LDP and WKB methods is justified in Bressloff and Faugeras (2017). People may believe that the scenario with multiple gene states should share the same mathematical results as the case with only two gene states, and the multiple-state FR model has indeed already been applied to investigate the stochastic kinetics of lac operon (Ge et al. 2018), but the mathematical methods in Ge et al. (2015) cannot be generalized straightforwardly.

Besides the rigorous LDP for PDMPs, there are a variety of application techniques in solving the first-passage time problems of stochastic hybrid systems, which include

WKB approximations and matched asymptotics (Bressloff and Newby 2014b; Keener and Newby 2011; Newby 2012; Newby and Keener 2011; Newby et al. 2013), and path-integrals (Bressloff and Newby 2014a; Bressloff 2015). But the calculation of the LDP rate function by these techniques is still difficult for the multiple-state case even numerically. That is why most applications of the PDMP so far are still restricted to the case of two or three gene states.

We focus on the LDP rate function of the stationary distribution, which is a quasi-potential of the FR model and describes the leading order of the fluctuations of the protein abundance. We found out that its derivative with respect to the continuous variable can be formulated as the unique nontrivial solution of a dominant generalised eigenvalue problem in the case with arbitrary finite number of gene states, which generalises the results in Ge et al. (2015). Main mathematical tools are the famous Perron–Frobenius theorem (Frobenius 1912) and the convexity of the dominant eigenvalue of an essentially nonnegative matrix on diagonal elements (Deutsch and Neumann 1984). Such a detailed investigation of dominant generalised eigenvalue problems promotes a new numerical algorithm for obtaining the LDP rate function of steady-state distribution in the FR models. We further prove the Lyapunov property of the LDP rate function with respect to the DMFD, based on the above analysis. Whereas the results of Faggionato et al. (2009) are restricted to the case with unique fixed point, our result of the Lyapunov property is general. The prefactor of the FR model, which provides the next order of fluctuations, is also proved to be continuous and positive.

We use a tristable example to numerically show that the rate function of the FR model correctly predicts the transition rates between phenotypic states, i.e. different attractors of DMFD, in the intermediate region based on the Freidlin–Wentzell-type LDP. Moreover, the local fluctuations, i.e. asymptotic variance, of each phenotypic state in the intermediate region of gene-state switchings are highly different from those in the rapid region, even if their DMFDs are the same.

This paper is organised as follows. Heuristic derivation of the FR model with multiple gene states as well as the associated LDP rate function is given in Sect. 2. Mathematically rigorous proof of the LDP rate function as the unique nontrivial solution of a dominant generalised eigenvalue problem and the Lyapunov property with respect to DMFD are given in Sect. 3. In the same section, we propose a new numerical algorithm for the calculation of LDP rate function. In Sect. 4, a tristable example is analysed in detail to further justify the main results. The conclusions and remarks are presented in Sect. 5.

## 2 Approximate the full CME model with multiple gene states

We briefly describe the full CME model of protein syntheses in Sect. 2.1. In Sect. 2.2, we reduce the full CME model to the FR model through the rapid limit of protein synthesis. Then as the gene-state switching rates further approach infinity, the DMFD of the FR model and the LDP for its stationary distribution are given in Sects. 2.3 and 2.4, respectively. We briefly introduce the reduced CME model in Sect. 2.5, which approximates the full CME model through the rapid limit of gene-state switchings.

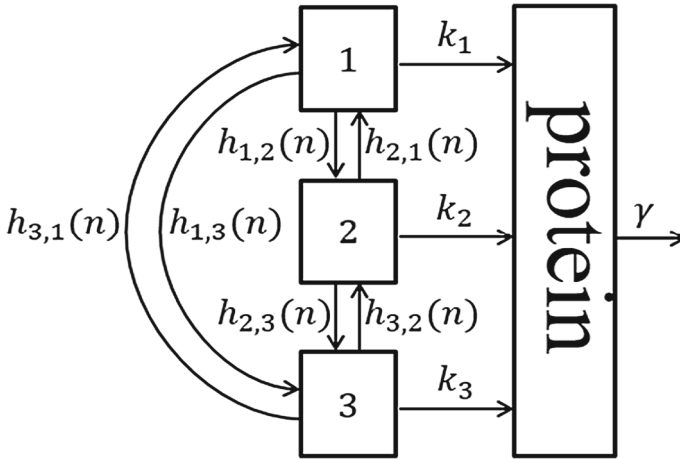


Fig. 1 Diagram of an example of the full CME model with three gene states

**2.1 Full CME model**

We assume the total number of gene states is  $G$ . A gene switches from state  $i$  to state  $j$  by the rate  $h_{i,j}(n)$ , which depends on the protein copy number  $n$ . A gene in state  $i$  synthesises proteins by the rate  $k_i$ , and a protein degrades by rate  $\gamma$ . Without loss of generality, assume that  $k_1 > k_2 > \dots > k_G$ . Figure 1 is the diagram of an example of the full CME model with three gene states. The state of a single cell is characterized by the gene state and copy number of protein molecules. Define  $p_i(n, t)$  as the probability of the cell state in which the gene state is  $i$  and there are  $n$  protein molecules at the moment  $t$ . The CME is (Delbrück 1940; Grima et al. 2012; Gillespie 1976, 1977)

$$\begin{aligned} \partial_t p_i(n, t) = & k_i p_i(n - 1, t) + \gamma(n + 1) p_i(n + 1, t) + \sum_{j \neq i} h_{j,i}(n) p_j(n, t) \\ & - k_i p_i(n, t) - \gamma n p_i(n, t) - \sum_{j \neq i} h_{i,j}(n) p_i(n, t). \end{aligned} \tag{1}$$

**2.2 FR model**

Let  $h_{i,j}(n)$  and  $\gamma$  be fixed, and define  $n_{\max} = \frac{k_1}{\gamma}$ . Denote  $k_i = n_{\max} k_i^0$  with  $k_i^0$  being fixed, in which  $k_1^0 = \gamma$ . We give a heuristic derivation of the FR model which approximates the full CME model as  $n_{\max} \rightarrow \infty$ . The FR model is actually a PDMP. Rigorous definition and proof of the convergence of the full CME model to PDMP as  $n_{\max} \rightarrow \infty$  are given by Theorem 3.1 in Crudu et al. (2012).

Define  $x := n/n_{\max}$ ,  $\tilde{p}_i(x, t) := p_i(n_{\max} \cdot x, t)$  and  $\tilde{h}_{i,j}(x) := h_{i,j}(n_{\max} \cdot x)$ . Then Eq. (1) becomes

$$\begin{aligned} \partial_t \tilde{p}_i(x, t) &= n_{\max} k_i^0 \tilde{p}_i\left(x - \frac{1}{n_{\max}}, t\right) + \gamma n_{\max} \left(x + \frac{1}{n_{\max}}\right) \tilde{p}_i\left(x + \frac{1}{n_{\max}}, t\right) \\ &+ \sum_{j \neq i} \tilde{h}_{j,i}(x) \tilde{p}_j(x, t) - n_{\max} k_i^0 \tilde{p}_i(x, t) - \gamma n_{\max} x \tilde{p}_i(x, t) - \sum_{j \neq i} \tilde{h}_{i,j}(x) \tilde{p}_i(x, t). \end{aligned} \tag{2}$$

Substitute

$$\begin{aligned} \tilde{p}_i\left(x - \frac{1}{n_{\max}}, t\right) &\approx \tilde{p}_i(x, t) - \frac{1}{n_{\max}} \partial_x \tilde{p}_i(x, t), \\ \left(x + \frac{1}{n_{\max}}\right) \tilde{p}_i\left(x + \frac{1}{n_{\max}}, t\right) &\approx x \tilde{p}_i(x, t) + \frac{1}{n_{\max}} \partial_x [x \tilde{p}_i(x, t)], \end{aligned}$$

into Eq. (2), resulting in

$$\begin{aligned} \partial_t \tilde{p}_i(x, t) &= -\partial_x \left\{ \left[ k_i^0 - \gamma x \right] \tilde{p}_i(x, t) \right\} \\ &+ \sum_{j \neq i} \tilde{h}_{j,i}(x) \tilde{p}_j(x, t) - \sum_{j \neq i} \tilde{h}_{i,j}(x) \tilde{p}_i(x, t). \end{aligned} \tag{3}$$

It is exactly the Fokker–Planck equation of a PDMP process, called FR model for single-cell dynamics (Ge et al. 2015). The gene switches stochastically among different states, while at each gene state, the fluctuation of the protein kinetics is eliminated by taking the limit of the large active synthesis rate of proteins, leaving the rescaled protein number  $x$  to follow a deterministic dynamics.

### 2.3 DMFD of the FR model

Define the negative transition rate matrix as

$$\tilde{H}(x) := \begin{pmatrix} \tilde{h}_{1,\cdot}(x) & -\tilde{h}_{2,1}(x) & \cdots & -\tilde{h}_{G,1}(x) \\ -\tilde{h}_{1,2}(x) & \tilde{h}_{2,\cdot}(x) & \cdots & -\tilde{h}_{G,2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{h}_{1,G}(x) & -\tilde{h}_{2,G}(x) & \cdots & \tilde{h}_{G,\cdot}(x) \end{pmatrix},$$

where  $\tilde{h}_{i,\cdot}(x) := \sum_{j \neq i} \tilde{h}_{i,j}(x)$ . In the FR model of Eq. (3), let  $k_i^0$  and  $\gamma$  be fixed, and  $\tilde{H}(x) = \mathcal{H} \hat{H}(x) (\tilde{h}_{i,j}(x) = \mathcal{H} \hat{h}_{i,j}(x))$  with  $\hat{H}(x)$  being fixed and  $\mathcal{H} \rightarrow \infty$ . If  $\hat{H}(x)$  is irreducible, the Markov chain with the transition rate matrix  $-\mathcal{H} \hat{H}(x)$  has the unique stationary probability  $\tilde{\zeta}_i(x)$  for gene state  $i$ . We give a heuristic derivation of the DMFD of the FR model. The rigorous counterpart is the averaging principle of PDMP given by Theorem 2.2 in Faggionato et al. (2010).

As  $\mathcal{H} \rightarrow \infty$ , the characteristic time-scale of the gene switching becomes much faster than the dynamics of the protein abundance  $x$  in the FR model, and  $\tilde{p}_i(x, t) \approx \tilde{\zeta}_i(x)$  approximately. Thus, we define

$$g(x) := \sum_{i=1}^G \tilde{\zeta}_i(x) k_i^0. \tag{4}$$

Let  $\tilde{p}(x, t) := \sum_{i=1}^G \tilde{p}_i(x, t)$ . Then summing Eq. (3) over  $i$ , and substituting Eq. (4), we have

$$\partial_t \tilde{p}(x, t) = -\partial_x \{ [g(x) - \gamma x] \tilde{p}(x, t) \}. \tag{5}$$

Equation (5) is exactly the Louville equation, describing the evolution of probability along the DMFD

$$\frac{dx}{dt} = g(x) - \gamma x. \tag{6}$$

### 2.4 LDP for the stationary distribution of the FR model

Except for the particular cases in Propositions 3.4 and 3.5 of Faggionato et al. (2009), obtaining the exact stationary solution of the FR model with more than two gene states is generally difficult. We alternatively give a heuristic derivation of the LDP rate function for the stationary distribution of the FR model.

Assume Eq. (3) has the unique stationary distribution  $\tilde{p}_i^{ss}(x)$ , which is absolutely continuous w.r.t. the Lebesgue measure. Then

$$0 = -\partial_x \left\{ \left[ k_i^0 - \gamma x \right] \tilde{p}_i^{ss}(x) \right\} + \sum_{j \neq i} \mathcal{H} \widehat{h}_{j,i}(x) \tilde{p}_j^{ss}(x) - \sum_{j \neq i} \mathcal{H} \widehat{h}_{i,j}(x) \tilde{p}_i^{ss}(x). \tag{7}$$

The pathwise LDP has been proved for many mesoscopic models approaching the macroscopic ones (including diffusion processes) (Freidlin and Wentzell 2014; Feng and Kurtz 2015; Olivieri and Vares 2005; Touchette 2009). For PDMP, the LDP of the path density is proved by Theorem 2.3 in Faggionato et al. (2010). Then by the same classic techniques used in Freidlin–Wentzell theory (Freidlin and Wentzell 2014), one can prove that the limit  $\lim_{\mathcal{H} \rightarrow \infty} -\frac{1}{\mathcal{H}} \log \sum_i \tilde{p}_i^{ss}(x)$  exists. We further assume that the limit  $\lim_{\mathcal{H} \rightarrow \infty} -\frac{1}{\mathcal{H}} \log \tilde{p}_i^{ss}(x)$  exists for each  $i$  and is independent of  $i$ , which is denoted by  $\widehat{\Phi}(x)$ .

In other words, the stationary distribution of the FR model satisfies  $\tilde{p}_i^{ss}(x) = C_i(x, \mathcal{H}) \exp(-\mathcal{H} \widehat{\Phi}(x))$  scaled by the gene switching intensity  $\mathcal{H}$ , where  $C_i(x, \mathcal{H})$  is the prefactor satisfying  $\lim_{\mathcal{H} \rightarrow \infty} -\frac{1}{\mathcal{H}} \log C_i(x, \mathcal{H}) = 0$ .

We have assumed that  $\widehat{\Phi}_i(x) = \lim_{\mathcal{H} \rightarrow \infty} -\frac{1}{\mathcal{H}} \log \tilde{p}_i^{ss}(x)$  is independent of  $i$ . This is based on the following heuristic argument. Substitute  $\tilde{p}_i^{ss}(x) = C_i(x, \mathcal{H}) \exp(-\mathcal{H} \widehat{\Phi}_i(x))$  into Eq. (7), and divide both sides by  $\exp(-\mathcal{H} \widehat{\Phi}_i(x))$ .

$$\begin{aligned}
 0 = & -\partial_x \left\{ \left[ k_i^0 - \gamma x \right] C_i(x, \mathcal{H}) \right\} + \mathcal{H} \left[ k_i^0 - \gamma x \right] C_i(x, \mathcal{H}) \partial_x \widehat{\Phi}_i(x) \\
 & + \sum_{j \neq i} \mathcal{H} \widehat{h}_{j,i}(x) C_j(x, \mathcal{H}) \exp \left[ -\mathcal{H} \left( \widehat{\Phi}_j(x) - \widehat{\Phi}_i(x) \right) \right] \\
 & - \sum_{j \neq i} \mathcal{H} \widehat{h}_{i,j}(x) C_i(x, \mathcal{H}). \tag{8}
 \end{aligned}$$

The third term on the right hand side of Eq. (8) will be exponentially large unless its exponential parts vanish. Thus,

$$\widehat{\Phi}_1(x) = \widehat{\Phi}_2(x) = \widehat{\Phi}_3(x) = \dots = \widehat{\Phi}_G(x) := \widehat{\Phi}(x).$$

Define  $C(x, \mathcal{H}) := [C_1(x, \mathcal{H}), C_2(x, \mathcal{H}), \dots, C_G(x, \mathcal{H})]^T$ . Assume that

$$\begin{aligned}
 \lim_{\mathcal{H} \rightarrow +\infty} C(x, \mathcal{H}) / \|C(x, \mathcal{H})\|_2 &= \widetilde{C}(x), \\
 \lim_{\mathcal{H} \rightarrow \infty} \frac{1}{\mathcal{H}} \partial_x [\log C_i(x, \mathcal{H})] &= 0.
 \end{aligned}$$

Then divide both sides of Eq. (8) by  $\mathcal{H} \|C(x, \mathcal{H})\|_2$ , neglect the first term in Eq. (8) as  $\mathcal{H} \rightarrow \infty$ , and substitute  $x_i^* := k_i^0 / \gamma$ , we have

$$0 = \gamma (x_i^* - x) \widetilde{C}_i(x) \partial_x \widehat{\Phi}(x) + \sum_{j \neq i} \widehat{h}_{j,i}(x) \widetilde{C}_j(x) - \sum_{j \neq i} \widehat{h}_{i,j}(x) \widetilde{C}_i(x). \tag{9}$$

Define

$$A(x) := \begin{pmatrix} \gamma (x_1^* - x) & 0 & \dots & 0 \\ 0 & \gamma (x_2^* - x) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \gamma (x_G^* - x) \end{pmatrix}, \tag{10}$$

then Eq. (9) can be rewritten in the matrix form:

$$\left[ \widehat{H}(x) - \partial_x \widehat{\Phi}(x) A(x) \right] \widetilde{C}(x) = 0. \tag{11}$$

$\exists \widetilde{C}(x) \neq 0$  satisfying Eq. (11) iff

$$\det \left[ \widehat{H}(x) - \partial_x \widehat{\Phi}(x) A(x) \right] = 0. \tag{12}$$

What we have used is the WKB method. The connections between LDP and less rigorous WKB methods for Markov chains have been established for a long time (Dykman et al. 1994; Hanggi et al. 1984; Knessl et al. 1985; Vellela and Qian 2008). Similar results for PDMP are recently proved in Bressloff and Faugeras (2017). Such a



LDP rate function also provides the transition-rate formula between different attractors of DMFD, which takes the general Arrhenius/Kramers form (Keener and Newby 2011; Ge et al. 2015). In literature, people sometimes call the LDP rate function  $\widehat{\Phi}(x)$  the nonequilibrium landscape function (NLF) (Feng et al. 2011; Ge et al. 2015), which is an analog of the equilibrium landscape function established in the field of protein folding (Frauenfelder et al. 1991; Onuchic et al. 1997).

### 2.5 Reduced CME model

Let  $k_i$  and  $\gamma$  be fixed, and  $H(n) = \mathcal{H}\widehat{H}(n)$  with  $\widehat{H}(n)$  being fixed and  $\mathcal{H} \rightarrow \infty$ . Resembling Ge et al. (2015), we give a heuristic derivation of the reduced CME model which averages the full CME model over the stationary distribution of the gene state as  $\mathcal{H} \rightarrow \infty$ . The rigorous averaging principle of the full CME model is given by Theorem 5.1 in Kang and Kurtz (2013).

Define  $p(n, t) := \sum_{i=1}^G p_i(n, t)$ . Let  $\zeta_i(n)$  be the stationary probability for gene state  $i$  of the Markov chain with transition rate matrix  $-\mathcal{H}\widehat{H}(n)$ . As  $\mathcal{H} \rightarrow \infty$ ,  $p_i(n, t) \propto \zeta_i(n)$  approximately. Define  $\bar{k}(n) := \sum_{i=1}^G \zeta_i(n) k_i$ . Sum Eq. (1) over  $i$ .

$$\begin{aligned} \partial_t p(n, t) &= \bar{k}(n-1) p(n-1, t) \gamma(n+1) p(n+1, t) \\ &\quad - \bar{k}(n) p(n, t) - \gamma n p(n, t). \end{aligned} \tag{13}$$

Equation (13) is called the reduced CME model.

Note that the reduced CME is actually a full CME with only one gene state. Thus, the FR approximation of it as in Sect. 2.2 directly leads to the DMFD. Use the same rate scales as in Sect. 2.2, and define  $\widetilde{k}(x) := \bar{k}(n)$ . By Eq. (4),  $\widetilde{k}(x) = n_{\max} g(x)$ . Then Eq. (13) becomes

$$\begin{aligned} \partial_t \widetilde{p}(x, t) &= n_{\max} g\left(x - \frac{1}{n_{\max}}\right) \widetilde{p}\left(x - \frac{1}{n_{\max}}, t\right) \\ &\quad + \gamma n_{\max} \left(x + \frac{1}{n_{\max}}\right) \widetilde{p}\left(x + \frac{1}{n_{\max}}, t\right) \\ &\quad - n_{\max} g(x) \widetilde{p}(x, t) - \gamma n_{\max} x \widetilde{p}(x, t). \end{aligned} \tag{14}$$

Substitute

$$\begin{aligned} g\left(x - \frac{1}{n_{\max}}\right) \widetilde{p}\left(x - \frac{1}{n_{\max}}, t\right) &\approx g(x) \widetilde{p}(x, t) - \frac{1}{n_{\max}} \partial_x [g(x) \widetilde{p}(x, t)], \\ \left(x + \frac{1}{n_{\max}}\right) \widetilde{p}\left(x + \frac{1}{n_{\max}}, t\right) &\approx \widetilde{p}(x, t) + \frac{1}{n_{\max}} \partial_x [x \widetilde{p}(x, t)], \end{aligned}$$

into Eq. (14), one arrives at

$$\partial_t \widetilde{p}(x, t) = -\partial_x \{ [g(x) - \gamma x] \widetilde{p}(x, t) \},$$

which is exactly Eq. (5). Therefore, the FR model and the reduced CME model share the same DMFD in Eq. (6).

