Transient Probability Distributions of Gene Regulatory Networks with Slow Promoter Kinetics

M. Ali Al-Radhawi¹

¹Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115.

January 7, 2019

1 Introduction

In a previous work [1], we have shown that, under a suitable time-scale separation hypothesis, the stationary probability mass functions of gene regulatory networks can be written as mixtures of Poisson distribution.

In this work, we study the time-evolution of the probability mass function and show that it can be approximated via similar techniques.

First, we review the relevant material from [1].

1.1 The Reaction Network Structure

In this paper, a GRN is formally defined as a set of nodes (genes) that are connected with each other through regulatory interactions via the proteins that the genes express. The regulatory proteins are called *transcription factors* (TFs). A TF regulates the expression of a gene by reversibly binding to the gene's promoter and by either enhancing expression or repressing it.

The formalism we employ in order to describe GRNs at the elementary level is that of Chemical Reaction Networks (CRNs) [2]. A CRN consists of *species* and *reactions*, which we describe below.

Species: The species in our context consist of promoter configurations for the various genes participating in the network, together with the respective TFs expressed from these genes and some of their multimers. A configuration of a promoter is characterized by the possible locations and number of TFs bound to the promoter at a given time. If a promoter is expressed constitutively, then there are two configurations specifying the expression activity state, active or inactive. A multimer is a compound consisting of a protein binding to itself several times. For instance, dimers and trimers are 2-mers and 3-mers, respectively. If a

protein forms an n^{th} -order multimer then we say that it has a cooperativity index of n . If species is denoted by X , then its copy number is denoted by X . The set of all species is $\mathcal{S}$.

For simplicity we assume the following:

- (A1) Each promoter can have up to two TFs binding to it.;
- (A2) Each TF is a single protein that has a fixed cooperativity index, i.e, it cannot act as a TF with two different cooperativity indices;
- (A3) Each gene is present with only a *single copy*.

All the above assumptions can be relaxed.

Consider the i^{th} promoter. The expression rate of a gene is dependent on the current configuration of its promoter. We call the set of all possible such configurations the *binding-site set* B_i . Each member of B_i corresponds to a configuration that translates into a specific species $D_j^i, j \in B_i$. If a promoter has just one or no regulatory binding sites, then we let $B_i = \{0, 1\}$. Hence, the promoter configuration can be represented by *two species*: the unbound species D_0^i and the bound species D_1^i . If the promoter has no binding sites then the promoter configuration species are interpreted as the inactive and active configurations, respectively. On the other hand, if the promoter has *two binding* sites then $B_i = \{00, 01, 10, 11\}^1$. The first digit in a member of B_i specifies whether the first binding site is occupied, and the second digit specifies the occupancy of the second binding site. Hence, the promoter configuration can be represented by four species $D_{00}^i, D_{10}^i, D_{01}^i, D_{11}^i$. Note that in general we need to define 2^κ species for a promoter with κ binding sites.

The species that denotes the protein produced by the i^{th} gene is X_i . A protein's multimer is denoted by X_{ic} . If protein X_i does not form a multimer then $X_{ic} := X_i$.

Reactions: In our context, the reactions consist of TFs binding and unbinding with promoters and the respective protein expression (with transcription and translation combined in one step), decay, and n -merization. For each gene, we define a *gene expression block*. Each block consists of a set of *gene reactions* and a set of *protein reactions* as shown in Figure 1.

If the promoter is constitutive, i.e. it switches between two configurations autonomously without an explicitly modeled TF-promoter binding, then $B_i = \{0, 1\}$ and the gene reactions block consists of:

We refer to D_0^i and D_1^i as the *inactive* and *active* configurations, respectively. If the promoter has one binding site, then also $B_i = \{0, 1\}$ and the gene reactions block consists of just two reactions:

where D_0^i and D_1^i are the promoter configurations when *unbound* and *bound* to the TF, respectively. Note that we did not designate a specific species as the active one since it depends on whether the TF is an activator or a repressor. Specifically, when TF is an activator, D_1^i will be the active configuration and D_0^i will be the inactive configuration, and

¹We interpret the elements of the binding set as integers in binary representation.

