Part III

Stochastic transcription and traversal times of Markov chains

Chapter 8

Introduction

A fundamental organizational part of the complex, highly organized biological processes we see around us takes place at a molecular level, as randomly diffusing proteins interact with individual gene loci to trigger the production of other molecular signals. Networks of such interactions can generate precise, reliable behaviors, despite the often large dependence of individual reactions on molecular concentrations and the inherent stochasticity of each interaction. Evolution and natural selection have produced many such robust, precise networks, but it is not known what characteristics of these networks lead to this emergent reliability.

Some of the best studied examples of emergent reliability and precision among regulatory networks of gene expression come from the dorsal-ventral (DV) patterning system in *Drosophila* [313, 312]. Patterning in this system is driven by localized, ventral activation of a uniformly expressed embryonic protein called Dorsal. The activated molecules diffuse dorsally and shuttle into the yet undifferentiated nuclei. Binding of Dorsal to regulatory DNA sequences in the nuclei activates a network of other transcription factors and signaling proteins. Five or more distinct domains of gene expression are delineated by this network, giving rise to distinct cell fates [151]. The boundaries between different cell types are determined to single cell precision and are reliably made at each of the hundred odd cell-junctions along the length of each boundary, despite variations in environmental gene dose, and protein concentration.

For a gene to be transcribed into RNA, first a RNA polymerase enzyme must bind to the DNA strand, a process which in eukaryotes requires the presence (or absence) of a

complex collection of other proteins, which we call generically *transcription factors*[1] and often abbreviate as "TF". As different proteins bind or dissociate to the complex molecular structure (the *regulatory complex*) near the gene to be expressed, the structure of the regulatory complex changes, so that in any given state of the regulatory complex, only certain other proteins my be allowed to bind or dissociate. Furthermore, the RNA polymerase may be *paused* at various points in the transcription process (even after elongation begins) by the absence of some required transcription factor, so that it does not finish transcription until additional molecular signals arrive.

There are many aspects of the process that leads to DNA transcription that can affect the speed and timing of transcription and the robustness to random fluctuation in transcription factor concentrations. One such aspect that might be expected to be important is the *topology* of the network describing the dependencies between various transcription factors, by which we mean the abstract "structure" of the relationships. Suppose for instance that there are two TFs **A** and **B** that must bind to the DNA molecule before transcription can begin. There are three most basic topologies: (a) **A** must bind before **B**; (b) **B** must bind before **A**; and (c) **A** and **B** bind independently of each other. The effect of network topology can be discussed meaningfully if, for instance, the network displays a certain property over a wide range of parameter values. In the above example, we can say that if the binding rates are the same in each case, topology (c) is invariably *faster* (has a smaller average time until transcription) than topologies (a) or (b), a fact which should be clear intuitively.

In this work we present a Markov model of regulatory complex assembly and transcription, and develop various analytical tools for the purpose of studying the effect of network topology on robustness properties of the transcription process, There are certainly generalizations to be made about when certain classes of networks are faster, less noisy, or more robust than others, but it seems likely that such generalizations will not tell the whole story, and that detailed analysis will shed light on specific examples. We have focused on developing tools that can be used for computing properties of the (random) time to transcription. We present general techniques that can be used to make the computation of the distribution (via a Laplace transform) of the time to transcription feasible in larger networks, and find even simpler methods for computing the moments of the time to

[1] Although this term is sometimes used in a more specific sense.

transcription.

Our tools are standard for the analysis of continuous-time Markov chains, but as these techniques are intended to be of practical use to non-specialists, we have attempted to give a level of detail and specificity that will avoid misunderstanding and error in implementation, and may seem overly pedantic to some.

In Section 8.1 we summarize previous work done on stochasticity in gene expression. In Chapter 9 we introduce our general framework for models of transcription and some basic tools from continuous-time Markov chain theory, discuss some measures of stochasticity or noisiness, and briefly analyze a simple model. In Chapter 10 we introduce a more detailed model of transcription for which the tools of Chapter 9 are less feasible, and motivate our strategies for simplifying the calculations, which we develop in Chapters 11 and 12. The general strategy is to decompose the larger Markov chain into smaller pieces and combine suitably chosen quantities computed for each smaller piece to find the desired quantity for the larger chain. Our main results are those of Chapter 11, in which we discuss how to compute quantities related to the transcription time if the model can be broken into sequentially arranged pieces in a manner similar to a birth-death chain. In Chapter 12 we discuss a similar strategy if instead the model can be seen as several noninteracting pieces working in parallel.

8.1 Previous work

The proteins produced by one gene may regulate the actions of other genes, so considered together a collection of such interrelated genes form a *regulatory network* [227], or, if the relationships are one-directional, a *regulatory cascade*. A first approximation to such a system is the ODE in which if $x(k)$ is the concentration of the k^{th} compound of interest, the time derivative of $x(k)$ is given as a function of all the concentrations, or

$$\dot{x}(k) = f_k(x). \tag{8.1.1}$$

However, for some of the quantities involved, it is more appropriate to model the *number* of molecules of a certain type, rather than the concentrations, since the actual numbers of each molecule present ("copy number") may be too small to justify the continuous-state approximation of (8.1.1). Such intercellular variation in protein concentration was first observed by Novick & Weiner in 1957 [236], but it was not until 2002

that Elowitz et al [87] demonstrated experimentally that stochasticity in gene expression can contribute significantly to variation in cellular responses. Interest in the topic jumped sharply after McAdams & Arkin [227] observed theoretically that stochastic effects in gene expression were likely at the observed concentration of regulatory factors, and could provide an important mechanism for phenotypic variation, even amongst genetically identical individuals, a phenomenon observed in bacteria and other clonal organisms [262].

Consider the simple model of gene transcription in which there is a fixed concentration of RNA polymerase, and a RNA polymerase molecule will bind to the DNA if it floats near enough, transcribing a copy of the gene. The production times of the transcribed RNA copies can be thought of as the departure times of a M/G/∞ queue [286], and under reasonable assumptions about the concentration of RNA polymerase in the cell, the arrival times of RNA polymerase form a Poisson process, so at stationarity the times at which transcription is completed form a Poisson process as well. If the copies decay at a fixed rate, the number of transcribed RNA copies at any time would follow Poisson statistics. However, it was quickly observed that this does not usually fit the data [262], and that the actual transcription times are clustered relative to the Poisson process one would find in this simplest, two-state, memoryless system. One suggestion for interpretation of this fact is that rather than producing a single transcript and starting again from scratch, a gene through some mechanism generally produces more than one transcript during each transcription event. Various models of this "bursting" phenomenon have been studied, for instance by Paulsson and coauthors [240, 241], usually compound Poisson models in which each Poisson arrival leads to a fixed or random number of transcript copies,

Since the appearance of McAdams & Arkin's 1997 paper [227], there has been a moderate amount of theoretical work done on the subject of stochasticity in gene expression. In their paper and in most subsequent papers, the models considered have been first-order perturbations about the equilibrium of Equation (8.1.1), (referred to in this framework as the *low noise approximation* [177]) using Fourier transform methods and Langevin dynamics to analyze the effects of small (linear) perturbations with different types of noise. Steps have been made to consider gradually more general models with these methods by Paulsson and others [241]. Tkačik and co-authors [291] have used these techniques to look at regulatory cascades as information channels, for instance computing the maximal (Shannon) conditional information transmittable by the (time-varying) concentration of the final product of expression. Thattai & vanOudenaarden [289] used these techniques to demonstrate that

by using "thresholding" steps in a regulatory cascade, the propagation of noise in a long cascade could be bounded even as the length of the cascade grows.

Some attention has been paid in the literature to *which stage* in the gene expression process produces the observed variation in output protein concentration. Elowitz et. al. [87] distinguished between "intrinsic" (within-cell sources of variation) and "extrinsic" (between-cell sources of variation) noise. Tkačik et. al. [292] further distinguish two types of intrinsic noise: "input" noise coming from stochasticity in gene regulation and "output" noise coming from stochasticity in both the (post-regulation) transcription and the translation processes.

A final note is in order. Most of the work mentioned above studies stochasticity in the entire transcription–translation process. In the present work we study only the transcription step, which corresponds most closely to recent experimental data [25].

Acknowledgements: This part has been enjoyable joint work with Alistair Boettiger and Steve Evans.

Chapter 9

Markov models of protein-DNA interactions

Our model of transcription will be a fairly general continuous-time Markov chain. In this chapter, we clarify our assumptions that lead to this model and Markov chain notation, introduce general analytical techniques for computing properties of the time until transcription, and discuss some possible measures of "noisiness". Finally, we give an example of these methods applied to a simple, concrete model of transcription.

9.1 Model construction

In constructing our general model of transcription, we introduce three assumptions:

1. Brownian molecular collisions and electrostatic interactions mediate all TF binding events.

2. The statistics of these interactions are stationary in time.

3. The TF concentrations are large enough that binding an individual factor to a gene locus does not decrease the binding rate at other loci.

These assumptions are satisfied if concentrations do not change significantly on the time scale we are interested in, and TFs are sufficiently well mixed in the nucleus so that the local concentration does not fluctuate in a manner inconsistent with Brownian dynamics.

We will consider events at only a single gene locus, neglecting, for instance, the corresponding locus on the homologous chromosome. Our assumptions of constant concentrations and Brownian dynamics imply that the process is memoryless, so the times between TF arrival events are exponentially distributed. Therefore, we may model the state of the regulatory complex at a locus as a continuous-time Markov chain whose state records the current intermediary complex present at the locus. More specifically, we enumerate the set of all possible intermediary complexes — for instance, if there are k different TFs, and each exists in one of m states (e.g. *bound* and *unbound* if $m = 2$), then the set of possible intermediary complexes could be enumerated as $1, \ldots, m^k$. The state of the complex at time t is denoted X_t, and $X = (X_t)_{t \geq 0}$ is a time-homogeneous Markov process on the set of intermediary complexes. There is some "final" state f that corresponds to successful transcription; the time it takes the chain to reach this state is the *first passage time to* f or *transcription time*.

In this model, the time it takes the process X to leave state i is exponentially distributed with rate r_i, where r_i is the sum of the reaction rates for all reactions that lead out of that state. If there are N states in total and the transition rate between state i and state j is q_{ij}, then

$$r_i = \sum_{j=1}^{N} q_{ij}. \tag{9.1.1}$$

Once the system leaves state i, it changes to state j with probability q_{ij}/r_i.