Define  $p^{ss}(n)$  as the stationary distribution of  $p(n, t)$ . By Eq. (13), for  $n \geq 1$ ,

$$0 = \bar{k}(n - 1) p^{ss}(n - 1) + \gamma(n + 1) p^{ss}(n + 1) - \bar{k}(n) p^{ss}(n) - \gamma n p^{ss}(n).$$

$0 = \gamma p^{ss}(1) - \bar{k}(0) p^{ss}(0)$  for  $n = 0$ . Therefore,

$$\begin{aligned} p^{ss}(xn_{\max}) &= p^{ss}(0) \prod_{i=0}^{xn_{\max}-1} \frac{\bar{k}(i)}{\gamma(i+1)} = p^{ss}(0) \exp \left[ \sum_{i=0}^{xn_{\max}-1} \log \left( \frac{\bar{k}(i)}{\gamma(i+1)} \right) \right] \\ &= p^{ss}(0) \exp \left[ n_{\max} \sum_{i=0}^{xn_{\max}-1} \frac{1}{n_{\max}} \log \left( \frac{\bar{k}(i)}{\gamma(i+1)} \right) \right] \\ &= p^{ss}(0) \exp \left[ n_{\max} \sum_{i=0}^{xn_{\max}-1} \frac{1}{n_{\max}} \log \left( \frac{g\left(\frac{i}{n_{\max}}\right)}{\gamma \frac{i+1}{n_{\max}}} \right) \right] \\ &\approx p^{ss}(0) \exp \left[ n_{\max} \int_0^x \log \left( \frac{g(y)}{\gamma y} \right) dy \right]. \end{aligned} \tag{15}$$

Define  $\widehat{\Phi}^R(x) := \lim_{n_{\max} \rightarrow \infty} -\frac{1}{n_{\max}} \log [p^{ss}(xn_{\max})] = -\int_0^x \log \left( \frac{g(y)}{\gamma y} \right) dy$ . Then

$$\partial_x \widehat{\Phi}^R(x) = -\log \left( \frac{g(x)}{\gamma x} \right). \tag{16}$$

By Eqs. (6) and (16), we have the Lyapunov property of  $\widehat{\Phi}^R(x)$ .

$$\frac{d\widehat{\Phi}^R(x)}{dt} = \partial_x \widehat{\Phi}^R(x) \frac{dx}{dt} = -\log \left( \frac{g(x)}{\gamma x} \right) [g(x) - \gamma x] \leq 0.$$

### 3 Rigorous analysis of the rate function and the prefactor of the FR model

Based on the heuristic arguments in the previous section, we propose several basic assumptions for the FR models:

**Assumption 1** The stationary distribution  $\widetilde{p}_i^{ss}(x)$  of the FR model can be expressed as

$$\widetilde{p}_i^{ss}(x) = C_i(x, \mathcal{H}) \exp \left( -\mathcal{H} \widehat{\Phi}(x) \right),$$

in which  $\lim_{\mathcal{H} \rightarrow \infty} -\frac{1}{\mathcal{H}} \log C_i(x, \mathcal{H}) = 0$  and  $\lim_{\mathcal{H} \rightarrow \infty} \frac{1}{\mathcal{H}} \partial_x [\log C_i(x, \mathcal{H})] = 0$ .

Furthermore,  $\widehat{\Phi}(x)$  is continuously-differentiable, and  $\partial_x \widehat{\Phi}(x)$  satisfies Eq. (12).

**Assumption 2**  $C(x, \mathcal{H})$  in Assumption 1 satisfies  $\lim_{\mathcal{H} \rightarrow +\infty} C(x, \mathcal{H}) / \|C(x, \mathcal{H})\|_2 = \tilde{C}(x)$ , and  $\tilde{C}(x)$  satisfies Eq. (11).

**Assumption 3** The negative transition matrixes  $\widehat{H}(x)$  are irreducible and continuous in  $(x_G^*, x_1^*)$ .

Generalised eigenvalue (GE) proposed in previous works (Chu 1987; Ikramov 1993) and dominant generalised eigenvalue (DGE) proposed in this work are the two most important concepts in this section. The classical eigenvalue problem of a matrix  $Z$  is to solve the equation  $\det(\lambda I - Z) = 0$ . Replacing the identity matrix  $I$  by another matrix  $D$ , roots of  $\det(\lambda D - Z) = 0$  are called GEs of the matrix  $Z$  on  $D$ , which form the generalised spectrum. Assume  $Z$  is a  $Z$ -matrix, i.e. all the off-diagonal elements are nonpositive, then we can define the classical dominant eigenvalue problem for  $Z$ . For  $\lambda$  large enough,  $\lambda I - Z$  is a nonnegative matrix with a real positive dominant eigenvalue  $\lambda_0$  due to the Perron–Frobenius theorem. Then  $\lambda - \lambda_0$  is a real eigenvalue with the smallest real part among all eigenvalues of  $Z$ , which is called the classic dominant eigenvalue of  $Z$  and denoted by  $r(Z)$ . Obviously, such a definition does not depend on the choice of  $\lambda$ . Note that  $Z - r(Z)I$  is still a  $Z$ -matrix, and the eigenvalue of  $Z - r(Z)I$  with the smallest real part is exactly 0. Thus,  $r(Z - r(Z)I) = 0$ . Replacing  $I$  by another diagonal matrix  $D$ , any solution of  $r(Z - \lambda D) = 0$  (see Definition A1) is called a DGE of  $Z$  on  $D$ .

By Assumption 1 and Definition A1,  $\partial_x \widehat{\Phi}(x)$  is a GE of  $\widehat{H}(x)$  on  $A(x)$ . By Assumption 3, Corollary A1 and Theorem A1,  $\widehat{H}(x)$  has at most two DGEs on  $A(x)$  for  $x \in (x_G^*, x_1^*)$ . One is always 0. The other, if exists, is defined as the nontrivial DGE  $\mu(\widehat{H}|A)(x)$ . Otherwise, define  $\mu(\widehat{H}|A)(x) := 0$  (see Definition A4). For simplicity, define  $\mu(x) := \mu(\widehat{H}|A)(x)$  for  $x \in (x_G^*, x_1^*)$ .

### 3.1 The relation between $\partial_x \widehat{\Phi}(x)$ and $\mu(x)$

First, we justify the continuity of  $\mu(x)$ , especially at those  $x_i^*$ 's with degenerated  $A(x)$ . For  $x \neq x_i^*$ ,  $\det[\widehat{H}(x) - \lambda A(x)] = 0$  is a  $\lambda$ -polynomial of order  $G$  and has  $G$  roots (GEs). However,  $\det[\widehat{H}(x_i^*) - \lambda A(x_i^*)] = 0$  is of order  $G - 1$  and has only  $G - 1$  roots (GEs). The following theorem shows that one of the  $G$  GEs for  $x \neq x_i^*$ , denoted by  $\lambda_\infty(x)$ , diverges at  $x_i^*$ , and the other  $G - 1$  GEs for  $x \neq x_i^*$  tend to the  $G - 1$  GEs at  $x_i^*$ . Furthermore, for  $x \neq x_i^*$ , the nontrivial DGE is not  $\lambda_\infty(x)$  near  $x_i^*$ , and converges at  $x_i^*$  to the nontrivial DGE at  $x_i^*$ .

**Theorem 1** Under Assumption 3, we have for  $i \in [2, G - 1]$

1. There exists a single root  $\lambda_\infty(x)$  of  $\det[\widehat{H}(x) - \lambda A(x)] = 0$  near  $x_i^*$  such that  $\widehat{h}_{i,\cdot}(x_i^*) = \lim_{x \rightarrow x_i^*} \gamma(x_i^* - x) \lambda_\infty(x)$ .

2.  $\forall \lambda,$

$$\lim_{x \rightarrow x_i^*} p(x)(\lambda) = \det \left[ \widehat{H}(x_i^*) - \lambda A(x_i^*) \right] / \left[ \widehat{h}_{i,\cdot}(x_i^*) \prod_{i' \neq i} \gamma(x_i^* - x_{i'}^*) \right],$$

where for  $x \in \bigcup_{i'=1}^{G-1} (x_{i'+1}^*, x_{i'}^*),$

$$p(x)(\lambda) := \det \left[ \widehat{H}(x) - \lambda A(x) \right] / \left\{ [\lambda - \lambda_\infty(x)] \prod_{i'=1}^G \gamma(x - x_{i'}^*) \right\}. \tag{17}$$

3.  $\lim_{x \rightarrow x_i^*} \mu(x) = \mu(x_i^*).$

**Proof** For  $x \in \bigcup_{i'=1}^{G-1} (x_{i'+1}^*, x_{i'}^*),$   $A(x)$  is nonsingular, thereby

$$\det \left[ \widehat{H}(x) - \lambda A(x) \right] = \det \left[ \widehat{H}(x) A^{-1}(x) - \lambda I \right] \det [A(x)],$$

and the GEs of  $\widehat{H}(x)$  on  $A(x)$  are the eigenvalues of  $\widehat{H}(x) A^{-1}(x).$   $\forall \lambda,$

$$\lim_{x \rightarrow x_i^*} \det \left[ \gamma(x_i^* - x) \widehat{H}(x) A^{-1}(x) - \lambda I \right] = (-1)^G \lambda^{G-1} \left[ \lambda - \widehat{h}_{i,\cdot}(x_i^*) \right],$$

so as  $x \rightarrow x_i^*,$   $G - 1$  eigenvalues (count multiplicity) of  $\gamma(x_i^* - x) \widehat{H}(x) A^{-1}(x)$  tend to 0, and the remaining one tends to  $\widehat{h}_{i,\cdot}(x_i^*).$  The eigenvalues of  $\gamma(x_i^* - x) \widehat{H}(x) A^{-1}(x)$  are just those of  $\widehat{H}(x) A^{-1}(x)$  multiplying  $\gamma(x_i^* - x),$  so we have the statement 1.

Multiply both sides of Eq. (17) by  $[\lambda - \lambda_\infty(x)] \prod_{i'=1}^G \gamma(x - x_{i'}^*).$

$$\left[ \prod_{i' \neq i} \gamma(x - x_{i'}^*) \right] \gamma(x - x_i^*) [\lambda - \lambda_\infty(x)] p(x)(\lambda) = \det \left[ \widehat{H}(x) - \lambda A(x) \right].$$

Because  $\lim_{x \rightarrow x_i^*} \gamma(x_i^* - x) \lambda_\infty(x) = \widehat{h}_{i,\cdot}(x_i^*),$  for any fixed  $\lambda,$

$$\begin{aligned} & \widehat{h}_{i,\cdot}(x_i^*) \left[ \prod_{i' \neq i} \gamma(x_i^* - x_{i'}^*) \right] \lim_{x \rightarrow x_i^*} p(x)(\lambda) \\ &= \lim_{x \rightarrow x_i^*} \left[ \prod_{i' \neq i} \gamma(x - x_{i'}^*) \right] \gamma(x - x_i^*) [\lambda - \lambda_\infty(x)] p(x)(\lambda) \\ &= \lim_{x \rightarrow x_i^*} \det \left[ \widehat{H}(x) - \lambda A(x) \right] = \lim_{x \rightarrow x_i^*} \det \left[ \widehat{H}(x_i^*) - \lambda A(x_i^*) \right]. \end{aligned}$$

This is the statement 2.

Because  $\lim_{x \rightarrow x_i^*} \gamma(x_i^* - x) \lambda_\infty(x) = \widehat{h}_{i,\cdot}(x_i^*) > 0$ , we have

$$\lim_{x \rightarrow x_i^* -} \lambda_\infty(x) = +\infty, \quad \lim_{x \rightarrow x_i^* +} \lambda_\infty(x) = -\infty.$$

Denote the number of positive and negative diagonal elements of  $A(x)$  by  $m[A(x)]$  and  $n[A(x)]$ , respectively. Because  $m[A(x_i^*)], n[A(x_i^*)] > 0$ , some diagonal element of  $\widehat{H}(x) - \lambda_\infty(x)A(x)$  tends to  $-\infty$  as  $x \rightarrow x_i^*$ . By the Gershgorin circle theorem (Gershgorin 1931),  $\lim_{x \rightarrow x_i^*} r[\widehat{H}(x) - \lambda_\infty(x)A(x)] = -\infty$ . Thus,  $\mu(x) \neq \lambda_\infty(x)$  for  $x$  sufficiently close to  $x_i^*$ . By statement 2, the roots of  $\det[\widehat{H}(x) - \lambda A(x)] = 0$  except  $\lambda_\infty(x)$  tend to the  $G - 1$  roots of  $\det[\widehat{H}(x_i^*) - \lambda A(x_i^*)] = 0$  as  $x \rightarrow x_i^*$ , thereby continuous at  $x_i^*$  and bounded near  $x_i^*$ ; thus,  $-\infty < \underline{\lim}_{x \rightarrow x_i^*} \mu(x) \leq \overline{\lim}_{x \rightarrow x_i^*} \mu(x) < +\infty$ . Let  $\lim_{j \rightarrow +\infty} y_j = x_i^*$  be a sequence such that  $\lim_{j \rightarrow +\infty} \mu(y_j) = \overline{\lim}_{x \rightarrow x_i^*} \mu(x)$ .  $\widehat{h}_{i,\cdot}(y_j) > 0$  because  $\widehat{H}(x)$  is irreducible by Assumption 3. Then

$$r \left[ \widehat{H}(x_i^*) - \overline{\lim}_{x \rightarrow x_i^*} \mu(x) A(x_i^*) \right] = \lim_{j \rightarrow +\infty} r \left[ \widehat{H}(y_j) - \mu(y_j) A(y_j) \right] = 0.$$

Thus,  $\overline{\lim}_{x \rightarrow x_i^*} \mu(x)$  is a DGE of  $\widehat{H}(x_i^*)$  on  $A(x_i^*)$ , and so is  $\underline{\lim}_{x \rightarrow x_i^*} \mu(x)$ .

Now we are going to prove that  $\overline{\lim}_{x \rightarrow x_i^*} \mu(x) = \underline{\lim}_{x \rightarrow x_i^*} \mu(x)$  by contradiction. If not so, then  $\overline{\lim}_{x \rightarrow x_i^*} \mu(x)$  and  $\underline{\lim}_{x \rightarrow x_i^*} \mu(x)$  are the only two DGEs of  $\widehat{H}(x_i^*)$  on  $A(x_i^*)$  by Corollary A1, and one must be 0 by Theorem A1. Assume without loss of generality that  $0 = \overline{\lim}_{x \rightarrow x_i^*} \mu(x) > \underline{\lim}_{x \rightarrow x_i^*} \mu(x)$ . Let  $\bar{\mu} := \underline{\lim}_{x \rightarrow x_i^*} \mu(x) / 2 < 0$ . By Corollary A1,  $r[\widehat{H}(x_i^*) - \bar{\mu}A(x_i^*)] > 0$  since  $\underline{\lim}_{x \rightarrow x_i^*} \mu(x) < \bar{\mu} < \overline{\lim}_{x \rightarrow x_i^*} \mu(x)$ . For  $j$  large enough,  $\bar{\mu} < \mu(y_j)$ . Then  $r[\widehat{H}(y_j) - \bar{\mu}A(y_j)] < 0$  by Corollary A1 since  $\bar{\mu} < 0$  and  $\bar{\mu} < \mu(y_j)$ . As  $j \rightarrow \infty$ ,  $r[\widehat{H}(x_i^*) - \bar{\mu}A(x_i^*)] \leq 0$ , conflicts. Therefore,  $\overline{\lim}_{x \rightarrow x_i^*} \mu(x) = \underline{\lim}_{x \rightarrow x_i^*} \mu(x)$ .

Now we show that  $\lim_{x \rightarrow x_i^*} \mu(x) = \mu(x_i^*)$ .

1. If  $\mu(x_i^*) = 0$ , then by Definition A4, 0 is the only DGE of  $\widehat{H}(x_i^*)$  on  $A(x_i^*)$ . Then  $\lim_{x \rightarrow x_i^*} \mu(x) = 0 = \mu(x_i^*)$  because  $\lim_{x \rightarrow x_i^*} \mu(x)$  is a DGE.
2. If  $\mu(x_i^*) > 0$ , then following Definition A3, denote the number (count multiplicity) of GEs with positive (negative) real parts of  $\widehat{H}(x)$  on  $A(x)$  by  $m(\widehat{H}|A)(x)$  ( $n(\widehat{H}|A)(x)$ ). By symmetry, it is enough to prove  $\lim_{x \rightarrow x_i^* -} \mu(x) = \mu(x_i^*)$ .

By Theorem A1,  $m(\widehat{H}|A)(x_i^*) = m[A(x_i^*)] = i - 1$ . By continuity,  $\exists \delta > 0$  such

that for  $x \in (x_i^* - \delta, x_i^*)$ , the  $i - 1$  roots of  $\det [\widehat{H}(x_i^*) - \lambda A(x_i^*)] = 0$  with positive real parts at  $x_i^*$  still have positive real parts at  $x$ . By  $\lim_{x \rightarrow x_i^* -} \lambda_\infty(x) = +\infty$ , it is possible to decrease  $\delta$  such that  $\lambda_\infty(x) > 0$ , thereby  $m(\widehat{H}|A)(x) \geq i$  for  $x \in (x_i^* - \delta, x_i^*)$ . Conversely, for  $x \in (x_{i+1}^*, x_i^*)$ ,  $m(\widehat{H}|A)(x) \leq m[A(x)] = i$  by Theorem A1. Therefore,  $m(\widehat{H}|A)(x) = m[A(x)] = i$  for  $x \in (x_i^* - \delta, x_i^*)$ . By Theorem A1,  $\mu(x) > 0$  for  $x \in (x_i^* - \delta, x_i^*)$ . Denote the roots of  $\det [\widehat{H}(x) - \lambda A(x)] = 0$  other than  $\lambda_\infty(x)$  and  $\mu(x)$  by  $\{\lambda_j(x)\}_{j=1}^{G-2}$ . Assume without loss of generality that  $\{\lambda_j(x)\}_{j=1}^{i-2}$  have positive real parts in  $(x_i^* - \delta, x_i^*)$ . By statement 2,  $\mu(x)$  and  $\{\lambda_j(x)\}_{j=1}^{G-2}$  tend to the roots of  $\det [\widehat{H}(x_i^*) - \lambda A(x_i^*)] = 0$ . If  $\lim_{x \rightarrow x_i^* -} \mu(x) = \mu(x_i^*)$  does not hold, then  $\lim_{x \rightarrow x_i^* -} \mu(x) = 0$ . Also, since  $\{\lambda_j\}_{j=i-1}^{G-2}$  have nonpositive real parts in  $(x_i^* - \delta, x_i^*)$ ,  $\lim_{x \rightarrow x_i^* -} \Re[\lambda_j(x)] \leq 0$  for  $j \in [i - 1, G - 2]$ . Thus,  $m(\widehat{H}|A)(x_i^*) \leq i - 2 < i - 1 = m[A(x_i^*)]$ , which conflicts Theorem A1.

3. The case for  $\mu(x_i^*) < 0$  can be proved similarly as if  $\mu(x_i^*) > 0$ .

□

Theorems A2 and 1 imply Corollary 1.

**Corollary 1** Under Assumption 3,  $\mu(x)$  is continuous in  $(x_G^*, x_1^*)$ .

**Remark 1** A generalisation of Theorem 1 is given in Theorem A3 under a weaker assumption than Assumption 3 that for  $i \in [2, G - 1]$ ,  $\widehat{H}(x)$  is continuous at  $x_i^*$ ,  $\lim_{x \rightarrow x_i^*} \frac{\widehat{h}_{i \cdot}(x)}{|x - x_i^*|} = +\infty$ , and  $\lim_{x \rightarrow x_i^*} \frac{\widehat{h}_{i,j}(x)}{\widehat{h}_{i \cdot}(x)}$  exists  $\forall j \neq i$ .

