

Explicitly Solvable Continuous-time Inference for Partially Observed Markov Processes

Daniel Chen, Alexander G. Strang, Andrew W. Eckford *Senior Member, IEEE*, Peter J. Thomas

Abstract—Many natural and engineered systems can be modeled as discrete state Markov processes. Often, only a subset of states are directly observable. Inferring the conditional probability that a system occupies a particular hidden state, given the partial observation, is a problem with broad application. In this paper, we introduce a continuous-time formulation of the sum-product algorithm, which is a well-known discrete-time method for finding the hidden states’ conditional probabilities, given a set of finite, discrete-time observations. From our new formulation, we can explicitly solve for the conditional probability of occupying any state, given the transition rates and observations within a finite time window. We apply our algorithm to a realistic model of the cystic fibrosis transmembrane conductance regulator (CFTR) protein for exact inference of the conditional occupancy probability, given a finite time series of partial observations.

I. INTRODUCTION

Markov processes—dynamic processes whose future behavior depends only on their present state—approximate a wide variety of natural and engineered systems. Despite rapid advances in high-throughput data acquisition and data processing, many systems of interest contain important degrees of freedom that cannot be directly observed. Inferring the conditional probability that such a *partially observed Markov process* occupies specific hidden states, given the available observations, is a ubiquitous problem in science and engineering. Examples appear in robotics [1], ecology [2], neuroscience [3], and algorithmic text analysis [4].

We are motivated by biological examples in the present paper. Ion channels in excitable membranes, such as the sodium (Na^+) and potassium (K^+) channels described in Hodgkin and Huxley’s quantitative model for action potential generation [5], [6], [7], provide an early example. Discrete state Markov models based on Hodgkin and Huxley’s K^+ channel contain five states, only one of which conducts an ionic current; the other states are “silent” and cannot be distinguished by direct electrophysiological observation. Similarly,

the Na^+ channel has eight states: seven with zero conductance and one with nonzero conductance. Colquhoun and Hawkes introduced maximum likelihood methods for inferring the rate constants of a partially observed Markov process representing the nicotinic Acetylcholine receptor [8], [9], [10], [11], but did not address the question of inferring microscopic state occupancy from observable conductance time series. More recently, research into the molecular biology of cystic fibrosis (CF) has focused on the CF transmembrane conductance regulator (CFTR), which can be modeled as a 7-state system with two conducting states and five nonconducting states [12] (detailed below in Section V). Beyond these biological examples, problems of inferring or estimating hidden states from incomplete observations are widely studied in the signal processing literature [13].

The literature contains several approaches to approximating the behavior of hidden states of partially observed Markov processes. Sampling provides one common technique for approximate inference [14]. As an example, recent work by Fang *et al.* demonstrated an efficient algorithm for simulating stochastic reaction networks with multiple separated time scales using particle filters [15]. In general, Markov Chain Monte Carlo is a widely employed sampling technique used to infer hidden states [16], [17], [that has also been applied to ion channels \[18\]](#). For partially observed Bayesian networks operating in discrete time, message passing algorithms on factor graphs provide an efficient and exact inference method [19]. The factor-graph formalism is highly flexible. Algorithms based on the message passing concept have been extended to applications in localization [20], compressed sensing [21], and decision fusion [22]. [Factor graphs are not limited to models with a discrete number of variables. For example, Gaussian message passing in linear Gaussian models \(e.g. Kalman filtering and smoothing\) has been developed for continuous-time models with discrete-time observations \[23\], \[24\], \[25\]. However, the state reconstruction problem for continuous-time finite-state hidden Markov models has not been addressed in the literature, to be best of our knowledge.](#)

[In this work, we extend the message-passing algorithm in order to analytically interpolate state-occupancy probabilities of a continuous-time system, given a discretely sampled time series.](#) That is, we show how to infer the time-dependent conditional probabilities of latent states for continuous-time discrete-state [homogeneous](#) Markov processes given a set of partial observations over a finite time window. We derive an equivalent formulation of the sum-product algorithm in continuous time that allows one to find an explicit analytic solution for the state occupancy probability. Having

This work was supported in part by National Institutes of Health BRAIN Initiative grant R01 NS118606 and a National Science Foundation grant DMS-2052109 to PJT, as well as research support from the Oberlin College Libraries, and an NSERC Discovery grant to AWE.