Specifying the set of transition rates q_{ij} and a starting state fully specifies the dynamics of the system. If we introduce the *generator matrix* G whose elements are given by $G_{ij} = q_{ij}$ for $i \neq j$ and $G_{ii} = -r_i$, then the probability that the system is in state j at time t, given that it started in i at time 0, can be written as

$$\mathbb{P}\{X_t = j | X_0 = i\} = (e^{tG})_{ij} = \sum_{k=0}^{\infty} \frac{t^k (G^k)_{ij}}{k!}. \tag{9.1.2}$$

From this formulation we can compute properties of the transcription time. For instance, if we have a population of duplicate systems (cells), the mean transcription time corresponds to the average delay between induction and expression, while the variance of the transcription time corresponds to variance among the systems, and so indicates the degree of asynchrony in initial expression. The total number of times the system makes the specified transition in a finite window of time corresponds to the number of mRNA molecules made during that time period.

9.2 Computing first-passage times of Markov chains

Let X be a continuous-time Markov chain on a finite state space with generator matrix G, and suppose that it starts in state s at time 0, and with probability 1 will eventually reach the final state f. Let τ denote the time that X first reaches f. The probability distribution of τ can be computed using the modified transition matrix

$$G^\star_{ij} = \begin{cases} G_{ij}, & \text{if } i \neq f, \\ 0, & \text{if } i = f. \end{cases}$$

This is the generator matrix for the *killed* chain $\widetilde{X}$, defined as $\widetilde{X}_t = X_{\min(t,\tau)}$, and is a Markov chain that moves the same as X up until it first hits state f, at which time it stops.

The Laplace transform of the random variable τ is defined as

$$\mathbb{E}[e^{-\lambda\tau}] = \int_0^\infty e^{-\lambda\tau}\mathbb{P}\{\tau \in dt\}. \tag{9.2.1}$$

Since $\widetilde{X}$ stops when it hits state f,

$$\mathbb{P}\{\tau \leq t\} = \mathbb{P}\{\widetilde{X}_t = f\}, \tag{9.2.2}$$

so after an integration by parts, we can find an expression for the Laplace transform of τ:

$$\begin{aligned} \mathbb{E}[e^{-\lambda\tau}] &= \mathbb{P}\{\tau \leq t\}e^{-\lambda t}|_0^\infty + \lambda\int_0^\infty \mathbb{P}\{\tau \leq t\}e^{-\lambda t}dt & (9.2.3)\\ &= \lambda\int_0^\infty \mathbb{P}\{\tau \leq t\}e^{-\lambda t}dt & (9.2.4)\\ &= \int_0^\infty \lambda e^{-\lambda t}\mathbb{P}\{\widetilde{X}_t = f\}dt & (9.2.5)\\ &= \int_0^\infty \lambda e^{-\lambda t}(e^{tG^\star})_{s,f}dt & (9.2.6)\\ &= \lambda[(\lambda - G^\star)^{-1}]_{s,f}. & (9.2.7) \end{aligned}$$

Inverting this Laplace transform gives the probability density function of the first passage time from the starting state s to the final state f. In general, this can only be done numerically and will require an assignment of specific parameter values for the evaluation. However, by our assumptions G_{-f}, the submatrix of G obtained by removing the row and column corresponding to f, can be inverted, and so the nth moment can be computed analytically as (see Lemma 11.5.1 in Section 11.5 for a proof):

$$\mathbb{E}[\tau^n] = (-1)^n \frac{d^n}{d\lambda^n}\lambda[(\lambda - G^\star)^{-1}]_{s,f}\Big|_{\lambda=0} = n!\sum_y (-G_{-f})^{-(n+1)}_{sy} G^\star_{yf} \tag{9.2.8}$$

which which may be more computationally feasible than inverting the Laplace transform. If our measures of noisiness only involve the first few moments of τ, we will not need to invert the Laplace transform at all.

9.3 Measuring noisiness of assembly

To compare noisiness between different models of transcription or gene expression, we must decide what to measure the noisiness of, and we must have a measure of noisiness. There are (at least) two natural quantities to study: the amount of transcript produced by (fixed) time t, which we denote by $N(t)$; or the amount of time necessary before a (fixed) number or concentration of transcripts are produced, which we denote by S_n.

As reviewed in the introduction, most of the existing theoretical work on stochasticity in gene expression (e.g. [241, 87, 289]) works at a larger scale, and focuses on the "output" protein concentration at stationarity. (in their models, stationarity is well-defined) The stochasticity, or noisiness, of this quantity is often explicitly or implicitly defined as the standard deviation divided by the mean, a natural quantity that is generally known as the *coefficient of variation*. It is often more convenient to avoid square roots by using the square of this value, which we call the *squared coefficient of variation*, and denote for an arbitrary random variable X by

$$\eta[X] := \frac{\mathrm{var}[X]}{\mathbb{E}[X]^2}.$$

Another quantity that is somewhat less frequently studied is the variance of $N(t)$ divided by its mean, again at stationarity [83]. This quantity goes by a number of names, including *index of dispersion*, *variance-to-mean ratio*, or *Fano factor*. In this context, it is particularly useful as a measurement of deviation from the Poisson distribution, which has index of dispersion equal to one.

Our focus is on a somewhat different scale than the previous work — we study the distribution of τ, the random time necessary to make a single transcript. If $\tau_1, \tau_2, \ldots$, are independent copies of τ, and $S_n = \tau_1 + \cdots + \tau_n$, then S_n is the amount of time necessary to make n transcripts, and the number of products as a function of time is

$$N(t) = \max\{n \geq 0 : S_n \leq t\}.$$

In this model, $N(t)$ does not have a stationary version, but it is natural to consider a normalized measure of noisiness of either S_n or $N(t)$.

The squared coefficient of variation of S_n is just

$$\eta[S_n] = \frac{n \operatorname{var}[\tau]}{n^2 \mathbb{E}[\tau]^2} = \frac{1}{n}\eta[\tau].$$

On the other hand, the squared coefficient of variation of $N(t)$ is not in general easily expressible. However, the process N is a *renewal process*, about which there is a significant literature. The exact computations of moments of $N(t)$ is in general a difficult problem, but there are recent methods developed in Chaudry [46] that could be applied to our case. We will content ourselves with the well-known observation [112] that under mild conditions, as $t \to \infty$,

$$\begin{aligned} \mathbb{E}[N(t)] &\approx \frac{t}{\mathbb{E}[\tau]}, \quad \text{and} \\ \operatorname{var}[N(t)] &\approx \frac{\operatorname{var}[\tau] t}{\mathbb{E}[\tau]^3}, \quad \text{so} \\ \eta[N(t)] &\approx \frac{\mu}{t}\eta[\tau] = \frac{\operatorname{var}[\tau]}{t\mathbb{E}[\tau]}. \end{aligned} \tag{9.3.1}$$

We are led to either the squared coefficient of variation of τ, or the index of dispersion of τ. We develop methods to compute moments of τ, which are therefore applicable to either of the measures of noisiness discussed above.

In realistic models of transcription, the sequence of times $\tau_1, \tau_2, \ldots$ between successive transcription events are not identically distributed, as we assumed above — at least the first time has a different distribution. It could also be of interest to study the noisiness of τ_1, the time until the *first* transcript, if for instance the cellular mechanisms are sensitive enough to respond to a single transcript. Furthermore, this is the quantity that has been measured in certain cellular systems [25]. The distinction between τ and τ_1, however, is merely that of the "starting distribution" for our Markov chain, and so our methods are equally applicable to either case.

9.4 Application to the effects of gene regulation on noise

In this section we apply the general analytical framework presented above to study the properties of a simplified model of gene expression.

We propose the following simplified description of transcription, applicable to the gene regulatory mechanisms of inducible genes involved in embryonic patterning. There are two necessary steps: the *promoter* must be assembled, and the *enhancer* must be active.

We model the promoter assembly process as a three-state system: a *closed promoter* unassociated with any transcription factors; an *open promoter* with a loaded polymerase (RNA Pol II) ready to transcribe; and an *actively transcribing gene*. The promoter may switch back and forth between the closed and open state, depending on the arrival and stable binding of the appropriate transcription factors. Once in the actively transcribing state it may stall or pause but will only return to the previous states by successful completion of transcription. Having completed transcription, the system returns to the closed promoter state and polymerase loading can occur again. The corresponding three state Markov chain is shown in figure 9.1.

Regulation of this gene expression cascade depends on the state of an *enhancer*, which we add to our description of the system as the state of a second Markov chain, and vary the topology by making the transition rates of the promoter chain depend on the state of the enhancer chain. In this first model we allow the enhancer to take only two states, A or B. The enhancer modifies the behavior of the promoter chain by requiring that the enhancer be in state B for the promoter chain to make a certain transition step. We then vary the identity of this *paused* step and compare the resulting transcription time distributions.

If the step from *closed* to *open* (transition $1 \rightarrow 2$ in the Figure 9.1) is regulated by the enhancer chain, we say the process is *initiation regulated*, and require that the enhancer chain must be in state B for the promoter chain to leave the *closed* state, and conversely that while the promoter chain is *open*, the enhancer chain cannot leave state B.

On the other hand, if the step from *open* to *actively transcribing* (transition $2 \rightarrow 3$ in the Figure 9.1) is regulated by the enhancer chain, we say the process is *elongation regulated*, and require that the enhancer chain must be in state B for the promoter chain to move from the *open* state to the *actively transcribing* state. In this case, the enhancer chain is unconstrained by the promoter chain.

We can specify all the transition rates independently of the regulatory mechanism. In other words, in either model, the transition rates of any allowed transition are identical, so no parameters are introduced other than the identity of the paused transition in the promoter chain. Thus, we will not need to assume anything about the specific kinetics and can make predictions about changes that occur purely on the basis of the topological rearrangement of regulation. The two methods of coupling the enhancer and promoter chains define the two new Markov chains shown in figure 9.1.