The following lemma is essential for the relation between  $\partial_x \widehat{\Phi}(x)$  and  $\mu(x)$ .

**Lemma 1** Under Assumption 3, any real continuous GE  $\lambda(x)$  in  $(x_G^*, x_1^*)$  satisfies  $\lambda(x) \in \{0, \mu(x)\}$ ,  $\forall x \in (x_G^*, x_1^*)$ .

**Proof** By Corollary 1,  $\mu(x)$  is continuous in  $(x_G^*, x_1^*)$ . By Lemma A1,  $\mu(x)$  and 0 is not a single root of  $\det [\widehat{H}(x) - \lambda A(x)] = 0$  iff  $\mu(x) = 0$ . So  $\mu(x)$  and 0 do not intersect any other GE of  $\widehat{H}(x)$  on  $A(x)$ . Thus,  $\exists x \in (x_G^*, x_1^*)$  such that  $\lambda(x) \in \{0, \mu(x)\}$  implies  $\lambda(x) \in \{0, \mu(x)\} \forall x \in (x_G^*, x_1^*)$ . So it is enough to prove the former. By Lemma A6,  $0 < \lambda(x)$  for  $x \in (x_G^*, x_{G-1}^*)$ , and  $0 > \lambda(x)$  for  $x \in (x_2^*, x_1^*)$ . By continuity of  $\lambda(x)$ ,  $\exists x \in (x_G^*, x_1^*)$  such that  $\lambda(x) = 0 \in \{0, \mu(x)\}$ . □

By Assumption 1 and Lemma 1,  $\partial_x \widehat{\Phi}(x) \in \{0, \mu(x)\}$ . However, we still can not uniquely determine  $\partial_x \widehat{\Phi}$ . It is because if  $\mu(x)$  has two roots  $x_G^* < x' < x'' < x_1^*$ , one can let  $\partial_x \widehat{\Phi}(x) = 0$  in  $(x', x'')$  and  $\partial_x \widehat{\Phi}(x) = \mu(x)$  in  $(x_G^*, x') \cup (x'', x_1^*)$ .

Then  $\partial_x \widehat{\Phi}(x)$  is still continuous and does not violate Assumption 1 and Lemma 1. Therefore, we have to add another assumption in order to not consider those trivial scenarios.

**Assumption 4** If  $\partial_x \widehat{\Phi}(x) \in \{0, \mu(x)\}$ , then  $\partial_x \widehat{\Phi}(x) = \mu(x)$ .

### 3.2 The Lyapunov property of $\widehat{\Phi}(x)$ for DMFD

Theorem 2 proves the Lyapunov property of  $\widehat{\Phi}(x)$ .

**Theorem 2** Under Assumptions 1, 3 and 4, in  $(x_G^*, x_1^*)$ :

1.  $\mu(x) = 0$  iff  $g(x) - \gamma x = 0$ ;
2.  $\widehat{\Phi}(x)$  satisfies

$$\frac{d\widehat{\Phi}(x)}{dt} = \partial_x \widehat{\Phi}(x) \frac{dx}{dt} = \mu(x) [g(x) - \gamma x] \leq 0, \tag{18}$$

and the equality holds iff  $g(x) - \gamma x = 0$ .

**Proof** By Eq. (4) and Lemma A8,

$$g(x) - \gamma x = \frac{\sum_{j=1}^G \det \left[ \widehat{H}^{\bar{j}}(x) \right] \gamma(x_j^* - x)}{\sum_{j=1}^G \det \left[ \widehat{H}^{\bar{j}}(x) \right]}. \tag{19}$$

$\sum_{j=1}^G \det \left[ \widehat{H}^{\bar{j}}(x) \right] > 0$  because  $\widehat{H}(x)$  is irreducible. So

$$g(x) - \gamma x = 0 \Leftrightarrow \sum_{j=1}^G \det \left[ \widehat{H}^{\bar{j}}(x) \right] \gamma(x_j^* - x) = 0, \tag{20}$$

where  $\widehat{H}^{\bar{j}}(x)$  is the submatrix of  $\widehat{H}(x)$  deleting the  $j$ th row and column. For  $x \in \bigcup_{j=1}^{G-1} (x_{j+1}^*, x_j^*)$ , denote the  $G$  roots of  $\det \left[ \widehat{H}(x) - \lambda A(x) \right]$  by  $\{\lambda_j\}_{j=1}^G$ , and for  $x = x_i^*$  for some  $i \in [2, G - 1]$ , denote the  $G - 1$  roots by  $\{\lambda_j\}_{j=1}^{G-1}$ . Assume without loss of generality that  $\lambda_1(x) \equiv 0$ . Observing the coefficients of  $\det \left[ \widehat{H}(x) - \lambda A(x) \right]$ , we have for  $x \in \bigcup_{j=1}^{G-1} (x_{j+1}^*, x_j^*)$ ,

$$\frac{\sum_{j=1}^G \det \left[ \widehat{H}^{\bar{j}}(x) \right] \gamma(x_j^* - x)}{\prod_{j=1}^G \gamma(x_j^* - x)} = \sum_{j=1}^G \prod_{j' \neq j} \lambda_{j'}(x) = \prod_{j=2}^G \lambda_j(x), \tag{21}$$

and for  $x = x_i^*$ ,

$$\frac{\sum_{j=1}^G \det \left[ \widehat{H}^{\bar{j}}(x_i^*) \right] \gamma(x_j^* - x_i^*)}{\widehat{h}_{i, \cdot}(x_i^*) \prod_{j \neq i} \gamma(x_j^* - x_i^*)} = \sum_{j=1}^{G-1} \prod_{j' \neq j} \lambda_{j'}(x_i^*) = \prod_{j=2}^{G-1} \lambda_j(x_i^*).$$

By Lemma A1,  $\mu(x)$  or 0 is not a single root of  $\det \left[ \widehat{H}(x) - \lambda A(x) \right] = 0$  iff  $\mu(x) = 0$ . In other words,  $\prod_{j=2}^G \lambda_j(x) = 0$  (or  $\prod_{j=2}^{G-1} \lambda_j(x) = 0$  for  $x = x_i^*$ ) is equivalent to  $\mu(x) = 0$ , thereby  $\sum_{j=1}^G \det \left[ \widehat{H}^{\bar{j}}(x) \right] \gamma(x_j^* - x) = 0$ . This is the statement 1.

By Eqs. (19) and (21), and Assumption 4, for  $x \in \bigcup_{j=1}^{G-1} (x_{j+1}^*, x_j^*)$ ,

$$\begin{aligned} \frac{d\widehat{\Phi}(x)}{dt} &= \partial_x \widehat{\Phi}(x) \frac{dx}{dt} = \mu(x) [g(x) - \gamma x] \\ &= \frac{\mu(x)}{\sum_{j=1}^G \det \left[ \widehat{H}^{\bar{j}}(x) \right]} \left[ \prod_{j=2}^G \lambda_j(x) \right] \prod_{j=1}^G \gamma(x_j^* - x). \end{aligned}$$

By Lemma A6, exactly one of the following occurs.

1.  $m \left( \widehat{H}|A \right)(x) = m[A(x)]$ ,  $n \left( \widehat{H}|A \right)(x) = n[A(x)] - 1$ ,  $\mu(x) > 0$ .
2.  $m \left( \widehat{H}|A \right)(x) = m[A(x)] - 1$ ,  $n \left( \widehat{H}|A \right)(x) = n[A(x)]$ ,  $\mu(x) < 0$ .
3.  $m \left( \widehat{H}|A \right)(x) = m[A(x)] - 1$ ,  $n \left( \widehat{H}|A \right)(x) = n[A(x)] - 1$ ,  $\mu(x) = 0$ .

The signs of  $\left[ \prod_{j=2}^G \lambda_j(x) \right]$  and  $\prod_{j=1}^G \gamma(x_j^* - x)$  are determined by  $n \left( \widehat{H}|A \right)(x)$  and  $n[A(x)]$ , respectively. In case 1, they have opposite signs, and  $\mu(x) > 0$ . In case 2, they have the same sign, and  $\mu(x) < 0$ . In case 3,  $\mu(x) = 0$ . Since  $\sum_{i=1}^G \det \left[ \widehat{H}^{\bar{i}}(x) \right] > 0$ ,  $\frac{d\widehat{\Phi}(x)}{dt} \leq 0$  in all cases. Because  $\mu(x)$  and  $g(x)$  are continuous in  $(x_G^*, x_1^*)$ ,  $\frac{d\widehat{\Phi}(x)}{dt}$  is also continuous by Eq. (18). Thus,  $\frac{d\widehat{\Phi}(x)}{dt} \leq 0$  for  $x = x_i^*$  with  $i \in [2, G - 1]$ . The necessary and sufficient condition for the equality is implied by the statement 1. Hence the statement 2 is proved.  $\square$

**Proposition 1** Under Assumptions 2, 3 and 4,  $\widetilde{C}_i(x)$  is continuous and positive in  $(x_G^*, x_1^*)$ .

**Proof** By definition,  $r \left[ \widehat{H}(x) - \mu(x) A(x) \right] = 0$  for  $x \neq x_i^*$ . By continuity,  $r \left[ \widehat{H}(x) - \mu(x) A(x) \right] = 0 \forall x \in (x_G^*, x_1^*)$ . By Assumption 4,  $\partial_x \widehat{\Phi}(x) = \mu(x)$ . Then by Eq. (11),  $\widetilde{C}(x)$  is the eigenvector of  $\widehat{H}(x) - \mu(x) A(x)$  corresponding to the dominant eigenvalue  $r \left[ \widehat{H}(x) - \mu(x) A(x) \right] = 0$ . Because  $\widehat{H}(x) - \mu(x) A(x)$  is



irreducible, by the Perron–Frobenius theorem,  $\tilde{C}_i(x) > 0 \forall i$ , and  $\tilde{C}_i(x)$  is continuous since  $r \left[ \widehat{H}(x) - \mu(x) A(x) \right] = 0$  is a single eigenvalue.  $\square$

**Proposition 2** *Under Assumptions 1, 3 and 4, for  $\epsilon > 0$ ,*

$$\int_{x_G^*}^{x_G^* + \epsilon} \partial_x \widehat{\Phi}(x) dx = -\infty, \quad \int_{x_1^* - \epsilon}^{x_1^*} \partial_x \widehat{\Phi}(x) dx = +\infty.$$

**Proof** Statement 1 in Theorem 1 is also true for  $x_1^*$  and  $x_G^*$ . So

$$\int_{x_G^*}^{x_G^* + \epsilon} \lambda_\infty(x) dx = -\infty, \quad \int_{x_1^* - \epsilon}^{x_1^*} \lambda_\infty(x) dx = +\infty.$$

By Assumption 4,  $\partial_x \widehat{\Phi}(x) = \mu(x)$ . It is enough to prove that  $\mu(x) = \lambda_\infty(x)$  in  $(x_G^*, x_G^* + \epsilon)$  and  $(x_1^* - \epsilon, x_1^*)$  for  $\epsilon$  small enough. Since  $\lim_{x \rightarrow x_G^*+} \lambda_\infty(x) = -\infty$ , for  $\epsilon < x_{G-1}^* - x_G^*$  small enough,  $\lambda_\infty(x) < 0$  in  $(x_G^*, x_G^* + \epsilon)$ . Thus,  $n \left( \widehat{H} | A \right) (x) \geq 1 = n[A(x)]$ . By Theorem A1,  $n \left( \widehat{H} | A \right) (x) = n[A(x)]$  ( $\lambda_\infty(x)$  is the only GE with negative real part) and  $\mu(x) < 0$ . Thus,  $\mu(x) = \lambda_\infty(x)$  in  $(x_G^*, x_G^* + \epsilon)$ . The case for  $x_1^*$  can be proved similarly.  $\square$

By Proposition 2,

$$\lim_{x \rightarrow x_G^*+} \widehat{\Phi}(x) = +\infty, \quad \lim_{x \rightarrow x_1^*-} \widehat{\Phi}(x) = +\infty.$$

It is appropriate to define  $\widehat{\Phi}(x) = +\infty$  for  $x \in [0, x_G^*] \cup [x_1^*, +\infty)$  because the stationary distribution of the FR model vanishes outside  $(x_G^*, x_1^*)$ . In fact, once  $x$  entering  $(x_G^*, x_1^*)$ , it will never leave based on the piecewise-deterministic dynamics of the FR models.

### 3.3 The numerical algorithm for $\widehat{\Phi}(x)$

We need the following proposition to support our numerical algorithm.

**Proposition 3** *Under Assumption 3, if  $\mu(x) \neq 0$ , then no real GE of  $\widehat{H}(x)$  on  $A(x)$  locates strictly between  $\mu(x)$  and 0.*

**Proof** Assume without loss of generality that  $\mu(x) > 0$ . By Corollary A1,  $r \left[ \widehat{H}(x) - \lambda A(x) \right] > 0$  for  $0 < \lambda < \mu(x)$ . Recall that  $r \left[ \widehat{H}(x) - \lambda A(x) \right]$  is the eigenvalue of  $\widehat{H}(x) - \lambda A(x)$  with the smallest real part. Thus, all eigenvalues of  $\widehat{H}(x) - \lambda A(x)$  have positive real parts. Therefore,  $\det \left[ \widehat{H}(x) - \lambda A(x) \right] > 0$ . So any  $\lambda \in (0, \mu(x))$  cannot be a GE.  $\square$

Proposition 3 together with Theorem A1 can be used to calculate  $\mu(x)$  in  $(x_G^*, x_1^*)$  numerically, thereby  $\partial_{(x)} \widehat{\Phi}(x)$  under Assumption 4. Then based on the fact that the global minimum of the LDP rate function is always zero (Touchette 2009), one can numerically obtain  $\widehat{\Phi}(x)$ .

The rest of this subsection is the numerical algorithm for obtaining the nontrivial DGE at any fixed  $x \in \bigcup_{i=1}^{G-1} (x_{i+1}^*, x_i^*)$ . Since  $A$  is nonsingular, the GEs of  $\widehat{H}$  on  $A$  are the eigenvalues of  $\widehat{H}A^{-1}$ , which can be solved by existing numerical methods. We first assume the ideal case that the numerically solved spectrum is completely accurate. Then  $\mu$  can be obtained based on Theorem A1 and Proposition 3, following the four steps:

1. Calculate  $m(\widehat{H}|A)$  and  $n(\widehat{H}|A)$  by the  $G$  eigenvalues of  $\widehat{H}A^{-1}$ .
2. If  $m(\widehat{H}|A) = m(A)$ , let  $\mu$  be the smallest real positive eigenvalue in the spectrum.
3. If  $n(\widehat{H}|A) = n(A)$ , let  $\mu$  be the largest real negative eigenvalue in the spectrum.
4. If  $m(\widehat{H}|A) = m(A) - 1$  and  $n(\widehat{H}|A) = n(A) - 1$ , let  $\mu = 0$ .

However, in real applications, the solved eigenvalues have round-off error. So we have to modify the above procedure as follows:

1. Set the eigenvalue with the smallest absolute value as 0.
2. Calculate  $m(\widehat{H}|A)$  and  $n(\widehat{H}|A)$  by the  $G$  eigenvalues.
3. If  $m(\widehat{H}|A) \geq m(A)$ , then
  - find the eigenvalues with the smallest absolute imaginary part among all the eigenvalues with positive real parts;
  - let  $\mu$  be the smallest real part among the eigenvalues found in the above step.
4. If  $n(\widehat{H}|A) \geq n(A)$ , do symmetrically as the case when  $m(\widehat{H}|A) \geq m(A)$ .
5. Otherwise, let  $\mu = 0$ .

One may expect that the above modification is able to obtain an accurate  $\mu$  if the round-off error is small. Nevertheless, no theory promises its correctness. Thus, one still needs to check the obtained  $\mu$ . Corollary A1 provides a possible method, which implies that

1. If  $r(\widehat{H} - \lambda A) \geq 0$ , then both  $(-\infty, \lambda]$  and  $[\lambda, +\infty)$  contain DGE.
2. If  $\lambda > 0$  and  $r(\widehat{H} - \lambda A) \leq 0$ , then  $(\lambda, +\infty)$  contains no DGE.
3. If  $\lambda < 0$  and  $r(\widehat{H} - \lambda A) \leq 0$ , then  $(-\infty, \lambda)$  contains no DGE.

We assume that the above numerical algorithm is accurate enough such that the following statements are valid for some small error tolerances  $\tau > \epsilon > 0$ :

1. If  $r(\widehat{H} - \lambda A) \geq 0$ , then both  $(-\infty, \lambda + \epsilon]$  and  $[\lambda - \epsilon, +\infty)$  contain DGE.
2. If  $\lambda \geq \tau$  and  $r(\widehat{H} - \lambda A) \leq 0$ , then  $(\lambda - \epsilon, +\infty)$  contains no DGE.

3. If  $\lambda \leq \tau$  and  $r(\widehat{H} - \lambda A) \leq 0$ , then  $(-\infty, \lambda + \epsilon)$  contains no DGE.

Denote the accurate value of  $\mu$  by  $\mu^*$ . The following procedure checks whether  $\mu$  is a good approximation of  $\mu^*$ . If not so, such a check procedure still can further narrow the possible range for the searching of  $\mu^*$ .

1. If  $|\mu| < 2\tau$ , then

- If  $r(\widehat{H} + 2\tau A) \leq 0$  and  $r(\widehat{H} - 2\tau A) \leq 0$ , then  $\mu^* \in [-2\tau - \epsilon, 2\tau + \epsilon]$ , thereby  $\mu$  is a good approximation of  $\mu^*$ .
- If  $r(\widehat{H} + 2\tau A) > 0$  and  $r(\widehat{H} - 2\tau A) \leq 0$ , then  $(-\infty, -2\tau + \epsilon]$  contains DGE. Since  $0 \notin (-\infty, -2\tau + \epsilon]$ ,  $\mu^* \in (-\infty, -2\tau + \epsilon]$ .
- If  $r(\widehat{H} + 2\tau A) \leq 0$  and  $r(\widehat{H} - 2\tau A) > 0$ , then  $\mu^* \in [2\tau - \epsilon, +\infty)$  symmetrically.
- If  $r(\widehat{H} + 2\tau A) > 0$  and  $r(\widehat{H} - 2\tau A) > 0$ , then  $\mu^* \in (-\infty, -2\tau + \epsilon] \cap [2\tau - \epsilon, +\infty) = \emptyset$ , conflicts.

2. If  $\mu \geq 2\tau$ , then

- If  $r(\widehat{H} - (\mu - \tau)A) \geq 0$  and  $r(\widehat{H} - (\mu + \tau)A) \leq 0$ , then  $\mu^* \in [\mu - \tau - \epsilon, \mu + \tau + \epsilon]$ , thereby  $\mu$  is a good approximation of  $\mu^*$ .
- If  $r(\widehat{H} - (\mu - \tau)A) < 0$  and  $r(\widehat{H} - (\mu + \tau)A) \leq 0$ , then  $\mu^* \in (-\infty, \mu - \tau + \epsilon]$ .
- If  $r(\widehat{H} - (\mu - \tau)A) \geq 0$  and  $r(\widehat{H} - (\mu + \tau)A) > 0$ , then  $\mu^* \in [\mu + \tau - \epsilon, +\infty)$ .
- If  $r(\widehat{H} - (\mu - \tau)A) < 0$  and  $r(\widehat{H} - (\mu + \tau)A) > 0$ , then  $\mu^* \in (-\infty, \mu - \tau + \epsilon] \cap [\mu + \tau - \epsilon, +\infty) = \emptyset$ , conflicts.

3. If  $\mu \leq -2\tau$ , do symmetrically as the case when  $\mu \geq 2\tau$ .

Unless  $\mu$  is already a good approximation of  $\mu^*$ , one should

1. remove eigenvalues whose real parts equal to  $\mu$ , or do not belong to the possible range of  $\mu^*$  obtained from the above steps;
2. find one eigenvalue with the smallest absolute imaginary part among the remaining eigenvalues, and let  $\mu$  be the real part of it;
3. check whether the new  $\mu$  is a good approximation of  $\mu^*$ , and if not so, further narrow the possible range of  $\mu^*$ .