D. Chen and P. J. Thomas are with the Department of Mathematics, Applied Mathematics, and Statistics; Department of Electrical, Control and Systems Engineering; Department of Computer and Data Science; Department of Biology, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: txc461/pjthomas@case.edu).

A. G. Strang is with Department of Statistics, University of Chicago, Chicago, IL 60637 USA (email: alexstrang@uchicago.edu).

A. W. Eckford is with Department of Electrical Engineering and Computer Science, York University, Toronto, ON M3J 1P3 Canada (e-mail: aeckford@yorku.ca).

explicit solutions lowers the computational cost and, unlike sampling-based approximate methods, does not sacrifice accuracy. Furthermore, the continuous-time formalism leads to elegant simplification of the analytic solutions. For certain systems—like the three-state systems shown in Figure 2—the conditional probability obeys a second-order inhomogeneous linear ordinary differential equation. Finally, we demonstrate the practical functionality of the algorithm with the 7-state model for CFTR using simulated data.

The paper is organized as follows. Section II reviews the message-passing algorithm for Markov processes with binary observations. Section III displays the main result of the paper: a continuous-time formulation of the sum-product algorithm. We present the derivation of the conditional probability using the message-passing approach in continuous time, and the corresponding sum-product algorithm. Section IV discusses the implications of the continuous-time formulation further. We give examples of small systems to display how analytic solutions may be found. In Section V we demonstrate the value of the algorithm for a larger, realistic system.

II. THEORY OF THE SUM-PRODUCT ALGORITHM

The sum-product algorithm can be used generally for inference on probabilistic models that can be written as factor graphs [19]. There are many variants of the sum-product algorithm, each suitable for accomplishing a different task. For our purposes, we will focus on the forward/backward algorithm for inference on hidden Markov models.

We consider a continuous-time, discrete-state homogeneous Markov process on a finite state space Ω . Given a discrete, uniformly spaced sampling interval, the continuous-time process induces a discrete-time Markov process specified by some column-stochastic transition matrix P that is invariant in time. Let $\mathcal{S} \subset \Omega$ be a subset of states, and let $S_t \in \Omega$ be the state of the system at time t . We assume an observer can only see whether S_t is in $\mathcal{S}$ or not. Accordingly, let $\mathcal{Y}_t = m(S_t)$ represent the observable where $m(s) : \Omega \rightarrow \{0, 1\}$ is the indicator function for the set $\mathcal{S}$. The goal of the algorithm is to infer the conditional probability of being at a particular state, $i \in \Omega$, given the binary observation.

The algorithm involves three vector-valued quantities: a forward message α_t , a backward message β_t , and the observation message χ_t . One can interpret the forward message as the probability of arriving at a certain state from time 0 to time t and the backward message as the likelihood of occupying a certain state at time t conditioned on ending up in a given target state, or a given target set of states, at the end of the measurement T . On the other hand, the observation message χ_t is an indicator function of the possible states given the observation. For instance, if at time t , the observation of the total system were $\mathcal{Y}_t = 1$, then χ_t would be a vector with 1's on the states in the observation set $\mathcal{S}$ and 0 otherwise. Using superscripts to denote vector indices, *i.e.* $v^{(i)}$ denotes the i -th element of vector v , we present pseudo-code for the sum-product algorithm in Algorithm 1. Note that, the normalization constant $Z = \sum_k \alpha_t^{(k)} \beta_t^{(k)}$ in Line 10 of Algorithm 1 is time invariant [26].

Algorithm 1 Forward/Backward Algorithm

Input: The transition matrix $P \in \mathbb{R}^{n \times n}$, the observation message χ_t for $t \in \{1, \dots, T\}$.