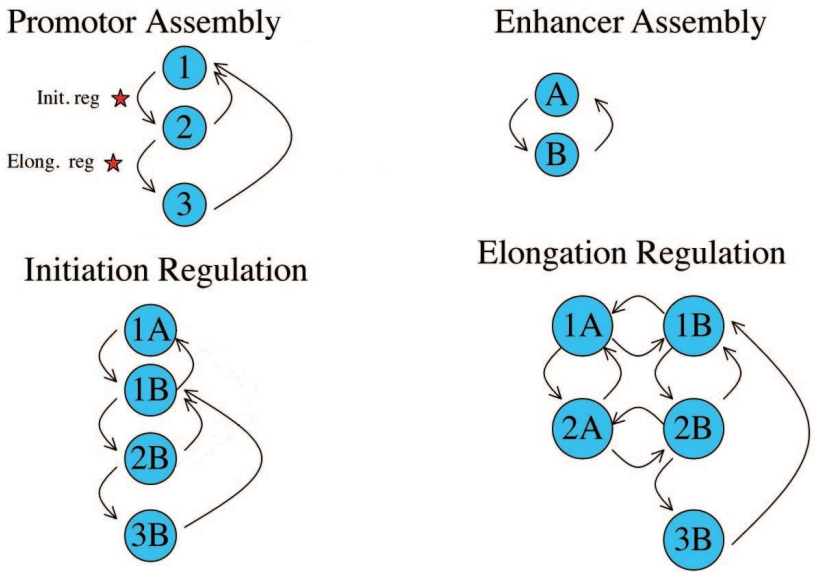

Figure 9.1: Two topologies for one process: Initiation controlled vs. elongation controlled regulation of gene expression

Having defined the system we can now compute the distribution of first passage times from a state with naked DNA to a state where the first mRNA is transcribed. The transition matrix for the initiation regulated model is

$$G_I^\star = \begin{array}{c} \\ 1A \\ 1B \\ 2B \\ 3B \end{array} \begin{array}{c} \begin{array}{cccc} 1A & 1B & 2B & 3B \end{array} \\ \begin{pmatrix} * & k_{ab} & 0 & 0 \\ k_{ba} & * & k_{12} & 0 \\ 0 & k_{21} & * & k_{23} \\ 0 & 0 & 0 & * \end{pmatrix} \end{array}, \tag{9.4.1}$$

where $*$ denotes the appropriate quantity so that the rows sum to zero. The elongation regulated model has transition matrix

$$G_E^\star = \begin{array}{c} \\ 1A \\ 2A \\ 1B \\ 2B \\ 3B \end{array} \begin{array}{c} \begin{array}{ccccc} 1A & 2A & 1B & 2B & 3B \end{array} \\ \begin{pmatrix} * & k_{12} & k_{ab} & 0 & 0 \\ k_{21} & * & 0 & k_{ab} & 0 \\ k_{ba} & 0 & * & k_{12} & 0 \\ 0 & k_{ba} & k_{21} & * & k_{23} \\ 0 & 0 & 0 & 0 & * \end{pmatrix} \end{array}. \tag{9.4.2}$$

We are interested in the time it takes to go from state $1A$ (enhancer in state A and promoter in state 1) to state $3B$ (enhancer in active state B and promoter in fully assembled state 3). Note that this corresponds to the *first* time to transcription — the time *between* successive transcripts would have the chains beginning in state $1B$. Write τ_I for the time to transcription in the initiation-regulated chain, and τ_E for the time to transcription in the enhancer-regulated chain.

If we apply Equation (9.2.8) to this situation, with the help of Mathematica, we find that regardless of the values of the rate parameters,

$$\mathbb{E}[\tau_I] > \mathbb{E}[\tau_E]$$
$$\text{and} \qquad \eta[\tau_E] > \eta[\tau_I].$$

In fact, both differences are monotone, separately, in each of the rate parameters, since it turns out that both $\mathbb{E}[\tau_I] - \mathbb{E}[\tau_E]$ and $\mathbb{E}[\tau_I^2]\mathbb{E}[\tau_E]^2 - \mathbb{E}[\tau_E^2]\mathbb{E}[\tau_I]^2$ are rational polynomials in the rate coefficients with strictly positive coefficients. (note that $\eta[\tau_E] < \eta[\tau_I]$ if and only if the latter expression is positive) We spare the reader the lengthy equations.

We may conclude that in this simplified model, elongation regulation is faster, but noisier, which could lead to interesting evolutionary tradeoffs. However, the strict monotonicity observed here is not expected to be a general property.

Chapter 10

Detailed regulatory models

The process of polymerase loading has been shown to require the independent binding of several complexes and thus is not best described by a simple exponentially distributed loading event as in three–state promoter chain of the previous chapter. In this chapter, we discuss more explicit models for both polymerase loading and enhancer assembly. In principle, the same computations described in the previous section apply, but they quickly become infeasible as the number of component TFs grows. We then discuss several decomposition techniques that could be used to simplify the computations and analytical comparisons to compare the network behavior.

In the present work we only develop analytical tools, and do not present results on these models, which are for the present only for motivation. Numerical and analytical results will be presented in a later work.

Figure 10.1A shows the major steps of the transcriptional initiation cascade, adapted from [241]. Some intermediate steps which don't greatly change the topology of the cascade have been suppressed, such as sequential binding of TFIIA and B or variable binding of mediating factors after TFIIB binding. Figure 10.1B illustrates the corresponding Markov chain, with step numbers as noted in 10.1A. Branches occur where alternate binding orders are observed. As in the previous chapter, this model could be either *initiation regulated* or *elongation regulated*, which corresponds to requiring that an auxiliary *enhancer chain* is in a specified state before the transition marked with a star is allowed.

We also consider two more complex enhancer chain structures: *fully cooperative* assembly where the transcription factors must associate with the complex in a specific order, and *fully non-cooperative* binding, where each of the transcription factors associates with

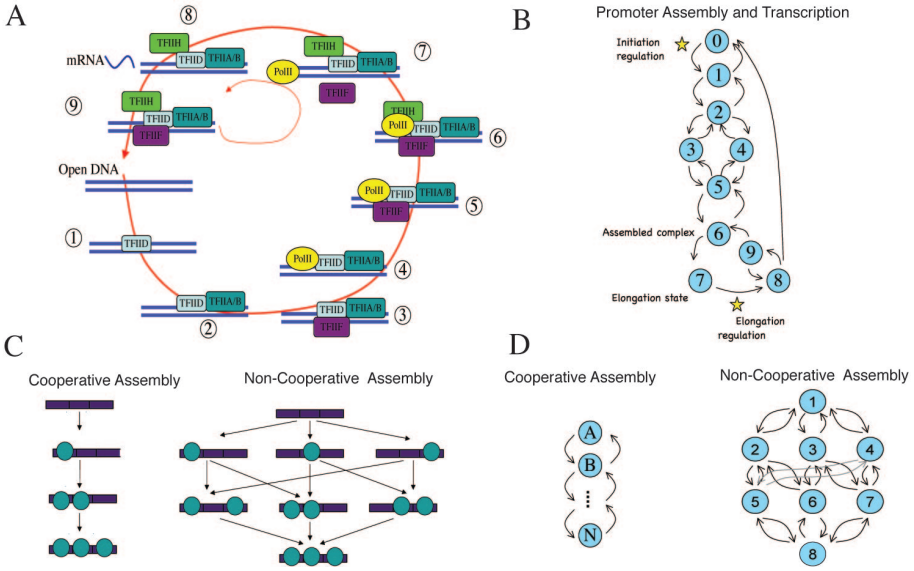

Figure 10.1: **Detailed Model of Gene Expression:** **A** Schematic representation of the molecular interactions of transcription at the promoter element. **B** Markov chain corresponding to the transcription process illustrated in A. **C** Schematic representation of two models of enhancer assembly. **D** Corresponding Markov chain representations for the models shown in **C**.

the complex independently. The fully cooperative model is a birth–death chain, while the fully non–cooperative model is a product of independent two–state chains.

Rather than write out a transition matrix for the each of these joint systems (which has dimension equal to the product of the number of states in the enhancer chain and the number of states in the promoter chain), we describe first how this chain can be decomposed into more readily analyzed pieces.

If we are only interested in the first hitting time of the state labeled 8 in the promoter chain of Figure 10.1B, we may ignore the arrows leading out of that state (and thus ignore state 9 as well). Using these transitions to model the time between successive transcripts presents no additional difficulties, but for now we consider the time until the first transcript. Without these arrows, we can separate the promoter chain into three "modules" by breaking it at states 1 and 6, as depicted schematically in Figure 10.2. These modules have the property that they are arranged in a linear order, and that the chain can only move between adjacent modules. This observation suggests a decomposition: if we know the way the chain moves within each module, and how it moves between modules, then we know everything about the chain. The chain moves between modules in a simple way — it is a birth–death chain, which is analytically tractable and well–studied. Within each module, we have to only analyze a smaller Markov chain, which we may do with the methods of the previous chapter. This subject of decomposition into sequential modules is the subject of Chapter 11.

We have not yet addressed the inclusion of the enhancer chain. Depending on how we choose to implement the regulation, we may still be able to decompose the entire chain into modules, as described above. This will be the case if, for instance, the step $X \to Y$ is regulated, and once the promoter chain reaches state X, the enhancer chain begins moving, and the promoter chain is stuck in X until the enhancer chain reaches its final state.

However, this is not the most commonly accepted mode of regulation. A better model might be for the promoter and enhancer chains to move independently, with the exception that the promoter chain cannot take the regulated transition unless the enhancer is in its final state. We refer to this sort of structure — in which we wait for two independent chains to simultaneously occupy some prespecified pair of states — as a *parallel* construction, to contrast with the case of *sequential* modules above.

It turns out that the decomposition techniques for parallel chains do not give such nice results as we find for sequentially arranged chains, but in Chapter 12 we discuss some

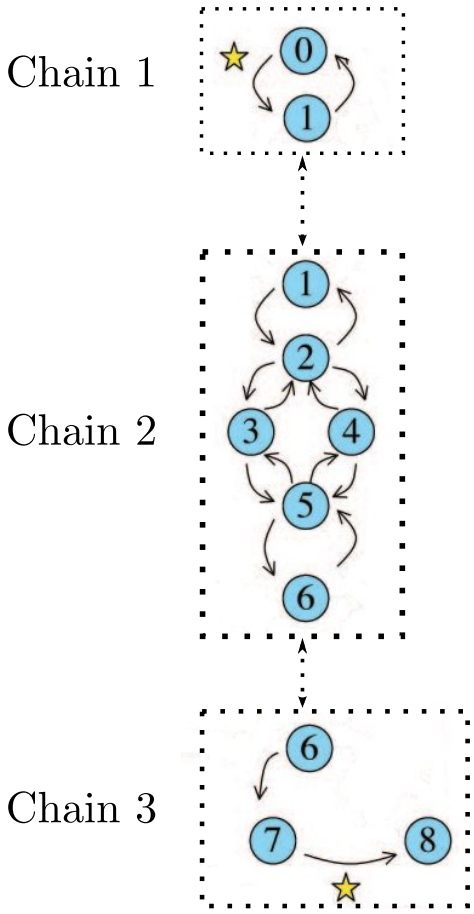

Figure 10.2: The promoter chain of Figure 10.1B broken into sequential "modules". Transitions after transcription (state 8) have been suppressed.

techniques that make their computation feasible, as well as some results on the contrasting enhancer chains described above.

Chapter 11

Decomposition into sequential modules

In this chapter we present and prove analytical tools for the decomposition of a detailed transcription model into sequential modules connected in a sequential manner. We present the framework in an abstract manner, but its connection will soon become clear.