Repeat the above procedure until one of the following cases is encountered.

1. If  $\mu$  is a good approximation of  $\mu^*$ , then return  $\mu$ .
2. Otherwise, if the possible range of  $\mu^*$  is small enough, then return the mid-point of the range.
3. If all eigenvalues are removed, and the possible range of  $\mu^*$  is still not small enough, then apply a modified dichotomic search in the possible range.

In summary, we have Algorithm 3.1, where  $\Re$  and  $\Im$  mean the real and imaginary parts, respectively, and  $[a, b]$  is the possible range of  $\mu^*$ . The modified dichotomic

search converges because  $b - a \leq (b' - a')/2 + \tau$ , where  $[a', b']$  is the possible range of  $\mu^*$  in the previous step.  $\mu$  at  $x_i^*$  for  $i \in [2, G - 1]$  can be obtained by continuity.

#### 4 Local fluctuations and transition rates between phenotypic states in a tristable example

We apply Algorithm 3.1 and Eq. (16) to a tristable example. The LDP rate functions in the intermediate and rapid regions of gene-state switchings are calculated, and the local fluctuations obtained from the rate function are compared. Here, local fluctuations are just the second derivative of the LDP rate function at each local minimum, which corresponds to each steady fixed point of the DMFD. To be more precise, we can expand the rate function  $\phi(x)$  near any local minimum  $x^*$ , i.e.

$$\phi(x) = \phi(x^*) + \frac{\phi''(x^*)}{2} (x - x^*)^2 + \dots \quad (22)$$

Then the asymptotic Gaussian variance of  $p^{ss}(x)$  close to  $x^*$  is  $(\mathcal{H}\phi''(x^*))^{-1}$ .

In this section, we also show that the transition rates between phenotypic states in the intermediate region are correctly predicted by rate formula proposed in the FR model based on the Freidlin–Wentzell LDP (Freidlin and Wentzell 2014), and a direct comparison with the stationary distribution of the full CME model further appreciates the validity of the numerically obtained rate function of the FR model.

##### 4.1 Tristable example

Let  $G = 3$ . By the definitions of  $g(x)$  and  $A(x)$  in Eqs. (4) and (10), the fixed points (phenotypic states) of the DMFD in Eq. (6) are the roots of

$$\begin{aligned} \det \left[ \widehat{H}^{\bar{1}}(x) \right] \gamma(x_1^* - x) + \det \left[ \widehat{H}^{\bar{2}}(x) \right] \gamma(x_2^* - x) \\ + \det \left[ \widehat{H}^{\bar{3}}(x) \right] \gamma(x_3^* - x) = 0, \end{aligned} \quad (23)$$

where  $\widehat{H}^{\bar{j}}(x)$  is the submatrix of  $\widehat{H}(x)$  deleting the  $j$ th row and column. Let

$$\begin{aligned} \widehat{h}_{1,3}(x) \equiv \widehat{h}_{3,1}(x) \equiv 0, \quad \widehat{h}_{1,2}(x) := \widehat{h}_{1,2}^c \equiv 1, \quad \widehat{h}_{2,3}(x) := \widehat{h}_{2,3}^c > 0, \\ \widehat{h}_{2,1}(x) := \widehat{h}_{2,1}^c x^2 > 0, \quad \widehat{h}_{3,2}(x) := \widehat{h}_{3,2}^c x^2 > 0. \end{aligned} \quad (24)$$

Then,

$$\det \left[ \widehat{H}^{\bar{1}}(x) \right] = \widehat{h}_{2,1}(x) \widehat{h}_{3,2}(x),$$

**Algorithm 3.1** The numerical algorithm for  $\mu$

**Require:**  $\widehat{H}$ ,  $A$ ,  $\tau > \epsilon > 0$ .  
 Calculate the eigenvalues  $\mathbb{E} := \{\lambda_i\}_{i=1}^G$  of  $\widehat{H}A^{-1}$ .  $i^* = \arg \min_{1 \leq i \leq G} |\lambda_i|$ .  $\lambda_{i^*} = 0$ .  
 $m(\widehat{H}|A) = \sum_{i=1}^G \delta_{\Re(\lambda_i) > 0}$  and  $n(\widehat{H}|A) = \sum_{i=1}^G \delta_{\Re(\lambda_i) < 0}$ .  
 if  $m(\widehat{H}|A) \geq m(A)$  then  
      $\mathbb{I} = \{i \in [1, G] | \Re(\lambda_i) > 0\}$ .  $b = \min_{i \in \mathbb{I}} |\Im(\lambda_i)|$ .  $\mu = \min_{i \in \mathbb{I}: |\Im(\lambda_i)| = b} \Re(\lambda_i)$ .  
 else if  $n(\widehat{H}|A) \geq n(A)$  then  
     Do symmetrically as the case when  $m(\widehat{H}|A) \geq m(A)$ .  
 else  $\mu = 0$ .  
 end if  
 $a = -\infty$ ,  $b = +\infty$ . flag = false.  
**repeat**  
   if  $|\mu| < 2\tau$  then  
     if  $r(\widehat{H} + 2\tau A) \leq 0$  and  $r(\widehat{H} - 2\tau A) \leq 0$  then return  $\mu$ .  
     else if  $r(\widehat{H} + 2\tau A) > 0$  and  $r(\widehat{H} - 2\tau A) \leq 0$  then  $b = \min(b, -2\tau + \epsilon)$ .  
     else if  $r(\widehat{H} + 2\tau A) \leq 0$  and  $r(\widehat{H} - 2\tau A) > 0$  then  $a = \max(a, 2\tau - \epsilon)$ .  
     else return "The problem is too singular to be solved".  
   end if  
   else if  $\mu \geq 2\tau$  then  
     if  $r(\widehat{H} - (\mu - \tau)A) \geq 0$  and  $r(\widehat{H} - (\mu + \tau)A) \leq 0$  then return  $\mu$ .  
     else if  $r(\widehat{H} - (\mu - \tau)A) < 0$  and  $r(\widehat{H} - (\mu + \tau)A) \leq 0$  then  
        $b = \min(b, \mu - \tau + \epsilon)$ .  
     else if  $r(\widehat{H} - (\mu - \tau)A) \geq 0$  and  $r(\widehat{H} - (\mu + \tau)A) > 0$  then  
        $a = \max(a, \mu + \tau - \epsilon)$ .  
     else return "The problem is too singular to be solved".  
   end if  
   else  
     Do symmetrically as the case of  $\mu \geq 2\tau$ .  
   end if  
   if  $b < a$  is empty then return "The problem is too singular to be solved".  
   else if  $b - a \leq 4\tau + 2\epsilon$  then return  $(a + b)/2$ .  
   end if  
   for  $\lambda \in \mathbb{E}$  do  
     if  $\Re(\lambda) \notin [a, b]$  or  $\Re(\lambda) = \mu$  then Remove  $\lambda$  from  $\mathbb{E}$ .  
   end if  
   if  $\mathbb{E}$  is not empty then Find  $\lambda \in \mathbb{E}$  minimizing  $|\Im(\lambda)|$ .  $\mu = \Re(\lambda)$ .  
   else flag = true.  
   end if  
**until** flag  
 if  $b = +\infty$  then  
    $K = \max(\tau, 1, a)$ .  
   while  $r(\widehat{H} - KA) > 0$  do  
      $a = \max(a, K - \epsilon)$ .  $K = 2K$ .  
   end while  
    $b = K + \epsilon$ .  
 else if  $a = -\infty$  then Do symmetrically.  
 end if  
 while  $b - a > 4\tau + 2\epsilon$  do  
    $\mu = (a + b)/2$ .  
   if  $|\mu| < \tau$  then  
     if  $\tau - \mu \leq \mu + \tau$  then  $\mu = \tau$ .  
     else  $\mu = -\tau$ .  
   end if  
   end if  
   if  $\mu < 0$  then  
     if  $r(\widehat{H} - \mu A) \leq 0$  then  $a = \mu - \epsilon$ .  
     else  $b = \mu + \epsilon$ .  
   end if  
   else Do symmetrically.  
   end if  
 end while  
 return  $(a + b)/2$ .

$$\begin{aligned} \det \left[ \widehat{H}^2(x) \right] &= \widehat{h}_{1,2}(x) \widehat{h}_{3,2}(x), \\ \det \left[ \widehat{H}^3(x) \right] &= \widehat{h}_{1,2}(x) \widehat{h}_{2,3}(x). \end{aligned} \tag{25}$$

Eq. (24) together with  $k_1 > k_2 > k_3$  describes a positive feedback regulation with proteins attaching to the gene as dimers.

Substitute Eqs. (24) and (25) into Eq. (23).

$$\widehat{h}_{2,1}^c \widehat{h}_{3,2}^c (k_1^0 - \gamma x) x^4 + \widehat{h}_{1,2}^c \widehat{h}_{3,2}^c (k_2^0 - \gamma x) x^2 + \widehat{h}_{1,2}^c \widehat{h}_{2,3}^c (k_3^0 - \gamma x) = 0. \tag{26}$$

Define  $u_1 := \widehat{h}_{1,2}^c / \widehat{h}_{2,1}^c$ ,  $u_2 := \widehat{h}_{2,3}^c / \widehat{h}_{3,2}^c$ , and divide Eq. (26) by  $\widehat{h}_{2,1}^c \widehat{h}_{3,2}^c$ .

$$f(x, u_1, u_2) := (k_1^0 - \gamma x) x^4 + u_1 (k_2^0 - \gamma x) x^2 + u_1 u_2 (k_3^0 - \gamma x) = 0.$$

We plot a curve

$$\begin{cases} f(x, u_1, u_2) = 0, \\ \partial_x f(x, u_1, u_2) = 0, \end{cases} \tag{27}$$

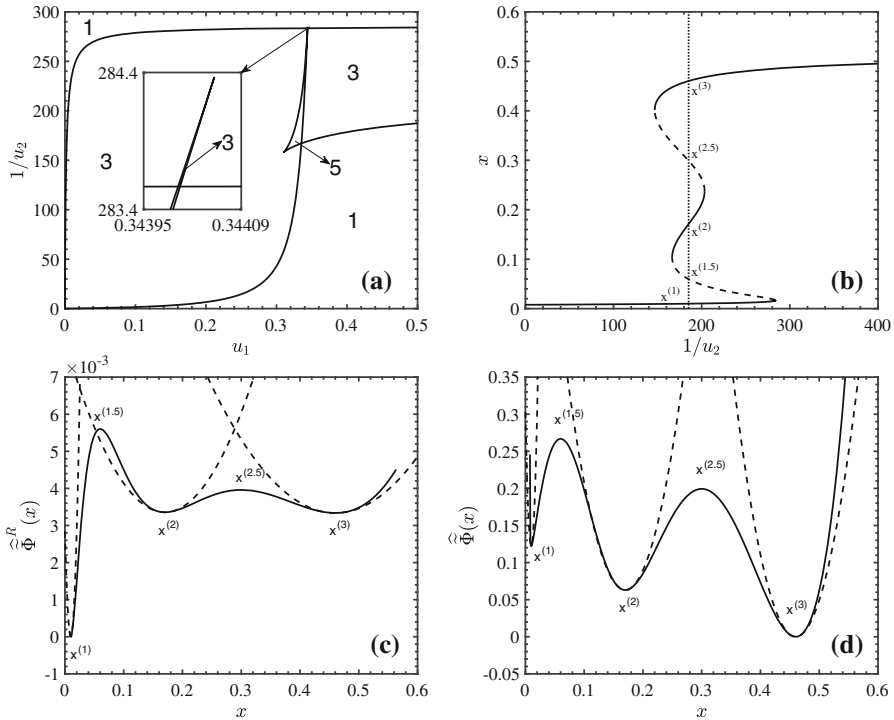
in the phase plane of  $u_1$  and  $1/u_2$  in Fig. 2a for  $k_2^0/k_1^0 = 0.1283$  and  $k_3^0/k_1^0 = 0.0078$  (remind that  $k_1^0 \equiv \gamma$ ).

The root number in each region divided by the curve is labeled in Fig. 2a. In one region,  $f(x, u_1, u_2) = 0$  has five roots. Fixing  $u_1 = 0.3329$ , we plot  $1/u_2$  as a function of  $x$  by  $f(x, u_1, u_2) = 0$  in Fig. 2b. As  $u_2 = 0.0054$  (vertical dotted line),  $f(x, u_1, u_2) = 0$  has five roots  $x^{(1)}, x^{(1.5)}, x^{(2)}, x^{(2.5)}, x^{(3)}$ . These roots are the fixed points of the DMFD (an ODE). Three stable fixed points  $x^{(1)}, x^{(2)}$  and  $x^{(3)}$  are interlaced by two unstable ones  $x^{(1.5)}$  and  $x^{(2.5)}$ , which indicates that the FR model is tristable.

### 4.2 FR model exhibits different local fluctuations from those of the reduced CME model

In our tristable example,  $\partial_x \widehat{\Phi}(x^{(i)}) = \partial_x \widehat{\Phi}^R(x^{(i)}) = 0$  at the stable fixed point  $x^{(i)}$  of the DMFD by Theorem 2 and Eq. (16). Therefore, around  $x^{(i)}$ , we have the Gaussian approximations

$$\begin{aligned} \widehat{\Phi}(x) &\approx \widehat{\Phi}(x^{(i)}) + \frac{\partial_x^2 \widehat{\Phi}(x^{(i)})}{2} (x - x^{(i)})^2, \\ \widehat{\Phi}^R(x) &\approx \widehat{\Phi}^R(x^{(i)}) + \frac{\partial_x^2 \widehat{\Phi}^R(x^{(i)})}{2} (x - x^{(i)})^2. \end{aligned}$$



**Fig. 2** **a** Curve in the phase plane of  $u_1$  and  $1/u_2$  by Eq. (27) with  $G = 3, k_2^0/k_1^0 = 0.1283, k_3^0/k_1^0 = 0.0078$ . As a function of  $x$ , the number of roots of  $f(x, u_1, u_2) = 0$  changes with  $u_1$  and  $u_2$ . In each region divided by the curve, we label the number of roots of  $f(x, u_1, u_2) = 0$  (i.e. the fixed points of the DMFD). **b** Phase diagram with the same parameters as in **a** and  $u_1 = 0.3329$ . It shows the fixed points of the DMFD vary with  $u_2$ . Dash lines are the unstable fixed points indicated by  $x^{(1.5)}$  and  $x^{(2.5)}$ , while real lines are those stable ones indicated by  $x^{(1)}, x^{(2)}$  and  $x^{(3)}$ . **c** The rate function  $\widehat{\Phi}^R(x)$  of the reduced CME model, and the Gaussian approximations around  $x^{(i)}$  with  $\partial_x^2 \widehat{\Phi}^R(x^{(1)}) : \partial_x^2 \widehat{\Phi}^R(x^{(2)}) : \partial_x^2 \widehat{\Phi}^R(x^{(3)}) = 1 : 0.0056 : 0.0027$ . **d** The rate function  $\widehat{\Phi}(x)$  of the FR model, and the Gaussian approximations around  $x^{(i)}$  with  $\partial_x^2 \widehat{\Phi}(x^{(1)}) : \partial_x^2 \widehat{\Phi}(x^{(2)}) : \partial_x^2 \widehat{\Phi}(x^{(3)}) = 1 : 0.0163 : 0.0170$ . In **c, d**,  $G = 3, k_2^0/k_1^0 = 0.1283, k_3^0/k_1^0 = 0.0078, u_1 = 0.3329, u_2 = 0.0054$ . In **d**,  $\widehat{h}_{3,2}^c/\widehat{h}_{2,1}^c = 1$  and  $k_1^0 \equiv \gamma = 0.01$

The local fluctuations around the fixed point  $x^{(i)}$  for the FR model and the reduced CME model are indicated by  $\partial_x^2 \widehat{\Phi}(x^{(i)})$  and  $\partial_x^2 \widehat{\Phi}^R(x^{(i)})$ , respectively. The larger the second derivative is, the weaker the local fluctuation will be, illustrated by the ratio  $\partial_x^2 \widehat{\Phi}(x^{(1)}) : \partial_x^2 \widehat{\Phi}(x^{(2)}) : \partial_x^2 \widehat{\Phi}(x^{(3)})$  (or  $\partial_x^2 \widehat{\Phi}^R(x^{(1)}) : \partial_x^2 \widehat{\Phi}^R(x^{(2)}) : \partial_x^2 \widehat{\Phi}^R(x^{(3)})$ ) of the local fluctuations over all fixed points.

In Fig. 2c, we plot  $\widehat{\Phi}^R(x)$  of the tristable example in Sect. 4.1. Further fixing  $\widehat{h}_{3,2}^c/\widehat{h}_{2,1}^c = 1$  and  $k_1^0 \equiv \gamma = 0.01$ , we plot  $\widehat{\Phi}(x)$  of the same example in Fig. 2d. The Gaussian approximations Eq. (28) around  $x^{(i)}$  are plotted in dash line in Fig. 2c, d. The local fluctuations of  $\widehat{\Phi}(x)$  are different from those of  $\widehat{\Phi}^R(x)$  even if they share

the same DMFD. Actually, we have

$$\begin{aligned} \partial_x^2 \widehat{\Phi}^R(x^{(1)}) : \partial_x^2 \widehat{\Phi}^R(x^{(2)}) : \partial_x^2 \widehat{\Phi}^R(x^{(3)}) &= 1 : 0.0056 : 0.0027, \\ \partial_x^2 \widehat{\Phi}(x^{(1)}) : \partial_x^2 \widehat{\Phi}(x^{(2)}) : \partial_x^2 \widehat{\Phi}(x^{(3)}) &= 1 : 0.0163 : 0.0170. \end{aligned}$$

Thus, the local fluctuations at  $x^{(2)}$  and  $x^{(3)}$  are stronger in the rapid regime (the reduced CME model) of gene-state switchings than those in the intermediate one (the FR model).

### 4.3 Transition rates between phenotypic states

The mean first-passage time  $T_{ij}^c$  from  $x^{(i)}$  to  $x^{(j)}$  of the full CME model is numerically calculated by the conventional first-passage theory (Redner 2007). In Fig. 3,  $T_{ij}^c$  is plotted logarithmically as functions of  $\mathcal{H}$  in  $\circ, \times, +$  for different  $n_{\max}$ .

$T_{ij}^c$  is insensitive to  $n_{\max}$  in the intermediate region  $n_{\max}k_i \gg \mathcal{H}h_{i,j}(x) \gg \gamma$  of gene-state switchings. The FR model is simulated by the Doob–Gillespie method to obtain the mean first-passage time  $T_{ij}^f$  (Gillespie 1976, 1977). In Fig. 3,  $T_{ij}^f$  (real line) agrees well with  $T_{ij}^c$  for different  $n_{\max}$  ( $\circ, \times, +$ ).

