Folding kinetics of proteins with multiple domains: inference of transition rates from extinction times

Yingxiang Zhou and Pak-Wing Fok

Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, United States of America

E-mail: yxzhou@udel.edu and pakwing@udel.edu

Keywords: extinction times, stochastic particle dynamics, protein folding kinetics, activated escape, Kramers problem

Abstract

In this paper we are concerned with the folding kinetics of large proteins that may have multiple binding/unbinding domains. The theory underlying such systems have received much less attention compared to small, ‘two-state’ (folded/unfolded) proteins. To analyze this problem, we model the reaction coordinate using a birth-death chain, a well-known stochastic process. Specifically, we consider a protein with $N$ domains where every domain that unfolds (folds) increases (decreases) the reaction coordinate by an integer so that the state-space of the reaction coordinate is the set of integers \{0, 1, \ldots, N\}. As input data, our method uses (i) the extinction time of trajectories, starting from when the protein leaves the 0 state for the first time and finishing when the protein re-enters state 0 for the first time and (ii) the maximum number of unfolded domains in the said trajectory. Since the maximum value $n$ reached by each trajectory is known, we use the proportion of trajectories that do not exceed $n$ and corresponding mean extinction times (ET) to recover the birth-death rates sequentially from 1 to $N$. In general, the initial error will propagate with the site number exponentially. However, with sufficient amount of input data, we can recover the rates with relatively small error. For instance, given 50 million ETs of an 11-site BDP, we can recover the rates with a relative error of about 3%.

1. Introduction

The prediction of the native conformation of a protein, given its amino acid sequence is one of the great open problems in biophysics \[1–5\]. Since the 1990s \[6, 7\], techniques such as Atomic Force Microscopy (AFM) \[8\] and Förster Resonance Energy Transfer (FRET) \[9, 10\] have allowed experimentalists to explore the relationship between macromolecular structure and folding/unfolding rates.

A computational approach to protein folding is to implement large-scale all-atom molecular dynamics (MD) simulations \[11\]. The folding of a macromolecule is represented by tracking the positions of every single atom in the molecule and the result of the simulation is a trajectory in a high-dimensional state-space. A more conceptual way to understand protein folding is to introduce a reaction coordinate \[12\] which effectively maps the high-dimensional space onto a single scalar that measures the progress of the folding. The progress along the reaction coordinate is modulated by a free energy ‘landscape’ \[13\] that may exhibit multiple minima, corresponding to multiple metastable configurations. Inferring the shape of these landscapes from quantities such as extinction times \[14\], rupture forces \[15\] or time-displacement trajectories \[16\] remains a challenging theoretical problem.

One of the most common ways of probing the energy landscape is through AFM \[17\]. In AFM experiments, one end of the molecule is tethered to a movable platform and the other is attached to a cantilever tip: see figure 1(a). Small deflections of the cantilever are detected using a laser-photodiode setup. The AFM can operate in several ways. One protocol is ‘force-ramp’ mode where the platform lowers at a constant speed. As a protein domain is coercively unfolded, the cantilever deflection increases until a critical platform position is reached. Beyond this point, the cantilever quickly relaxes, corresponding to domain rupture. The resulting force-extension
curve allows quantification of the ‘entropic elasticity’ of that particular domain. The procedure can also be performed sequentially if multiple domains are present in a large protein [18, 19]. The discrete event corresponding to rupture is interesting both physically and mathematically. Experiments show that rupture does not always occur at the same force. Furthermore, the rupture force distribution shifts towards higher values when the platform speed is larger. Both of these observations are in stark contrast to mechanical bonds that break at a single yield stress. They point towards a model of bond-breaking that is based on ‘thermally activated escape,’ i.e. a theory of random walks.

Besides force-ramp mode, another protocol is to keep the AFM platform stationary and operate in ‘force-clamp’ mode. Under this mode of operation, one focuses on deflections of the cantilever which essentially provide the reaction coordinate as a function of time. Some possible time traces are shown in figures 1(b) and (c). The protein spends most of its time in metastable configurations (when the reaction coordinate is an integer) and very little time in-between these states. Figure 1(b) shows the trace for a simple protein with a single folding domain that can be in one of two states: folded or unfolded. The kinetics in this case are well-described by two exponential distributions [20]; on the 0 → 1 transition and the other for the 0 → 1 transition. The half-lives or rate constants associated with each exponential distribution can easily be inferred from the time trace in figure 1(b). For proteins with multiple domains the traces can be more complex and could resemble figure 1(c). If one assumes exponential kinetics as before (i.e. the transition between states always follows an exponential distribution, but the parameters of the distribution could be state-dependent) the resulting stochastic process is a birth-death chain. These chains correspond to energy landscapes with multiple metastable states: see figure 2.

An extinction time is the time taken for the reaction coordinate to start at 1 and reach 0 for the first time, corresponding to the state where all domains are folded. In figure 1(c), the measurement of τ starts when the reaction coordinate reaches 1 for the first time. For each excursion, one can also define the maximal site nj that is reached before extinction occurs. After suitable processing of the signal, a single time trace generates many pairs \((τ_j, n_j)\), \(j = 1, \ldots, M\) and \(M \gg 1\). In this paper, we show how to recover all the transition rates of the birth-death chain from \(\{τ_j\}\) and \(\{n_j\}\). How to detect jumps in the reaction coordinate, that define when transitions between states have occurred, is a separate issue and beyond the scope of this paper.

1.1. Energy landscape inference: existing methods

In this section, we briefly describe some of the existing methods used to interpret AFM data. One of the earliest methods is Bell’s method [21] which assumes a force-dependent rate ansatz (of Arrhenius form) for the state of a domain or chemical bond. Providing that the force on the protein depends on the instantaneous deflection in the cantilever, Bell’s method predicts rupture force distributions and survival probabilities for a chemical bond under the action of a force ramp. The method is used to solve the ‘forward’ problem in the sense that rupture distributions are predicted from a potential well whose shape is known. In contrast, Dudko and co-authors [22]
and Freund [23] essentially solved the inverse problem by inferring rate constants, features of the potential well and other related parameters from rupture force distributions.

Chang and co-authors [16] utilized a path integral method that takes trajectory data (time traces) as input rather than force distributions. Based on non-parametric Bayesian estimation, the method makes very few assumptions about the underlying energy landscape and is able to simultaneously infer the energy landscape and an effective spatially-dependent diffusivity for the reaction coordinate.

Most of the above methods are mainly concerned with the AFM operating in force-ramp mode. However, in works such as [14, 24], the authors develop methodology to extract energy landscapes from data generated by AFMs in force-clamp mode. They treat the reaction coordinate using a Smoluchowski framework to infer features of the energy landscape; this type of analysis dates back to Kramers’s classic transition state theory [25] for chemical reactions. Finally, it is worth mentioning that if the reaction coordinate is treated as a Brownian random walker on an energy landscape, estimating the parameters of the resulting stochastic process from sample paths is a classic problem in statistics and control theory [26, 27].

The method described in this paper is different because from the outset, we assume that the underlying stochastic process is a birth-death chain (and subsequently, transitions are always exponentially distributed). Estimation of parameters from sample paths of a birth-death chain is a classic problem [28, 29]. However, our goal is to estimate transition rates ‘mainly’ from extinction times. Unfortunately, as discussed below using only extinction times results in a severely ill-posed problem. Having access to maximal site data turns out to render the inference problem much better-posed.

The inclusion of maximal sites is important. The calculation of transition rates in a birth-death chain purely from extinction times essentially reduces to finding the best-fit coefficients and exponents to a given extinction time distribution. Such problems are highly ill-posed: a small amount of noise added to the curve can lead to a large change in the best-fit coefficients/exponents. Nevertheless, because fitting exponential modes to given data is one of the most commonly-occurring inverse problems, it has a long history and has been investigated by many researchers: see for example [30–33] and references within.

2. Governing equations for the birth death process

We now present a method to compute the transition rates of a birth death process from its extinction times. Our stochastic model takes the form of a random walker on a finite integer lattice: \(0, 1, \ldots, N\). The birth and death rates at site \(n\) are \(\lambda_n\) and \(\mu_n\), respectively. We assume the particle starts at site \(n = 1\) and the process terminates (goes extinct) when it reaches site 0. For each trajectory, we record the extinction time and the maximal site reached (i.e. the largest site number it attains before reaching site 0). From this data, we wish to infer the \(2N - 1\) rates \(\lambda_0, \lambda_1, \ldots, \lambda_{N-1}\) and \(\mu_1, \ldots, \mu_N\).