Figure 1: **A Gene Expression Block.** A GRN that consists of gene expression blocks. A block consists of a gene reactions block and a protein reactions block. The gene reactions are described in the text. **TF** is a vector of TFs which can be monomers, dimers, or higher order multimers. D^i is a vector whose components consist of the D_j^i 's. The dimension of **TF** is equal to the number of binding sites of the gene.

vice versa when TF is a repressor. Finally, if the promoter has two TFs binding to it, then they can bind *independently*, *competitively*, or *cooperatively*. Cooperative binding is discussed in SI-§4.3.1. If they bind independently, then the promoter has two binding sites. Hence, $B_i = \{00, 01, 10, 11\}$ and the gene block contains the following *gene reactions*:

The activity of each configuration species is dependent on whether the TFs are activators or repressors, and on how they behave jointly. This can be characterized fully by assigning a production rate for each configuration as will be explained below. In the case of competitive binding, two different TFs compete to bind to the same location. This can be modeled similarly to the previous case except that the transitions to D_{11}^i , i.e. the configuration where both TFs are bound, are not allowed. Hence, the gene reactions block will have only the first and third, and the binding set reduces to $B_i = \{00, 01, 10\}$.

We assume that RNA polymerase and ribosomes are available in high copy numbers, and that we can lump transcription and translation into one simplified “production” reaction. The rate of production is dependent on the promoter’s configuration. So for each

configuration $D_j^i, j \in B_i$ the *production* reaction is:

, where the kinetic constant k_{ij} is a non-negative number. The case $k_{ij} = 0$ means that when the promoter configuration is D_j^i there is no protein production, and hence D_j^i is an inactive configuration. The promoter configuration can be ranked from the most active to the least active by ranking the corresponding production kinetic rate constants.

Consequently, the character of a TF is manifested as follows: if the maximal protein production occurs at a configuration with the TF being bound we say that the TF is *activating*, and if the reverse holds it is *repressing*. And, if the production is maximal with multiple configurations such that the TF is bound in some of them and unbound in others then the TF is neither repressing nor activating.

We model decay and/or dilution as a single reaction:

The expressed proteins can act as TFs. They may combine to form dimers or higher order multimers before acting as TFs. The numbers of copies of the TF needed to form a multi-mer is called the *the cooperativity index* and we denote it by n . Hence, we model the cooperativity reactions as given in Figure 1 as follows (called the *n-imerization* reactions):

If the cooperativity index of X_i is 1, then the species $X_{ic} := X_i$, and the multimerization reaction becomes empty.

1.2 The Chemical Master Equation

A generic reaction $R_j \in \mathcal{R}$ takes the form:

$$R_j : \sum_{i=1}^{|\mathcal{S}|} \alpha_{ij} Z_i \rightarrow \sum_{i=1}^{|\mathcal{S}|} \beta_{ij} Z_i, \quad (1)$$

where $Z_i \in \mathcal{S}$, α_{ij} , and β_{ij} are positive integers. The reactions that we consider are limited to at most two reactants. The reverse reaction of R_j is the reaction in which the products and reactants are interchanged. If the network contains both the reaction and its reverse then we use the short-hand notation to denote both of them as

$$R_j : \sum_{i=1}^{|\mathcal{S}|} \alpha_{ij} Z_i \rightleftharpoons \sum_{i=1}^{|\mathcal{S}|} \beta_{ij} Z_i. \quad (2)$$

The stoichiometry of a CRN can be summarized by a *stoichiometry matrix* Γ which is defined element-wise as follows:

$$[\Gamma]_{ij} = \beta_{ij} - \alpha_{ij}.$$

The columns of the stoichiometry matrix $\gamma_1, \dots, \gamma_{|\mathcal{R}|}$ are known as the stoichiometry vectors. We say that a nonzero nonnegative vector d gives a *conservation law* for the stoichiometry if $d^T \Gamma = 0$.

Kinetics:

The kinetics of the network quantify the speed of transformation from reactants into products whenever a reaction occurs. In order to keep track of molecule counts, each species $Z_i \in \mathcal{S}$ is associated with a copy number $z_i \in \mathbb{Z}_{\geq 0}$.

To each reaction R_j one associates a propensity function R_j . Assuming a homogeneous well-stirred isothermal medium with a fixed volume, the most common model of propensities, which we use, is the *Mass-Action Kinetics* which is derived from the principle that the likelihood of two reactant molecules colliding and reacting is proportional to their copy numbers. If R_j has a single reactant species Z_i with stoichiometry coefficient α_{ij} , then [3]:

$$R_j(z_i) = k_j \binom{z_i}{\alpha_{ij}} = k_j \frac{z_i(z_i - 1) \dots (z_i - \alpha_{ij} + 1)}{\alpha_{ij}!},$$

where k_j is a kinetic rate constant. Note if $\alpha_{ij} = 1$, then $R_j(z_i) = k_j z_i$.