Suppose we have a sequence of continuous-time Markov chains X^k on a sequence of state spaces $\mathcal{X}^k$, for $k \in \{1, 2, \ldots, n\}$. Suppose that each state space $\mathcal{X}^k$ has two distinguished (and distinct) states s_k and f_k. Each Markov chain X^k represents a single "stage" of the transcription factor assembly. We assume that f_k is accessible from any state in $\mathcal{X}^k$, for each k. The entire process of transcription is modeled by a Markov chain X which is constructed by stringing the state spaces sequentially, identifying s_k with f_{k-1} for $2 \leq k \leq n$, and leaving the infinitesimal transition rates the same. We call the state $f_k = s_{k+1}$ the k^{th} *pinch point*, and denote it by p_k.

For some state $b \in \mathcal{X}^k$ and a Markov chain Y on $\mathcal{X}^k$, define

$$\tau_b(Y) = \inf\{t > 0 : Y(t) = b\},$$

the time it takes the chain Y to first arrive at b.

Once X^k leaves s_k, there are several possible behaviors, and we need to introduce chains that act as X^k conditioned on each behavior. For each k, let $\nu_s^k(\cdot)$ denote the distribution of X^k after the first jump from s_k, namely, if T is the time of the first jump, then

$$\nu_s^k(i) = \mathbb{P}\{X_T = i | X_0 = s\}.$$

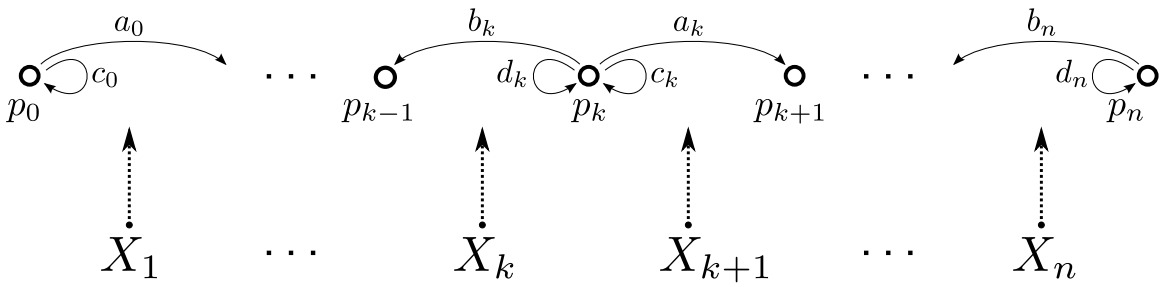

Figure 11.1: A schematic of the decomposition. The probabilities a_k, b_k, c_k, and d_k depend only on the distributions of both adjacent chains X_k and X_{k+1}, while the behavior of X between pinch points p_{k-1} and p_k only depends on the distribution of X_k.

Similarly, let $\nu_f^k(\cdot)$ denote the distribution of X^k after the first jump from f_k. Let $X^k_{\rightarrow}$ be a Markov chain on $\mathcal{X}^k$ that has the distribution of X^k begun with distribution ν_s^k and conditioned to hit f_k before returning to s_k; also let $X^k_{\circlearrowleft}$ have the distribution of X^k begun with distribution ν_s^k and conditioned to return to s_k before hitting f_k. Define $X^k_{\leftarrow}$ and $X^k_{\circlearrowright}$ similarly but with the roles of s_k and f_k reversed. Then define the four random traversal times

$$\begin{aligned}
\tau^k_{\rightarrow} &= \tau_{f_k}(X^k_{\rightarrow}) \\
\tau^k_{\leftarrow} &= \tau_{s_k}(X^k_{\leftarrow}) \\
\tau^k_{\circlearrowleft} &= \tau_{s_k}(X^k_{\circlearrowleft}) \\
\tau^k_{\circlearrowright} &= \tau_{f_k}(X^k_{\circlearrowright}).
\end{aligned} \tag{11.0.1}$$

Denote the pinch points $p_0, \ldots, p_n$, where $p_0 = s_1$, $p_n = f_n$, and $p_k = f_k$, which has been identified with s_{k+1}, for $1 \leq k < n$. If the chain X is at the k^{th} pinch point p_k, for $1 \leq k < n$, then it has four options with the following corresponding probabilities:

$$\begin{aligned}
a_k &= \mathbb{P}\{ \text{ hit } p_{k+1} \text{ without returning to } p_k \text{ }\}, \\
b_k &= \mathbb{P}\{ \text{ hit } p_{k-1} \text{ without returning to } p_k\}, \\
c_k &= \mathbb{P}\{ \text{ move into } \mathcal{X}^{k+1} \text{ but return to } p_k \text{ before hitting } p_{k+1} \text{ }\}, \\
\text{or} d_k &= \mathbb{P}\{ \text{ move into } \mathcal{X}^k \text{ but return to } p_k \text{ before hitting } p_{k-1} \text{ }\}.
\end{aligned} \tag{11.0.2}$$

If X is at either of the pinch points p_0 or p_n it has only two of these options. Once we choose one of these options, X then moves like a conditioned X^k chain until it hits a pinch point again. For instance, if the event with probability a_k happens, then p_{k+1} will be the

next pinch point hit, and until p_{k+1} is hit, X moves like the chain $X^k_{\rightarrow}$. The respective probabilities are computed in Chapter 11.3.

When the chain leaves a pinch point and returns, it could have done so in either direction, so we combine $\tau_{\circlearrowleft}$ and $\tau_{\circlearrowright}$ to form an additional traversal time. For each $1 \leq k \leq n$ let $\tau^k_{\circ}$ be a mixture of $\tau^{k+1}_{\circlearrowleft}$ and $\tau^k_{\circlearrowleft}$, defined by

$$\tau^k_{\circ} = \begin{cases} \tau^{k+1}_{\circlearrowleft} & \text{with probability } \frac{c_k}{c_k+d_k}, \\ \tau^k_{\circlearrowleft} & \text{with probability } \frac{d_k}{c_k+d_k}. \end{cases} \tag{11.0.3}$$

If $c_k = 0$ and $d_k > 0$ then $\tau^k_{\circ} = \tau^{k+1}_{\circlearrowleft}$, if $c_k > 0$ and $d_k = 0$ then $\tau^k_{\circ} = \tau^k_{\circlearrowleft}$, and if $c_k = d_k = 0$ then we define $\tau^k_{\circ} = 0$ (although this will not enter into the computations).

The glue that joins the modules together is the "pinch chain" Z, defined to be the discrete-time Markov chain that records the order in which X visits the pinch points. Formally, Z is a Markov chain on $\{0, 1, \ldots, n\}$ that at each step either moves up by one, down by one, or stays put, and the transition probabilities are for $0 \leq k \leq n$

$$\mathbb{P}\{Z_{k+1} = j | Z_k = i\} = \begin{cases} a_i, & \text{if } j = i+1 \leq n, \\ b_i, & \text{if } j = i-1 \geq 0, \\ c_i + d_i, & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{11.0.4}$$

We define P to be the transition matrix for the chain Z stopped upon hitting n, so that

$$\begin{aligned} P_{ij} &= \mathbb{P}\{Z_1 = j | Z_0 = i\} \quad \text{for } 0 \leq i < n,\ 0 \leq j \leq n, \\ P_{nj} &= 0 \quad \text{for } 0 \leq j < n, \\ \text{and} \quad P_{nn} &= 1. \end{aligned} \tag{11.0.5}$$

We discuss computation of P in Section 11.3.

Finally, for each pinch point $0 \leq k < n$, define an independent random variable S^k with exponential distribution

$$\mathbb{P}\{S^k > t\} = \exp\left\{-t(r^k(f) + r^{k+1}(s))\right\}, \tag{11.0.6}$$

where $r^k(f)$ is the jump rate out of f_k for X^k, and $r^{k+1}(s)$ is the jump rate out of s_{k+1} for X^{k+1}. S^k has the distribution of the amount of time X spends at p_k before moving.

11.1 Computing system noise properties

Now we are ready to state our main theorems. First we give the form of the Laplace transform and the moments of the assembly time in terms of the transition probabilities between modules and the distributions of the traversal times. In Chapter 11.3 we discuss how to compute the transition probabilities, and in Chapter 11.4 we discuss how to compute the relevant quantities for the traversal times.

Theorem 11.1.1. *Let P be as defined in* (11.0.5) *and define for $0 \leq j < n$ and $0 \leq k \leq n$*

$$\tau_{jk} = \begin{cases} \tau_{\leftarrow}^{j} + S^{j} & \text{if } k = j-1 \text{ and } P_{jk} > 0, \\ \tau_{\circ}^{j} + S^{j} & \text{if } k = j \text{ and } P_{jk} > 0, \\ \tau_{\rightarrow}^{j+1} + S^{j} & \text{if } k = j+1 \text{ and } P_{jk} > 0, \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

where the random variables are as defined as in (11.0.1), (11.0.3) *and* (11.0.6). *Furthermore, define*

$$\phi_{jk}(\lambda) = \mathbb{E}[\exp(-\lambda \tau_{jk})],$$

and define the $n \times n$ matrix Z having entries $Z_{jk} = \phi_{jk}(\lambda) P_{jk}$ for $0 \leq j, k \leq n-1$. Let $v(\lambda)$ be a vector with $v_n(\lambda) = 1$ and

$$v_i(\lambda) = \sum_{j=0}^{n-1} (I - Z)_{ij}^{-1} \phi_{jn}(\lambda) P_{jn},$$

for $0 \leq i < n$.

Then the total time to assembly, $\tau = \inf\{t > 0 : X_t = f_n\}$, has Laplace transform

$$\mathbb{E}[\exp(-\lambda \tau) \mid X_0 = p_i] = v_i(\lambda),$$

for $0 \leq i \leq n$.

Corollary 11.1.1. *Define the matrices M, σ, and R by*

$$M_{ij} = P_{ij} \mathbb{E}[\tau_{ij}] \quad 0 \leq i < n,\ 0 \leq j \leq n \tag{11.1.1}$$

$$\Sigma_{ij} = P_{ij} \mathbb{E}[\tau_{ij}^2], \quad 0 \leq i < n,\ 0 \leq j \leq n \tag{11.1.2}$$

$$\text{and} \quad R_{ij} = \begin{cases} (I - P_{-n})_{ij}^{-1} & 0 \leq i < n,\ 0 \leq j < n \\ 0 & i = n,\ 0 \leq j < n\ . \end{cases} \tag{11.1.3}$$

Then the first and second moments of the total assembly time τ can be written as follows. For $0 \le i < n$,

$$\mathbb{E}[\tau \mid X_0 = p_i] = (RM\mathbf{1})_i, \quad \textit{and} \tag{11.1.4}$$

$$\mathbb{E}[\tau^2 \mid X_0 = p_i] = \left(R\Sigma\mathbf{1} + 2(RM)^2\mathbf{1}\right)_i. \tag{11.1.5}$$

To prove the theorem, we decompose the path of X by first looking at the order in which X traverses the pinch points — the sample path of the pinch chain Z — and then depending on the path that Z takes, add in the appropriate random amounts of time for each step. The Laplace transform of the assembly time will be put together from two pieces: the joint probability generating function of the transition counts of the pinch chain; and the Laplace transforms of the relevant traversal times.