Output: The inferred probability p_t for $t \in \{1, \dots, T\}$.

```

1: Initialize the forward message  $\alpha_0 = \pi$ , the stationary
   distribution of  $P$ 
2: Initialize the backward message  $\beta_T^{(i)} = 1$ , for all  $i$ 
3: for  $t$  from  $2, \dots, T$  do
4:    $\alpha_t = \text{diag}(\chi_t) P \alpha_{t-1}$ 
5: end for
6: for  $s$  from  $T-1, \dots, 1$  do
7:    $\beta_s = P^T \text{diag}(\chi_{s+1}) \beta_{s+1}$ 
8: end for
9: for  $t$  from  $1, \dots, T$  do
10:   $p_t = \frac{\alpha_t \odot \beta_t}{\sum_k \alpha_t^{(k)} \beta_t^{(k)}}$ 
11: end for
```

Fig. 1 illustrates the sum-product algorithm's application to a time series of discretely sampled observations. Consider a three-state-chain with symmetric transition rates as shown in the top-left of Figure 1. If the system takes states 1 or 2, "0" will be observed, and "1" will be observed otherwise. Using the sum-product algorithm, we can find the probability of the system occupying each state given the discrete-time observations. As shown in the bottom row, as the sampling time step decreases, the conditional probabilities appear to converge to a smooth curve within each interval with a fixed observation (either "0" or "1"). Intuitively, there should exist a continuous, perhaps piecewise differentiable, representation of the conditional probability of a partially observed process. We formalize this intuition below.

III. CONTINUOUS-TIME MESSAGE PASSING

In this section, we present the main result, namely the derivation of the continuous-time message passing algorithm. In this new formalism, messages are passed in the form of linear differential equations on possible states given the observable system. In order to guarantee the existence of the continuous sum-product algorithm, we assume the following conditions.

Assumptions:

- A1 The continuous-time process $\{S(t); t \in [0, T]\}$ takes values in a finite state space $\Omega = \{1, 2, \dots, N\}$.
- A2 $S(t)$ has the Markov property, and has exponentially distributed waiting times parameterized by a rate matrix W , with w_{ji} specifying the transition rate from state i to state j . Note that W is constant in the interval $[0, T]$.
- A3 There is a distinguished subset $\mathcal{S} \subset \Omega$ such that the observable process $\mathcal{Y}(t)$ satisfies

$$\mathcal{Y}(t) = \begin{cases} 1 & \text{if } S(t) \in \mathcal{S}, \\ 0 & \text{otherwise.} \end{cases}$$

Further details appear in Appendix A.

Under these assumptions, we obtain a continuous-time version of the sum-product algorithm by executing the following steps (made rigorous in the proof of Theorem 1 below). Write