Dynamics:

The dynamics of the network refers to the manner in which the *state* evolves in time, where the state $Z(t) \in \mathbf{Z} \subset \mathbb{Z}_{\geq 0}^{|\mathcal{S}|}$ is the vector of copy numbers of the species of the network at time t . Since the collision of molecules is random in nature, the time-evolution of states is described mathematically by a stochastic process. The standard stochastic model for a CRN is that of a continuous Markov chain. Let $\mathbf{Z}$ denote the state space. Consider a time t and let the state be $Z(t) = z \in \mathbf{Z}$. Then, the probability that the j^{th} reaction fires in an interval $[t, t + \delta]$ is $R_j(z)\delta + o(\delta)$. If R_j fires, then the states changes from z to $z + \gamma_j$, where γ_j is the corresponding stoichiometric vector.

As Z is a stochastic process we are interested in characterizing its qualitative behavior given by the joint probability distribution $p_z(t) = \Pr[Z(t) = z | Z(0) = z_0]$ for any given initial condition z_0 . The time-evolution of the probability distribution can be shown [3] to be given by a system of linear ordinary differential equations known as the *forward Kolmogorov equation* or the *Chemical Master Equation*, given by:

$$\dot{p}_z(t) = \sum_{j=1}^{|\mathcal{R}|} R_j(z - \gamma_j) p_{z-\gamma_j}(t) - R_j(z) p_z(t), z \in \mathbf{Z}, \quad (3)$$

where $\gamma_1, \dots, \gamma_{|\mathcal{R}|}$ are the columns of the stoichiometry matrix.

Since our species are either gene species or protein species, we split the stochastic process $Z(t)$ into two subprocesses: *the gene process* $D(t)$ and *the protein process* $X(t)$, as explained below.

Consider the i^{th} gene. For each configuration species $D_j^i \in B_i$, let $D_j^i(t) \in \{0, 1\}$ denote its occupancy, i.e. if $D_j^i(t) = 1$, then at time t the i^{th} gene is in a configuration $j \in B_i$. It can be seen from gene reactions that the network always has a conservation law supported on $\{D_j^i, j \in B_i\}$, so that:

$$\sum_{j \in B_i} D_j^i(t) = 1,$$

which reflects the physical constraint that the promoter can be in only one configuration at any given time.

This conservation law enables us to introduce an equivalent reduced representation. For each gene we define one process D_i such that $D_i(t) \in B_i$. $D_i(t) = j$ if and only if $D_j^i(t) = 1$. Collecting these into a vector, define the gene process $D(t) := [D_1(t), \dots, D_N(t)]^T$ where $D(t) \in \prod_{i=1}^N B_i$. The i^{th} gene can be represented by $|B_i|$ states, so $L := \prod_{i=1}^N |B_i|$ is the total number of promoter configurations in the GRN. With abuse of notation, we write also $D(t) \in \{0, \dots, L-1\}$ in the sense of the bijection between $\{0, \dots, L-1\}$ and $\prod_{i=1}^N B_i$ defined by interpreting $D_1 \dots D_N$ as a binary representation of an integer. Hence, $d \in \{0, \dots, L-1\}$ corresponds to $(d_1, \dots, d_N) \in B_1 \times \dots \times B_N$ and we write $d = (d_1, \dots, d_N)$.

Since each gene expresses a corresponding protein, we define $X_{i1}(t) \in \mathbb{Z}_{\geq 0}$, $i = 1, \dots, N$ protein processes. If the multimerized version of the i^{th} protein participates in the network as an activator or repressor then we define $X_{ic}(t)$ as the corresponding multimerized protein process, and we denote $X_i(t) := [X_{i1}(t), X_{ic}(t)]^T$. If there is no multimerization reaction then we define $X_i(t) := X_{i1}(t)$. Since not all proteins are necessarily multimerized, the total number of protein processes is $N \leq M \leq 2N$. Hence, the *protein process* is $X(t) = [X_1^T(t), \dots, X_N^T(t)]^T \in \mathbb{Z}_{\geq 0}^M$ and the state space can be written as $\mathbf{Z} = \mathbb{Z}_{\geq 0}^M \times \prod_{i=1}^N B_i$.