Consider a birth-death process on a lattice with sites labeled \(0, 1, 2, \ldots, N\), where \(N\) is finite but unknown: see figure 3. A particle starts at site 1 and executes a random walk which we write as \(X(t)\): \(X(t)\) is a random walk on the non-negative integers. At site \(i\), the rightward (leftward) hopping rate is \(\lambda_i (\mu_i)\). When the particle reaches site 0, we record the time of exit. When this experiment is repeated many times, we may use the resulting data to compute the extinction time distribution \(W(t)\).
The matrix that determines this process is defined as

\[
A^{(N)} = \begin{bmatrix}
-(\lambda_1 + \mu_1) & \mu_2 & 0 & \cdots & 0 \\
\lambda_1 & -(\lambda_2 + \mu_2) & \mu_3 & \cdots & 0 \\
0 & \lambda_2 & -(\lambda_3 + \mu_3) & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & -(\lambda_N + \mu_N)
\end{bmatrix}
\]  

and \(\lambda_N = 0\). If \(P_k(t) = P[X(t) = k]\), then the \(P = [P_0, ..., P_N]^T\) satisfy the forward master equation

\[
P = A^{(N)}P, \quad P(0) = [1, 0, ..., 0]^T,
\]

and it is a simple matter to find \(W(t)\) from \(P_k(t)\) (see section 2.1).

If \(N\) is known, finding all the transition rates \(\{\lambda_i, \mu_i\}_{i \leq N}\) from \(W(t)\) amounts to an exponential fitting problem, which is very ill-conditioned \[34\]. To overcome this difficulty, we assume that for every trajectory \(X_i(t)\) we also record the maximal site of the particle, \(n_i\). The maximal site for the random walker is simply the largest site number that it attains before exiting. By grouping trajectories according to their maximal site and computing the statistics of the extinction times of each group, we are able to accurately infer transition rates for birth death chains of length 8–11 from about \(5 \times 10^5\) extinction times with relative error of a few percent.

To be more precise, if \(X_i(t)\) is the \(j\)th trajectory, then \(\tau_j\) is the \(j\)th extinction time defined as

\[
\text{inf}\{\tau: X_i(\tau) = 0, n_i = \max_{0 \leq l \leq l_i} X_i(t)\} \text{ is the maximal site of the } j \text{th trajectory and } S_n = \{\tau_i; n_i \leq n\} \text{ is the set of extinction times corresponding to trajectories whose maximal site does not exceed } n. \text{ Then for a finite sample of trajectories, we have } S_1 \subseteq S_2 \subseteq \cdots \subseteq S_n.\]

The algorithm that we propose takes as input \(|S_n|\) and \(S_n\) (cardinality and mean of \(S_n\)) to infer \(\lambda_n\) and \(\mu_n\) for each \(n\).

Random walkers that exit from the dashed box in figure 3 can exit at site 0 or \(n + 1\). We informally call the times that correspond to exit at site 0 (\(n + 1\)) ‘left’ (‘right’) extinction times. Then, the distribution of extinction times for the birth–death process, conditioned on trajectories not exceeding site \(n\) is identical to the left extinction time distribution out of the sublattice \(\{1, ..., n\}\). By finding finding analytical expressions for the moments of this distribution and matching them to the observed moments, we may infer the transition rates on the lattice. This forms the basis of our method.

### 2.1. Extinction times and probability fluxes

We assume that sites 0 and \(n + 1\) are absorbing in the sense that if the particle reaches site 0 or \(n + 1\) (‘exits’), it stays at these sites for all time. We use the superscript \(n\) to distinguish the subproblem from the entire chain.

Define the probability that the random walker is at site \(k\) at time \(t\) as

\[
P_k^{(n)}(t) = P[X(t) = k].
\]

If we take the \(n \times n\) leading principal submatrix of \(A^{(N)}\), then it follows that the conditional probabilities \(P = [P_1, P_2, ..., P_{n+1}]^T\) satisfy the forward master equations

\[
P^{(n)} = A^{(n)}P^{(n)}, \quad P^{(n)}(0) = e_1^{(n)} = [1, 0, ..., 0]^T,
\]

for \(1 \leq n \leq N\). Equation (2.4) is the starting point for the reconstruction process.

Now we introduce the two random variables

- \(E^{(n)} \in \{0, n + 1\}\), a binary random variable which represents the exit site of the random walker:
  \[
P[E^{(n)} = 0] = \Pi_0^{(n)},
  \]
  \[
P[E^{(n)} = n + 1] = \Pi_n^{(n)},
\]
  and \(\Pi_0^{(n)} + \Pi_n^{(n)} = 1\).
The extinction time of the random walker, defined to be the time at which the walker arrives either at site 0 or \( n + 1 \) for the first time. Conditioning on \( E^{(n)} \), let the density of extinction times \( T^{(n)} \) be defined as:

\[
P(t \leq T^{(n)} \leq t + dt|E^{(n)} = 0) = w_L^{(n)}(t) dt,
\]

and we let \( W_L^{(n)}(t) \) and \( W_R^{(n)}(t) \) be the corresponding CDFs.

By equation (2.5), \( \Pi^{(n)} \) is strictly increasing with \( n \). Now we show how extinction times are related to probability fluxes. The probability of the particle being at site 0 at time \( t + dt \) is given by

\[
P_{01}^{(n)}(t + dt) = P_1^{(n)}(t) \mu_1 dt + P_0^{(n)}(t) \times 1.
\]

Note that the probability of hopping left in time \( dt \) from site 1 is \( \mu_1 dt \) and the probability of staying at site 0 in time \( dt \) is 1 since site 0 is absorbing. Equation (2.9) implies that

\[
\frac{dP_0^{(n)}(t)}{dt} = \mu_1 P_1^{(n)}(t).
\]

Lemma 1. The flux out of site 1, \( \mu_1 P_1^{(n)}(t) \), and the left extinction time density \( w_L^{(n)}(t) \) are related through

\[
\mu_1 P_1^{(n)}(t) = w_L^{(n)}(t) \Pi^{(n)},
\]

where \( \mu_1 \) is the death rate from site 1 and \( \Pi^{(n)} \) is defined in equation (2.5).

Proof. If the random walker is at site 0 at time \( t \), it must have arrived there either at \( t \) or before. Then

\[
P_0^{(n)}(t) = P[T^{(n)} \leq t, E^{(n)} = 0],
\]

\[
\Rightarrow P_0^{(n)}(t) = P[T \leq t|E^{(n)} = 0] \cdot P[E^{(n)} = 0],
\]

\[
\Rightarrow P_0^{(n)}(t) = W_0^{(n)}(t) \Pi^{(n)},
\]

\[
\Rightarrow \frac{dP_0^{(n)}(t)}{dt} = w_L^{(n)}(t) \Pi^{(n)},
\]

\[
\Rightarrow \mu_1 P_1^{(n)}(t) = w_L^{(n)}(t) \Pi^{(n)},
\]

using equation (2.10).

3. Algorithm for reconstructing transition rates

Our algorithm for transition rate reconstruction requires the following as input: for each \( n \), the fraction of random walks that exit and whose maximal site does not exceed \( n \) and the mean extinction time for these conditional random walks. For each \( n \geq 1 \), \( \Pi^{(n)} \) and \( \mathbb{E}[T^{(n)}|E^{(n)} = 0] \equiv M^{(n)} \) yield \( \{\lambda_n, \mu_n\} \): see figure 4.

3.1. Inference of \( \mu_1 \) and \( \lambda_1 \)

In the first step, we recover the birth and death rates at site 1. Note that \( P^{(1)}(t) \) only contains a single element, and the forward master equation can be written as

\[
P^{(1)} = A^{(1)} P^{(1)},
\]

with

\[
A^{(1)} = -(\lambda_1 + \mu_1) \quad \text{and} \quad P^{(1)}(0) = 1.
\]
This simple ODE has solution
\[ P^{(1)}(t) = e^{-(\lambda_1 + \mu_1)t}. \] (3.3)

Suppose we only consider left extinction times, with \( n = 1 \), generated by all trajectories that directly arrive at site 0 from site 1. These extinction times are exponentially distributed with parameter \( \lambda_1 + \mu_1 \):
\[ W^{(1)}_L(t) = P[T^{(1)} \leq t | E^{(1)} = 0] = 1 - e^{-(\lambda_1 + \mu_1)t}. \] (3.4)

It follows from the property of exponential distributions that
\[ \lambda_1 + \mu_1 = \frac{1}{E[T^{(1)}|E^{(1)} = 0]} = \frac{1}{M^{(1)}}. \] (3.5)

In the next step, we use (2.11) to get
\[ \mu_1 P^{(1)}_1(t) = w^{(1)}_L(t) \Pi^{(1)}, \] (3.6)
\[ \Rightarrow \mu_1 = \frac{\Pi^{(1)}}{\int_0^\infty p^{(1)}_1(t') dt'} \] (3.7)
\[ \Rightarrow \mu_1 = \Pi^{(1)}(\lambda_1 + \mu_1) = \frac{\Pi^{(1)}}{M^{(1)}} \] (3.8)
\[ \Rightarrow \lambda_1 = \frac{1}{M^{(1)}} - \Pi^{(1)}. \] (3.9)

We now have obtained the approximations for \( \mu_1 \) and \( \lambda_1 \).