Refer to the discussion on the links between large deviations and WKB for the PDMP (Bressloff and Faugeras 2017), the transition rate  $k_{ij}$  from  $x^{(i)}$  to  $x^{(j)}$  is approximated by the general Arrhenius form

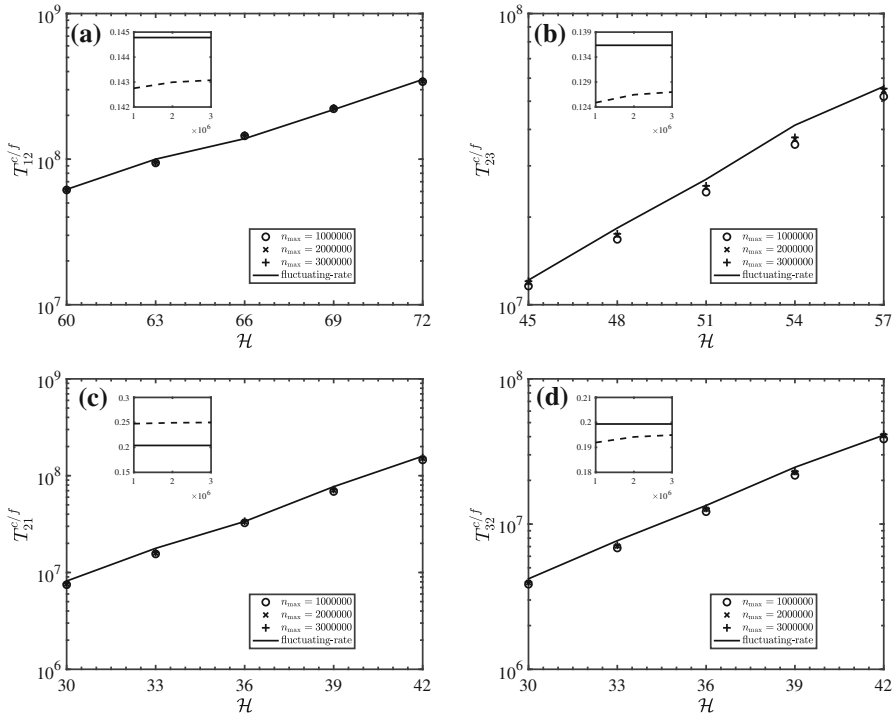
$$k_{ij} \approx k_{ij}^0 \exp(-\mathcal{H}\Delta\widehat{\Phi}_{ij}), \quad (28)$$

where  $\Delta\widehat{\Phi}_{ij} := \widehat{\Phi}\left[x^{\left(\frac{i+j}{2}\right)}\right] - \widehat{\Phi}[x^{(i)}]$  is the barrier height from  $x^{(i)}$  to  $x^{(j)}$ , and  $k_{ij}^0$  is a prefactor such that  $\lim_{\mathcal{H} \rightarrow \infty} -\frac{1}{\mathcal{H}} \log k_{ij}^0 = 0$ . Substitute  $k_{ij} = 1/T_{ij}^c$  into Eq. (28).

$$\log(T_{ij}^c) \approx \log(1/k_{ij}^0) + \mathcal{H}\Delta\widehat{\Phi}_{ij}. \quad (29)$$

The linear relationship between  $\log(T_{ij}^c)$  and  $\mathcal{H}$  in Eq. (29) is observed in Fig. 3, by which we fit and plot the slope  $\Delta\widehat{\Phi}_{ij}$  for different  $n_{\max}$  in the insets of Fig. 3 by dash lines. The real lines, which are calculated directly by  $\Delta\widehat{\Phi}_{ij} := \widehat{\Phi}\left[x^{\left(\frac{i+j}{2}\right)}\right] - \widehat{\Phi}[x^{(i)}]$ , match the dash lines. In conclusion, the LDP of the FR model correctly predicts the transition rates between phenotypic states in the intermediate region of gene-state switchings.





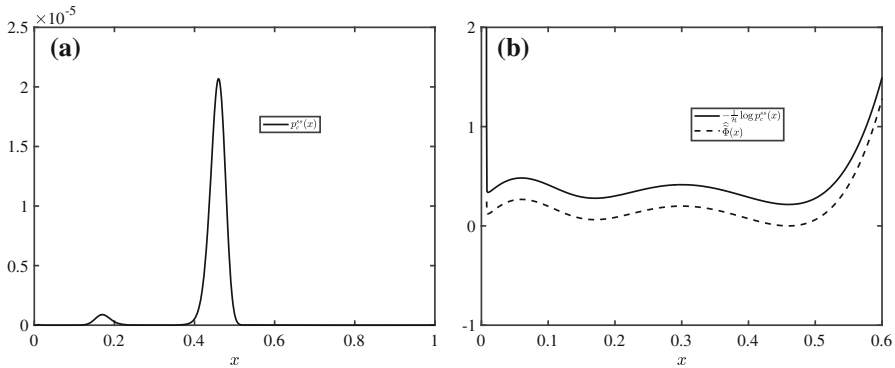
**Fig. 3** Logarithmic mean first-passage time of the full CME model as functions of  $\mathcal{H}$  for different  $n_{\max}$  ( $\circ$ ,  $\times$ ,  $+$ ), and that of the FR model (real lines).  $k_1^0 \equiv \gamma = 0.01$ ,  $u_1 = 0.3329$ ,  $u_2 = 0.0054$ ,  $\hat{h}_{3,2}^c/\hat{h}_{2,1}^c = 1$ ,  $k_1^0 : k_2^0 : k_3^0 = 1 : 0.1283 : 0.0078$ . The insets show the barrier heights  $\Delta\hat{\Phi}_{ij}$  calculated by fitting Eq. (29) for different  $n_{\max}$  (dash lines), and those directly calculated by  $\Delta\hat{\Phi}_{ij} = \hat{\Phi}\left[x\left(\frac{i+j}{2}\right)\right] - \hat{\Phi}\left[x^{(i)}\right]$  (real lines)

**4.4 Comparison with the stationary distribution of the full CME model**

By the same parameters as in Fig. 2d, we plot the stationary distribution  $p_c^{ss}(x)$  ( $x = n/n_{\max}$ ) of the full CME model for  $n_{\max} = 1000000$  and  $\mathcal{H} = 50$  in Fig. 4a. As expected,  $-\frac{1}{\mathcal{H}} \log p_c^{ss}(x)$  (real line) and  $\hat{\Phi}(x)$  (dash line) are quite close (Fig. 4b).

**5 Conclusions and remarks**

A living cell usually has multiple phenotypic states to face fluctuating environments (Kussell and Leibler 2005; Acar et al. 2008). The FR model proposed in Ge et al. (2015) for single-cell dynamics with two gene states quantitatively describes the stabilities of the phenotypic states and the transition rates among them in the intermediate regime, which is the case in *E. coli*, but rarely studied before. Considering that genes often have more than two states due to the combinatorial nature of transcriptional regulations (Zhu



**Fig. 4** Parameters are the same as in Fig. 2d. **a** Stationary distribution  $p_c^{SS}(x)$  ( $x = n/n_{\max}$ ) of the full CME model for  $n_{\max} = 1,000,000$  and  $\mathcal{H} = 50$ . **b**  $-\frac{1}{\mathcal{H}} \log p_c^{SS}(x)$  (real line) and  $\widehat{\Phi}(x)$  obtained by our numerical algorithm (dash line)

et al. 2005), we generalise the FR model together with the LDP rate function of steady state to the scenario with more than two gene states. We found that the derivative of the rate function is just the unique nontrivial DGE of the negative gene-state transition-rate matrix on the protein birth–death matrix.

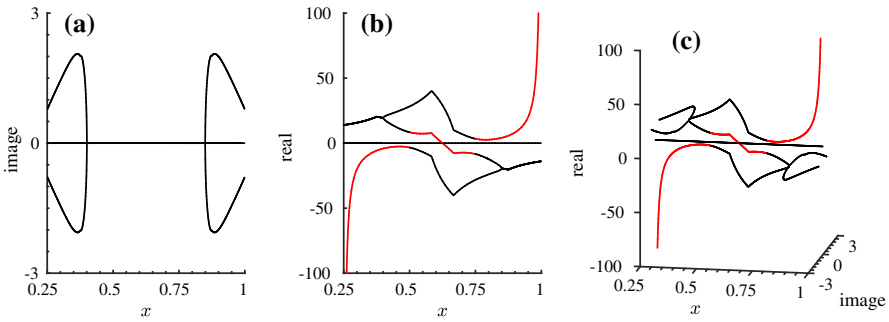
Given that several GEs exist, a method that determines the rate function from GEs is necessary. Under appropriate assumptions, we prove that the nontrivial DGE is continuous, the derivative of the rate function equals the nontrivial DGE, and the rate function satisfies the Lyapunov property of the DMFD. The numerical results of a tristable example support the equivalence of the nontrivial DGE and the LDP rate function.

The result in Corollary 1 is interesting because under Assumption 3 the matrix  $\widehat{H}(x)A^{-1}(x)$  is discontinuous at each  $x_i^*$  but always has a continuous nontrivial real eigenvalue  $\mu(x)$  in the whole interval  $(x_G^*, x_1^*)$ . On the contrary, Example 1 will show that even if  $\widehat{H}(x)A^{-1}(x)$  is continuous in  $(x_G^*, x_1^*)$ , it may have no continuous nontrivial real eigenvalue on the whole interval.

**Example 1** Let  $x_i^* = (5 - i) / 4$  for  $i \in [1, 4]$ ,  $\gamma = 1$ ,

$$J := \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad F := \begin{pmatrix} -11 & 8 & 0 & 8 \\ 4 & -16 & 10 & 1 \\ 5 & 4 & -17 & 0 \\ 2 & 4 & 7 & -9 \end{pmatrix},$$

$$E_2(x) := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 64(x - 3/4)^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E_3(x) := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 64(x - 1/2)^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$



**Fig. 5** **a–c** GEs of  $\widehat{H}(x)$  on  $A(x)$ , or equivalently, eigenvalues of  $\widehat{H}(x)A^{-1}(x)$  in Example 1 for  $x \in (x_4^*, x_1^*)$ . **a, b** Real and imaginary parts of GEs. **c** Corresponding 3D plot

$$\widehat{H}(x) = \begin{cases} F|A|(3/8), & x \in (0, 3/8], \\ F|A|(3/8)E_3(x), & x \in (3/8, 7/12], \\ 12(2/3 - x)F|A|(3/8)E_3(7/12) \\ + 12(x - 7/12)JFJ|A|(7/8)E_2(2/3), & x \in (7/12, 2/3], \\ JFJ|A|(7/8)E_2(x), & x \in (2/3, 7/8], \\ JFJ|A|(7/8), & x \in (7/8, +\infty). \end{cases}$$

It is not hard to see that the limit of  $\widehat{H}(x)A^{-1}(x)$  at each  $x_i^*$  is finite, hence the matrix  $\widehat{H}(x)A^{-1}(x)$  is continuous in the whole  $(x_4^*, x_1^*)$ . The eigenvalues of  $\widehat{H}(x)A^{-1}(x)$  given in Fig. 5a–c are also continuous in  $(x_4^*, x_1^*)$  as expected, but none of them is always real.

The realness of the nontrivial DGE is automatically promised by Definition A1, whereas nondominant GEs may be complex. The nontrivial DGE  $\mu(x)$  of  $\widehat{H}(x)$  on  $A(x)$  indicated by red lines in Fig. 5b, c, which should be continuous if Assumption 3 holds, is now discontinuous at  $x_3^*$  and  $x_2^*$  because  $\widehat{H}(x)$  becomes reducible.

In this paper, we have assumed a linear  $A(x)$  in Eq. (10). However, this is not necessary. Generally, denote the  $i$ th diagonal element of  $A(x)$  by  $A_i(x)$ . If no two diagonal elements  $A_i(x)$  vanish at a same  $x$ , then simply replacing  $\gamma(x_i^* - x)$  by  $A_i(x)$ , the proof of Theorem 1 is still valid. After slight modification, other main results in this paper are kept.

For now, we only considered the one-dimensional case (single self-regulatory gene). For high dimensional cases, Eq. (12) becomes  $\det[\widehat{H}(x) - \mathbb{F}(x)] = 0$ , where

$$\mathbb{F}(x) := \begin{pmatrix} \nabla\widehat{\Phi}(x) \cdot \mathcal{F}_1(x) & 0 & \cdots & 0 \\ 0 & \nabla\widehat{\Phi}(x) \cdot \mathcal{F}_2(x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \nabla\widehat{\Phi}(x) \cdot \mathcal{F}_G(x) \end{pmatrix}, \quad (30)$$

and  $\mathcal{F}_i(x)$  is the vector field at gene state  $i$ . To apply the methods in this paper to high dimensional cases, many concepts must be generalised since the derivative  $\nabla\widehat{\Phi}(x)$

is now a vector but not a scalar. The most crucial step should be to properly regard the difference between  $\nabla \widehat{\Phi}(\mathbf{x})$  and other solutions of the underdetermined system  $\det [\widehat{H}(\mathbf{x}) - \mathbb{F}(\mathbf{x})] = 0$ . In one-dimensional cases, it is the unique nontrivial DGE.

**Acknowledgements** This work was supported by Natural Science Foundation of China, Grant No. 11971037 and 11622101 (Hao Ge), and Natural Science Foundation of China, Grant No. 11271083 (Yunxin Zhang).

## Appendix A: Results related to dominant generalised eigenvalue problem

- $G > 0$ : matrix dimension/size.
- $\mathcal{G} := \{1, 2, 3, \dots, G\}$ .
- $B = (b_{i,j})_{G \times G}$ : general  $G \times G$  matrix.
- $B^{\mathcal{S}_1, \mathcal{S}_2}$  for  $\mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{G}$ : principal submatrix of  $B$  formed by rows in  $\mathcal{S}_1$  and columns in  $\mathcal{S}_2$ . Particularly,  $B^{\mathcal{S}, \mathcal{S}} := B^{\mathcal{S}, \mathcal{S}}$  for  $\mathcal{S} \subset \mathcal{G}$ .
- $\bar{i} := \mathcal{G} \setminus \{i\}$ .
- $|b_{\cdot, j}| := \sum_{i \neq j} |b_{i, j}|$ .
- $Z$ :  $G \times G$   $Z$ -matrix (real matrix such that off-diagonal elements are nonpositive) in the normal form (block upper-triangular form with each diagonal block irreducible).
- $\mathcal{I}$ : indices of irreducible diagonal blocks of  $Z$ .
- $Z_i$  for  $i \in \mathcal{I}$ : irreducible diagonal blocks of  $Z$ .
- $M$ :  $G \times G$   $M$ -matrix ( $Z$ -matrix such that all eigenvalues have nonnegative real parts) in the normal form. Note that we allow singular  $M$ -matrix.
- $D$ :  $G \times G$  real diagonal matrix.
- $D_i$  for  $i \in \mathcal{I}$ : diagonal blocks of  $D$  corresponding to  $Z_i$ .
- $d_i$  for  $i \in \mathcal{G}$ : the  $i$ th diagonal element of  $D$ .
- $\mathcal{G}_0 := \{i \in \mathcal{G} | d_i = 0\}$ .  $\mathcal{G}_+ := \{i \in \mathcal{G} | d_i > 0\}$ .  $\mathcal{G}_- := \{i \in \mathcal{G} | d_i < 0\}$ .
- $r(Z)$ : dominant eigenvalue (the eigenvalue with smallest real part, which must be real by the Perron–Frobenius theorem) of  $Z$ .
- $m(D)$  and  $n(D)$ : the number of positive and negative diagonal elements of  $D$ .  $D \neq 0$  means  $m(D) + n(D) > 0$ .
- $O(o, \rho) := \{z \in \mathbb{C} | |z - o| \leq \rho\}$ , where  $o \in \mathbb{C}$  and  $\rho \geq 0$ .

In this “Appendix”, we state results without proofs. Their proofs are given in “Appendix B”. We always assume  $D \neq 0$ .

**Definition A1** Denote the real part of  $\lambda \in \mathbb{C}$  by  $\Re(\lambda)$ . Define

$$\begin{aligned} \Lambda(B|D) &:= \{\lambda \in \mathbb{C} | \det(B - \lambda D) = 0\}, \\ \Lambda_+(B|D) &:= \{\lambda \in \Lambda(B|D) | \Re(\lambda) > 0\}, \\ \Lambda_-(B|D) &:= \{\lambda \in \Lambda(B|D) | \Re(\lambda) < 0\}, \\ \Lambda_0(B|D) &:= \{\lambda \in \Lambda(B|D) | \Re(\lambda) = 0\}. \end{aligned}$$

$\lambda \in \Lambda(B|D)$  is called the generalised eigenvalue (GE) of  $B$  on  $D$  (Chu 1987; Ikramov 1993). For  $\lambda_0 \in \Lambda(B|D)$ , let  $c(\lambda_0|B|D)$  be the multiplicity of  $\lambda_0$  as a root of  $\det(B - \lambda D) = 0$ . We define  $R(Z|D) := \{\lambda \in \mathbb{R} | r(Z - \lambda D) = 0\}$ , and call  $\lambda \in R(Z|D)$  the dominant generalised eigenvalue (DGE) of  $Z$  on  $D$ . Define  $\mathcal{E}(Z - \lambda D) := \det(Z - \lambda D) / r(Z - \lambda D)$ , which is analytic even for  $\lambda \in R(Z|D)$  because they are removable singular points.

**Lemma A1** *Assume that  $M$  is irreducible.*

1. *If  $n(D) = 0$ , then  $R(M|D) = \{\mu_+\}$  with  $\mu_+ \geq 0$  and  $c(\mu_+|M|D) = 1$ .  $r(M - \lambda D) > 0$  for  $\lambda < \mu_+$  and  $r(M - \lambda D) < 0$  for  $\lambda > \mu_+$ .*
2. *If  $m(D) = 0$ , then  $R(M|D) = \{\mu_-\}$  with  $\mu_- \leq 0$  and  $c(\mu_-|M|D) = 1$ .  $r(M - \lambda D) > 0$  for  $\lambda > \mu_-$  and  $r(M - \lambda D) < 0$  for  $\lambda < \mu_-$ .*
3. *Otherwise,  $R(M|D) = \{\mu_-, \mu_+\}$  with  $\mu_+ \geq 0 \geq \mu_-$ .  $c(\mu_\pm|M|D) = 1$  unless  $\mu_\pm = 0$ , at which  $c(\mu_\pm|M|D) = c(0|M|D) = 2$ .  $r(M - \lambda D) > 0$  for  $\mu_- < \lambda < \mu_+$  and  $r(M - \lambda D) < 0$  for  $\lambda < \mu_-$  or  $\lambda > \mu_+$ .*

Lemma A1 justifies Definition A2.

**Definition A2** For irreducible  $M$ , define  $\mu_+(M|D) := \max_{\lambda \in R(M|D)} \lambda$  for  $m(D) > 0$ , and  $\mu_-(M|D) := \min_{\lambda \in R(M|D)} \lambda$  for  $n(D) > 0$ .

Definition A2 can be explained as follows. By Lemma A1, for  $m(D) > 0$  and  $n(D) = 0$ , there is only a nonnegative DGE for  $M$  on  $D$ , which is denoted by  $\mu_+(M|D)$ . Similarly, for  $m(D) = 0$  and  $n(D) > 0$ , there is only a nonpositive DGE, which is denoted by  $\mu_-(M|D)$ . If  $m(D) > 0$  and  $n(D) > 0$ , then there are both a nonnegative DGE and a nonpositive DGE, which are respectively denoted by  $\mu_+(M|D)$  and  $\mu_-(M|D)$ .

Since  $r(M - \lambda D) = \min_{i \in \mathcal{G}} r(M_i - \lambda D_i)$ , the results in Lemma A1 can be generalised to reducible  $M$  (see Corollary A1) if for any irreducible diagonal block  $i \in \mathcal{I}$  such that  $D_i = 0$ ,  $M_i$  is nonsingular. Lemma A2 provides a sufficient condition for this.

**Lemma A2**  $\mathcal{K}_0^{\text{sin}} := \{i \in \mathcal{I} | m(D_i) = n(D_i) = 0, \det(M_i) = 0\} = \emptyset$  if and only if  $\det(M^{G_0}) \neq 0$ .

**Corollary A1** *Assume that  $\det(M^{G_0}) \neq 0$ .*

1. *If  $n(D) = 0$ , then  $R(M|D) = \{\mu_+\}$  with*

$$\mu_+ = \mu_+(M|D) := \min_{i:m(D_i)>0} \mu_+(M_i|D_i),$$

*$r(M - \lambda D) < 0$  for  $\lambda > \mu_+$ , and  $r(M - \lambda D) > 0$  for  $\lambda < \mu_+$ .*

2. *If  $m(D) = 0$ , then  $R(M|D) = \{\mu_-\}$  with*

$$\mu_- = \mu_-(M|D) := \max_{i:n(D_i)>0} \mu_-(M_i|D_i),$$

*$r(M - \lambda D) < 0$  for  $\lambda < \mu_-$ , and  $r(M - \lambda D) > 0$  for  $\lambda > \mu_-$ .*

3. Otherwise,  $R(M|D) = \{\mu_-, \mu_+\}$  with

$$\begin{aligned} \mu_+ &= \mu_+(M|D) := \min_{i:m(D_i)>0} \mu_+(M_i|D_i), \\ \mu_- &= \mu_-(M|D) := \max_{i:n(D_i)>0} \mu_-(M_i|D_i), \end{aligned}$$

$$r(M - \lambda D) < 0 \text{ for } \lambda < \mu_- \text{ or } \lambda > \mu_+, \text{ and } r(M - \lambda D) > 0 \text{ for } \mu_- < \lambda < \mu_+.$$

Corollary A1 extends Definition A2 to reducible  $M$ , and gives Lemma A3.