Consider the joint probability distribution:

$$p_{d,x}(t) = \Pr[X(t) = x, D(t) = d], \quad (4)$$

which represents the probability at time t that the protein process X takes the value $x \in \mathbb{Z}_+^M$ and the gene process D takes the value $d \in \{0, \dots, L-1\}$. Recall that x is a vector of copy numbers for the protein processes while d encodes the configuration of each promoter in the network. Then, we can define for each fixed d :

$$p_d(t) := [p_{dx_0}(t), p_{dx_1}(t), \dots]^T, \quad (5)$$

representing the vector enumerating the probabilities (4) for all values of x and for a fixed d , where $x_0, x_1, \dots$ is an indexing of $\mathbb{Z}_{\geq 0}^M$. Note that $p_d(t)$ can be thought of as an infinite vector with respect to the aforementioned indexing. Finally, let

$$p(t) := [p_0(t)^T, \dots, p_{L-1}(t)^T]^T \quad (6)$$

representing a concatenation of the vectors (5) for $d = 0, \dots, L-1$. Note that $p(t)$ is a finite concatenation of infinite vectors.

A gene regulatory network: Consider a set of N genes, binding sets $\{B_i\}_{i=1}^N$, and kinetic constants k_j 's. A *gene expression block*, as shown in Figure 1, is a set of gene reactions and protein reactions as defined above. Each gene block has an output that is either the protein or its n -mer, and it is designated by X_{ic} . The input to each gene expression block is a subset of the set of the outputs of all blocks. Then, a GRN is an arbitrary interconnection of gene expression blocks (Figure 1). SI-§4 defines a more general class of network that we can study. A directed graph can be associated with a GRN as follows. Each vertex corresponds to a gene expression block. There is a directed edge from vertex A to vertex B if the output of A is an input to B. For simplicity, we assume the following: **(A4)** The graph of gene expression blocks is connected. Note that if A4 is violated, our analysis can be applied to each connected component.

Time-Scale Separation: We assume that the gene reactions are considerably slower than the protein reactions. In order to model this assumption, we write the kinetic rates of gene reactions in the form εk_j , where $0 < \varepsilon \ll 1$ and assume that all other kinetic rates (for protein production, decay and multi-merization) are ε^{-1} -times faster.

1.3 Decomposition of the CME

It is crucial to our analysis to represent the linear system of differential equations given by the CME as an interconnection of weakly coupled linear systems. To this end, we present the appropriate notation in this subsection.

Consider the joint PMF: $p_{d,x}(t) = \Pr[X(t) = x, D(t) = d]$, which represents the probability at time t that the protein process X takes the value $x \in \mathbb{Z}_+^M$ and the gene process D takes the value $d \in \{0, \dots, L-1\}$. Recall that x is a vector of copy numbers for the protein processes while d encodes the configuration of each promoter in the network. Then, we can define for each fixed d : $p_d(t) := [p_{dx_0}(t), p_{dx_1}(t), \dots]^T$, representing the vector enumerating the joint probabilities for all values of x and for a fixed d , where $x_0, x_1, \dots$ is an indexing of $\mathbb{Z}_{\geq 0}^M$. Note that $p_d(t)$ can be thought of as an infinite vector with respect to the aforementioned indexing. Finally, let

$$p(t) := [p_0(t)^T, \dots, p_{L-1}(t)^T]^T, \quad (7)$$

representing a concatenation of the vectors $p_{d,x}(t)$ for $d = 0, \dots, L-1$. Note that $p(t)$ is a finite concatenation of infinite vectors. The joint stationary PMF $\bar{\pi}$ is defined as the following limit, which we assume to exist and is independent of the initial PMF: $\bar{\pi} = \lim_{t \rightarrow \infty} p(t)$. Note that $\bar{\pi}$ depends on ε .

Consider a given GRN. The CME is defined over a countable state space $\mathbf{Z}$. Hence, the CME can be interpreted as an infinite system of differential equations with an infinite infinitesimal generator matrix Λ which contains the reaction rates.