Let $Z_0 = 0$, and let T be the first time that Z hits n, after which Z stays fixed. Define for each j, k the transition counts

$$N_{jk} = \#\{1 \le i \le T : Z_{i-1} = j \text{ and } Z_i = k\},$$

the number of times Z moves from j to k. If k is not one of $j-1$, j, or $j+1$, then N_{jk} will be zero.

The following lemma giving the form of the joint probability generating function of the transition counts is proved in Section 11.5.

Lemma 11.1.1. *Let $\{s_{ij}\}$ be a set of dummy variables with $s_{ij} \in [0,1]$ for all $0 \le i, j \le n$. Define the matrix $P(s)_{jk} = s_{jk}P_{jk}$ and define the vector $v(s)$ by $v(s)_n = 1$ and*

$$v(s)_i = \sum_{j=0}^{n-1} \left((I - P(s)_{-n})^{-1}\right)_{ij} P(s)_{jn} \quad \textit{for } 0 \le i < n, \tag{11.1.6}$$

where $P(s)_{-n}$ is the matrix $P(s)$ with the last row and column removed. Then

$$\mathbb{E}\left[\prod_{jk} s_{jk}^{N_{jk}} \,\middle|\, Z_0 = i\right] = v_i(s).$$

Suppose $Z_i = k$, indicating that X is at p_k. The amount of time before X leaves p_k has the distribution of S^k, so we need to add an independent copy of S^k. If $Z_{i+1} = k+1$, then X will hit p_{k+1} before returning to p_k. By construction, the amount of time this takes has the same distribution as $\tau_{\rightarrow}^{k+1}$, so we need to add on a copy of $\tau_{\rightarrow}^{k+1}$, whose value is

independent of everything else. Similarly, if $Z_i = k$ and $Z_{i+1} = k-1$, we need to add on a copy of $\tau^k_{\leftarrow}$. If $Z_i = Z_{i+1} = k$, then this corresponds to a single excursion from the k^{th} pinch point, which could have been in either direction. In this case, we need to add a random time $\tau^k_{\circ}$ that is a mixture of the distribution of $\tau^{k+1}_{\circlearrowleft}$ with probability $c_k/(c_k+d_k)$ and the distribution of $\tau^k_{\circlearrowleft}$ with probability $d_k/(c_k+d_k)$, as defined in (11.0.3).

Let the total time to assembly be denoted τ, and for each k let $\tau^k_{\rightarrow,1}, \tau^k_{\rightarrow,2}, \ldots$ be an infinite sequence of independent copies of $\tau^k_{\rightarrow}$; define $\tau^k_{\leftarrow,m}$ and $\tau^k_{\circ,m}$ for $m \geq 1$ similarly. Also let $S^k_{1,m}$, $S^k_{2,m}$, and $S^k_{3,m}$ be three infinite sequences of independent copies of S^k. Our decomposition by the path of Z tells us that τ is distributed as

$$\sum_{k=0}^{n}\left(\sum_{m=1}^{N_{k,k-1}}\left(\tau^k_{\leftarrow,m}+S^k_{1,m}\right)+\sum_{m=1}^{N_{k,k}}\left(\tau^k_{\circ,m}+S^k_{2,m}\right)+\sum_{m=1}^{N_{k,k+1}}\left(\tau^{k+1}_{\rightarrow,m}+S^k_{3,m}\right)\right).$$

Therefore, by conditioning on Z, we get that

$$\begin{aligned}\mathbb{E}\left[e^{-\lambda\tau}\right] &= \mathbb{E}\left[\prod_{k=0}^{n}\left(\prod_{m=1}^{N_{k,k-1}} e^{-\lambda(\tau^k_{\leftarrow,m}+S^k_{1,m})}\prod_{m=1}^{N_{k,k}} e^{-\lambda(\tau^k_{\circ,m}+S^k_{2,m})}\prod_{m=1}^{N_{k,k+1}} e^{-\lambda(\tau^{k+1}_{\rightarrow,m}+S^k_{3,m})}\right)\right]\\ &= \mathbb{E}\left[\prod_{k=0}^{n}\left(\mathbb{E}\big[e^{-\lambda(\tau^k_{\leftarrow}+S^k)}\big]^{N_{k,k-1}}\mathbb{E}\big[e^{-\lambda(\tau^k_{\circ}+S_k)}\big]^{N_{k,k}}\mathbb{E}\big[e^{-\lambda(\tau^{k+1}_{\rightarrow}+S_k)}\big]^{N_{k,k+1}}\right)\right].\end{aligned}$$

This proves Theorem 11.1.1.

Several notes are in order. One is that since, for instance, S^k and $\tau^k_{\circlearrowleft}$ are independent, their Laplace transforms and moments may be computed separately. Furthermore,

$$\mathbb{E}\left[e^{-\lambda\tau^k_{\circ}}\right] = \frac{1}{c_k+d_k}\left(c_k\mathbb{E}\left[e^{-\lambda\tau^{k+1}_{\circlearrowleft}}\right]+d_k\mathbb{E}\left[e^{-\lambda\tau^k_{\circlearrowleft}}\right]\right). \tag{11.1.7}$$

We now prove the corollary. We may, without loss of generality, take $X_0 = p_0$, and leave this implicit, writing for instance $\mathbb{E}[\tau] = \mathbb{E}[\tau|X_0 = p_0]$. First note that by differentiating Lemma 11.1.1, we get that

$$\begin{aligned}\mathbb{E}[\tau] &= \sum_{j=0}^{n-1}\sum_{k=0}^{n}\mathbb{E}[\tau_{jk}]\partial_{s_{jk}}v_0(1), \quad \text{and}\\ \mathbb{E}[\tau^2] &= \sum_{j=0}^{n-1}\sum_{k=0}^{n}\mathbb{E}[\tau^2_{jk}]\partial_{s_{jk}}v_0(1)+\sum_{j=0}^{n-1}\sum_{k=0}^{n}\sum_{\ell=0}^{n-1}\sum_{m=0}^{n}\mathbb{E}[\tau_{jk}]\mathbb{E}[\tau_{\ell m}]\partial_{s_{jk}}\partial_{s_{\ell m}}v_0(1).\end{aligned} \tag{11.1.8}$$

First we compute the derivatives of v at $s=1$. Write $\partial_{s_{jk}}$ as ∂_{jk}. Since we know that $P(s)v(s) = v(s)$,

$$\partial_{jk}v(s) = (\partial_{jk}P(s))v(s)+P(s)\partial_{jk}v(s).$$

Now, since $v(1) = (1, 1, 1, \ldots, 1)^T$ and

$$(\partial_{jk} P(s))_{qr} = \begin{cases} P_{jk} & \text{if } q = j, \text{ and } r = k, \\ 0 & \text{otherwise,} \end{cases}$$

we get that $\partial_{jk} v(1)$ solves the set of linear equations $(I - P)\partial_{jk} v(1) = P_{jk} e_j$, where e_j is the j^{th} standard basis vector, or more explicitly,

$$\partial_{jk} v(1)_i - \sum_r P_{ir} \partial_{jk} v(1)_r = \begin{cases} P_{jk}, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{11.1.9}$$

Since we require that $v(s)_n = 1$, we have that $\partial_{jk} v(s)_n = 0$.

For the moment write I_m as the identity matrix of order m. Since $(I_{n+1} - P)$ is the transition matrix for an irreducible continuous-time Markov chain stopped upon reaching n, we know that $(I_n - P_{-n})$ is invertible, and so $(I_n - P_{-n})^{-1}$ exists. Define R to be $(I_n - P_{-n})^{-1}$ with an extra row of zeros at the bottom, as in the statement of the corollary, and let w be the j^{th} column of R. It is easy to check that $(I - P)w = e_j$, and that this solution is unique up to scaling by a constant, so that

$$\partial_{jk} v(1)_i = R_{ij} P_{jk}$$

for $0 \leq i < n$ and $0 \leq j, k \leq n$.

If we differentiate the identity a second time, we get that

$$\partial_{jk}\partial_{lm} v(s) = (\partial^2_{jk} P(s))v(s) + (\partial_{jk} P(s))\partial_{lm} v(s) + (\partial_{lm} P(s))\partial_{jk} v(s) + P(s)\partial^2_{jk} v(s). \tag{11.1.10}$$

Now let $u = \partial_{jk}\partial_{lm} v(1)$. Since $\partial_{jk}\partial_{lm} P(s) = 0$, we have that u satisfies $(I_{n+1} - P)u = P_{jk}\partial_{lm} v(1)_k e_j + P_{lm}\partial_{jk} v(1)_m e_l$, or (using our solution for the first derivative):

$$u_i - \sum_r P_{ir} u_r = \begin{cases} P_{jk} P_{lm} R_{kl}, & \text{if } i = j, \\ P_{lm} P_{jk} R_{mj}, & \text{if } i = l, \\ 0, & \text{otherwise.} \end{cases} \tag{11.1.11}$$

By linearity, we can use our solution from above to solve this, so that

$$\partial_{jk}\partial_{lm} v(s)_i = P_{jk} P_{lm} \left(R_{il} R_{mj} + R_{ij} R_{kl} \right). \tag{11.1.12}$$

Evaluating the sums in (11.1.8) gives (11.1.4).

11.2 Transition counts for the pinch-chain

In this section we prove Lemma 11.1.1.

Recall that Z is "X watched only at the pinch points," and so is a discrete-time Markov chain on $\{0, \ldots, n\}$ that at each step either moves up by one, down by one, or stays put — also known as a birth-death chain.

Let $Z_0 = 0$, and let T be the first time that Z hits n, at which time we stop. Define for each j and k

$$N_{jk} = \#\{1 \leq i \leq T : Z_{i-1} = j \text{ and } Z_i = k\},$$

the number of times Z moves from j to k. If k is not one of $j-1$, j, or $j+1$, then N_{jk} will be zero.