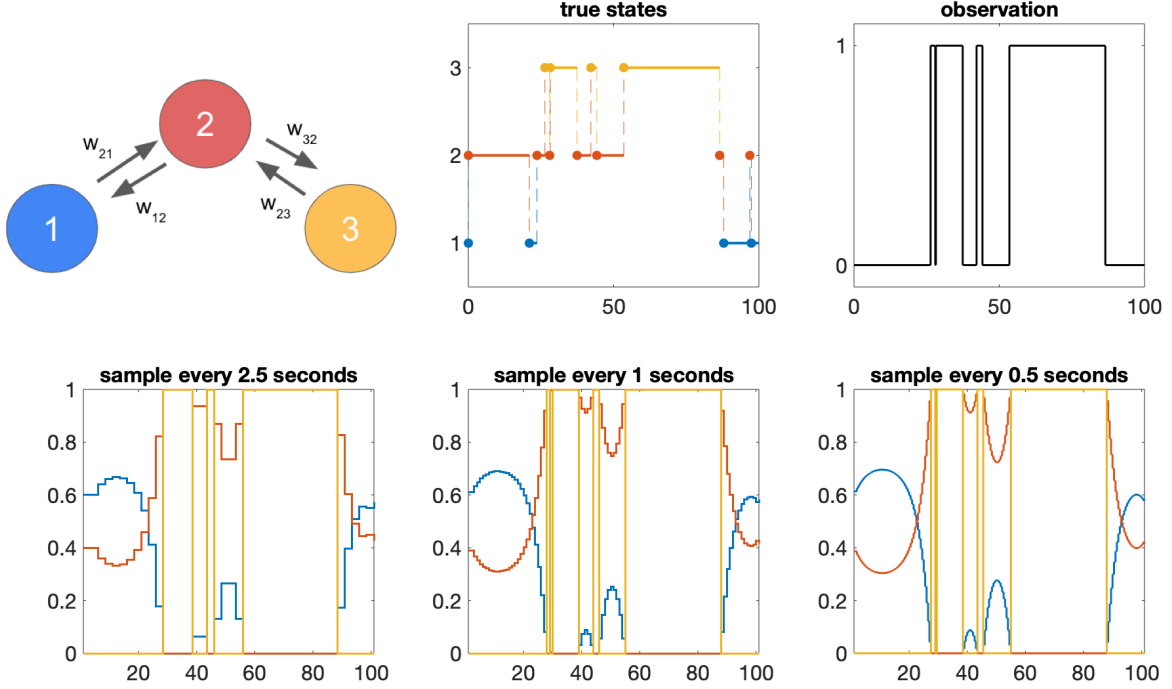


Fig. 1: Illustrating convergence of conditional state occupancy probabilities to a differentiable function for a three-state model. **(Top left)** The state diagram. **(Top middle)** True simulated states. **(Top right)** Binary observation derived from true states. **(Bottom row)** Inference of hidden states via the sum-product algorithm with time steps 2.5 sec **(Bottom Left)**, 1.0 sec **(Bottom Middle)** and 0.5 sec **(Bottom Right)**.

out the matrix multiplication of the discrete-time algorithm element-wise. Focus on one sojourn where the observation doesn't change. Within that time interval, take the limit as the time step goes to zero to derive the continuous-time dynamics of the conditional probabilities. Extend the solution to the full time interval via appropriate boundary conditions at the transition between each sojourn. The main result is stated in the theorem below.

Theorem 1. Suppose processes $S(t)$ and the associated process $\mathcal{Y}(t)$ satisfies assumptions A1, A2, and A3 above. Then, given a realization of the process $\mathcal{Y}(t)$, the conditional probability $\mathbf{p}(t) = \Pr[S(t)|\mathcal{Y}(t)]$ exists, is piecewise smooth (C^∞), and is C^∞ on all intervals where $\mathcal{Y}(t)$ is constant. In particular, $\mathbf{p}(t)$ takes the form

$$\mathbf{p}(t) = \frac{\boldsymbol{\rho}(t)}{\sum_k \rho^{(k)}(t)} = \frac{\boldsymbol{\alpha}(t) \odot \boldsymbol{\beta}(t)}{\sum_k \alpha^{(k)}(t) \odot \beta^{(k)}(t)} \quad (1)$$

where $\odot$ denotes the element-wise product, and the quantities $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}(t)$ are functions of time that follow the linear ordinary differential equations

$$\frac{d\alpha^{(i)}(t)}{dt} = \sum_{k \in S \setminus \{i\}} w_{ki} \alpha^{(k)}(t) - \sum_{l \neq i} w_{il} \alpha^{(i)}(t), \quad (2)$$

$$\frac{d\beta^{(j)}(t)}{dt} = - \sum_{k \in S \setminus \{j\}} w_{jk} \beta^{(k)}(t) + \sum_{l \neq j} w_{jl} \beta^{(j)}(t). \quad (3)$$

Proof. Without loss of generality, focus on the case where $\mathcal{Y}(0) = \mathcal{Y}(T) = 0$, and $\mathcal{Y}(t) = 1$ for $0 < t < T$. We use $q(t)$ to denote a quantity, q , evolving in continuous time on the interval $[0, T]$, and q_t to denote the same process sampled at discrete times.

Let the time interval $[0, T]$ be discretized with a step size $\Delta t = T/n$ for some integer $n \gg 1$. At each time step, $S(t)$ is sampled. Then, the sum-product algorithm can be used to solve for $\mathbf{p}_t$. Writing the matrix multiplication out yields the following set of equations for the forward and backward messages in discrete time:

$$\alpha_{t+\Delta t}^{(i)} = \sum_k \Pr[s_{t+\Delta t} = i | s_t = k] \alpha_t^{(k)} \chi_t^{(i)}, \quad (4)$$

$$\beta_t^{(i)} = \sum_k \Pr[s_{t+\Delta t} = k | s_t = i] \beta_{t+\Delta t}^{(k)} \chi_{t+\Delta t}^{(i)}. \quad (5)$$