Consider partitioning the PMF vector as in (7). Recall that reactions have been divided into two sets: slow gene reactions and fast protein reactions. This allows us to write Λ as a sum of a slow matrix $\varepsilon \hat{\Lambda}$ and a fast matrix $\tilde{\Lambda}$, which we call a fast-slow decomposition. Furthermore, $\tilde{\Lambda}$ can be written as a block diagonal matrix with L diagonal blocks which correspond to conditioning the MC on a specific gene state d . This is stated in the following basic proposition (see [1] for the proof):

Proposition 1. *Given a GRN. Its CME can be written as*

$$\dot{p}(t) = \Lambda_\varepsilon p(t) = (\tilde{\Lambda} + \varepsilon \hat{\Lambda}) p(t), \quad (8)$$

where $p(t) = [p_0^T(t), \dots, p_{L-1}^T(t)]^T$, and

$$\tilde{\Lambda} = \text{diag}[\Lambda_0, \dots, \Lambda_{L-1}] \quad (9)$$

where $\tilde{\Lambda}$ is the fast matrix, $\hat{\Lambda}$ is the slow matrix, and $\Lambda_0, \dots, \Lambda_{L-1}$ are stochastic matrices.

1.4 Conditional MCs

For each d , consider modifying the MC $Z(t)$ defined in the previous section by replacing the stochastic process $D(t)$ by a deterministic constant process $D(t) = d$. This means that the resulting MC does not describe the gene process dynamics, it only describes the protein process dynamics *conditioned on d* . Henceforth, we refer to the resulting MC as the *MC conditioned on d* . The infinitesimal generator of a MC conditioned on d is denoted by Λ_d , and is identical to the corresponding block on the diagonal of $\tilde{\Lambda}$ as given in (9). In other words, fixing $D(t) = d \in \{0, \dots, L-1\}$, the dynamics of the network can be described by a CME:

$$\dot{p}_{X|d} = \Lambda_d p_{X|d}, \quad (10)$$

$\dot{p}_{X|d} = \Lambda_d p_{X|d}$, where $p_{X|d}$ is a vector that enumerates the conditional probabilities $p_{x|d} = \Pr[X(t) = x | D(t) = d]$ for a given d . The conditional stationary PMF is denoted by: $\pi_{X|d}^{(J)} = \lim_{t \rightarrow \infty} p_{X|d}(t)$, where (J) refers to the fact that it is joint in the protein and multimerized protein processes. Note that $\pi_{X|d}^{(J)}$ is independent of ε . This notion of a conditional MC is useful since, at the SPK limit, $D(t)$ stays constant. It can be noted from (9) that when $\varepsilon = 0$ the dynamics of p_d decouples and becomes independent of $p_{\tilde{d}}$, $\tilde{d} = 0, \dots, L-1$, $\tilde{d} \neq d$.

We show below that each conditional MC has a simple structure. Fixing the promoter configuration $D(t) = d = (d_1, \dots, d_N)$, the network consists of *uncoupled* birth-death processes. So for each d_i , the protein reactions of production and dimerization corresponding to the i^{th} promoter can be written as follows without multimerization: $\emptyset \xrightleftharpoons[k_{-i}]{k_{id_i}} X_i$, where the subscript id_i refers to the production kinetic constant corresponding to the configuration species $D_{d_i}^i$, or, if there is a multimerization reaction, it takes the form: $\emptyset \xrightleftharpoons[k_{-i}]{k_{id_i}} X_i$, $n_i X_i \xrightleftharpoons[\beta_{-i}]{\beta_i} X_{ic}$. Note that the stochastic processes $X_i(t)$, $i = 1, \dots, N$ conditioned on $D(t) = d$ are independent of each other. Hence, the conditional stationary PMF $\pi_{X|d}^{(J)}$ can be written as a product of stationary PMFs and the individual stationary PMFs have Poisson expressions. The following proposition gives the analytic expression of the conditional stationary PMFs: (see [1] for proof)