Let P be the transition matrix for the chain Z *stopped* (not killed) upon hitting n. For a collection of dummy variables $\{s_{jk}\} \in [0,1]^{n^2}$, define the matrix $P(s)_{jk} = s_{jk}P_{jk}$. Then it is easy to see that the joint probability-generating function for the N_{jk} is given by

$$\mathbb{E}\left[\prod_{jk} s_{jk}^{N_{jk}}\right] = \lim_{m\to\infty} (P^m(s))_{0,n}.$$

For this to be nonzero, we must take $s_{nn} = 1$. Also, since $P(s)$ is substochastic and the last row of $P(s)$ is zero except with 1 on the diagonal, $P(s)$ has a single eigenvalue of value 1, with left eigenvector $\pi = (0, 0, \ldots, 1)$. It is known (*Matrix Analysis*, Horn and Johnson [152]) that any tridiagonal matrix (a_{ij}) satisfying $a_{i,i+1}a_{i+1,1} > 0$ for all i is similar to a Hermitian matrix, and in particular has a full complement real eigenvalues with corresponding left and right eigenvectors. Therefore, we know that there is a unique right eigenvector with eigenvalue 1, which we call $v(s)$, and normalize so that $v_n(s) = 1$. The other eigenvalues are strictly less than one, so

$$\lim_{m\to\infty} (P^m(s))_{jk} = \pi_k v_j(s)$$

whence

$$\mathbb{E}\left[\prod_{jk} s_{jk}^{N_{jk}}\right] = v_0(s).$$

Now we need a solution to $P(s)v(s) = v(s)$, normalized so that $v_n(s) = 1$. By Lemma

11.5.1, the unique solution is given by

$$v(s)_i = \sum_{j=0}^{n-1} (I - P_{-n})^{-1}_{ij} P_{jn}.$$

Note that it is also easy to see by writing down the one-step recursion that $v_i(s) = \mathbb{E}\left[\prod_{jk} s_{jk}^{N_{jk}} \,\middle|\, Z_0 = i\right]$ is a solution to $P(s)v(s) = v(s)$.

11.3 Transition probabilities between modules

Now we compute the transition probabilities (a_k, b_k, c_k, d_k), defined in (11.0.2).

For reference, the matrix P is tridiagonal with entries:

$$P = \begin{bmatrix} 1-a_0 & a_0 & & & & \\ b_1 & 1-(b_1+a_1) & a_1 & & & \\ & b_2 & 1-(b_2+a_2) & a_2 & & \\ & & \ddots & & & \\ & & b_{n-1} & 1-(b_{n-1}+a_{n-1}) & a_{n-1} \\ & & & 0 & 1 \end{bmatrix}. \tag{11.3.1}$$

Suppose X^k jumps out of f_k at rate r_f^k, X^{k+1} jumps out of s_{k+1} at rate r_s^{k+1}, and the probability that X^{k+1} begun at s_{k+1} reaches f_{k+1} before returning to s_{k+1} is q. Then

$$a_k = q r_s^{k+1} / (r_f^k + r_s^{k+1}). \tag{11.3.2}$$

Similarly, if the probability that X^k begun at f_k reaches s_k before returning to f_k is q', then

$$b_k = q' r_f^k / (r_f^k + r_s^{k+1}).$$

Also $c_k = (1-q) r_s^{k+1} / (r_f^k + r_s^{k+1})$ and $d_k = (1-q') r_f^k / (r_f^k + r_s^{k+1})$. We need only find a method to compute q and q' for each component chain X^k separately.

For the remainder of this section, let Y be an irreducible continuous-time Markov chain with distinguished states s and f, and matrix of transition rates G, with $G_{ii} = -\sum_{j \neq i} G_{ij}$ as usual.

Define G^{**} to be the matrix of transition rates for Y stopped at both s and f, e.g.

$$G_{ij}^{**} = \begin{cases} G_{ij} & \text{if } i \neq s, f \\ 0 & \text{otherwise.} \end{cases}$$

Note that if we define

$$x_i = \mathbb{P}\{Y \text{ hits } f \text{ before } s \mid Y_0 = i\}, \tag{11.3.3}$$

then it is well-known that [111]

$$\sum_j G^{**}_{ij} x_j = 0 \quad \text{for all } i.$$

In other words, x is the unique right eigenvector of G^{**} corresponding to the zero eigenvalue that satisfies the boundary conditions $x_s = 0$ and $x_f = 1$.

In fact, we know by Lemma 11.5.1 that if we denote by G_{-sf} the submatrix obtained from G by removing rows and columns corresponding to both s and f, then for $i \notin \{s, f\}$,

$$x_i = \sum_{j \notin \{s,f\}} (-G_{-sf})^{-1}_{ij} G_{jf}. \tag{11.3.4}$$

What we really need is $\mathbb{P}\{Y \text{ hits } f \text{ before } s \mid Y_0 = s\}$, i.e. the probability q in (11.3.2). To find this, we condition on the first step. Let S be the time that Y first leaves s, which is exponentially distributed with rate $-G_{ss} = \sum_{j \neq s} G_{sj}$.

$$\begin{aligned}\mathbb{P}\{Y \text{ hits } f \text{ before } s \mid Y_0 = s\} &= \sum_i \mathbb{P}\{Y_S = i\} \\ &\qquad \times \mathbb{P}\{Y \text{ hits } f \text{ before } s \mid Y_0 = i\} \\ &= \sum_i \frac{G_{si}}{-G_{ss}} x_i.\end{aligned}$$

We now summarize the above computations. Let (G^k, x^k) correspond to the k^{th} chain and let (a_k, b_k, c_k, d_k) be defined as in (11.0.2). Then $b_0 = d_0 = a_n = c_n = 0$, and if we take $G^0_{ff} = G^{n+1}_{ss} = 0$,

$$\begin{aligned} a_k &= \frac{G^{k+1}_{ss}}{G^{k+1}_{ss} + G^k_{ff}} \sum_{i \neq s} \frac{G^{k+1}_{si}}{-G^{k+1}_{ss}} x^{k+1}(i) \\ &= -\sum_{i \neq s} \frac{G^{k+1}_{si}}{G^{k+1}_{ss} + G^k_{ff}} x^{k+1}(i) \qquad \text{for } 0 \leq k < n \\ &= \frac{1}{-G^{k+1}_{ss} - G^k_{ff}} \left(G^{k+1}_{sf} + \sum_{i \notin \{s,f\}} \sum_{j \notin \{s,f\}} G^{k+1}_{si} (G^{k+1}_{-sf})^{-1}_{ij} G^{k+1}_{jf} \right) \end{aligned} \tag{11.3.5}$$

and similarly

$$b_k = -\sum_{i \neq f} \frac{G^k_{fi}}{G^{k+1}_{ss} + G^k_{ff}}(1 - x^k(i)) \qquad \text{for } 0 < k \leq n \tag{11.3.6}$$

$$c_k = -\sum_{i \neq s} \frac{G^{k+1}_{si}}{G^{k+1}_{ss} + G^k_{ff}}(1 - x^{k+1}(i)) \qquad \text{for } 0 \leq k < n \tag{11.3.7}$$

$$d_k = -\sum_{i \neq f} \frac{G^k_{fi}}{G^{k+1}_{ss} + G^k_{ff}} x^k(i) \qquad \text{for } 0 < k \leq n. \tag{11.3.8}$$

Note that we can also express

$$1 - x^k_i = \sum_{j \notin \{s,f\}} (-G_{-sf})^{-1}_{ij} G_{js}.$$

11.4 Traversal times within modules

Here we show how to compute quantities related to the traversal times. Again let Y be an irreducible Markov chain with transition matrix G, and let G^{**} be the transition matrix for Y stopped upon hitting either s or f, so that $G^{**}_{ij} = G_{ij}$ for $i \notin \{s, f\}$ and $G_{sj} = G_{fj} = 0$.

Lemma 11.4.1. *Let $\tau_{\rightarrow}$, $\tau_{\leftarrow}$, $\tau_{\circlearrowright}$, and $\tau_{\circlearrowleft}$ be defined as in* (11.0.1) *for a chain with matrix of transition probabilities G. Let $x_s = 0$, $x_f = 1$, and*

$$x_i = \sum_{j \notin \{s,f\}} (-G_{-sf})^{-1}_{ij} G_{jf}. \tag{11.4.1}$$

The Laplace transforms are then given by

$$\mathbb{E}[e^{-\lambda \tau_{\rightarrow}}] = \sum_{i \notin \{s,f\}: x_i > 0} \frac{G_{si}}{-G_{ss}} \frac{\lambda \left((\lambda - G^{**})^{-1}\right)_{if}}{x_i}, \tag{11.4.2}$$

$$\mathbb{E}[e^{-\lambda \tau_{\leftarrow}}] = \sum_{i \notin \{s,f\}: x_i < 1} \frac{G_{fi}}{-G_{ff}} \frac{\lambda \left((\lambda - G^{**})^{-1}\right)_{is}}{(1 - x_i)}, \tag{11.4.3}$$

$$\mathbb{E}[e^{-\lambda \tau_{\circlearrowright}}] = \sum_{i \notin \{s,f\}: x_i < 1} \frac{G_{si}}{-G_{ss}} \frac{\lambda \left((\lambda - G^{**})^{-1}\right)_{is}}{(1 - x_i)}, \text{ and} \tag{11.4.4}$$

$$\mathbb{E}[e^{-\lambda \tau_{\circlearrowleft}}] = \sum_{i \notin \{s,f\}: x_i > 0} \frac{G_{fi}}{-G_{ff}} \frac{\lambda \left((\lambda - G^{**})^{-1}\right)_{if}}{x_i}. \tag{11.4.5}$$

Corollary 11.4.1. *The moments of the traversal times are*

$$\begin{aligned}\mathbb{E}[\tau_{\rightarrow}^m] &= m! \sum_{i,j\notin\{s,f\}:x_i>0} \frac{G_{si}(-G_{-sf})^{-(m+1)}_{ij}G_{jf}}{(-G_{ss})x_i} \\ \mathbb{E}[\tau_{\circlearrowright}^m] &= m! \sum_{i,j\notin\{s,f\}:x_i<1} \frac{G_{si}(-G_{-sf})^{-(m+1)}_{ij}G_{js}}{(-G_{ss})(1-x_i)}\end{aligned} \tag{11.4.6}$$

and the moments of $\tau_{\leftarrow}$ and $\tau_{\circlearrowleft}$ can be computed by exchanging the roles of s and f (which also exchanges x_i and $(1-x_i)$).

Note that $\mathbb{E}[e^{-\lambda\tau_{\circ}^k}]$ can be evaluated by substituting these into (11.1.7).

Note 4. To use these in Theorem 11.1.1 we need to translate the τ_{jk} defined there into combinations of the above traversal times. For convenience, we record here which entries of the matrices ϕ, M or Σ depend on which traversal times. The following $(n+1)\times(n+1)$ matrix is tridiagonal, and the $(j,k)^{\text{th}}$ entry contains the random variables on whose distributions the $(j,k)^{\text{th}}$ entry of ϕ, M, or Σ depend. The empty entries are those that do not need to be computed, as they are identically zero.

$$\phi, M, \Sigma \quad \text{depend on} \quad \begin{bmatrix} \tau_{\circlearrowright}^1 & \tau_{\rightarrow}^1 & & & \\ \tau_{\leftarrow}^1 & (\tau_{\circlearrowright}^1, \tau_{\circlearrowright}^2) & \tau_{\rightarrow}^2 & & \\ & \tau_{\leftarrow}^2 & (\tau_{\circlearrowright}^2, \tau_{\circlearrowright}^3) & \tau_{\rightarrow}^3 & \\ & & \ddots & & \\ & & \tau_{\leftarrow}^{n-1} & (\tau_{\circlearrowright}^{n-1}, \tau_{\circlearrowright}^n) & \tau_{\rightarrow}^n \\ & & & 0 & 0 \end{bmatrix}. \tag{11.4.7}$$

Compare this to (11.3.1).