We neglect states not in S because only the probability conditioned on the observations is of interest. Then, for any state $i \in S$, we argue that the corresponding forward message, $\alpha^{(i)}(t)$, and backward message, $\beta^{(i)}(t)$ in continuous time can be written as solutions of systems of differential equations, upon taking limits as $\Delta t \rightarrow 0$. For notational simplicity, for

$t > \tau$ we let $\mathcal{P}_{t,\tau}^{(i,j)} = \Pr[S(t) = i | S(\tau) = j]$.

$$\begin{aligned} & \frac{d\alpha^{(i)}(t)}{dt} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\alpha^{(i)}(t + \Delta t) - \alpha^{(i)}(t)}{\Delta t} \end{aligned} \quad (6)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[\sum_k \mathcal{P}_{t+\Delta t,t}^{(i,k)} \alpha^{(k)}(t) \chi^{(i)}(t) - \alpha^{(i)}(t) \right] \quad (7)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[\sum_{k \in S \setminus \{i\}} (w_{ki} \Delta t + o(\Delta t)) \alpha^{(k)}(t) \dots \right. \\ \left. - \sum_{l \neq i} (w_{il} \Delta t + o(\Delta t)) \alpha^{(i)}(t) \right] \quad (8)$$

$$= \sum_{k \in S \setminus \{i\}} w_{ki} \alpha^{(k)}(t) - \sum_{l \neq i} w_{il} \alpha^{(i)}(t) \quad (9)$$

$$\begin{aligned} & \frac{d\beta^{(i)}(t)}{dt} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\beta^{(i)}(t + \Delta t) - \beta^{(i)}(t)}{\Delta t} \end{aligned} \quad (10)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[\beta^{(i)}(t + \Delta t) \dots \right. \\ \left. - \sum_k \mathcal{P}_{t+\Delta t,t}^{(k,i)} \beta^{(k)}(t + \Delta t) \chi^{(i)}(t + \Delta t) \right] \quad (11)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[- \sum_{k \in S \setminus \{i\}} (w_{ik} \Delta t + o(\Delta t)) \beta^{(k)}(t + \Delta t) \right. \\ \left. + \sum_{l \neq i} (w_{il} \Delta t + o(\Delta t)) \beta^{(i)}(t + \Delta t) \right] \quad (12)$$

$$= - \sum_{k \in S \setminus \{i\}} w_{ik} \beta^{(k)}(t) + \sum_{l \neq i} w_{il} \beta^{(i)}(t) \quad (13)$$

where $\lim_{\Delta t \rightarrow 0} o(\Delta t)/\Delta t = 0$. Readers could refer to Appendix A or consult existing literature such as [27] for the relationship between transition probabilities and transition rates of a Markov jump process.

The conditional probability can be found by solving the differential equations, taking the component-wise product of $\alpha(t)$ and $\beta(t)$ for every $t \in [0, T]$, and normalizing so as to obtain a valid probability distribution.

The differential equation formulation is only applicable for the time intervals where the observation $\mathcal{Y}(t)$ is constant. When the observable $\mathcal{Y}(t)$ changes (when the systems transitions from a state in S to a state not in S), the probability with respect to time might not be differentiable; in some cases, it is not even continuous. Therefore, we must specify boundary conditions to connect the probabilities from one sojourn to the next. The observable may change either by the system entering S or else leaving S . Suppose a transition occurred within time $(t_* - \Delta t, t_*]$ such that, for $t < t_* - \Delta t$, $S(t) \in S$, and $S(t) \notin S$ for $t \geq t_*$. Call this event E . Then, we obtain

the following transition rule for the forward message.

$$\begin{aligned} & \Pr[S(t_*) = j | E] \\ &= \frac{\sum_{i \in S} \mathcal{P}_{t_*, t_* - \Delta t}^{j,i} \Pr[S(t_* - \Delta t) = i]}{\sum_{k \notin S} \sum_{i \in S} \mathcal{P}_{t_*, t_* - \Delta t}^{k,i} \Pr[S(t_* - \Delta t) = i]} \end{aligned} \quad (14)$$

$$= \frac{\sum_{i \in S} (w_{ji} \Delta t + o(\Delta t)) \Pr[S(t_* - \Delta t) = i]}{\sum_{k \notin S} \sum_{i \in S} (w_{ki} \Delta t + o(\Delta t)) \Pr[S(t_* - \Delta t) = i]} \quad (15)$$

$$\rightarrow \frac{\sum_{i \in S} w_{ji} \Pr[S(t_*^-) = i]}{\sum_{k \notin S} \sum_{i \in S} w_{ki} \Pr[S(t_*^-) = i]} \quad (16)$$

as $\Delta t \rightarrow 0$. Here $\Pr[S(t_*^-) = i]$ is the probability of occupying state i the instant before the transition, which can be found by solving the differential equations introduced above.