3.2. Inference of \( \mu_2 \) and \( \lambda_2 \)

The forward master equations are for \( P^{(2)}(t) \) are
\[ P^{(2)} = A^{(2)} P^{(2)} \] (3.10)
where
\[ A^{(2)} = \begin{bmatrix} -(\lambda_2 + \mu_2) & \mu_2 \\ \lambda_2 & -(\lambda_1 + \mu_1) \end{bmatrix}. \] (3.11)

By (2.11), we have that
\[ \mu_1 P^{(2)}_1(t) = w^{(2)}_L(t) \Pi^{(2)}, \] (3.12)
\[ \Rightarrow \mu_1 \int_0^\infty P^{(2)}_1(t') dt' = \Pi^{(2)}, \] (3.13)
\[ \Rightarrow w^{(2)}_L(t) = \frac{P^{(2)}_1(t)}{\int_0^\infty P^{(2)}_1(t') dt'}. \] (3.14)

Now introduce the Laplace transform \( \mathcal{L}\{P(t)\} = \hat{P}(s) \). Then the Laplace transformed equation for \( P^{(2)}_1(t) \) satisfies
\[ [s^2 + \xi^{(2)}_1 s + \xi^{(2)}_1 \hat{P}^{(2)}_1(s) = s + \eta^{(2)}_1, \] (3.15)
where \( \xi^{(2)}_1 = \lambda_1 + \mu_1 + \lambda_2 + \mu_2, \xi^{(2)}_1 = \lambda_1 \lambda_2 + \mu_1 \mu_2 + \lambda_2 \mu_1 \) and \( \eta^{(2)}_1 = \lambda_2 + \mu_2 \). Taking derivatives with respect to \( s \), we have
\[ [2s + \xi^{(2)}_1 \hat{P}^{(2)}_1(s) + [s^2 + \xi^{(2)}_1 s + \xi^{(2)}_1] \frac{d\hat{P}^{(2)}_1(s)}{ds} = 1, \] (3.16)
and when \( s = 0 \),
\[ (\lambda_1 \lambda_2 + \mu_1 \mu_2 + \lambda_2 \mu_2) \hat{P}^{(2)}_1(0) = (\lambda_2 + \mu_2), \] (3.17)
\[ (\lambda_1 + \mu_1 + \lambda_2 + \mu_2) \hat{P}^{(2)}_1(0) + (\lambda_1 \lambda_2 + \mu_1 \mu_2 + \lambda_2 \mu_1) \frac{d\hat{P}^{(2)}_1(s)}{ds} \bigg|_{s=0} = 1. \] (3.18)

Equations (3.17) and (3.18) can be rewritten as
\[ \frac{\lambda_1 + \mu_1}{\mu_1} \Pi^{(2)} - 1 \lambda_2 + [\Pi^{(2)} - 1] \mu_2 = 0, \] (3.19)
\[ [1 - M^{(2)}(\lambda_1 + \mu_1)] \lambda_2 + [1 - M^{(2)} \mu_1] \mu_2 = \frac{\mu_1}{\Pi^{(2)}} - \lambda_1 - \mu_1. \] (3.20)
a linear system for λ₂, μ₂ where

\[
Π^{(2)} = μ₁ \bar{P}^{(2)}(0) \quad \text{and} \quad M^{(2)} = -\left. \frac{d \bar{P}^{(2)}(s)}{ds} \right|_{s=0}. \tag{3.21}
\]

Assuming λ₁ and μ₁ are known from the n = 1 case, solving equations (3.19) and (3.20) allows us to compute λ₂ and μ₂ from the conditional moments Π^{(2)} and M^{(2)}.

### 3.3. Inference of μₙ and λₙ for n ≥ 3

Now we consider the n-th site after computing the birth and death rates for the first n − 1 sites. The Laplace transformed ODEs of \( \bar{P}^{(n)}(s) = [\bar{P}^{(n)}(s), \ldots, \bar{P}^{(n)}(s)]^T \) can be represented in the following matrix form:

\[
(sI_n - A^{(n)}) \bar{P}^{(n)}(s) = \mathbf{e}^{(n)}
\]

where \( \mathbf{e}^{(n)} = [1, 0, \ldots, 0]^T \) has n elements and \( I_n \) is the identity matrix of size \( n \times n \). Whenever s is not an eigenvalue of \( A^{(n)} \), we have that \( \bar{P}^{(n)}(s) = (sI_n - A^{(n)})^{-1}\mathbf{e}^{(n)} \). Let the characteristic polynomial of \( A^{(n)} \) be

\[
q(s; λ₁, μ₁) = s^n + \xi^{(n)}_1 s^{n-1} + \cdots + \xi^{(n)}_n = \det(sI_n - A^{(n)}),
\]

where the \( \{\xi^{(n)}_i\} \) are the coefficients of the characteristic polynomial of \( A^{(n)} \), and they can be written as functions of birth and death rates:

\[
\xi^{(n)}_i = \xi^{(n)}_i(λ₀, \ldots, λₙ; μ₀, \ldots, μₙ),
\]

In addition, define

\[
\eta^{(n)}_i = \xi^{(n)}_i(λ₀ = 0, λ₂, \ldots, λₙ; μ₁ = 0, μ₂, \ldots, μₙ).
\]

Finally, we note that the coefficient of \( s^n \) in (3.23) is 1 and for notational convenience, define

\[
\begin{align*}
\xi^{(n)}_{n+1} &= 1; & \xi^{(n)}_{n+k} &= 0 \quad & \text{for} \quad k \geq 2, \\
\eta^{(n)}_{n+1} &= 1; & \eta^{(n)}_{n+k} &= 0 \quad & \text{for} \quad k \geq 2.
\end{align*}
\]

#### Lemma 2

The Laplace-transformed probability \( \hat{\bar{P}}^{(n)}(s) \), the first element in \( \bar{P}^{(n)}(s) \) from equation (3.22), is a rational function in s satisfying

\[
[s^n + \xi^{(n)}_1 s^{n-1} + \cdots + \xi^{(n)}_n] \xi^{(n)}_1(s) = s^{n-1} + \eta^{(n)}_1 s^{n-2} + \cdots + \eta^{(n)}_n s + \eta^{(n)}_2.
\]

for some constants \( \{\xi^{(n)}_i\}_{i=1}^n \) and \( \{\eta^{(n)}_i\}_{i=2}^n \).

**Proof.** Let \( \hat{A}^{(k-1)} \) be the \((k - 1) \times (k - 1)\) submatrix of \( A^{(k)} \) with the first row and first column removed, so that

\[
\hat{A}^{(k-1)} = \begin{bmatrix}
-(λ₂ + μ₂) & μ₃ & \cdots & μₖ \\
λ₂ & -(λ₃ + μ₃) & \cdots & μₖ \\
\vdots & \ddots & \ddots & \vdots \\
λ_{k-1} & \cdots & -(λₖ₋₁ + μₖ₋₁) & μₖ \\
λ_{k-1} & \cdots & \cdots & -(λₖ + μₖ)
\end{bmatrix}.
\]

By Cramer’s rule and the definition of the determinant in terms of its cofactor expansion, the first element of the solution vector \( \hat{P}^{(n)}(s) \) can be calculated as

\[
\hat{P}^{(n)}_1(s) = \frac{\det \begin{bmatrix}
1 & -μ₂ \xi^{(n-1)}_1 \\
0 & sI_{n-1} - \hat{A}^{(n-1)}
\end{bmatrix}}{\det(sI_n - A^{(n)})} = \frac{\det(sI_{n-1} - \hat{A}^{(n-1)})}{\det(sI_n - A^{(n)})}.
\]

Now introduce the polynomial

\[
\det(sI_{n-1} - \hat{A}^{(n-1)}) = s^{n-1} + c_{n-2}s^{n-2} + c_{n-3}s^{n-3} + \cdots + c₀,
\]

for some coefficients \( c_{n-2}, c_{n-3}, \ldots, c₀ \). Then it is clear that

\[
q(s; 0, 0) = s^n + \eta^{(n)}_1 s^{n-1} + \cdots + \eta^{(n)}_2 s + \eta^{(n)}_n,
\]

\[
= \det \begin{bmatrix}
s & -μ₂ \xi^{(n-1)}_1 \\
0 & sI_{n-1} - \hat{A}^{(n-1)}
\end{bmatrix} = s \det(sI_{n-1} - \hat{A}^{(n-1)}),
\]

from the definitions in equations (3.25) and the fact that the (1, 1) and (2, 1) entries of \( A^{(n)} \) are zero when \( λ₁ = μ₁ = 0 \). Because \( q(s; 0, λ₀ = 0, μ₁ = 0) = 0 \), it follows that \( η^{(n)}_1 = 0 \). Equations (3.32) and (3.33)
imply that
\[ s^n + \eta_1^{(n)} s^{n-1} + \cdots + \eta_3^{(n)} s^2 + \eta_2^{(n)} s + \frac{\eta_1^{(n)}}{s} = s^n + c_{n-1} s^{n-1} + c_{n-2} s^{n-2} + \cdots + c_2 s^2 + c_1 s. \]  \hfill (3.34)

Comparing coefficients, we find that \( \eta_1^{(n)} = c_{n-1}, \eta_2^{(n)} = c_{n-2}, \ldots, \eta_3^{(n)} = c_1 \), so that equation (3.31) becomes
\[
\det(sI - A^{(n-1)}) = s^{n-1} + \eta_1^{(n)} s^{n-2} + \cdots + \eta_3^{(n)} s + \eta_2^{(n)}.
\]  \hfill (3.35)

\[
\Rightarrow \tilde{P}_1^{(n)}(s) = \frac{s^{n-1} + \eta_1^{(n)} s^{n-2} + \cdots + \eta_3^{(n)} s + \eta_2^{(n)}}{s^n + c_{n} s^{n-1} + \cdots + c_2 s^2 + c_1 s}.
\]  \hfill (3.36)

Since the expression \( \frac{\Pi^{(n)}}{\mu_1} \) appears frequently in later analysis, we define the notation
\[
r^{(n)} = \frac{\Pi^{(n)}}{\mu_1}
\]  \hfill (3.37)

and use it in the analysis below.