Also recall that S^k is exponentially distributed with rate $-G_{ss}^{k+1} - G_{ff}^k$, and so

$$\begin{aligned}\mathbb{E}[e^{-\lambda S^k}] &= \frac{-G_{ss}^{k+1} - G_{ff}^k}{\lambda - G_{ss}^{k+1} - G_{ff}^k}, \\ \text{and} \quad \mathbb{E}[(S^k)^n] &= \frac{n!}{(-G_{ss}^{k+1} - G_{ff}^k)^n}.\end{aligned}$$

Note 5. At first sight, (11.4.2) doesn't appear to give the right answer at $\lambda = 0$. However, recall that G^{**} is not invertible. As is shown in Lemma 11.5.1 of Section 11.5,

$\left[\lim_{\lambda\to 0}\lambda(\lambda - G^{**})^{-1}\right]_{if} = x_i$, so

$$\begin{aligned}\lim_{\lambda\to 0}\mathbb{E}[e^{-\lambda\tau_{\rightarrow}}] &= \sum_{i\neq s}\frac{G_{si}}{-G_{ss}}\frac{x_i}{x_i} \\ &= 1.\end{aligned} \tag{11.4.8}$$

Proof. Let $Y^{\rightarrow}$ denote the chain Y conditioned to hit f before hitting s, and let Y^{**} be the chain Y stopped upon hitting either s or f. Denote by $A_{\rightarrow}$ the event that Y hits f before hitting s. We can compute for $i \notin \{s, f\}$:

$$\begin{aligned}\mathbb{P}\{Y_t^{\rightarrow} = j | Y_0^{\rightarrow} = i\} &= \frac{\mathbb{P}\{Y_t = j, A_{\rightarrow}|Y_0 = i\}}{\mathbb{P}\{A_{\rightarrow}|Y_0 = i\}} \\ &= \frac{\mathbb{P}\{Y_t^{**} = j|Y_0^{**} = i\}\mathbb{P}\{A_{\rightarrow}|Y_0 = j\}}{\mathbb{P}\{A_{\rightarrow}|Y_0 = i\}} \\ &= \left(e^{tG^{**}}\right)_{ij}\frac{x_j}{x_i}.\end{aligned}$$

Therefore, if $\tau_{\rightarrow}$ is the first time that $Y^{\rightarrow}$ hits f, and S is the first time that Y leaves s, then by conditioning on S and Y_S,

$$\begin{aligned}\mathbb{E}[e^{-\lambda\tau_{\rightarrow}}] &= \sum_{i\neq s}\frac{G_{si}}{-G_{ss}}\lambda\int_0^{\infty}\mathbb{P}\{Y_t^{\rightarrow} = f|Y_0^{\rightarrow} = i\}e^{-\lambda t}dt \\ &= \sum_{i\neq s}\frac{G_{si}}{-G_{ss}}\frac{\lambda\left((\lambda - G^{**})^{-1}\right)_{if}}{x_i} \\ &= \sum_{i\neq s}\frac{G_{si}}{-G_{ss}}\frac{\lambda\left((\lambda - G^{**})^{-1}\right)_{if}}{x_i}.\end{aligned} \tag{11.4.9}$$

Now note that by a quick computation with Lemma 11.5.1, if we define x by

$$x_i = \lim_{\lambda\to 0}\lambda(\lambda - G^{**})^{-1}_{if} \tag{11.4.10}$$

then x is the unique solution to $G^{**}x = 0$ with $x_f = 1$ and $x_s = 0$, and so coincides with our definition of x in (11.3.3).

Differentiating and using Lemma 11.5.1 gives (11.4.6).

□

11.5 Inverses and singular matrices

Lemma 11.5.1. *Let A be a block upper triangular matrix of the form*

$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & 0 \end{array}\right], \tag{11.5.1}$$

where the dimensions of A_{11}, A_{12} and A are respectively $m\times m$, $m\times k$ and $(m+k)\times(m+k)$, and suppose that $(\lambda - A_{11})$ is invertible for all $\lambda \in [0,\epsilon)$ for some $\epsilon > 0$. Then

$$\lim_{\lambda\to 0} \lambda(\lambda - A)^{-1} = \left[\begin{array}{c|c} 0 & -A_{11}^{-1}A_{12} \\ \hline 0 & I \end{array}\right], \tag{11.5.2}$$

and

$$(-1)^n \partial_\lambda^n \lambda(\lambda - A)^{-1}\Big|_{\lambda=0} = \left[\begin{array}{c|c} -n!(-A_{11})^{-n} & n!(-A_{11})^{-(n+1)}A_{12} \\ \hline 0 & 0 \end{array}\right]. \tag{11.5.3}$$

Furthermore, if c is a vector of length k, then the unique solution to

$$Ax = 0$$
$$\text{and} \quad (x_{m+1},\dots,x_{m+k}) = c$$

is

$$x = \left[\begin{array}{c} -A_{11}^{-1}A_{12}c \\ \hline c \end{array}\right]. \tag{11.5.4}$$

Proof. By the block inversion formula for a 2×2 block matrix,

$$(\lambda - A)^{-1} = \left[\begin{array}{c|c} (\lambda - A_{11})^{-1} & \frac{1}{\lambda}(\lambda - A_{11})^{-1}A_{12} \\ \hline 0 & \frac{1}{\lambda}I \end{array}\right].$$

Using the following identity for differentiating the inverse of a matrix

$$\partial_t B(t)^{-1} = -B(t)^{-1}(\partial_t B(t))B(t)^{-1},$$

and differentiating in each slot, we see that

$$(-1)^n \partial_\lambda^n \lambda(\lambda - A)^{-1} = \left[\begin{array}{c|c} n!(\lambda - A_{11})^{-(n+1)}A_{11} & n!(\lambda - A_{11})^{-(n+1)}A_{12} \\ \hline 0 & 0 \end{array}\right].$$

Since A_{11} is invertible, we can take the limit as $\lambda \to 0$ from above.

That x solves $Ax = 0$ is obvious; we need only justify that it is the unique solution. This follows since A_{11} is invertible, and so A has rank m.

□

Chapter 12

Noise properties in interacting parallel chains

12.1 The general case

In this chapter we introduce a few tools for simplifying the calculations needed to analyze the parallel composition of chains, such as independent enhancer and promoter transcription factor loading. We model this as a collection of noninteracting Markov chains, each of which must be in its final state for transcription to occur. In general, it does not seem possible to express the first two moments of the traversal time for the composite chain with only the first two moments of the component chains or similar quantities, as the example in Section 12.2 will show. It is still possible to compute quantities for the composite chain in terms of the component chains through a spectral decomposition, as we discuss in this section. In Section 12.2, we discuss a simple approximation for the case of a two–state enhancer chain, and in Section 12.3 we use the spectral decomposition technique to compute the moments for an arbitrary set of parallel two-state chains, as in the fully noncooperative enhancer assembly of Chapter 10.

Formally, we again have a sequence of continuous-time Markov chains X^k on a sequence of state spaces $\mathcal{X}^k$, for $k \in \{1, 2, \ldots, n\}$, each with two distinguished (and distinct) states s_k and f_k. We assume that each has at most one absorbing state, and so has a generator that can be diagonalized by an invertible matrix. The composite Markov chain X is just the product chain on the Cartesian product $\prod_{i=1}^{n} \mathcal{X}^k$ given by $X_t = (X_t^1, \ldots, X_t^n)$,

where $X^1, \ldots, X^n$ evolve independently. We denote by G the transition rate matrix of the full chain X, and write a state x for the product chain X as $x = (x_1, x_2, \ldots, x_n)$, where $x_k \in \mathcal{X}^k$ for each $1 \leq k \leq n$.

We suppose that transcription (or the jump to the next stage) occurs at rate ρ while X is in state $f = (f_1, \ldots, f_n)$. Thus, we are interested in the time until death of the chain X if it is killed at rate ρ while in state f. The Feynman-Kac formula gives a way to compute the Laplace transforms and moments of the killing times — for an excellent discussion, see Fitzsimmons & Yor [116]. If Π is the projection matrix with $\Pi_{ff} = 1$ and $\Pi_{ij} = 0$ otherwise, and if τ is the killing time, then

$$\mathbb{E}^x \left[\exp(-\lambda\tau)\right] = \rho \left(\lambda I - G + \rho\Pi\right)^{-1}_{xf} \tag{12.1.1}$$

This is equation (48) in [116] (but beware the differences in notation).

Since G is of product form, and we can find its spectral decomposition in terms of the spectral decompositions of the component chains, it would be nice to compute $(\lambda I - G + \rho\Pi)^{-1}$ in terms of $(\lambda I - G)^{-1}$. This turns out to be possible, thanks to the following lemma, which is a special case of the Matrix Inversion Lemma, also known as the Sherman-Morrison-Woodbury formula [145].

Lemma 12.1.1. *Let B be an invertible $m \times m$ matrix, and let u and v be m-dimensional vectors such that $v^t B u \neq -1/\rho$. Then*

$$(B + \rho u v^t)^{-1} = B^{-1} - \frac{\rho}{1 + \rho v^t B^{-1} u} B^{-1} u v^t B^{-1}. \tag{12.1.2}$$

Note 6. The lemma allows us to compute the inverse of a rank-one correction to B easily using only B^{-1} — if u and v are the i^{th} and j^{th} basis vectors respectively, then $(B^{-1} u v^t B^{-1})_{xy} = B^{-1}_{xi} B^{-1}_{jy}$, while $v^t B^{-1} u = B^{-1}_{ij}$.

Using this lemma, if we let $q = (\lambda I - G)^{-1}_{ff}$, we may write

$$(\lambda I - G + \rho\Pi)^{-1}_{xy} = (\lambda I - G)^{-1}_{xy} - \frac{\rho}{1 + \rho q} (\lambda I - G)^{-1}_{xf} (\lambda I - G)^{-1}_{fy},$$

and hence

$$\begin{aligned} \mathbb{E}^x \left[\exp(-\lambda\tau)\right] &= \rho (\lambda I - G)^{-1}_{xf} \left\{ 1 - \frac{\rho}{1 + \rho q} (\lambda I - G)^{-1}_{ff} \right\} \\ &= \frac{\rho}{1 + \rho (\lambda I - G)^{-1}_{ff}} (\lambda I - G)^{-1}_{xf}. \end{aligned} \tag{12.1.3}$$

Note 7. If we take $\rho \to \infty$, we get the expression for Laplace transform of the first hitting time of f (denoted here by τ_f) as the ratio of two terms in the resolvent:

$$\mathbb{E}^x[\exp(-\lambda\tau_f)] = \frac{(\lambda I - G)^{-1}_{xf}}{(\lambda I - G)^{-1}_{ff}}, \tag{12.1.4}$$

which is, of course, also obtainable by recognizing that $\lambda(\lambda I - G)^{-1} = \int_0^\infty \lambda e^{-\lambda t} P_t dt$ and using the strong Markov property.