**Lemma A3** Assume that  $M(\theta)$  is a continuous  $G \times G$   $M$ -matrix function of  $\theta$ ,  $D(\theta)$  is a continuous  $G \times G$  diagonal matrix function of  $\theta$  with constant  $\mathcal{G}_\pm$  and  $\mathcal{G}_0 \forall \theta$ , and  $\det[M^{\mathcal{G}_0}(\theta)] \neq 0 \forall \theta$ .

1. If  $m[D(\theta)] > 0$ , then  $\mu_+(\theta) := \mu_+(M|D)(\theta)$  is continuous.
2. If  $n[D(\theta)] > 0$ , then  $\mu_-(\theta) := \mu_-(M|D)(\theta)$  is continuous.

**Lemma A4** 1. Assume that  $|b_{j,j}| \geq |b_{\cdot,j}| \forall j \in \mathcal{G}_0$ ,  $B^{\mathcal{G}_0}$  is irreducible, and  $\exists j_0 \in \mathcal{G}_0$  such that  $|b_{j_0,j_0}| > \sum_{i \in \mathcal{G}_0 \setminus \{j_0\}} |b_{i,j_0}|$ . Then

$$\bigcup_{j \in \mathcal{G} \setminus \mathcal{G}_0} O(b_{j,j}/d_j, |b_{\cdot,j}|/|d_j|) =: \mathcal{O}(B|D) \supset \Lambda(B|D).$$

2. Under the assumptions in statement 1, further assume that  $b_{j,j} \geq 0$  for  $j \notin \mathcal{G}_0$  and  $B$  is diagonally dominant for all columns. Then

$$\sum_{\lambda \in [\Lambda \cap \mathcal{O}_+](B|D)} c(\lambda|B|D) \geq m(D), \quad \sum_{\lambda \in [\Lambda \cap \mathcal{O}_-](B|D)} c(\lambda|B|D) \geq n(D),$$

where  $\mathcal{O}_\pm(B|D) := \bigcup_{j \in \mathcal{G}_\pm} O(b_{j,j}/d_j, |b_{\cdot,j}|/|d_j|)$ .

Lemma A4 generalises the Gershgorin circle theorem to GEs.

**Definition A3** Let  $T = (t_{i,j})_{G \times G}$  be a  $Z$ -matrix with  $\sum_{i \in \mathcal{G}} t_{i,j} = 0 \forall j \in \mathcal{G}$ .  $T$  is called a negative transition rate matrix. Let  $m(T|D) := \sum_{\lambda \in \Lambda_+(T|D)} c(\lambda|T|D)$  and  $n(T|D) := \sum_{\lambda \in \Lambda_-(T|D)} c(\lambda|T|D)$ .

Lemmas A5 and A6, and Theorem A1 connect the signs of the real parts of GEs with the signatures of  $D$  for negative transition rate matrix  $T$  step by step. Since  $T$  is diagonally dominant, it is not hard to prove by the Gershgorin circle theorem that  $T$  is a singular  $M$ -matrix. Applying Lemma A4 to  $T$ , we have Lemma A5.

**Lemma A5** Assume that  $T^{\mathcal{G}_0}$  is irreducible, and  $\exists j_0 \in \mathcal{G}_0$  such that  $|t_{j_0,j_0}| > \sum_{i \in \mathcal{G}_0 \setminus \{j_0\}} |t_{i,j_0}|$ . Then  $\Lambda_0(T|D) \subset \{0\}$ , and exact one of the following happens.

1.  $m(T|D) = m(D)$ .
2.  $n(T|D) = n(D)$ .
3.  $m(T|D) < m(D)$ ,  $n(T|D) < n(D)$ .

If  $T$  is irreducible, Lemma A5 becomes Lemma A6 by Lemma A1.

**Lemma A6** Assume that  $T$  is irreducible,  $T^{\mathcal{G}_0}$  is irreducible, and  $\exists j_0 \in \mathcal{G}_0$  such that  $|t_{j_0, j_0}| > \sum_{i \in \mathcal{G}_0 \setminus \{j_0\}} |t_{i, j_0}|$ . Then exact one of the following happens.

1.  $m(T|D) = m(D)$ ,  $n(T|D) = n(D) - 1$ ,  $\mu_-(T|D) = 0$ ,  $c(\mu_-|T|D) = 1$ . If  $m(D) > 0$ ,  $\mu_+(T|D) > 0$  and  $c(\mu_+|T|D) = 1$ .
2.  $m(T|D) = m(D) - 1$ ,  $n(T|D) = n(D)$ ,  $\mu_+(T|D) = 0$ ,  $c(\mu_+|T|D) = 1$ . If  $n(D) > 0$ ,  $\mu_-(T|D) < 0$  and  $c(\mu_-|T|D) = 1$ .
3.  $m(T|D) = m(D) - 1$ ,  $n(T|D) = n(D) - 1$ ,  $\mu_{\pm}(T|D) = 0$ ,  $c(\mu_{\pm}|T|D) = 1$ .

Theorem A1 generalises Lemma A6 to reducible  $T$ , and concretizes Lemma A5.

**Theorem A1** Assume that  $T^{\mathcal{G}_0}$  is irreducible, and  $\exists j_0 \in \mathcal{G}_0$  such that  $|t_{j_0, j_0}| > \sum_{i \in \mathcal{G}_0 \setminus \{j_0\}} |t_{i, j_0}|$ . Then exact one of the following happens.

1.  $m(T|D) = m(D)$ ,  $\mu_-(T|D) = 0$ . If  $m(D) > 0$ ,  $\mu_+(T|D) > 0$ .
2.  $n(T|D) = n(D)$ ,  $\mu_+(T|D) = 0$ . If  $n(D) > 0$ ,  $\mu_-(T|D) < 0$ .
3.  $m(T|D) < m(D)$ ,  $n(T|D) < n(D)$ ,  $\mu_{\pm}(T|D) = 0$ .

**Definition A4**  $T^{\mathcal{G}_0}$  is irreducible.  $\exists j_0 \in \mathcal{G}_0$  such that  $|t_{j_0, j_0}| > \sum_{i \in \mathcal{G}_0 \setminus \{j_0\}} |t_{i, j_0}|$ . For  $m(D), n(D) > 0$ , define

$$\mu(T|D) := \begin{cases} \mu_+(T|D), & \mu_+(T|D) > 0, \\ \mu_-(T|D), & \mu_+(T|D) = 0, \end{cases}$$

as the nontrivial DGE of  $T$  on  $D$ .

By Theorem A1, at least one of  $\mu_{\pm}(T|D)$  is zero. Therefore, the nontrivial DGE in Definition A4 is obtained by always choosing the nonzero DGE if possible.

**Theorem A2** Assume that  $T = T(\theta)$  is a continuous  $G \times G$  negative transition rate matrix function of  $\theta$ ,  $D(\theta)$  is a continuous  $G \times G$  diagonal matrix function of  $\theta$  with constant  $\mathcal{G}_{\pm}$  and  $\mathcal{G}_0 \forall \theta$ , and  $\det [T^{\mathcal{G}_0}(\theta)] \neq 0 \forall \theta$ . Then  $\mu(\theta) := \mu(T|D)(\theta)$  is continuous.

Theorem A2 proves the continuity of  $\mu(T|D)$  by Lemma A3.

**Lemma A7** For  $\epsilon > 0$  and  $j_0 \in \mathcal{G}$ , let  $Z^{\epsilon} := (z_{i,j}^{\epsilon})_{G \times G}$  with  $z_{i,j_0}^{\epsilon} := \epsilon z_{i,j_0}$  and  $z_{i,j}^{\epsilon} := z_{i,j}$  for  $j \neq j_0$ . Then  $\text{sgn}[r(Z^{\epsilon})] \equiv \text{sgn}[r(Z)]$ , where  $\text{sgn}$  is the sign function.

**Lemma A8** Assume that  $T$  is irreducible. Then  $v = (v_j)_{G \times 1} \neq 0$  satisfies  $Tv = 0$  iff

$$\frac{v_1}{\det(T^{\bar{1}})} = \frac{v_2}{\det(T^{\bar{2}})} = \frac{v_3}{\det(T^{\bar{3}})} = \dots = \frac{v_G}{\det(T^{\bar{G}})}.$$

## Appendix B: Proofs of results in "Appendix A"

Proof of Lemma A1.

**Proof** Case 1: Since  $r(M - \lambda D)$  has the smallest real part among all eigenvalues of  $M - \lambda D$ , it is not hard to show by the Gershgorin circle theorem that  $\lim_{\lambda \rightarrow +\infty} r(M - \lambda D) = -\infty$ . By  $r(M) \geq 0, \exists \mu_+ \geq 0$  such that  $r(M - \mu_+ D) = 0$ . Because  $D > 0, \lambda < \mu_+ \Rightarrow M - \lambda D > M - \mu_+ D \Rightarrow r(M - \lambda D) > r(M - \mu_+ D) = 0$ .  $\frac{dr(M - \lambda D)}{d\lambda}|_{\lambda=\mu_+} < 0$  since otherwise, by  $\frac{d^2r(M - \lambda D)}{d\lambda^2} < 0$  (Deutsch and Neumann 1984),  $r(M - \lambda D) < 0$  for  $\lambda < \mu_+$ , conflicts. By  $\frac{d^2r(M - \lambda D)}{d\lambda^2} < 0, r(M - \lambda D) < 0$  for  $\lambda > \mu_+$ .  $\Xi(M - \mu_+ D) > 0$  because it is the product of nondominant eigenvalues of irreducible  $M$ -matrix  $M - \mu_+ D$ . Therefore,

$$\begin{aligned} \frac{d \det(M - \lambda D)}{d\lambda} \Big|_{\lambda=\mu_+} &= \frac{dr(M - \lambda D) \Xi(M - \lambda D)}{d\lambda} \Big|_{\lambda=\mu_+} \\ &= \frac{dr(M - \lambda D)}{d\lambda} \Xi(M - \lambda D) \Big|_{\lambda=\mu_+} \\ &\quad + r(M - \lambda D) \frac{d\Xi(M - \lambda D)}{d\lambda} \Big|_{\lambda=\mu_+} \\ &= \frac{dr(M - \lambda D)}{d\lambda} \Big|_{\lambda=\mu_+} \Xi(M - \mu_+ D) < 0, \quad (\text{A1}) \end{aligned}$$

i.e.  $c(\mu_+ | M | D) = 1$ .

Case 2: Similar as case 1.

Case 3: Resembling case 1,  $\lim_{\lambda \rightarrow \pm\infty} r(M - \lambda D) = -\infty$ . If  $r(M) > 0, \exists \mu_+ > 0 > \mu_-$  such that  $r(M - \mu_{\pm} D) = 0$ . Because  $\frac{d^2r(M - \lambda D)}{d\lambda^2} < 0$ , we have

$$\frac{dr(M - \lambda D)}{d\lambda} \Big|_{\lambda=\mu_+} < 0 < \frac{dr(M - \lambda D)}{d\lambda} \Big|_{\lambda=\mu_-},$$

$r(M - \lambda D) > 0$  for  $\mu_- < \lambda < \mu_+$ , and  $r(M - \lambda D) < 0$  for  $\lambda < \mu_-$  or  $\lambda > \mu_+$ . By Eq. (A1),  $\frac{d \det(M - \lambda D)}{d\lambda} \Big|_{\lambda=\mu_+} < 0 < \frac{d \det(M - \lambda D)}{d\lambda} \Big|_{\lambda=\mu_-}$ . Thus,  $c(\mu_{\pm} | M | D) = 1$ .

If  $r(M) = 0$  and  $\frac{dr(M - \lambda D)}{d\lambda} \Big|_{\lambda=0} > 0$  ( $\frac{dr(M - \lambda D)}{d\lambda} \Big|_{\lambda=0} < 0$ ),  $\exists \mu_- = 0 < \mu_+$  ( $\mu_- < 0 = \mu_+$ ) such that  $r(M - \mu_{\pm} D) = 0$  since  $\lim_{\lambda \rightarrow \pm\infty} r(M - \lambda D) = -\infty$ . The subsequent proofs are the same as  $r(M) > 0$ . If  $\frac{dr(M - \lambda D)}{d\lambda} \Big|_{\lambda=0} = 0$ , since  $\frac{d^2r(M - \lambda D)}{d\lambda^2} < 0, r(M - \lambda D) < 0$  for  $\lambda \neq 0$ , and  $c(0 | M | D) = 2$ .  $\square$

Proof of Lemma A2.

**Proof** Sufficiency: Define  $\mathcal{K}_0^{\text{non}} := \{i \in \mathcal{I} | m(D_i) = n(D_i) = 0, \det(M_i) \neq 0\}$ .  $\forall i \in \mathcal{K}_0^{\text{sin}} \cup \mathcal{K}_0^{\text{non}}, M_i$  is a principal submatrix of the nonsingular  $M$ -matrix  $M^{G_0}$ , thereby nonsingular. Thus,  $i \in \mathcal{K}_0^{\text{non}}$ , which means  $\mathcal{K}_0^{\text{sin}} = \emptyset$ .

Necessity: Assume without loss of generality that  $M^{G_0}$  is in the normal form. Any irreducible diagonal block  $M_j^{G_0}$  of  $M^{G_0}$  is either an irreducible diagonal block  $M_i$  of



$M$ , or a principal submatrix of that. For the former,  $M_j^{\mathcal{G}_0}$  is nonsingular since  $\mathcal{K}_0^{\text{sin}} = \emptyset$ . For the latter,  $M_j^{\mathcal{G}_0}$  is nonsingular since  $M_i$  is an irreducible  $M$ -matrix. In summary,  $M^{\mathcal{G}_0}$  is nonsingular.  $\square$

Proof of Lemma A3.

**Proof** Case 1: Prove by contradiction that  $\lim_{\theta \rightarrow \theta_0} \mu_+(\theta) = \mu_+(\theta_0)$ . Otherwise, since  $\mu_+(\theta)$  is bounded near  $\theta_0$ ,  $+\infty > \overline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta) > \underline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta) > -\infty$ . Let  $\lim_{i \rightarrow +\infty} \theta_i = \theta_0$  be a sequence such that

$$\lim_{i \rightarrow +\infty} \mu_+(\theta_i) = \underline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta).$$

By continuity of  $r(M - \lambda D)$ ,

$$\begin{aligned} 0 &= \lim_{i \rightarrow +\infty} r[M(\theta_i) - \mu_+(\theta_i) D(\theta_i)] = r\left[M(\theta_0) - \lim_{i \rightarrow +\infty} \mu_+(\theta_i) D(\theta_0)\right] \\ &= r\left[M(\theta_0) - \underline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta) D(\theta_0)\right]. \end{aligned}$$

So  $\underline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta) \in R(M|D)(\theta_0)$ . Similarly,  $\overline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta) \in R(M|D)(\theta_0)$ . Since  $\underline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta) \neq \overline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta)$ , the only possibility is that

$$\mu_+(\theta_0) = \overline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta) > \underline{\lim}_{\theta \rightarrow \theta_0} \mu_+(\theta) = \lim_{i \rightarrow +\infty} \mu_+(\theta_i) = \mu_-(\theta_0).$$

For  $i$  large enough,  $\mu_+(\theta_i) < [\mu_+(\theta_0) + \mu_-(\theta_0)]/2 =: \bar{\mu}(\theta_0)$ . By Corollary A1,  $r[M(\theta_0) - \bar{\mu}(\theta_0) D(\theta_0)] > 0$  since  $\mu_-(\theta_0) < \bar{\mu}(\theta_0) < \mu_+(\theta_0)$ , and for  $i$  large enough,  $r[M(\theta_i) - \bar{\mu}(\theta_0) D(\theta_i)] < 0$  since  $\bar{\mu}(\theta_0) > \mu_+(\theta_i)$ . By the continuity of  $r(M - \lambda D)$ ,

$$0 \geq \lim_{i \rightarrow +\infty} r[M(\theta_i) - \bar{\mu}(\theta_0) D(\theta_i)] = r[M(\theta_0) - \bar{\mu}(\theta_0) D(\theta_0)] > 0,$$

conflicts. Thus,  $\mu_+(\theta)$  is continuous, and similarly,  $\mu_-(\theta)$  is continuous.

Case 2: Similar as case 1.  $\square$

Proof of Lemma A4.

**Proof**  $\forall \lambda \in \Lambda(B|D)$ ,  $\exists v = (v_1, v_2, \dots, v_G)$  such that  $v(B - \lambda D) = 0$ . Prove by contradiction that  $\exists i_1 \notin \mathcal{G}_0$  such that  $|v_{i_1}| = \max_{i' \in \mathcal{G}} |v_{i'}|$ . Otherwise,  $|v_{i_2}| = \max_{i' \in \mathcal{G}} |v_{i'}|$  for some  $i_2 \in \mathcal{G}_0$ . Because  $B^{\mathcal{G}_0}$  is irreducible, there exists a path  $i_2 = \hat{i}_1, \hat{i}_2, \dots, \hat{i}_a = j_0 \in \mathcal{G}_0$  such that  $b_{\hat{i}_{a'+1}, \hat{i}_{a'}} > 0 \forall a' \in [1, a - 1]$ . Because  $d_j = 0$  and  $|b_{j,j}| \geq |b_{\cdot,j}| \forall j \in \mathcal{G}_0$ , by consecutively vanishing components  $\hat{i}_1, \hat{i}_2, \dots, \hat{i}_a$  of  $v(B - \lambda D)$ ,  $|b_{\hat{i}_{a'}, \hat{i}_{a'}}| = |b_{\cdot, \hat{i}_{a'}}|$  and  $|v_{\hat{i}_{a'}}| = \max_{i' \in \mathcal{G}} |v_{i'}| \forall a' \in [1, a]$ . Since  $|b_{j_0, j_0}| > \sum_{i \in \mathcal{G}_0 \setminus \{j_0\}} |b_{i, j_0}|$ ,  $\exists i_1 \notin \mathcal{G}_0$  such that  $b_{i_1, j_0} \neq 0$ . So  $|v_{i_1}| = \max_{i' \in \mathcal{G}} |v_{i'}|$ , conflicts.

Since component  $i_1$  of  $v(B - \lambda D)$  vanishes,  $v_{i_1}(\lambda d_{i_1} - b_{i_1, i_1}) = \sum_{i \neq i_1} v_i b_{i, i_1}$ . Thus,

$$\begin{aligned} |\lambda - b_{i_1, i_1}/d_{i_1}| &= \left| \left( \sum_{i \neq i_1} \frac{v_i}{v_{i_1}} b_{i, i_1} \right) / d_{i_1} \right| \leq \left( \sum_{i \neq i_1} \left| \frac{v_i}{v_{i_1}} b_{i, i_1} \right| \right) / |d_{i_1}| \\ &\leq \left( \sum_{i \neq i_1} |b_{i, i_1}| \right) / |d_{i_1}| = |b_{\cdot, i_1}| / |d_{i_1}|. \end{aligned}$$

Therefore,  $\lambda \in \mathcal{O}(b_{i_1, i_1}/d_{i_1}, |b_{\cdot, i_1}|/|d_{i_1}|) \subset \mathcal{O}(B|D)$ . This is statement 1.