**Lemma 3.** [Moment and Exit Probability Relations] Let \( M^{(n)} \equiv \mathbb{E}[T^{(n)}|E^{(n)} = 0] \), then
\[
\Pi^{(n)} = \mu_1 \tilde{P}_1^{(n)}(0),
\]  \hfill (3.38)

\[
M^{(n)} = -\frac{1}{r^{(n)}} \left. \frac{d\tilde{P}_1^{(n)}(s)}{ds} \right|_{s=0}.
\]  \hfill (3.39)

**Proof.** The Laplace transform of \( P_1^{(n)}(t) \) is \( \tilde{P}_1^{(n)}(s) = \int_0^\infty e^{-st}P_1^{(n)}(t)dt \). Integrating both sides of (2.11), we have that
\[
\Pi^{(n)} = \mu_1 \int_0^\infty \tilde{P}_1^{(n)}(t')dt' \Rightarrow w_L^{(n)}(t) = \frac{P_1^{(n)}(t)}{\int_0^\infty P_1^{(n)}(t')dt'}.
\]

Therefore, equation (3.38) holds:
\[
\tilde{P}_1^{(n)}(0) = \int_0^\infty P_1^{(n)}(t)dt = r^{(n)}.
\]

If we differentiate the Laplace transform \( \tilde{P}_1^{(n)}(s) \), then (3.39) is validated by
\[
\left. \frac{d\tilde{P}_1^{(n)}(s)}{ds} \right|_{s=0} = -\int_0^\infty te^{-st}P_1^{(n)}(t)dt
\]
\[
\Rightarrow \left. \frac{d\tilde{P}_1^{(n)}(s)}{ds} \right|_{s=0} = -\int_0^\infty tP_1^{(n)}(t)dt
\]
\[
= -r^{(n)} \int_0^\infty tw_L^{(n)}(t)dt = -r^{(n)}M^{(n)},
\]
using equation (2.11).

By setting \( s = 0 \) in (3.28) and its derivative equation, we find two constraints involving the exit probability and the first moment \( M^{(n)} \) of the conditional extinction times using (3.38) (3.39):
\[
\xi_1^{(n)} r^{(n)} = \eta_2^{(n)},
\]  \hfill (3.40)

\[
(\xi_2^{(n)} - \xi_1^{(n)} M^{(n)}) r^{(n)} = \eta_3^{(n)}.
\]  \hfill (3.41)

**Lemma 4.** (Recurrence Relations.) The coefficients \( \xi_1^{(n)}, \xi_2^{(n)} \) and \( \eta_2^{(n)}, \eta_3^{(n)} \) are linear in \( \lambda_n \) and \( \mu_n \), satisfying
\[
\xi_1^{(n)} = (\lambda_n + \mu_n)\xi_1^{(n-1)} - \mu_n \lambda_n - \xi_1^{(n-2)},
\]  \hfill (3.42)

\[
\xi_2^{(n)} = \xi_1^{(n-1)} + (\lambda_n + \mu_n)\xi_2^{(n-1)} - \mu_n \lambda_n - \xi_2^{(n-2)},
\]  \hfill (3.43)

and
\[
\eta_2^{(n)} = (\lambda_n + \mu_n)\eta_2^{(n-1)} - \mu_n \lambda_n - \eta_2^{(n-2)},
\]  \hfill (3.44)
If the matrix $\sigma_n$ and a cofactor expansion, we have

$$
\det(sA - A^{(n)}) = \det \left[ \begin{array}{cccc}
\sigma_1 - p_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sigma_n - p_n
\end{array} \right],
$$

where $p_i = \lambda_i + \mu_i$. It is clear from this recurrence that

$$
\lambda_n = \sigma(I - A^{(n)}) = \mu_n \lambda_{n-1} - \sigma(I - A^{(n-1)}),
$$

By equating coefficients at $O(1)$ and $O(\epsilon)$, we can establish a recurrence relation between the coefficients of the characteristic polynomial for the $n$-site subproblem and the $n - 1$ and $n - 2$ site subproblems:

$$
\xi_1^{(n)} = (\lambda_n + \mu_n)\xi_1^{(n-1)} - \mu_n \lambda_{n-1}\xi_2^{(n-1)},
$$

It is clear from this recurrence that $\xi_1^{(n)}$ and $\xi_2^{(n)}$ are linear in the transition rates $\lambda_n$ and $\mu_n$. Instead of $\xi_1^{(n-1)}$ only depends on $\lambda_1$, $\lambda_2$, $\mu_2$, and $\xi_2^{(n)}$ only depends on $\lambda_1$, $\lambda_2$, $\mu_1$, $\mu_2$, $\ldots$. A similar argument applied to $\det(sA - A^{(n-1)})$ shows that $\xi_1^{(n)}$ and $\xi_2^{(n)}$ are linear in $\lambda_n$ and $\mu_n$ also:

$$
\eta_1^{(n)} = (\lambda_n + \mu_n)\eta_1^{(n-1)} - \mu_n \lambda_{n-1}\eta_2^{(n-1)},
$$

The way of computing $(\lambda_n, \mu_n)$ for the $n$-site subproblem is to rewrite equations (3.40), (3.41) and (3.42)–(3.45):

$$
V_1^{(n)} \left[ \begin{array}{c}
\xi_1^{(n)} \\
\xi_2^{(n)} \\
\eta_1^{(n)} \\
\eta_2^{(n)}
\end{array} \right] = \left[ \begin{array}{c}
0 \\
0 \\
\sigma_n \\
\mu_n
\end{array} \right] V_2^{(n)} \left[ \begin{array}{c}
\sigma_n \\
\mu_n
\end{array} \right] = \left[ \begin{array}{c}
0 \\
0 \\
0 \\
0
\end{array} \right] V_2^{(n-1)} \left[ \begin{array}{c}
\xi_1^{(n-1)} \\
\xi_2^{(n-1)} \\
\eta_1^{(n-1)} \\
\eta_2^{(n-1)}
\end{array} \right],
$$

and $\sigma_n = \lambda_n + \mu_n$. Eliminating the vector $\left[ \begin{array}{c}
\xi_1^{(n)} \\
\xi_2^{(n)} \\
\eta_1^{(n)} \\
\eta_2^{(n)}
\end{array} \right]$, we find that

$$
V_1^{(n)} V_2^{(n)} \left[ \begin{array}{c}
\sigma_n \\
\mu_n
\end{array} \right] = \left[ \begin{array}{c}
0 \\
0 \\
0 \\
0
\end{array} \right] V_2^{(n-1)} \left[ \begin{array}{c}
\xi_1^{(n-1)} \\
\xi_2^{(n-1)} \\
\eta_1^{(n-1)} \\
\eta_2^{(n-1)}
\end{array} \right].
$$

If the matrix $V_1^{(n)} V_2^{(n)}$ is invertible, $\sigma_n$ and $\mu_n$ are uniquely determined, and so $\lambda_n$ and $\mu_n$ are uniquely determined.

**Theorem 1.** Given exact data $\{\Pi^{(n)}, M^{(n)}\}$, $n = 1, 2, \ldots, N$ generated by some underlying birth-death process, the rates $(\lambda_n, \mu_n)$, $n = 1, \ldots, N$ are uniquely determined. In particular, the matrix

$$
F^{(2)} = \begin{bmatrix}
(\lambda_1 + \mu_1)r^{(2)} - 1 & \Pi^{(2)} - 1 \\
1 - M^{(2)}(\lambda_1 + \mu_1) & 1 - M^{(2)} \mu_1
\end{bmatrix}
$$

from equations (3.19) and (3.20) is invertible and the $2 \times 2$ matrix $V_1^{(n)} V_2^{(n)}$ (where $V_1^{(n)}$ and $V_2^{(n)}$ are defined in equation (3.51)) is invertible for $n \geq 3$.