Proposition 2. Fix $d \in \{0, \dots, L-1\}$. Consider (10), then there exists a conditional stationary PMF $\pi_{X|d}^{(J)}$ and it is given by

$$\pi_{X|d}^{(J)}(x) = \prod_{i=1}^N \pi_{X|di}(x_i), \quad (11)$$

where

$$\pi_{X|di}^{(J)}(x_i) = \begin{cases} \mathbf{P}\left(x_{i1}, x_{i2}; \frac{k_{id_i}}{k_{-i}}, \frac{k_{id_i}^{n_i} \beta_i}{n_i! k_{-i}^{n_i} \beta_{-i}}\right) : n_i > 1 \\ \mathbf{P}\left(x_i; \frac{k_{id_i}}{k_{-i}}\right) : n_i = 1, \end{cases} \quad (12)$$

where (J) refers to the joint PMF in multimerized and non-multimerized processes, x_{i1} refers to the copy number of X_i , while x_{i2} refers to the copy number of X_{ic} , $\mathbf{P}(x; a) := \frac{a^x}{x!} e^{-a}$, $\mathbf{P}(x_1, x_2; a_1, a_2) := \frac{a_1^{x_1}}{x_1!} \frac{a_2^{x_2}}{x_2!} e^{-a_1 - a_2}$.

Remark 1. The conditional PMF in (11) is a joint PMF in the protein and multimerized protein processes. If we want to compute a marginal stationary PMF for the protein process only, then we average over the multimerized protein processes $X_{ic}, i = 1, \dots, N$ to get a joint Poisson in N variables. Hence, the formulae (11)-(12) can be replaced by:

$$\pi_{X|d}(x) := \sum_{i=1}^{M-N} \sum_{x_{i2}=0}^{\infty} \pi_{X|d}^{(J)}(x) = \prod_{i=1}^N \mathbf{P} \left(x_i; \frac{k_{id_i}}{k_{-i}} \right), \quad (13)$$

where $M - N$ is the number of n -merized protein processes, and $\pi_{X|d}$ is the marginal stationary PMF for the protein process.

Recall the conditional Markov chains with the associated infinitesimal generators as in Eq. (4). By the assumptions, for each $d \in \{0, \dots, L-1\}$ there exists a unique $\pi_{X|d}$ such that: $\Lambda_d \pi_{X|d} = 0$, $\pi_{X|d} > 0$, and $\sum_x \pi_{X|d}(x) = 1$. Recall that $\pi_{X|d}$ is the stationary distribution of the Markov chain conditioned on $D(t) = d$.

Defining the extended conditional distributions for $d = 0, \dots, L-1$ as:

$$\bar{\pi}_{X|d} := [\overbrace{\mathbf{0}^T \dots \mathbf{0}^T}^{d-1} \pi_{X|d}^T \overbrace{\mathbf{0}^T \dots \mathbf{0}^T}^{L-d}]^T. \quad (14)$$

The stationary distribution above can be interpreted as a function as follows: $\bar{\pi}_{X|d}(x, d) = \pi_{X|d}(x)$, and $\bar{\pi}_{X|d}(x, d') = 0$ when $d' \neq d$.

1.5 Stationary Distribution at the limit of slow promoter kinetics

Recall the slow-fast decomposition of the CME in (8) and the joint stationary PMF $\bar{\pi}$. In order to emphasize the dependence on ε we denote $\bar{\pi}^\varepsilon := \bar{\pi}(\varepsilon)$. Hence, $\bar{\pi}^\varepsilon$ is the unique stationary PMF that satisfies $\Lambda_\varepsilon \bar{\pi}^\varepsilon = 0$, $\bar{\pi}^\varepsilon > 0$, and $\sum_z \bar{\pi}_z^\varepsilon = 1$, where the subscript denotes the value of the stationary PMF at z .

Our aim is to characterize the stationary PMF as $\varepsilon \rightarrow 0$. Writing $\bar{\pi}_\varepsilon$ as an asymptotic expansion to first order in terms of ε , we have

$$\bar{\pi}^\varepsilon = \bar{\pi}^{(0)} + \bar{\pi}^{(1)}\varepsilon + o(\varepsilon). \quad (15)$$

Singular perturbation techniques can be used to derive the following theorem [1]:

Theorem 3. Consider a given GRN with L promoter states with the CME (8). Writing (15), then the joint stationary PMF $\bar{\pi} := \lim_{\varepsilon \rightarrow 0^+} \bar{\pi}^\varepsilon$ can be written as: $\bar{\pi}(x, d) = \sum_{d=0}^{L-1} \lambda_d \bar{\pi}_{X|d}(x, d)$, where $\lambda = [\lambda_0, \dots, \lambda_{L-1}]^T$ is the principal normalized eigenvector of:

$$\Lambda_r := \text{diag}[\mathbf{1}^T, \dots, \mathbf{1}^T] \hat{\Lambda} [\bar{\pi}_{X|0} \ \bar{\pi}_{X|1} \ \dots \ \bar{\pi}_{X|L-1}], \quad (16)$$

where $\bar{\pi}_{X|0}, \dots, \bar{\pi}_{X|L-1}$ are the extended conditional stationary PMFs defined as: $\bar{\pi}_{X|d}(x, d) = \pi_{X|d}(x)$, $\bar{\pi}_{X|d}(x, d') = 0$ when $d' \neq d$.