Now we discuss how $(\lambda I - G)^{-1}$ can be computed using information about only the component chains. Let $G^1, \ldots, G^n$ be the rate matrices of $X^1, \ldots, X^n$, and suppose each G^k has eigenvalues λ_i^k with corresponding left and right eigenvectors ℓ_i^k and r_i^k, for $1 \le i \le m_k$. If an eigenspace has dimension greater than one (as will be the case if f_k is absorbing) then any choice of of eigenvectors may be made as long as ℓ_i^k is orthogonal to r_j^k for $i \neq j$. If for each G^k we pick some right eigenvector, and form a vector in the product space in the natural way, then the resulting product vector will be an right eigenvector of G with eigenvalue equal to the product of the respective eigenvalues. It is well-known that under our assumptions, all right eigenvectors of G are formed in this way, and that they span the product space. Formally, we know that for each $i_1, \ldots, i_n$, with $1 \le i_k \le m_k$, $\lambda_{i_1,\ldots,i_n} = \prod \lambda_{i_j}$ is an eigenvalue for G, with corresponding left and right eigenvectors $\ell_{i_1,\ldots,i_n}(x) = \prod_k \ell^k_{i_k}(x_k)$ and $r_{i_1,\ldots,i_n}(x) = \prod_k r^k_{i_k}(x_k)$, where $x = (x_1, \ldots, x_n)$. To be clear about notation: $\ell = \ell_{i_1,\ldots,i_n}$ is a vector in the product space $\prod \mathcal{X}^k$, so it is indexed by elements $x \in \prod \mathcal{X}^k$, which have the form $x = (x_1, \ldots, x_n)$. We form the product vector ℓ be saying that $\ell_{i_1,\ldots,i_n}(x) = \prod_k \ell^k_{i_k}(x_k)$. Furthermore, this provides a spectral decomposition of G:

$$(\lambda I - G)^{-1}_{xy} = \sum_{i_1,\ldots,i_n} (\lambda - \prod_k \lambda^k_{i_k})^{-1} \prod_k r^k_{i_k}(x_k)\ell^k_{i_k}(y_k), \tag{12.1.5}$$

This allows us to compute an explicit expression for $\mathbb{E}[e^{-\lambda\tau}]$, if we have a spectral decomposition of each product chain.

12.2 A simple approximation for a parallel two-state enhancer

Now we consider a special case. Let X be a Markov chain with distinguished states s and f, and further assume that once X is in the final state f, it does not leave. Let Y be an independent two-state chain Y that takes values in $\{0, 1\}$, that moves from 0 to 1 at

rate α and moves from 1 to 0 with rate β. The transition probabilities for Y can be exactly computed:

$$\mathbb{P}\{Y_t = 1 | Y_0 = 0\} = \frac{\alpha}{\alpha+\beta}(1 - e^{-(\alpha+\beta)t}).$$

We construct a chain on the product space by saying that transcription occurs once X is in state f and Y is in state 1. Let τ be the first time this occurs,

$$\tau = \inf\{t \geq 0 : X_t = f \text{ and } Y_t = 1\}.$$

Let τ_X be the first time that X hits the state f, and let W be an independent Exponential(α) random variable. Then we see that

$$\tau \stackrel{d}{=} \tau_X + (1 - Y_{\tau_X})W \tag{12.2.1}$$

whence

$$\begin{aligned}\mathbb{E}[\exp(-\lambda\tau)] = &\left(\frac{\alpha}{\alpha+\beta} + \frac{\beta}{(\alpha+\beta)(\alpha+\lambda)}\right)\mathbb{E}[\exp(-\lambda\tau_X)] \\ &- \frac{\alpha}{\alpha+\beta}\left(1 + \frac{1}{\alpha+\lambda}\right)\mathbb{E}[\exp(-(\lambda+\alpha+\beta)\tau_X)]\end{aligned} \tag{12.2.2}$$

and (working directly from (12.2.1))

$$\begin{aligned}\mathbb{E}[\tau] &= \mathbb{E}[\tau_X] + \frac{1}{\alpha+\beta}\mathbb{E}\left[1 - e^{-(\alpha+\beta)\tau_X}\right] \\ \mathbb{E}[\tau^2] &= \mathbb{E}[\tau_X^2] + 2\frac{1}{\alpha+\beta}\mathbb{E}\left[\tau_X\left(1 - e^{-(\alpha+\beta)\tau_X}\right)\right] + \frac{2}{\alpha(\alpha+\beta)}\mathbb{E}\left[1 - e^{-(\alpha+\beta)\tau_X}\right].\end{aligned} \tag{12.2.3}$$

Therefore, it seems that computing the moments of τ requires the full Laplace transform of τ_X. However, if τ_X is reasonably large relative to $\alpha + \beta$, then a good approximation is

$$\begin{aligned}\mathbb{E}[\tau] &\approx \mathbb{E}[\tau_X] + \frac{1}{\alpha+\beta} \\ \mathbb{E}[\tau^2] &\approx \mathbb{E}[\tau_X^2] + 2\frac{1}{\alpha+\beta}\mathbb{E}[\tau_X] + \frac{2}{\alpha(\alpha+\beta)}.\end{aligned} \tag{12.2.4}$$

12.3 Moments for parallel two-state chains

The transition matrix for the two-state chain with transition rates α and β has generator matrix

$$\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix},$$

which has eigenvalues 0 and $-(\alpha+\beta)$, left eigenvectors $(\beta/(\alpha+\beta), \alpha/(\alpha+\beta))$ and $(-1,1)$, and right eigenvectors $(1,1)$ and $(-\alpha/(\alpha+\beta), \beta/(\alpha+\beta))$.

Assume we have n two-state chains in parallel with rates α_k, β_k, for $1 \le k \le n$. Reparametrize by setting $c_k = \alpha_k + \beta_k$ and $r_k = \alpha_k/\beta_k$. Then, if G is the generator of the product chain,

$$(\lambda - G)^{-1}_{ff} = \left(\prod_k \frac{\beta_k}{\alpha_k+\beta_k}\right) \sum_\eta \frac{(-1)^{|\eta|}}{\prod_k c_k^{\eta(k)} - \lambda(-1)^{|\eta|}}, \tag{12.3.1}$$

and

$$(\lambda - G)^{-1}_{sf} = \left(\prod_k \frac{\beta_k}{\alpha_k+\beta_k}\right) \sum_\eta \frac{\prod_k r_k^{\eta(k)}}{\prod_k c_k^{\eta(k)} - \lambda(-1)^{|\eta|}}, \tag{12.3.2}$$

where the sum is over all maps η from $\{0,1,\ldots,n\}$ to $\{0,1\}$, and $|\eta| = \sum_k \eta(k)$.

If we define

$$P(\eta) = \prod_k c_k^{-\eta(k)}/Z, \tag{12.3.3}$$

where

$$Z = \sum_\eta P(\eta)$$

then after some algebra with the above,

$$\begin{aligned}\mathbb{E}[\tau] = &\left(\sum_\eta P(\eta) \prod_k (-r_k/c_k)^{\eta(k)}\right)\left(\sum_\eta P(\eta)\prod_k (-1)^{\eta(k)}\right) \\ &- \left(\sum_\eta P(\eta)\prod_k r_k^{\eta(k)}\right)\left(\sum_\eta P(\eta)\prod_k (-1/c_k)^{\eta(k)}\right)\end{aligned} \tag{12.3.4}$$

and

$$\begin{aligned}\mathbb{E}[\tau^2] = &\left(\sum_\eta P(\eta) \prod_k (r_k/c_k^2)^{\eta(k)}\right)\left(\sum_\eta P(\eta)\prod_k (-1)^{\eta(k)}\right)^2 \\ &+ \left(\sum_\eta P(\eta)\prod_k r_k^{\eta(k)}\right)\left(\sum_\eta P(\eta)\prod_k (-1/c_k)^{\eta(k)}\right)^2 \\ &- \left(\sum_\eta P(\eta) \prod_k (-r_k/c_k)^{\eta(k)}\right)\left(\sum_\eta P(\eta)\prod_k (-1)^{\eta(k)}\right)\left(\sum_\eta P(\eta)\prod_k (-1/c_k)^{\eta(k)}\right) \\ &- \left(\sum_\eta P(\eta) \prod_k (r_k)^{\eta(k)}\right)\left(\sum_\eta P(\eta)\prod_k (-1)^{\eta(k)}\right)\left(\sum_\eta P(\eta)\prod_k (1/c_k^2)^{\eta(k)}\right)\end{aligned} \tag{12.3.5}$$

Since $P(\eta)$ is the probability distribution of n independent Bernoulli random variables, the k^{th} of which is 1 with probability $1/(1+c_k)$, we can decompose the sums into products, getting that

$$\begin{aligned}\mathbb{E}[\tau] = &\prod_k \left[1 - \frac{1}{1+c_k}\left(1 + \frac{r_k}{c_k}\right)\right]\left[1 - 2\frac{1}{1+c_k}\right] \\ &- \prod_k \left[1 - \frac{1}{1+c_k}(1-r_k)\right]\left[1 - \frac{1}{c_k}\right]\end{aligned} \tag{12.3.6}$$

and

$$\begin{aligned}\mathbb{E}[\tau^2] = &\prod_k \left[1 - \frac{1}{1+c_k}\left(1 - \frac{r_k}{c_k^2}\right)\right]\left[1 - 2\frac{1}{1+c_k}\right]^2 \\ &+ \prod_k \left[1 - \frac{1}{1+c_k}(1-r_k)\right]\left[1 - \frac{1}{c_k}\right]^2 \\ &- \prod_k \left[1 - \frac{1}{1+c_k}\left(1 + \frac{r_k}{c_k}\right)\right]\left[1 - 2\frac{1}{1+c_k}\right]\left[1 - \frac{1}{c_k}\right] \\ &- \prod_k \left[1 - \frac{1}{1+c_k}(1-r_k)\right]\left[1 - 2\frac{1}{1+c_k}\right]\left[1 - \frac{c_k - 1}{c_k^2}\right].\end{aligned} \tag{12.3.7}$$