**Proof.** Consider the case $n = 2$. Then, we see that

$$
\det F^{(2)} = \lambda_1 (r^{(n)} - M^{(2)}),
$$
\[ \frac{\lambda_2^2 \mu_2}{(\lambda_2 + \mu_2)(\lambda_1 \lambda_2 + \mu_1 \lambda_2 + \mu_2 \mu_1)} < 0, \quad (3.55) \]

where we used the relations
\[ \Pi^{(2)} = \frac{\mu_1(\lambda_2 + \mu_2)}{\lambda_1 \lambda_2 + \mu_1 \mu_2 + \lambda_1 \mu_2}, \quad (3.56) \]
\[ M^{(2)} = \frac{\lambda_2^2 + 2 \lambda_2 \mu_2 + \lambda_1 \mu_2 + \mu_2^2}{(\lambda_2 + \mu_2)(\lambda_1 \lambda_2 + \mu_1 \mu_2 + \mu_2 \mu_1)} \quad (3.57) \]

Therefore, all parts. First, we show that \( \det(V_1^{(n)})V_2^{(n)} \) is non-zero. Then, because \( \det(A^{(n)}) \neq 0 \). It is well-known that the eigenvalues of the infinitesimal generator matrix (and its submatrices) of a birth-death process are all negative [35], and the matrix \( A^{(n)} \) is the transpose of a submatrix of such an infinitesimal generator. It remains to show that this expression is non-zero for \( n \geq 3 \).

**Expression for Determinant.** The determinant depends on \( r^{(n)} \) and \( M^{(n)} \), which are determined by the (perfect) data. Using the recurrence relations (3.42)–(3.45), we note that
\[ r^{(n)} = \frac{\eta^{(n)}_1}{\xi^{(n)}_1} = \frac{\sigma_n \eta^{(n-1)}_2 - \gamma_n \lambda_n - \gamma^{(n-2)}_1}{\sigma_n \xi^{(n-1)}_1 - \gamma_n \lambda_n - \xi^{(n-2)}_1}, \quad (3.59) \]

and
\[ -M^{(n)}r^{(n)} = \frac{dR^{(n)}_1}{ds} = \frac{(\eta^{(n-1)}_2 + \sigma_n \eta^{(n-1)}_2 - \gamma_n \lambda_n - \gamma^{(n-2)}_1)(\sigma_n \xi^{(n-1)}_1 - \gamma_n \lambda_n - \xi^{(n-2)}_1) - (\sigma_n \xi^{(n-1)}_2 - \gamma_n \lambda_n - \xi^{(n-2)}_1)(\sigma_n \xi^{(n-1)}_1 - \gamma_n \lambda_n - \xi^{(n-2)}_1)^2}{(\sigma_n \xi^{(n-1)}_2 - \gamma_n \lambda_n - \xi^{(n-2)}_1)^2}, \]

Therefore, after some algebra
\[ \det(V_1^{(n)}V_2^{(n)}) = \frac{\lambda_n \gamma_2 (\eta^{(n-1)}_2 \xi^{(n-2)}_1 - \eta^{(n-2)}_2 \xi^{(n-1)}_1)^2}{\xi^{(n)}_1}. \quad (3.60) \]

The denominator is always nonzero because \( \xi^{(n)}_1 = (-1)^n \det(A^{(n)}) \neq 0 \). It is well-known that the eigenvalues of the infinitesimal generator matrix (and its submatrices) of a birth-death process are all negative [35], and the matrix \( A^{(n)} \) is the transpose of a submatrix of such an infinitesimal generator. It remains to show that this expression is non-zero for \( n \geq 3 \).

**Induction Proof.** First, we show that \( \det(V_1^{(3)}V_2^{(3)}) \) is non-zero. Because
\[ \xi^{(2)}_1 = \lambda_2 \lambda_1 + \mu_1 \mu_2 + \lambda_2 \mu_1, \quad (3.61) \]
\[ \eta^{(2)}_2 = \lambda_2 + \mu_2, \quad (3.62) \]
\[ \xi^{(1)}_1 = \lambda_1 + \mu_1, \quad (3.63) \]
\[ \eta^{(1)}_2 = 1, \quad (3.64) \]

\[ \Rightarrow \det(V_1^{(3)}V_2^{(3)}) = \frac{\lambda_3 \lambda_1 \mu_2^2}{\xi^{(2)}_1} > 0. \quad (3.65) \]

Now assume that \( \det(V_1^{(n)}V_2^{(n)}) \) is non-zero. Then
\[ \eta^{(n-1)}_2 \xi^{(n-2)}_1 - \eta^{(n-2)}_2 \xi^{(n-1)}_1 \neq 0. \quad (3.66) \]
It suffices to show that \( \eta_2^{(n)} \xi_1^{(n-1)} - \eta_2^{(n-1)} \xi_1^{(n)} = 0 \). Using (3.42) and (3.44),
\[
\eta_2^{(n)} \xi_1^{(n-1)} - \eta_2^{(n-1)} \xi_1^{(n)} = (\sigma_n \xi_1^{(n-1)} - \mu_n \lambda_{n-1} \psi_{2}^{(n-2)}) \xi_1^{(n-1)} - \eta_2^{(n-1)} (\sigma_n \xi_1^{(n-1)} - \mu_n \lambda_{n-1} \xi_1^{(n-2)}),
\]
\[
= \mu_n \lambda_{n-1} (\eta_2^{(n-1)} \xi_1^{(n-2)} - \eta_2^{(n-2)} \xi_1^{(n-1)}),
\]
\[
\neq 0.
\]
Therefore \( \det(V_1^{(n)} V_2^{(n)}) \) is non-zero for \( n \geq 3 \). Therefore \( V_1^{(n)} V_2^{(n)} \) is invertible for \( n \geq 3 \). From (3.52), \( \sigma_n \) and \( \mu_n \) are uniquely determined and so \( (\lambda_n, \mu_n) \) are also uniquely determined.

In summary, at step \( n \geq 3 \), we must solve the linear system
\[
V_1^{(n)} V_2^{(n)} \begin{bmatrix} \sigma_n \\ \mu_n \end{bmatrix} = \begin{bmatrix} 0 \\ \eta_2^{(n-1)} - r^{(n)} \xi_1^{(n-1)} \end{bmatrix},
\]
which we may write as
\[
F^{(n)} \nu_n = G^{(n)},
\]
where
\[
F^{(n)} = V_1^{(n)} V_2^{(n)} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \nu_n = \begin{bmatrix} \lambda_n \\ \mu_n \end{bmatrix}, \quad G^{(n)} = \begin{bmatrix} 0 \\ \eta_2^{(n-1)} - r^{(n)} \xi_1^{(n-1)} \end{bmatrix}.
\]

The \( F^{(n)} \) and \( G^{(n)} \) depend on the previous transition rates \( \nu_1, \ldots, \nu_{n-1} \) and from theorem 1, \( F^{(n)} \) is invertible. For reference, the entries of \( F^{(n)} \) are:
\[
F_{11}^{(n)} = r^{(n)} \xi_1^{(n-1)} - \eta_2^{(n-1)},
\]
\[
F_{12}^{(n)} = \eta_2^{(n-1)}(\xi_1^{(n-1)} - \lambda_n \xi_1^{(n-2)}) - \eta_2^{(n-2)},
\]
\[
F_{21}^{(n)} = \eta_2^{(n-1)}(\xi_2^{(n-1)} - \lambda_n \xi_2^{(n-2)}) - \eta_2^{(n-2)},
\]
\[
F_{22}^{(n)} = \eta_2^{(n)} \xi_1^{(n-1)} - \lambda_n \xi_1^{(n-2)} - \lambda_n \xi_2^{(n-2)}.
\]

Let's define the following notations
\[
x^* = (\lambda^*_1, \ldots, \lambda^*_n, \mu^*_1, \ldots, \mu^*_n, \Pi^{(1)}, \ldots, \Pi^{(n)}, M^{(1)}, \ldots, M^{(n)}),
\]
\[
\delta x = (\delta \lambda_1, \ldots, \delta \lambda_n, \delta \mu_1, \ldots, \delta \mu_n, \delta \Pi^{(1)}, \ldots, \delta \Pi^{(n)}, \delta M^{(1)}, \ldots, \delta M^{(n)}),
\]
where the birth-death rates with asterisks stand for exact rates and \( \Pi^{(n)}, M^{(n)} \) are exact data. In contrast, the elements in \( \delta x \) are perturbations to the corresponding elements. Therefore the transition rates at site \( n \) depend on the rates at sites 1, 2, \ldots, \( n - 1 \):
\[
\lambda_n = f_1^{(n)}(\lambda_1, \ldots, \lambda_{n-1}, \mu_1, \ldots, \mu_{n-1}, \Pi^{(1)}, \ldots, \Pi^{(n-1)}, M^{(1)}, \ldots, M^{(n-1)}),
\]
\[
\mu_n = f_2^{(n)}(\lambda_1, \ldots, \lambda_{n-1}, \mu_1, \ldots, \mu_{n-1}, \Pi^{(1)}, \ldots, \Pi^{(n-1)}, M^{(1)}, \ldots, M^{(n-1)}),
\]
where \( f_i^{(n)}: \mathbb{R}^{n-1} \to \mathbb{R} \) for \( i = 1, 2 \).