Let  $B^\epsilon = (b_{i, j}^\epsilon)_{G \times G}$  for  $\epsilon \in [0, 1]$ .  $b_{i, j}^\epsilon = \epsilon b_{i, j}$  for  $j \notin \mathcal{G}_0$  and  $i \neq j$ ; otherwise,  $b_{i, j}^\epsilon = b_{i, j}$ .  $B^\epsilon$  satisfies conditions in statement 1. Define  $\mathcal{G}_+^0 := \{j | j \in \mathcal{G}, d_j > 0, b_{j, j} = 0\}$  and  $\mathcal{G}_-^0 := \{j | j \in \mathcal{G}, d_j < 0, b_{j, j} = 0\}$ . For  $\epsilon = 0$ ,  $\Lambda(B^\epsilon|D) = \{b_{j, j}/d_j | j \notin \mathcal{G}_0\}$ . For  $\epsilon < 1$ ,

$$\begin{aligned} &\bigcup_{j \in \mathcal{G}_+^0 \cup \mathcal{G}_-^0} \mathcal{O}(b_{j, j}/d_j, \epsilon |b_{\cdot, j}|/|d_j|) \equiv \{0\}, \\ &\left[ \bigcup_{j \in \mathcal{G}_+ \setminus \mathcal{G}_+^0} \mathcal{O}(b_{j, j}/d_j, \epsilon |b_{\cdot, j}|/|d_j|) \right] \cap \left[ \bigcup_{j \in \mathcal{G}_- \setminus \mathcal{G}_-^0} \mathcal{O}(b_{j, j}/d_j, \epsilon |b_{\cdot, j}|/|d_j|) \right] = \emptyset. \end{aligned}$$

By statement 1 and the continuity of  $\Lambda(B^\epsilon|D)$  on  $\epsilon$ ,

$$\begin{aligned} m(D) + |\mathcal{G}_-^0| &\leq \sum_{\lambda \in [\Lambda \cap \mathcal{O}_+](B|D)} c(\lambda|B|D), \\ n(D) + |\mathcal{G}_+^0| &\leq \sum_{\lambda \in [\Lambda \cap \mathcal{O}_-](B|D)} c(\lambda|B|D). \end{aligned}$$

This is statement 2. □

Proof of Lemma A5.

**Proof** By Lemma A4,  $\sum_{\lambda \in [\Lambda \cap \mathcal{O}_+](T|D)} c(\lambda|T|D) \geq m(D)$ .  $[\Lambda_- \cap \mathcal{O}_+](T|D) = \emptyset$  and  $\sum_{\lambda \in \Lambda(T|D)} c(\lambda|T|D) = m(D) + n(D)$ , so  $n(T|D) \leq n(D)$ . Similarly,  $m(T|D) \leq m(D)$ . Thus, if both cases 1 and 2 fail, then  $m(T|D) < m(D)$  and  $n(T|D) < n(D)$ . So at least one of the three cases happens. Obviously, case 3 cannot happen simultaneously with any of cases 1 and 2. Because  $0 \in \Lambda(T|D)$ , cases 1 and 2 cannot happen simultaneously. In summary, exact one of the three cases happens. By Lemma A4,  $\Lambda_0(T|D) \subset \mathcal{O}(T|D)$ . So  $\Lambda_0(T|D) \subset \mathcal{O}(T|D) \cap \{z \in \mathbb{C} | \Re(z) = 0\} = \{0\}$ . □

Proof of Lemma A6.

**Proof** Case 1: Starting from case 1 in Lemma A5,  $m(T|D) = m(D)$ .  $T^{\mathcal{G}_0}$  is a nonsingular  $M$ -matrix. So  $\det(T^{\mathcal{G}_0}) > 0$ . For  $0 < \lambda < \min_{\lambda' \in \Lambda_+(T|D)} \Re(\lambda')$ ,

$$\begin{aligned} \det(T - \lambda D) &= \det\left(T^{\mathcal{G}_0}\right) \prod_{i \notin \mathcal{G}_0} (-d_i) (\lambda - \lambda_i) \\ &= \det\left(T^{\mathcal{G}_0}\right) \prod_{i \notin \mathcal{G}_0} (-1)^{m(D)} |d_i| (-1)^{m(T|D)} |\lambda - \lambda_i| \\ &= \det\left(T^{\mathcal{G}_0}\right) \prod_{i \notin \mathcal{G}_0} |d_i| |\lambda - \lambda_i| > 0, \end{aligned}$$

where  $\lambda_i$  are the roots of  $\det(T - \lambda D) = 0$ . Moreover, since  $T$  is an irreducible  $M$ -matrix,  $\mathcal{E}(T) > 0$  because it is the product of nondominant eigenvalues of  $M$ -matrix  $T$ . So for  $\lambda > 0$  small enough,  $\mathcal{E}(T - \lambda D) > 0$ . In conclusion,  $r(T - \lambda D) = \det(T - \lambda D) / \mathcal{E}(T - \lambda D) > 0$ . By Lemma A1, if  $m(D) > 0$ ,  $0 < \lambda < \mu_+(T|D)$ . Since  $r(T) = 0$ ,  $\mu_-(T|D) = 0$ . Finally,  $c(\mu_{\pm}|T|D) = 1$  by Lemma A1. Together with  $\Lambda_0 \subset \{0\}$  by Lemma A5, there is  $n(Z|D) = n(D) - 1$ ,

Case 2: Similar as case 1.

Case 3: Starting from case 3 in Lemma A5,  $m(T|D) < m(D)$  and  $n(T|D) < n(D)$ . Because  $c(0|T|D) \leq 2$  by Lemma A1, the only possibility is  $m(T|D) = m(D) - 1$ ,  $n(T|D) = n(D) - 1$ , and  $\mu_{\pm}(T|D) = 0$ . □

Proof of Theorem A1.

**Proof** Case 1: Starting from case 1 in Lemma A5,  $m(T|D) = m(D)$ .  $T_i$  is irreducible and diagonally dominant in columns, so  $T_i$  is singular iff it has vanishing column sums. Also, since  $T^{\mathcal{G}_0}$  is irreducible, it is either an irreducible block of  $T$ , or a principal submatrix of that. As a result, singular  $T_i$  satisfies all conditions in Lemma A6. So there are three types of singular  $T_i$ .

1.  $m(T_i|D_i) = m(D_i)$ ,  $n(T_i|D_i) = n(D_i) - 1$ ,  $\mu_-(T_i|D_i) = 0$ . If  $m(D_i) > 0$ ,  $\mu_+(T_i|D_i) > 0$ .
2.  $m(T_i|D_i) = m(D_i) - 1$ ,  $n(T_i|D_i) = n(D_i)$ ,  $\mu_+(T_i|D_i) = 0$ . If  $n(D_i) > 0$ ,  $\mu_-(T_i|D_i) < 0$ .
3.  $m(T_i|D_i) = m(D_i) - 1$ ,  $n(T_i|D_i) = n(D_i) - 1$ ,  $\mu_{\pm}(T_i|D_i) = 0$ .

Since  $m(T|D) = m(D)$ , only type 1 can happen. By Lemma A1,  $r(T_i - \lambda D_i) > 0$  for  $0 = \mu_-(T_i|D_i) < \lambda < \mu_+(T_i|D_i)$  if  $m(D_i) > 0$ ; otherwise,  $r(T_i - \lambda D_i) > 0$  for  $0 = \mu_-(T_i|D_i) < \lambda$ . On the other hand,  $r(T_i - \lambda D_i) > 0$  for nonsingular  $T_i$  and  $\lambda > 0$  small enough. In summary,  $r(T - \lambda D) > 0$  for  $\lambda > 0$  small enough. Because  $T^{\mathcal{G}_0}$  is nonsingular, by Corollary A1,  $\mu_+(T|D) > \lambda > 0$  if  $m(D) > 0$ . Finally,  $\mu_-(T|D) = 0$  since  $r(T) = 0$ .

Case 2: Similar as case 1.

Case 3: Starting from case 3 in Lemma A5,  $m(T|D) < m(D)$  and  $n(T|D) < n(D)$ . If there exists singular  $T_i$  of type 3, then by Lemma A1,  $r(T_i - \lambda D_i) < 0$  for  $\lambda \neq 0$ , thereby  $r(T - \lambda D) < 0$  for  $\lambda \neq 0$ . Thus,  $\mu_{\pm}(T|D) = 0$  by Corollary A1. Otherwise, by  $m(T|D) < m(D)$ , there exists singular  $T_i$  of type 2. By Lemma A1,

$r(T_i - \lambda D_i) < 0$  for  $\lambda > \mu_+(T_i|D_i) = 0$ , thereby  $r(T - \lambda D) < 0$  for  $\lambda > 0$ . Similarly, by  $n(T|D) < n(D)$ , there exists singular  $T_i$  of type 1, thereby  $r(T - \lambda D) < 0$  for  $\lambda < 0$ . In summary,  $r(T - \lambda D) < 0$  for  $\lambda \neq 0$ . Thus,  $\mu_{\pm}(T|D) = 0$  by Corollary A1. □

Proof of Theorem A2.

**Proof** By Lemma A3,  $\mu_{\pm}(\theta)$  are continuous. If  $\mu_+(\theta_0) > \mu_-(\theta_0)$ , by continuity,  $\exists \delta > 0$  such that  $\mu_+(\theta_1) > \mu_-(\theta_2) \forall \theta_1, \theta_2 \in (\theta_0 - \delta, \theta_0 + \delta)$ . Therefore, either  $\mu_+(\theta) \equiv 0$  or  $\mu_-(\theta) \equiv 0$  in  $(\theta_0 - \delta, \theta_0 + \delta)$ . Thus, either  $\mu(\theta) \equiv \mu_-(\theta)$  or  $\mu(\theta) \equiv \mu_+(\theta)$  in  $(\theta_0 - \delta, \theta_0 + \delta)$ . Then  $\mu(\theta)$  is continuous at  $\theta_0$ .

If  $\mu_{\pm}(\theta_0) = 0$ , then since  $\mu \in \{\mu_+, \mu_-\}$ ,

$$\mu(\theta_0) = \mu_{\pm}(\theta_0) = 0 = \lim_{\theta \rightarrow \theta_0} \mu_{\pm}(\theta) = \lim_{\theta \rightarrow \theta_0} \mu(\theta).$$

□

Proof of Lemma A7.

**Proof** First prove the case of irreducible  $Z$ .

1. Assume  $r(Z) = 0$ . Since  $Z$  is irreducible,  $\exists v > 0$  such that  $vZ = 0$ . Thus,  $vZ^\epsilon = 0$ . Because  $Z^\epsilon$  is irreducible and  $v > 0$ ,  $r(Z^\epsilon) = 0$ .
2. Assume  $r(Z) \neq 0$ . Prove by contradiction that  $\text{sgn}[r(Z)] = \text{sgn}[r(Z^\epsilon)]$ . Otherwise, by continuity,  $\exists \epsilon' > 0$  such that  $r(Z^{\epsilon'}) = 0$ . By case 1,  $r(Z) = 0$ , conflicts.

If  $Z$  is reducible,

$$\begin{aligned} \text{sgn}[r(Z)] &= \text{sgn}\left[\min_{i \in \mathcal{I}} r(Z_i)\right] = \min_{i \in \mathcal{I}} \text{sgn}[r(Z_i)] \\ &= \min_{i \in \mathcal{I}} \text{sgn}[r(Z_i^\epsilon)] = \text{sgn}\left[\min_{i \in \mathcal{I}} r(Z_i^\epsilon)\right] = \text{sgn}[r(Z^\epsilon)]. \end{aligned}$$

□

Proof of Lemma A8.

**Proof** Sufficiency: Because column sums of  $T$  vanish, the following three steps transform  $T^{\bar{i}, \bar{j}}$  to  $T^{\bar{j}, \bar{j}}$  for  $i > j$ .

1. Sum all rows of  $T^{\bar{i}, \bar{j}}$  except row  $j$  to row  $j$ .
2. Multiply row  $j$  by  $-1$ .
3. Exchange row  $j$  with rows  $j + 1, j + 2, \dots, i - 1$  one by one.

During which there are totally  $i - j$  changes of sign. So  $T^{\bar{j}, \bar{j}} = (-1)^{i-j} T^{\bar{i}, \bar{j}} = (-1)^{i+j} T^{\bar{i}, \bar{j}}$  for  $i > j$ . There are similar statements for  $i < j$  as well. Then component  $i$  of  $Tv$  satisfies

$$\sum_{j=1}^G t_{i,j} T^{\bar{j}, \bar{j}} = \sum_{j=1}^G (-1)^{i+j} t_{i,j} T^{\bar{i}, \bar{j}} = \det(T) = 0.$$

Necessity: Because  $T$  is irreducible, its dominant eigenvalue 0 is simple, and the corresponding eigenvector space is of dimension 1.  $\square$

### Appendix C: Extension of Theorem 1

**Theorem A3** For  $i \in [2, G - 1]$ , if  $\widehat{H}(x)$  is continuous at  $x_i^*$ ,  $\lim_{x \rightarrow x_i^*} \frac{\widehat{h}_{i,\cdot}(x)}{|x - x_i^*|} = +\infty$ , and  $\lim_{x \rightarrow x_i^*} \frac{\widehat{h}_{i,j}(x)}{\widehat{h}_{i,\cdot}(x)}$  exists  $\forall j \neq i$ , then:

1. There exists a single root  $\lambda_\infty(x)$  of  $\det[\widehat{H}(x) - \lambda A(x)] = 0$  near  $x_i^*$  such that  $\widehat{h}_{i,\cdot}(x_i^*) = \lim_{x \rightarrow x_i^*} \gamma(x_i^* - x) \lambda_\infty(x)$ .
2.  $\mathbb{H}_i := \lim_{x \rightarrow x_i^*} \widehat{H}(x) I_i [1/\widehat{h}_{i,\cdot}(x)]$  satisfies that  $\forall \lambda$ ,

$$\lim_{x \rightarrow x_i^*} p(x)(\lambda) = \det[\mathbb{H}_i - \lambda A(x_i^*)] / \left[ \prod_{i' \neq i} \gamma(x_i^* - x_{i'}^*) \right],$$

where  $I_i [1/\widehat{h}_{i,\cdot}(x)]$  is obtained by replacing the  $i$ th diagonal element of the identity matrix by  $1/\widehat{h}_{i,\cdot}(x)$ , and for  $x \in \bigcup_{i'=1}^{G-1} (x_{i'+1}^*, x_{i'}^*)$ ,

$$p(x)(\lambda) := \det[\widehat{H}(x) - \lambda A(x)] / \left\{ [\lambda - \lambda_\infty(x)] \prod_{i'=1}^G \gamma(x - x_{i'}^*) \right\}. \tag{A2}$$

3.  $\lim_{x \rightarrow x_i^*} \mu(x) = \mu_i := \mu[\mathbb{H}_i | A(x_i^*)]$ .

Under the assumption of Theorem A3, it is possible that  $\widehat{h}_{i,\cdot}(x_i^*) = 0$ . Then the characteristic polynomial of GEs degenerates, i.e.  $\det[\widehat{H}(x_i^*) - \lambda A(x_i^*)] \equiv 0 \forall \lambda \in \mathbb{C}$ . This explains the importance of irreducibility of  $\widehat{H}(x)$ , which promises  $\widehat{h}_{i,\cdot}(x_i^*) > 0$  in Theorem 1. By Theorem A1,  $\mu(\widehat{H}|A)(x)$  is still well-defined for  $x \in \bigcup_{i'=1}^{G-1} (x_{i'+1}^*, x_{i'}^*)$ , and we define  $\mu(x) := \mu(\widehat{H}|A)(x)$ . Nevertheless,  $\mu(x_i^*)$  no longer equals to  $\mu(\widehat{H}|A)(x_i^*)$ .

**Proof** For  $x \in \bigcup_{i'=1}^{G-1} (x_{i'+1}^*, x_{i'}^*)$ ,  $A(x)$  is nonsingular, thereby

$$\det[\widehat{H}(x) - \lambda A(x)] = \det[\widehat{H}(x) A^{-1}(x) - \lambda I] \det[A(x)],$$

and the GEs of  $\widehat{H}(x)$  on  $A(x)$  are the eigenvalues of  $\widehat{H}(x) A^{-1}(x)$ .  $\forall \lambda$ ,

$$\lim_{x \rightarrow x_i^*} \det \left[ \gamma (x_i^* - x) \widehat{H} (x) A^{-1} (x) - \lambda I \right] = (-1)^G \lambda^{G-1} \left[ \lambda - \widehat{h}_{i,\cdot} (x_i^*) \right],$$

so as  $x \rightarrow x_i^*$ ,  $G - 1$  eigenvalues (count multiplicity) of  $\gamma (x_i^* - x) \widehat{H} (x) A^{-1} (x)$  tend to 0, and one tends to  $\widehat{h}_{i,\cdot} (x_i^*)$ . The eigenvalues of  $\gamma (x_i^* - x) \widehat{H} (x) A^{-1} (x)$  are just those of  $\widehat{H} (x) A^{-1} (x)$  multiplying  $\gamma (x_i^* - x)$ , so we have statement 1.

Multiply both sides of Eq. (A2) by  $[\lambda - \lambda_\infty (x)] \prod_{i'=1}^G \gamma (x - x_{i'}^*) / \widehat{h}_{i,\cdot} (x)$ .

$$\begin{aligned} & \left[ \prod_{i' \neq i} \gamma (x - x_{i'}^*) \right] \gamma (x - x_i^*) [\lambda - \lambda_\infty (x)] / \widehat{h}_{i,\cdot} (x) P (x) (\lambda) \\ &= \det \left\{ \left[ \widehat{H} (x) - \lambda A (x) \right] I_i \left[ 1 / \widehat{h}_{i,\cdot} (x) \right] \right\}. \end{aligned}$$

Because  $\mathbb{H}_i := \lim_{x \rightarrow x_i^*} \widehat{H} (x) I_i \left[ 1 / \widehat{h}_{i,\cdot} (x) \right]$ ,  $\lim_{x \rightarrow x_i^*} \gamma (x_i^* - x) / \widehat{h}_{i,\cdot} (x) = 0$ , and  $\lim_{x \rightarrow x_i^*} \gamma (x_i^* - x) \lambda_\infty (x) = \widehat{h}_{i,\cdot} (x_i^*)$ , for any fixed  $\lambda$ ,

$$\begin{aligned} & \left[ \prod_{i' \neq i} \gamma (x_i^* - x_{i'}^*) \right] \lim_{x \rightarrow x_i^*} P (x) (\lambda) \\ &= \lim_{x \rightarrow x_i^*} \left[ \prod_{i' \neq i} \gamma (x - x_{i'}^*) \right] \gamma (x - x_i^*) [\lambda - \lambda_\infty (x)] / \widehat{h}_{i,\cdot} (x) P (x) (\lambda) \\ &= \lim_{x \rightarrow x_i^*} \det \left\{ \left[ \widehat{H} (x) - \lambda A (x) \right] I_i \left[ 1 / \widehat{h}_{i,\cdot} (x) \right] \right\} = \det [\mathbb{H}_i - \lambda A (x_i^*)]. \end{aligned}$$

This is statement 2.