We handle the boundary conditions at state transitions for the backward message similarly. Define E as above and let s^* be a particular goal state. Then:

$$\begin{aligned} & \Pr[S(T) = s^* | S(t_*) = j, E] \\ &= \sum_{i \notin S} \Pr[S(T) = s^*, S(t_*) = i | S(t_*) = j, E] \\ &= \frac{\sum_{i \notin S} \Pr[S(T) = s^* | S(t_*) = i] \mathcal{P}_{t_*, t_* - \Delta t}^{i,j}}{\sum_{k \in S} \sum_{i \notin S} \mathcal{P}_{T, t_*}^{s^*, i} \mathcal{P}_{t_*, t_* - \Delta t}^{i,k}} \end{aligned} \quad (17)$$

$$= \frac{\sum_{i \notin S} \Pr[S(T) = s^* | S(t_*) = i] (w_{ij} \Delta t + o(\Delta t))}{\sum_{k \in S} \sum_{i \notin S} \mathcal{P}_{T, t_*}^{s^*, i} (w_{ik} \Delta t + o(\Delta t))} \quad (18)$$

$$= \frac{\sum_{i \notin S} \Pr[S(T) = s^* | S(t_*) = i] (w_{ij} + \frac{o(\Delta t)}{\Delta t})}{\sum_{k \in S} \sum_{i \notin S} \mathcal{P}_{T, t_*}^{s^*, i} (w_{ik} + \frac{o(\Delta t)}{\Delta t})} \quad (19)$$

$$\rightarrow \frac{\sum_{i \notin S} w_{ij} \Pr[S(T) = s^* | S(t_*) = i]}{\sum_{k \in S} \sum_{i \notin S} w_{ik} \Pr[S(T) = s^* | S(t_*) = i]} \quad (20)$$

as $\Delta t \rightarrow 0$. Here, $\Pr[S(T) = s^* | S(t_*) = i]$ is a hitting probability associated with the backward message at time t_* , which can be found by solving the backward-message differential equation. These boundary conditions, together with the differential equations (2)-(3), give the continuous-time evolution of the conditional probability for any finite-length observations.

Note that the equations (2)-(3) extend to the case where $\mathcal{Y}(0) = \mathcal{Y}(T) = 1$ and $\mathcal{Y}(t) = 0$ for $0 < t < T$ by viewing $S \leftarrow \Omega \setminus S$. So, given a time series observation $\mathcal{Y}(t)$ where observation (the value of $\mathcal{Y}(t)$) changes at time $0 < t_1 < t_2 < \dots < t_m$, we can solve for the analytic solution at any interval with consistent observation $(t_i, t_{i+1}]$ using equation (2) and (3). Then, use the result to compute the initial condition for the next interval — namely, $(t_{i+1}, t_{i+2}]$ for the forward message and $(t_{i-1}, t_i]$ for the backward message — as specified in (16) and (20). **Thus, the statement of Theorem 1 holds.** $\square$

A general expression for the conditional probability can be obtained, but it is not of great utility in most systems. Yet, there are certain special cases that yield elegant solutions; we introduce several examples in Section IV.

The continuous-time sum-product algorithm follows directly from the derivation above, and is outlined in Algorithm 2.

The discrete algorithm passes information through matrix multiplication of a truncated transition matrix; the continuous-time algorithm does the same by solving a system of linear differential equations using the truncated rate matrix. Since the final conditional probability is only piecewise differentiable, the boundary condition must be applied whenever a transition in or out of the observable set $\mathcal{S}$ occurs.

Algorithm 2 Continuous-time Forward/Backward Algorithm

Input: The rate transition matrix $W \in \mathbb{R}^{n \times n}$, the observed process $\mathcal{Y}(t)$.

Output: The inferred probability $p(t)$ for $t \in [0, T]$.