**Theorem 2.** (Error Propagation with site number) Let \( \nu_n = (\lambda_n, \mu_n)^T \), \( D_n = (\Pi^{(n)}, M^{(n)})^T \) and let \( \nu_n^* \) be the exact transition rate at site \( n \). Suppose that all first derivatives of \( f_1^{(n)} \) and \( f_2^{(n)} \) in equations (3.76), (3.77) are bounded in a small neighborhood \( B(x^*, r) \) of \( x^* \) in equation (3.74), i.e. there exists \( r, R > 0 \) such that for any \( x \in B(x^*, r) \),
\[
\|\nabla f_m^{(n)}(x)\|_\infty \leq R/2,
\]
for \( 1 \leq n \leq N \) and \( m = 1, 2 \). If a sufficiently small error \( \delta D_i \) is introduced into the data \( \{D_k\}_{k=1}^n \) at each site such that \( D_k = D_k^* + \delta D_k \) and \( \|\delta D_k\|_\infty \leq r \), then the error of birth-death rates at site \( n \) satisfies
\[
\|\delta \nu_n\|_\infty \leq \sum_{j=1}^n R(1 + R)^{n-j} \|\delta D_j\|_\infty.
\]

**Proof.** The analysis is fairly standard and makes use of Taylor series expansions. The exact rates satisfy
\[
\lambda_n^* = f_1^{(n)}(x^*), \quad \mu_n^* = f_2^{(n)}(x^*)
\]
By the mean value theorem for multivariate functions, the errors at each site satisfy
\[
\lambda_n^* + \delta \lambda_n = f_1^{(n)}(x^* + \delta x) = f_1^{(n)}(x^*) + \nabla f_1^{(n)}(x^*) \cdot \delta x
\]
\[
\mu_n^* + \delta \mu_n = f_2^{(n)}(x^* + \delta x) = f_2^{(n)}(x^*) + \nabla f_2^{(n)}(x^*) \cdot \delta x
\]
for some \( z_i^{(n)} = x^* + c_{i1}^{(n)} \delta x \) with \( c_{i1}^{(n)} \in (0, 1), i = 1, 2 \), so that \( z_i^{(n)}, z_{2}^{(n)} \in B(x^*, r) \).

By equation (3.80), we have
\[
\delta \lambda_n = \sum_{i=1}^{n-1} \frac{\partial f_i^{(n)}}{\partial \lambda_i} \delta \lambda_i + \sum_{i=1}^{n-1} \frac{\partial f_i^{(n)}}{\partial \mu_i} \delta \mu_i + \sum_{i=1}^{n-1} \frac{\partial f_i^{(n)}}{\partial \Pi^{(i)}} \delta \Pi^{(i)} + \sum_{i=1}^{n-1} \frac{\partial f_i^{(n)}}{\partial M^{(i)}} \delta M^{(i)},
\]
and
\[
\delta \mu_n = \sum_{i=1}^{n-1} \frac{\partial f_i^{(n)}}{\partial \lambda_i} \delta \lambda_i + \sum_{i=1}^{n-1} \frac{\partial f_i^{(n)}}{\partial \mu_i} \delta \mu_i + \sum_{i=1}^{n-1} \frac{\partial f_i^{(n)}}{\partial \Pi^{(i)}} \delta \Pi^{(i)} + \sum_{i=1}^{n-1} \frac{\partial f_i^{(n)}}{\partial M^{(i)}} \delta M^{(i)}.
\]

Define the matrices
\[
R_k^{(n)} = \begin{bmatrix}
\frac{\partial f_1^{(n)}}{\partial \lambda_k} & \frac{\partial f_2^{(n)}}{\partial \lambda_k} \\
\frac{\partial f_1^{(n)}}{\partial \mu_k} & \frac{\partial f_2^{(n)}}{\partial \mu_k}
\end{bmatrix}
\]
and
\[
S_k^{(n)} = \begin{bmatrix}
\frac{\partial f_1^{(n)}}{\partial \Pi^{(k)}} & \frac{\partial f_2^{(n)}}{\partial \Pi^{(k)}} \\
\frac{\partial f_1^{(n)}}{\partial M^{(k)}} & \frac{\partial f_2^{(n)}}{\partial M^{(k)}}
\end{bmatrix}
\]

Therefore,
\[
\delta \nu_n = \sum_{k=1}^{n-1} R_k^{(n)} \delta \nu_k + \sum_{k=1}^{n-1} S_k^{(n)} \delta D_k.
\]

In general, we can repeatedly substitute to get \( \delta \nu_n \) in terms of \( \{\delta D_k\}_{k=1}^{n-1} \) only:
\[
\delta \nu_n = \sum_{j=1}^{n} \left( S_j^{(n)} + \sum_{j<k<n} S_j^{(k)} R_k^{(n)} + \sum_{j<k<n} S_j^{(k)} R_k^{(k)} R_l^{(n)} + \ldots + \sum_{j<k<\ldots<n} S_j^{(k)} R_k^{(k)} R_l^{(k)} R_{l_2}^{(n)} + \ldots + S_j^{(j)} R_{j+1}^{(j+1)} \cdots R_{n-1}^{(n)} \right) \delta D_j
\]

By equation (3.78) we have \( \| R_k^{(n)} \|_\infty \leq R \) and \( \| S_k^{(n)} \|_\infty \leq R \) for all \( 1 \leq i, j \leq N \). Then, a binomial expansion yields
\[
\| \delta \nu_n \|_\infty \leq \sum_{j=1}^{n} R (1 + R)^{n-j} \| \delta D_j \|_\infty
\]

Hence the errors of data introduced at each site will propagate exponentially in its subsequent sites. In other words, at each site \( n \), the error in the birth-death rates \( (\lambda_n, \mu_n) \) is the result of accumulating the exponential growth of errors from sites 1 to \( n - 1 \).

**Corollary 1.** Define a constant \( D \) such that \( D = \max_{1 \leq j \leq N} \| \delta D_j \|_\infty \). Then equation (3.79) becomes
\[
\| \delta \nu_n \|_\infty \leq D ((1 + R)^n - 1).
\]

In this special case where all errors in data are bounded by \( D \), we also expect to observe the error in birth-death rates grow exponentially.

### 3.4. Algorithm details
The implementation of the algorithm starts with the inference at site 1 and site 2. In order to keep track of the recurrence relation in Lemma 4, we only need to focus on the ‘feature vector’ defined as \( u_n = [\eta_1^{(n)}, \eta_2^{(n)}, \eta_1^{(n)}, \eta_2^{(n)}]^T \) at each site \( n \). For site 1, we have
For folded, the reaction coordinate is 0 and if all domains are unfolded, the reaction coordinate is 

Then we can define the matrix

where the \( j \)-th column is the feature vector at site \( j \).

At each step \( j \geq 3 \), we need to solve linear system defined by (3.68) and (3.69) to obtain \( [\lambda_j, \mu_j] \), and update the \( j \)-th column of \( Z \) by lemma 4. Details are listed in algorithm 1.

**Algorithm 1.** Inference of birth and death rates up to site \( N \) in a birth death chain

**Input:** An array of extinction times \( \{\tau_i\} \) along with maximal sites \( \{n_i\} \) from repeated simulation of a birth death process \( (j = 1, \ldots, L) \).
**Initialize:** For \( k = 1 \) to \( N \), find all \( \tau_i \) such that \( n_i = k \). Compute number of extinctions as a fraction of \( L (\Pi^{(\mu)}) \) and the mean of the extinction times (\( M^{(\mu)} \)). Initialize \( Z \) as a

4 \( \times \) \( N \) zero matrix.

1: At site 1, \( [\lambda_1, \mu_1] = \frac{(1 - \Pi^{(\mu)})/\Pi^{(\mu)}}{\Pi^{(\mu)}} \) as in (3.8), (3.9). Feature vector \( Z_{1} = u_1 \) is defined in (3.91).

2: **if** \( N = 1 \) **then** return \( \{\mu_1, \lambda_1\} \)

3: **end if**

4: At site 2, \( [\lambda_2, \mu_2] = \frac{\mu^2(\lambda_1 + \mu_1) - 1}{\Pi^{(\mu)} - 1} \begin{bmatrix} 0 & 0 \\ 1 - \Pi^{(\mu)}(\lambda_1 + \mu_1) & 1 - \Pi^{(\mu)} \mu_1 \end{bmatrix} \), see (3.19) (3.20). Feature vector \( Z_{2} = u_2 \) is defined as (3.92).

5: **for** \( j = 3 : N \) **do**

6: Compute \( V_j^{(\mu)} \) and \( V_j^{(\lambda)} \) as in (3.51).

7: Form the matrices \( F^{(\mu)} \) and \( G^{(\mu)} \) as in (3.68) and (3.69).

8: Solve \( [\lambda_j, \mu_j] = (F^{(\mu)})^{-1}G^{(\mu)} \).

9: Update \( Z_{j} = V_j^{(\mu)} \begin{bmatrix} \lambda_j + \mu_j \\ \mu_j \end{bmatrix} + \begin{bmatrix} Z_{j-1} \\ 0 \end{bmatrix} \)

10: **if** \( j = N \) **then**

11: \( \lambda_N = 0 \)

12: **end if**

13: **end for**

**Output:** \( \{\mu_1, \ldots, \mu_N\} \) and \( \{\lambda_0, \ldots, \lambda_{N-1}\} \)

4. Numerical results

Below we reconstruct the transition rates for proteins with \( N \) folding/unfolding domains. When all domains are folded, the reaction coordinate is 0 and if all domains are unfolded, the reaction coordinate is \( N \). Every domain that unfolds (folds) increases (decreases) the reaction coordinate by an integer.