The trace of  $\widehat{H} (x) A^{-1} (x)$  tends to  $+\infty$  as  $x \rightarrow x_i^*$  – because

$$\lim_{x \rightarrow x_i^*} \frac{\widehat{h}_{i,\cdot} (x)}{\gamma (x_i^* - x)} = +\infty.$$

Since the GEs are the eigenvalues of  $\widehat{H} (x) A^{-1} (x)$  for  $x \in \bigcup_{i'=1}^{G-1} (x_{i'+1}^*, x_i^*)$ , the sum of all GEs is the trace of  $\widehat{H} (x) A^{-1} (x)$ , thereby tending to  $+\infty$  as  $x \rightarrow x_i^*$ . By statement 2, GEs other than  $\lambda_\infty (x)$  have finite limits as  $x \rightarrow x_i^*$ . So  $\lim_{x \rightarrow x_i^*} \lambda_\infty (x) = +\infty$ . Similarly,  $\lim_{x \rightarrow x_i^*} \lambda_\infty (x) = -\infty$ . Because  $m[A(x_i^*)], n[A(x_i^*)] > 0$ , some diagonal element of  $\widehat{H} (x) - \lambda_\infty (x) A (x)$  tends to  $-\infty$  as  $x \rightarrow x_i^*$ . By the Gershgorin circle theorem (Gershgorin 1931),  $\lim_{x \rightarrow x_i^*} r \left[ \widehat{H} (x) - \lambda_\infty (x) A (x) \right] = -\infty$ . Thus,  $\mu (x) \neq \lambda_\infty (x)$  for  $x$  near  $x_i^*$ . By statement 2, the roots of  $\det \left[ \widehat{H} (x) - \lambda A (x) \right] = 0$  except  $\lambda_\infty (x)$  tend to the  $G - 1$  roots of  $\det [\mathbb{H}_i - \lambda A (x_i^*)]$  as  $x \rightarrow x_i^*$ , thereby continuous at  $x_i^*$  and bounded near  $x_i^*$ ; thus,  $-\infty < \underline{\lim}_{x \rightarrow x_i^*} \mu (x) \leq \overline{\lim}_{x \rightarrow x_i^*} \mu (x) <$

$+\infty$ . Let  $\lim_{j \rightarrow +\infty} y_j = x_i^*$  be a sequence such that

$$\lim_{j \rightarrow +\infty} \mu(y_j) = \overline{\lim}_{x \rightarrow x_i^*} \mu(x).$$

For  $j$  large enough,  $\widehat{h}_{i,\cdot}(y_j) > 0$  because  $\lim_{x \rightarrow x_i^*} \frac{\widehat{h}_{i,\cdot}(x)}{|x - x_i^*|} = +\infty$ . By Lemma A7,

$$\begin{aligned} r \left[ \widehat{H}(y_j) - \mu(y_j) A(y_j) \right] &= 0 \\ \Rightarrow r \left\{ \left[ \widehat{H}(y_j) - \mu(y_j) A(y_j) \right] I_i \left[ 1/\widehat{h}_{i,\cdot}(y_j) \right] \right\} &= 0. \end{aligned}$$

As a result,

$$\begin{aligned} r \left[ \mathbb{H}_i - \overline{\lim}_{x \rightarrow x_i^*} \mu(x) A(x_i^*) \right] \\ = \lim_{j \rightarrow +\infty} r \left\{ \left[ \widehat{H}(y_j) - \mu(y_j) A(y_j) \right] I_i \left[ 1/\widehat{h}_{i,\cdot}(y_j) \right] \right\} = 0. \end{aligned}$$

Thus,  $\overline{\lim}_{x \rightarrow x_i^*} \mu(x)$  is a DGE of  $\mathbb{H}_i$  on  $A(x_i^*)$ , and so is  $\underline{\lim}_{x \rightarrow x_i^*} \mu(x)$ . Prove by contradiction that  $\overline{\lim}_{x \rightarrow x_i^*} \mu(x) = \underline{\lim}_{x \rightarrow x_i^*} \mu(x)$ . Otherwise,  $\overline{\lim}_{x \rightarrow x_i^*} \mu(x)$  and  $\underline{\lim}_{x \rightarrow x_i^*} \mu(x)$  are the only two DGEs of  $\mathbb{H}_i$  on  $A(x_i^*)$  by Corollary A1, and one must be 0 by Theorem A1. Assume without loss of generality that  $0 = \overline{\lim}_{x \rightarrow x_i^*} \mu(x) > \underline{\lim}_{x \rightarrow x_i^*} \mu(x)$ . Let  $\bar{\mu} := \underline{\lim}_{x \rightarrow x_i^*} \mu(x) / 2 < 0$ . By Corollary A1,  $r[\mathbb{H}_i - \bar{\mu}A(x_i^*)] > 0$  since  $\underline{\lim}_{x \rightarrow x_i^*} \mu(x) < \bar{\mu} < \overline{\lim}_{x \rightarrow x_i^*} \mu(x)$ . For  $j$  large enough,  $\bar{\mu} < \mu(y_j)$ . Then

$$r \left\{ \left[ \widehat{H}(y_j) - \bar{\mu}A(y_j) \right] I_i \left[ 1/\widehat{h}_{i,\cdot}(y_j) \right] \right\} < 0$$

by Corollary A1 since  $\bar{\mu} < 0$  and  $\bar{\mu} < \mu(y_j)$ . As  $j \rightarrow +\infty$ ,  $r[\mathbb{H}_i - \bar{\mu}A(x_i^*)] \leq 0$ , conflicts. Therefore,

$$\overline{\lim}_{x \rightarrow x_i^*} \mu(x) = \underline{\lim}_{x \rightarrow x_i^*} \mu(x).$$

Define  $\mu(x_i^*) := \lim_{x \rightarrow x_i^*} \mu(x)$ . Then,  $\mu(x)$  is continuous at  $x_i^*$ .

Now we show that  $\mu(x_i^*) = \mu_i$ . If  $\mu_i = 0$ , then by Definition A4, 0 is the only DGE of  $\mathbb{H}_i$  on  $A(x_i^*)$ . Then  $\mu(x_i^*) = 0$  because  $\mu(x_i^*)$  is a DGE. Otherwise, assume without loss of generality that  $\mu_i > 0$ . By Theorem A1,  $m[\mathbb{H}_i | A(x_i^*)] = m[A(x_i^*)] = i - 1$ . By continuity,  $\exists \delta > 0$  such that for  $x \in (x_i^* - \delta, x_i^*)$ , the  $i - 1$  roots of  $\det[\mathbb{H}_i - \lambda A(x_i^*)] = 0$  with positive real parts at  $x_i^*$  still have positive real parts at  $x$ . By  $\lim_{x \rightarrow x_i^*} \lambda_\infty(x) = +\infty$ , it is possible to decrease  $\delta$  such that  $\lambda_\infty(x) > 0$ , thereby  $m(\widehat{H}|A)(x) \geq i$  for  $x \in (x_i^* - \delta, x_i^*)$ . Conversely,

for  $x \in (x_{i+1}^*, x_i^*)$ ,  $m(\widehat{H}|A)(x) \leq m[A(x)] = i$  by Theorem A1. Therefore,  $m(\widehat{H}|A)(x) = m[A(x)] = i$ . By Theorem A1,  $\mu(x) > 0$  for  $x \in (x_i^* - \delta, x_i^*)$ . Prove by contradiction that  $\mu(x_i^*) = \mu_i$ . Otherwise,  $\mu(x_i^*) = 0$ . By continuity, the  $i$  roots of  $\det[\widehat{H}(x) - \lambda A(x)] = 0$  with positive real parts in  $(x_i^* - \delta, x_i^*)$ , i.e. the  $i - 1$  roots of  $\det[\mathbb{H}_i - \lambda A(x_i^*)] = 0$  with positive real parts at  $x_i^*$  and  $\lambda_\infty(x)$ , are different from  $\mu(x)$  for  $x \in (x_i^* - \delta', x_i^*)$  with  $0 < \delta' < \delta$  small enough. Thus,  $m(\widehat{H}|A)(x) \geq i + 1 > i = m[A(x)]$ , conflicts. Thus, we have statement 3.  $\square$

Similar to Corollary 1, we have Corollary A2.

**Corollary A2** Assume  $\widehat{H}(x)$  is continuous in  $(x_G^*, x_1^*)$ ,  $\lim_{x \rightarrow x_i^*} \frac{\widehat{h}_{i,\cdot}(x)}{|x - x_i^*|} = +\infty \forall i \in [2, G - 1]$ , and  $\lim_{x \rightarrow x_i^*} \frac{\widehat{h}_{i,j}(x)}{\widehat{h}_{i,\cdot}(x)}$  exists  $\forall j \neq i$ . Let  $\mu(x_i^*) := \mu_i$  for  $i \in [2, G - 1]$ . Then,  $\mu(x)$  is continuous in  $(x_G^*, x_1^*)$ .

**Remark A1** Assume  $\widehat{h}_{i,\cdot}(x_i^*) > 0$ . Then the characteristic polynomial of GEs is nondegenerate. Since  $[\widehat{H}(x_i^*) - \lambda A(x_i^*)] I_i [1/\widehat{h}_{i,\cdot}(x)] = \mathbb{H}_i - \lambda A(x_i^*)$ , we have  $\det[\mathbb{H}_i - \lambda A(x_i^*)] = \det[\widehat{H}(x_i^*) - \lambda A(x_i^*)] / \widehat{h}_{i,\cdot}(x)$ . By Lemma A7,  $\forall \lambda, r$   $r[\widehat{H}(x_i^*) - \lambda A(x_i^*)] = 0$  iff  $r[\mathbb{H}_i - \lambda A(x_i^*)] = 0$ . Thus, GEs and DGEs of  $\widehat{H}(x_i^*)$  on  $A(x_i^*)$  are the same as those of  $\mathbb{H}_i$  on  $A(x_i^*)$ . So  $\mu(\widehat{H}|A)(x_i^*) = \mu[\mathbb{H}_i|A(x_i^*)] = \mu_i$ . In summary, the statements in Theorem A3 are equivalent to those in Theorem 1 for  $\widehat{h}_{i,\cdot}(x_i^*) > 0$ . Since  $\widehat{h}_{i,\cdot}(x_i^*) > 0$  for irreducible  $\widehat{H}(x_i^*)$ , Theorem A3 implies Theorem 1.

## References

- Acar M, Mettetal JT, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* 40:471–475
- Artyomov MN, Das J, Kardar M, Chakraborty AK (2007) Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. *Proc Natl Acad Sci USA* 104(48):18958–18963
- Babloyantz A, Sanglier M (1972) Chemical instabilities of “all-or-none” type in beta—galactosidase induction and active transport. *FEBS Lett* 23(3):364–366
- Berg OG (1978) A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol* 71:587–603
- Bressloff PC (2015) Path-integral methods for analyzing the effects of fluctuations in stochastic hybrid neural networks. *J Math Neurosci* 5(4):1–33
- Bressloff PC, Faugeras O (2017) On the Hamiltonian structure of large deviations in stochastic hybrid systems. *J Stat Mech Theory Exp* 2017:033206
- Bressloff PC, Newby JM (2014a) Path integrals and large deviations in stochastic hybrid systems. *Phys Rev E* 89(042):701
- Bressloff PC, Newby JM (2014b) Stochastic hybrid model of spontaneous dendritic NMDA spikes. *Phys Biol* 11(1):016006
- Choi PJ, Cai L, Frieda K, Xie XS (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322(5900):442–446



- Chu KWE (1987) Exclusion theorems and the perturbation analysis of the generalized eigenvalue problem. *SIAM J Numer Anal* 24(5):1114–1125
- Crudu A, Debussche A, Muller A, Radulescu O (2012) Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *Ann Appl Probab* 22(5):1822–1859
- Davis MHA (1984) Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. *J R Stat Soc Ser B (Methodol)* 46(3):353–388
- Davis MHA (1993) Markov models and optimization, monographs on statistics and applied probability, vol 49. Chapman and Hall, London
- Delbrück M (1940) Statistical fluctuations in autocatalytic reactions. *J Chem Phys* 8(1):120
- Deutsch E, Neumann M (1984) Derivatives of the Perron root at an essentially nonnegative matrix and the group inverse of an  $M$ -matrix. *J Math Anal Appl* 102(1):1–29
- Dürrenberger P, Gupta A, Khammash M (2019) A finite state projection method for steady-state sensitivity analysis of stochastic reaction networks. *J Chem Phys* 150(134):101
- Dykman MI, Mori E, Ross J, Hunt PM (1994) Large fluctuations and optimal paths in chemical kinetics. *J Chem Phys* 100:5735
- Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467:167–173
- Faggionato A, Gabrielli D, Crivellari MR (2009) Non-equilibrium thermodynamics of piecewise deterministic Markov processes. *J Stat Phys* 137:259–304
- Faggionato A, Gabrielli D, Crivellari MR (2010) Averaging and large deviation principles for fully-coupled piecewise deterministic Markov processes and applications to molecular motors. *Markov Process Relat Fields* 16(3):497–548
- Feng H, Han B, Wang J (2011) Adiabatic and non-adiabatic non-equilibrium stochastic dynamics of single regulating genes. *J Phys Chem* 115(5):1254–1261
- Feng J, Kurtz TG (2015) Large deviations for stochastic processes, mathematical surveys and monographs, vol 131. American Mathematical Society, Providence
- Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254(5038):1598–1603
- Freidlin MI, Wentzell AD (2014) Random perturbations of dynamical systems, Grundlehren der mathematischen Wissenschaften, vol 260, 3rd edn. Springer, Berlin
- Frobenius G (1912) Ueber matrixen aus nicht negativen elementen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pp 456–477
- Ge H, Qian H (2009) Thermodynamic limit of a nonequilibrium steady state: Maxwell-type construction for a bistable biochemical system. *Phys Rev Lett* 103(148):103
- Ge H, Qian H, Xie XS (2015) Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. *Phys Rev Lett* 114(078):101
- Ge H, Wu P, Qian H, Xie SX (2018) Relatively slow stochastic gene-state switching in the presence of positive feedback significantly broadens the region of bimodality through stabilizing the uninduced phenotypic state. *PLoS Comput Biol* 14(3):e1006051
- Gershgorin SA (1931) über die abgrenzung der eigenwerte einer matrix. *Bull l'Acad Sci l'URSS Classe Sci Math* 6:749–754
- Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22(4):403–434
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81(25):2340–2361
- Grima R, Schmidt DR, Newman TJ (2012) Steady-state fluctuations of a genetic feedback loop: An exact solution. *J Chem Phys* 137(3):035104
- Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146(4):633–644
- Gupta A, Mikelson J, Khammash M (2017) A finite state projection algorithm for the stationary solution of the chemical master equation. *J Chem Phys* 147(154):101
- Hanggi P, Grabert H, Talkner P, Thomas H (1984) Bistable systems: master equation versus Fokker–Planck modeling. *Phys Rev A* 29:371–378
- Hasenauer J, Wolf V, Kazerooni A, Theis FJ (2014) Method of conditional moments (MCM) for the chemical master equation. *J Math Biol* 69:687–735
- Hegland M, Hellander A, Lötstedt P (2008) Sparse grids and hybrid methods for the chemical master equation. *BIT Numer Math* 48:265–283

- Hornos JEM, Schultz D, Innocentini GCP, Wang J, Walczak AM, Onuchic JN, Wolynes PG (2005) Self-regulating gene: an exact solution. *Phys Rev E* 72(051):907
- Hufton PG, Lin YT, Galla T, McKane AJ (2016) Intrinsic noise in systems with switching environments. *Phys Rev E* 93(052):119
- Hufton PG, Lin YT, Galla T (2018) Phenotypic switching of populations of cells in a stochastic environment. *J Stat Mech Theory Exp* 023:501
- Hufton PG, Lin YT, Galla T (2019a) Classical stochastic systems with fast-switching environments: reduced master equations, their interpretation, and limits of validity. *Phys Rev E* 99(032):121
- Hufton PG, Lin YT, Galla T (2019b) Model reduction methods for population dynamics with fast-switching environments: reduced master equations, stochastic differential equations, and applications. *Phys Rev E* 99(032):122
- Ikramov KD (1993) Matrix pencils: theory, applications, and numerical methods. *J Sov Math* 64:783–853
- Jia C, Qian H, Chen M, Zhang MQ (2018) Relaxation rates of gene expression kinetics reveal the feedback signs of autoregulatory gene networks. *J Chem Phys* 148(9):095102
- Kang HW, Kurtz TG (2013) Separation of time-scales and model reduction for stochastic reaction networks. *Ann Appl Probab* 23(2):529–583
- Karmakar R, Bose I (2004) Graded and binary responses in stochastic gene expression. *Phys Biol* 1(4):197
- Kazeev V, adn Michael Nip MK, Schwab C (2014) Direct solution of the chemical master equation using quantized tensor trains. *PLOS Comput Biol* 10(3):e1003359
- Keener JP, Newby JM (2011) Perturbation analysis of spontaneous action potential initiation by stochastic ion channels. *Phys Rev E* 84(011):918
- Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys J* 81(6):3116–3136
- Kifer Y (2009) Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging. *Mem Am Math Soc* 201(944)
- Knessl C, Matkowsky BJ, Schuss Z, Tier C (1985) An asymptotic theory of large deviations for Markov jump processes. *SIAM J Appl Math* 45(6):1006–1028
- Kussell E, Leibler S (2005) Phenotypic diversity, population growth, and information in fluctuating environments. *Science* 309(5743):2075–2078
- Li GW, Xie XS (2011) Central dogma at the single-molecule level in living cells. *Nature* 475:308–315
- Lin YT, Doering CR (2016) Gene expression dynamics with stochastic bursts: construction and exact results for a coarse-grained model. *Phys Rev E* 93(022):409
- Lu M, Onuchic J, Ben-Jacob E (2014) Construction of an effective landscape for multistate genetic switches. *Phys Rev Lett* 113(078):102
- MacNamara S, Burrage K, Sidje RB (2008) Multiscale modeling of chemical kinetics via the master equation. *Multiscale Model Simul* 6(4):1146–1168
- Mateescu M, Wolf V, Didier F, Henzinger TA (2010) Fast adaptive uniformisation of the chemical master equation. *IET Syst Biol* 4(6):441–452
- Munskya B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* 124(4):044104
- Newby JM (2012) Isolating intrinsic noise sources in a stochastic genetic switch. *Phys Biol* 9(026):002
- Newby J (2015) Bistable switching asymptotics for the self regulating gene. *J Phys A Math Theor* 48(18):185001
- Newby J, Chapman J (2014) Metastable behavior in Markov processes with internal states. *J Math Biol* 69:941–976
- Newby JM, Keener JP (2011) An asymptotic analysis of the spatially inhomogeneous velocity-jump process. *Multiscale Model Simul* 9(2):735–765
- Newby JM, Bressloff PC, Keener JP (2013) Breakdown of fast–slow analysis in an excitable system with channel noise. *Phys Rev Lett* 111(128):101
- Ochab-Marcinek A, Tabaka M (2010) Bimodal gene expression in noncooperative regulatory systems. *Proc Natl Acad Sci USA* 107(51):22096–22101
- Olivieri E, Vares ME (2005) Large deviations and metastability, encyclopedia of mathematics and its applications, vol 100. Cambridge University Press, Cambridge
- Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Ann Rev Phys Chem* 48:545–600
- Ozbudak EM, Thattai M, Lim HN, Shraiman BI, van Oudenaarden A (2004) Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427:737–740

- Paulsson J (2005) Models of stochastic gene expression. *Phys Life Rev* 2(2):157–175
- Peleš S, Munsky B, Khammash M (2006) Reduction and solution of the chemical master equation using time scale separation and finite state projection. *J Chem Phys* 125(20):204104
- Qian H (2014) Fitness and entropy production in a cell population dynamics with epigenetic phenotype switching. *Quant Biol* 2(1):47–53
- Qian H, Shia PZ, Xing J (2009) Stochastic bifurcation, slow fluctuations, and bistability as an origin of biochemical complexity. *Phys Chem Chem Phys* 24(11):4861–4870
- Ramos AF, Innocentini GCP, Hornos JEM (2011) Exact time-dependent solutions for a self-regulating gene. *Phys Rev E* 83(062):902
- Redner S (2007) A guide to first-passage processes. Cambridge University Press, Cambridge
- Samad HE, Khammash M, Petzold L, Gillespie D (2005) Stochastic modelling of gene regulatory networks. *Int J Robust Nonlinear Control* 15(15):691–711
- Santillán M (2008) Bistable behavior in a model of the lac operon in *Escherichia coli* with variable growth rate. *Biophys J* 94(6):2065–2081
- Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329(5991):533–538
- Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98(15):8614–8619
- To TL, Maheshri N (2010) Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* 327(5969):1142–1145
- Touchette H (2009) The large deviation approach to statistical mechanics. *Phys Rep* 478:1–69
- Vellela M, Qian H (2008) Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlogl model revisited. *J R Soc Interface* 6(39):925–940
- Wang J, Xu L, Wang E, Huang S (2010) The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys J* 99(1):29–39
- Zhou JX, Aliyu MDS, Aurell E, Huang S (2012) Quasi-potential landscape in complex multi-stable systems. *J R Soc Interface* 9(77):3539–3553
- Zhu Z, Shendure J, Church GM (2005) Discovering functional transcription-factor combinations in the human cell cycle. *Genome Res* 15(6):848–855

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Jingwei Li<sup>1</sup> · Hao Ge<sup>2</sup>  · Yunxin Zhang<sup>3</sup>

✉ Hao Ge  
haoge@pku.edu.cn

Jingwei Li  
ljw2017@sjtu.edu.cn

Yunxin Zhang  
xyz@fudan.edu.cn

- <sup>1</sup> Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China
- <sup>2</sup> Beijing International Center for Mathematical Research (BICMR) and Biomedical Pioneering Innovation Center (BIOPIC), Peking University, Beijing 100871, China
- <sup>3</sup> Shanghai Key Laboratory for Contemporary Applied Mathematics, Centre for Computational Systems Biology, School of Mathematical Sciences, Fudan University, Shanghai 200433, China