2 Transient Dynamics of the CME

In [1], we have been interested in computing the analytical expression for the stationary PMF at the SPK limit. In this section, we use the same techniques to write the transient dynamics after an initial time interval (called *the boundary layer* in the singular perturbations literature [4].)

Recall that Proposition 1 provides a decomposition of the CME as follows:

$$\frac{dp}{d\tau} = (\tilde{\Lambda} + \varepsilon \hat{\Lambda})p, \quad p_0 = p(0). \quad (17)$$

The effect of the slow gene reactions matrix $\varepsilon \hat{\Lambda}$ in the model above takes a relatively long time to be active. Because of this, one calls τ the “fast” time-scale. So in order to for the effect of $\varepsilon \hat{\Lambda}$ to manifest sooner, we define the “slow time-scale” $t = \varepsilon \tau$. The Master Equation can be written as follows:

$$\varepsilon \frac{dp}{dt} = (\tilde{\Lambda} + \varepsilon \hat{\Lambda})p, \quad p(0) = p_0, \quad (18)$$

or equivalently as $dp/dt = (\tilde{\Lambda}/\varepsilon + \hat{\Lambda})p$.

Let $p_\varepsilon(t)$ be the solution of (18). Writing:

$$p_\varepsilon(t) = p_0(t) + \varepsilon p_1(t) + o(\varepsilon^2), \quad t \gg 0,$$

and substituting the above expansion in (18), we obtain the following equation from the zeroth-order term:

$$\tilde{\Lambda} p_0(t) \equiv 0.$$

Hence, the distribution p_0 will quickly approach the kernel space of the fast matrix, i.e $p_0(t) \in \ker \tilde{\Lambda}$, which is spanned by the extended conditional distributions $\bar{\pi}_{X|d}$, $d = 0, \dots, L$ as given in (14). Hence, we obtain:

$$p_0(t) = \sum_{d=0}^{L-1} \lambda_d(t) \bar{\pi}_{X|d}. \quad (19)$$

The first order term in this expression gives the following equation:

$$\dot{p}_0(t) - \tilde{\Lambda} p_1(t) = \hat{\Lambda} p_0(t).$$

Recall that $\tilde{\Lambda} = \text{diag}[\Lambda_0, \dots, \Lambda_{L-1}]$, and that each Λ_d is an infinitesimal generator which satisfies $\mathbf{1}^T \Lambda_d = 0$. Hence, if we pre-multiply the equation above by $\bar{\mathbf{1}} := \text{diag}[\mathbf{1}^T, \mathbf{1}^T, \dots, \mathbf{1}^T]$ and substitute (19) we obtain:

$$\sum_{d=0}^{L-1} \lambda_d(t) \mathbf{1}^T \pi_{X|d} = \bar{\mathbf{1}} \hat{\Lambda} \sum_{d=0}^{L-1} \lambda_d(t) \bar{\pi}_{X|d}.$$

Since $\mathbf{1}^T \pi_{X|d} = 1$ and recalling the definition of the reduced infinitesimal generator Λ_r (16), we get the following finite dimensional linear system:

$$\dot{\lambda} = \Lambda_r \lambda, \quad \lambda(0) = \bar{\mathbf{1}} p(0). \quad (20)$$

This is an autonomous linear system and the solution can be written as a matrix exponential as follows: $\lambda(t) = e^{\Lambda_r t} \lambda(0)$.

Hence, as $\varepsilon \rightarrow 0$ the solution of the master equation is approximated by the following expression after the boundary layer:

$$p_0(\tau) = \sum_{d=0}^{L-1} \lambda_d(\tau) \bar{\pi}_{X|d} = \Pi e^{\Lambda_r \tau} \bar{\mathbf{1}} p(0),$$

where $\Pi := [\bar{\pi}_{X|0}, \dots, \bar{\pi}_{X|d}]$.