**Example 1.** Protein with \( N = 4 \) domains. First we present a simple result of a birth-death chain with only 5 states and some pre-determined rates. All the rates are about the same order of magnitude, see figure 5.
The extinction time data is generated by simulating the birth-death process with these given rates, and we evaluate the reconstruction in comparison to them. With about $5 \times 10^6$ extinction times, we can infer the rates to very good accuracy—about 0.11% and 0.37% relative errors in $\lambda$ and $\mu$, respectively. First, the sum of birth and death rates $\lambda_1 + \mu_1$ follows from the mean extinction time conditioned on immediate exit through equation (3.5). The fraction of times corresponding to immediate exit then yields $\mu_1$ and $\lambda_1$ separately through (3.8) and (3.9). Next, $\lambda_2$ and $\mu_2$ are computed by (3.19) (3.20). Then for $n = 3, 4$, we compute the $n$-th columns of matrix $Z$ as in (3.93) and obtain $\{\lambda_n, \mu_n\}$ simultaneously. Notice that the last birth rate $\lambda_N$ is always assumed to be zero and no inference is necessary at that site.

Example 2. Protein with $N = 10$ domains. We now test our reconstruction algorithm on a longer, 11-site chain. In this result, $5 \times 10^7$ extinction times are simulated from an 11-site birth-death chain. Following the same steps, we have the inference results with a relative error in $\lambda$ and $\mu$ to be 3.29% and 3.71%, respectively: see figure 6.

Example 3. Protein with $N = 10$ domains and a single high activation energy barrier. This birth-death chain has a ‘bottleneck’ between states 3 and 4 so that the transition rates between these two configurations are much smaller than the others: see figure 7. The accuracy of the transition rates of sites after the bottleneck are more affected than the accuracy before the bottleneck, due to the fact that all inferences of sites after the bottleneck depend on

Figure 5. Bar plots of the inference results in a 5-site birth death chain. The top subplot (a) contains rates for $\mu_k$ and bottom subplot (b) for $\lambda_k$. The bars in dark blue represent numerically approximated rates, and yellow bars stand for exact rates. On top of each bar is the value associated with it.

Figure 6. Bar plots of the inference results in a 11-site birth death chain. The top subplot (a) contains rates for $\mu_k$ and bottom subplot (b) for $\lambda_k$. The bars in dark blue represent numerically approximated rates, and yellow bars stand for exact rates.
the extinction times generated by particles that have gone through the bottleneck. With about $5 \times 10^7$ extinction times, there are 1.17% and 1.24% relative errors in $\lambda$ and $\mu$, respectively. The effect of the bottleneck is evident from the data since we see an abrupt increase in $M_n$ as $n$ goes from 3 to 4.

Example 4. Protein with $N = 8$ domains and a single low activation energy barrier. This protein has a very low barrier between states 5 and 6; the transition rates between these sites are very large. From an AFM trace, it may be difficult to distinguish or separate these two configurations. However, given $5 \times 10^7$ simulated data, the birth and death rates can be inferred within a reasonable amount of accuracy, and the overall relative error is 2.21% in $\lambda$ and 2.24% in $\mu$. See figure 8.

Example 5. 11-site birth death chain error propagation. In this example, we show how error propagates with site number. The birth and death rates are all equal to 1, and $1 \times 10^8$ extinction times are generated from the Monte Carlo simulation. Bootstrap is done by taking random samples of size $5 \times 10^6$ from these extinction times and errors at each site are computed. This resampling procedure is repeated 50 times, and figures 9 and 10 display the mean error and 95% confidence intervals from the above 50 samples, where it is clear that errors for both $\lambda$ and $\mu$ increase exponentially with site number, as a result of theorem 2. Although this example uses noisy data, this exponential increase with site number is not surprising in light of theorem 2.

Figure 7. Bar plots of the inference results in a 9-site birth death chain with a bottleneck at site number 4. The top subplot (a) contains rates for $\mu_k$ and bottom subplot (b) for $\lambda_k$. The bars in dark blue represent numerically approximated rates, and yellow bars stand for exact rates.

Figure 8. Bar plots of the inference results in a 9-site birth death chain with extreme rates. The top subplot (a) contains rates for $\mu_k$ and bottom subplot (b) for $\lambda_k$. The bars in dark blue represent numerically approximated rates, and yellow bars stand for exact rates.
Example 6. Minimum number of extinction times required. Finally, we consider the minimum number of extinction times required for a birth death chain of length $N + 1$ such that relative error for both $\lambda_k$ and $\mu_k$ is below 15%. For each $N$, 50 bootstrap samples of a fixed size are selected and the average relative errors are computed from these samples. This procedure is then repeated for different sample sizes, from which we can estimate the minimum number of extinction times required for the chain of length $N + 1$ to have error below 15%. Note that this estimation applies to the chain where all birth-death rates are about the same order of magnitude. Detail is in Table 1, and Figure 11. It is obvious from the plot that minimum number of extinction times required is increasing exponentially with the length of chain.

5. Conclusions

In summary, we have presented a method for extracting the kinetic rates of large proteins, with multiple folding domains, from extinction times (when all domains have folded) and ‘maximal sites’ (the maximum number of unfolded domains before extinction). Both of these quantities can, in principle, be computed from AFM time traces.

The inference relies on the recurrence relation specified in Lemma 4 and starts with base cases when $n = 1, 2$, and inference of each subsequent site depends on its previous sites by solving a linear system. If the data $\Pi^{(n)}$ and
are exactly given, meaning that they correspond to the statistics of an underlying birth-death process, then
the birth-death rates are uniquely determined, and we proved that a small perturbation in site 1 will propagate
exponentially throughout the chain, given that the first derivatives of $f_{1}$ and $f_{2}$ in theorem 2 are bounded
near the exact solution. With sufficient data (about 50 million for a chain of length 8–12), the method can
compute these rates to very good accuracy and is capable of detecting bottlenecks or extreme values in the chain.
In general, the number of extinction times one needs to obtain reasonable results grows exponentially with the
total length of the chain.

There still remain many theoretical challenges to interpreting single-molecule AFM data. For example,
infereonce from extinction times in the ‘transmission’ problem where absorption/extinction is at site $N$ rather
than site 0 is severely ill-posed. Can some aspect of time-trace data be used to better-condition this problem?
Also, bifurcations or loops in the birth-death chain representing multiple pathways to a final unfolded state have
not yet been explored. Finally, it remains to be seen if our method can be adapted to force-ramp data, or how to
proceed if transition times between metastable configurations are not exponentially distributed.

Appendix

There is an alternative way to infer the birth death rates, instead of the recurrence relations. This method is
simple, yet more computationally intensive if number of sites is large.

In light of lemma 4, we can assume the following forms:

$$\xi^{(n)}_i = C^{(n)}_{\alpha \beta} \lambda_\alpha + C^{(n)}_{\alpha \beta} \mu_\beta + C^{(n)}_{\alpha \beta}, \quad i = 1, 2,$$

(A.1)
where \( C^{(n)} \) and \( \hat{C}^{(n)} \) are coefficients that depend on \( \lambda_0, \ldots, \lambda_{n-1}, \mu_1, \ldots, \mu_{n-1} \); we describe how to compute \( C^{(n)} \) and \( \hat{C}^{(n)} \) in section A.1. If we plug (A.1) and (A.2) into (3.40) and (3.41), and denote \( r^{(n)} = \Pi^{(n)}/\mu_t \), then we get a linear system in two variables \( \lambda_n, \mu_n \):

\[
F^{(n)} \begin{bmatrix} \lambda_n \\ \mu_n \end{bmatrix} = G^{(n)}
\]

where the \( 2 \times 2 \) matrix \( F^{(n)} \) and 2-vector \( G^{(n)} \) are defined as

\[
F^{(n)} = \begin{bmatrix} r^{(n)}(C_{11}^{(n)} - \hat{C}_{11}^{(n)}) & -r^{(n)}(C_{12}^{(n)} - \hat{C}_{12}^{(n)}) \\ r^{(n)}(C_{21}^{(n)} - \hat{C}_{21}^{(n)}) & r^{(n)}(C_{22}^{(n)} - \hat{C}_{22}^{(n)}) \end{bmatrix},
\]

\[
G^{(n)} = \begin{bmatrix} -r^{(n)}(C_{33}^{(n)} - \hat{C}_{33}^{(n)}) \\ r^{(n)}(C_{33}^{(n)} - \hat{C}_{33}^{(n)}) \end{bmatrix}.
\]

This matrix is consistent with equations (3.70)–(3.73). The rates at current site are therefore given by \( [\lambda_n, \mu_n]^T = (F^{(n)})^{-1}G^{(n)} \).

### A.1. Leverrier-faddeev algorithm

The method for computing the coefficient matrices \( C^{(n)} \) and \( \hat{C}^{(n)} \) is the Leverrier-Faddeev (L-F) algorithm [36]. The L-F algorithm computes all coefficients of the characteristic polynomial for a given \( n \times n \) square matrix with time complexity \( O(n^4) \). In this problem, it suffices to get only the last three coefficients \( \xi_2, \xi_1 \) and \( \xi_0 \). Given the expression in (A.1), we have

\[
\begin{align*}
C_{ii}^{(n)} &= \xi_i^{(n)}(\lambda_0, \ldots, \lambda_{n-1}, 1; \mu_1, \ldots, \mu_n) = \xi_i^{(n)}(\lambda_0, \ldots, \lambda_{n-1}, 0; \mu_1, \ldots, \mu_n) \\
C_{i2}^{(n)} &= \xi_i^{(n)}(\lambda_1, \ldots, \lambda_n; \mu_1, \ldots, \mu_{n-1}, 1) = \xi_i^{(n)}(\lambda_0, \ldots, \lambda_n; \mu_1, \ldots, \mu_{n-1}, 0) \\
C_{i3}^{(n)} &= \xi_i^{(n)}(\lambda_0, \lambda_n = 0; \mu_1, \ldots, \mu_n = 0)
\end{align*}
\]

for \( i = 1, 2 \). With (A.2), we also have

\[
\begin{align*}
\hat{C}_{i1}^{(n)} &= \hat{\eta}_i^{(n)}(\lambda_0, \ldots, \lambda_{n-1}, 1; \mu_1, \ldots, \mu_n) = \hat{\eta}_i^{(n)}(\lambda_0, \ldots, \lambda_{n-1}, 0; \mu_1, \ldots, \mu_n) \\
\hat{C}_{i2}^{(n)} &= \hat{\eta}_i^{(n)}(\lambda_1, \ldots, \lambda_n; \mu_1, \ldots, \mu_{n-1}, 1) = \hat{\eta}_i^{(n)}(\lambda_0, \ldots, \lambda_n; \mu_1, \ldots, \mu_{n-1}, 0) \\
\hat{C}_{i3}^{(n)} &= \hat{\eta}_i^{(n)}(\lambda_0, \lambda_n = 0; \mu_1, \ldots, \mu_n = 0)
\end{align*}
\]

for \( j = 2, 3 \). A concise and straightforward algorithm containing all steps is presented in algorithm 2.

### Algorithm 2. Inference of birth and death rates up to site \( N \) with L-F algorithm

**Input:** An array of extinction times \( T \) along with maximal site of repeated simulation of a birth death process.

**Initialize:** Compute the conditional probabilities of left exit \( \{\Pi^{(1)}, \ldots, \Pi^{(N)}\} \), and mean of conditional extinction times \( \{M^{(1)}, \ldots, M^{(N)}\} \).

1. At site 1, \( \mu_1 = \Pi^{(1)}/M^{(1)} \) and \( \lambda_0 = (1 - \Pi^{(1)})/M^{(1)} \).
2. for \( j = 2:N \) do
3.  Compute the coefficients of the characteristic polynomial of \( \lambda^{(j)} \) to obtain \( C \) and \( \hat{C} \), see section A.1;
4.  Form the constraint matrices \( F^{(j)} \) and \( G^{(j)} \) as in (A.4) and (A.5);
5.  Solve \([\lambda_j, \mu_j]^T = (F^{(j)})^{-1}G^{(j)}\).
6.  if \( j = N \) then
7.    \( \lambda_1 = 0 \)
8.  end if
9. end for

**Output:** \( \{\mu_1, \ldots, \mu_N\} \) and \( \{\lambda_1, \ldots, \lambda_{N-1}\} \)

---

**ORCID iDs**

Pak-Wing Fok  @  https://orcid.org/0000-0001-9655-614X
References

[1] Chen S-J and Dill K A 2000 RNA folding energy landscapes Proc. Natl Acad. Sci. 97 646–51
[2] Dobson C M 2003 Protein folding and misfolding Nature 426 884–90
[3] Baker D 2000 A surprising simplicity to protein folding Nature 405 39–42
[4] Bryngelson J D, Onuchic J N, Socci N D and Wolynes P G 1995 Funnels, pathways, and the energy landscape of protein folding: a synthesis, Proteins: structure, Function, and Bioinformatics 21 167–95
[5] Berendsen H J C 1998 A glimpse of the holy grail! Science 282 642–3
[6] Evans E and Ritchie K 1997 Dynamic strength of molecular adhesion bonds Biophys. J. 72 1541–55
[7] Evans E and Ritchie K 1999 Strength of a weak bond connecting flexible polymer chains Biophys. J. 76 2439–47
[8] Osterhelt F, Rief M and Gaub H E 1999 Single molecule force spectroscopy by afm indicates helical structure of poly(ethylene-glycol) in water New J. Phys. 1 6
[9] Lipman E A, Schuler B, Bakajin O and Eaton W A 2003 Single-molecule measurement of protein folding kinetics Science 301 1233–5
[10] Schuler B and Hofmann H 2013 Single-molecule spectroscopy of protein folding dynamics–expanding scope and timescales Current opinion in structural biology 23 36–47
[11] Schaeffer R D, Fersht A and Daggett V 2008 Combining experiment and simulation in protein folding: closing the gap for small model systems Current opinion in structural biology 18 4–9
[12] Best R P, Paci E, Hummer G and Dudko O K 2008 Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules The Journal of Physical Chemistry B 112 5968–76
[13] Woodsidge M T and Block S M 2007 High-resolution, single-molecule measurements of biomolecular motion Annu. Rev. Biophys. Biomol. Struct. 36 171–90
[14] Fok P-W and Chou T 2010 Reconstruction of potential energy profiles from multiple rupture time distributions Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 466 3479–99
[15] Suzuki Y and Dudko O K 2010 Single-molecule rupture dynamics on multidimensional landscapes Phys. Rev. Lett. 104 048101
[16] Chang J C, Fok P-W and Chou T 2015 Bayesian uncertainty quantification for bond energies and mobilities using path integral analysis Biophys. J. 109 966–74
[17] Ros R, Schweginger F, Anselmetti D, Kubon M, Schäfer R, Plückthun A and Tiefenauer L 1998 Antigen binding forces of individually addressed single-chain fv antibody molecules Proc. Natl Acad. Sci. 95 7402–5
[18] Oberhauser A F, Hansma P K, Carrion-Vazquez M and Fernandez JM 2001 Stepwise unfolding of titin under force–clamp atomic force microscopy Proc. Natl Acad. Sci. 98 468–72
[19] Rief M, Gault M, Schemmel A and Gaub H E 1998 The mechanical stability of immunoglobulin and fibronectin iii domains in the muscle protein titin measured by atomic force microscopy Biophys. J. 75 3008–14
[20] Rollins G C and Dill K A 2014 General mechanism of two-state protein folding kinetics J. Am. Chem. Soc. 136 11420–7
[21] Bell G I 1978 Models for the specific adhesion of cells to cells Science 200 618–27
[22] Dudko O K, Hummer G and Szabo A 2006 Intrinsic rates and activation free energies from pulling experiments Phys. Rev. Lett. 96 108101
[23] Freund L B 2009 Characterizing the resistance generated by a molecular bond as it is forcibly separated Proc. Natl Acad. Sci. 106 8818–23
[24] Manuel A P, Lambert J and Woodsidge M T 2015 Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories Proc. Natl Acad. Sci. 112 7183–8
[25] Kramers H A 1940 Brownian motion in a field of force and the diffusion model of chemical reactions Physica 7 284–304
[26] Flores-Rivas D 1989 Approximate discrete-time schemes for statistics of diffusion processes Statistics: A Journal of Theoretical and Applied Statistics 20 547–57
[27] Nielsen J N, Madsen H and Young P C 2000 Parameter estimation in stochastic differential equations: an overview Annual Reviews in Control 24 83–94
[28] Wolff R W 1965 Problems of statistical inference for birth and death queuing models Oper. Res. 13 343–57
[29] Keiding N 1975 Maximum likelihood estimation in the birth-and-death process Ann. Math. Stat. 46 1348–50
[30] Evans J W, Graig W B and LeVeque R J 1980 On least squares exponential sum approximation with positive coefficients Math. Comput. 34 203–26
[31] Osborne M R and Smyth G K 1991 A modified prony algorithm for fitting functions defined by difference equations SIAM J. Sci. Stat. Comput. 12 362–82
[32] Osborne M R and Smyth G K 1995 A modified Prony algorithm for exponential function fitting SIAM J. Sci. Comput. 16 119–38
[33] Kung S-Y, Arun K S and Bhaskar Rao D V 1983 State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem JOSA 73 1799–811
[34] Varah J M 1985 On fitting exponentials by nonlinear least squares SIAM J. Sci. Stat. Comput. 6 30–44
[35] Machihara F 1983 On the properties of some infinitesimal generator Oper. Res. Lett. 2 123–6
[36] Hou S-H 1998 Classroom notes: a simple proof of the Leverrier–Faddeev characteristic polynomial algorithm SIAM Rev. 40 706–9