In the slow time-scale, we can write

$$p_0(\tau) = \Pi e^{\varepsilon \Lambda_r \tau} \bar{\mathbf{1}} p(0), \quad (21)$$

which is consistent with the analysis carried out in [5] for the finite projection of the CME.

In order to estimate the initial boundary layer, let $\rho^* = \min\{|\mathbf{Re}(\rho)| : \rho \in \sigma(\tilde{\Lambda}), \rho \neq 0\}$, where $\sigma(\cdot)$ is the spectrum operator. Assuming that $\varepsilon < \rho^*$, it can be seen that the approximation $p(\tau) = p_0(\tau) + o(\varepsilon)$ is valid for $\tau > T$ where T is proportional to $-\ln \varepsilon / \rho^*$. In order to compare to the main results of [1], recall that Λ_r itself is an infinitesimal generator, hence $\lim_{t \rightarrow \infty} e^{\Lambda_r t} = \bar{\lambda}^T \mathbf{1}$, where $\bar{\lambda}$ is the principal eigenvector of Λ_r . Hence,

$$\lim_{t \rightarrow \infty} p_0(t) = \bar{\pi},$$

where $\bar{\pi}$ is the stationary PMF in the statement of Theorem 3. Hence, a generalization of Corollary 4 in [1], can be stated as:

$$\lim_{\varepsilon \rightarrow 0} \Pr[X(t) = x] = \sum_{d=0}^{L-1} \lambda_d(t) \prod_{i=1}^N \mathbf{P}\left(x_i; \frac{k_{id_i}}{k_{-i}}\right) + o(\varepsilon), \quad t > T.$$

Example: The Self-Repressing Gene

We now revisit the self-repressing gene:

which was studied in Figure 1 in [1].

Using (16), the reduced generator can be written as:

$$\Lambda_r = \begin{bmatrix} -\alpha k_0 / k_- & \alpha_- \\ \alpha k_0 / k_- & -\alpha_- \end{bmatrix}.$$

The stationary PMF is as follows:

$$\pi_1(x) = \frac{\alpha \rho_1}{\alpha_- + \alpha \rho_1} \mathbf{P}(x; k_1 / k_-) + \frac{\alpha_-}{\alpha_- + \alpha \rho_1} \mathbf{P}(x; k_0 / k_-), \quad (23)$$

Figure 2: **Time evolution of the PMF for a self-repressing gene with $\varepsilon = 0.1$.** (a) The full-state time evolution of the probability distribution for self-repressing gene with the initial condition $\Pr[X(0) = 0, D(0) = 0] = 1$. (b) The time evolution of $\Pr[X(t) = 40]$ compared between the finite state projection solution [5], and the singular perturbation approximation (21). The full trajectories have been generated by a finite projection numerical solution of the CME. Both figures use the network with parameters $\alpha = \varepsilon/200$, $\alpha_- = \varepsilon$, $k_0 = 40$, $k_- = 1$.

where

$$\rho_1 = \mathbb{E}[X|D = 0] = k_0/k_-.$$

We study the case $\varepsilon = 0.1$. Figure 2-a shows how the stationary distribution evolves in time. It can be seen that the distribution becomes bimodal at the two predicted modes at around $t = -(\ln \varepsilon)/\rho^* \approx 2.35$, so it approaches the kernel of $\tilde{\Lambda}$ quickly. Then, the slow dynamics given by (20) drive the time evolution of the weights of the two modes to their steady state values. Figure 2-b depicts evolution of the probability of $\Pr[X(t) = 40]$, and compares the exact and approximated solutions. It can be seen that the two solutions match after an initial boundary layer.

References

- [1] M Ali Al-Radhawi, Domitilla Del Vecchio, and Eduardo D Sontag. Multi-modality in gene regulatory networks with slow promoter kinetics. *arXiv preprint arXiv:1705.02330*, 2017.
- [2] Péter Érdi and János Tóth. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press, 1989.
- [3] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*. Springer, 2015.
- [4] R Phillips and P Kokotović. A singular perturbation approach to modeling and control of Markov chains. *IEEE Transactions on Automatic Control*, 26(5):1087–1094, 1981.

- [5] Slaven Peleš, Brian Munsky, and Mustafa Khammash. Reduction and solution of the chemical master equation using time scale separation and finite state projection. *The Journal of Chemical Physics*, 125(20):204104, 2006.