- 1: Let $[t_1, t_2, \dots, t_m]$ be a list of times where transitions occur
 - 2: $\tau \leftarrow 0$
 - 3: $\alpha^* \leftarrow \pi$, the stationary distribution
 - 4: **for** j from $1, \dots, m$ **do**
 - 5: $\alpha_j \leftarrow$ solution to the forward message differential equation (Equation 2) from τ to t_j with initial condition α^*
 - 6: $\tau \leftarrow t_j$
 - 7: $\alpha^* \leftarrow$ distribution specified according to Equation 16
 - 8: **end for**
 - 9: $\tau \leftarrow T$
 - 10: $\beta^* \leftarrow$ the uniform distribution
 - 11: **for** j from $m, \dots, 1$ **do**
 - 12: $\beta_j \leftarrow$ solution to the backward message differential equation (Equation 3) from τ to t_j with initial condition β^*
 - 13: $\tau \leftarrow t_j$
 - 14: $\beta^* \leftarrow$ distribution specified according to Equation 20
 - 15: **end for**
 - 16: $\alpha(t), \beta(t) \leftarrow$ concatenation α_j 's and β_j 's
 - 17: Compute $\rho(t) = \alpha(t) \odot \beta(t)$, the component-wise product between $\alpha(t)$ and $\beta(t)$ pointwise with respect to t
 - 18: Compute the conditional probability $p(t) = \frac{\rho(t)}{\sum_k \rho^{(k)}(t)}$
-

From a practical perspective, having the ability to solve for the conditional probabilities exactly through differential equations drastically lowers the computational cost of the forward/backward algorithm. Traditionally, the discrete-time algorithm propagates the forward and backward messages through matrix operations *at each time step*. For long time-series and/or high-dimensional systems, this is computationally prohibitive. Through our continuous-time formalism, we solve the differential equations analytically, which is an operation that is independent of the length of the time-series, to find the forward or backward message at any time point. This difference effectively reduces the asymptotic scaling from $\mathcal{O}(\Delta t^{-1})$ to $\mathcal{O}(1)$, with the later scaling only in the number of observable transitions. In scenarios where finding the appropriate boundary condition would require an iterative procedure of solving the forward and backward messages multiple times, our continuous time approach should be much more efficient than the traditional discrete-time method. We discuss the performance of the continuous-time message-

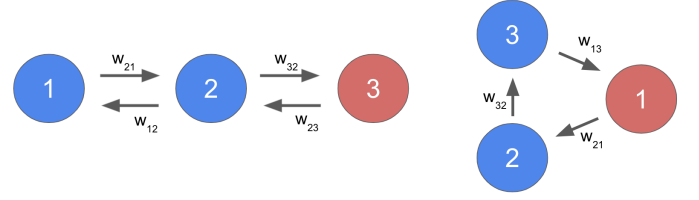


Fig. 2: State diagrams of two systems for which the continuous-time message passing algorithm exhibit analytic simplifications. **(Left)** 3-state chain with symmetric rates $w_{12} = w_{21}$. **(Right)** Irreversible 3-state loop. States marked in red return $\mathcal{Y}(t) = 1$ and blue return $\mathcal{Y}(t) = 0$.

passing algorithm further in Section V.

IV. ANALYTIC SOLUTION

Theorem 1 in the previous section shows that the conditional probability is always available analytically upon normalizing the component-wise product of the forward and backward messages, *i.e.* $p(t) = \rho(t)/Z$ where $Z = \sum_k \rho^{(k)}(t)$. As in the discrete-time case, the normalizing term Z is time-invariant in the continuous-time case as well. See Appendix B. The conditional probability may therefore be expressed in a particularly elegant form in certain cases, namely as the solution of a linear nonhomogeneous second-order differential equation. We begin this section by considering two examples. Following the examples, we consider extensions to higher dimensions.

A. Symmetric 3-State Chain

Consider the three-state chain depicted in the left panel of Figure 2, where states 1 and 2 are hidden. Assume that the per-capita transition rates within the hidden block are symmetric, *i.e.* $w_{12} = w_{21} > 0$, and assume $w_{13} = w_{31} = 0$. The rates $w_{23} > 0$ and $w_{32} > 0$ may be arbitrary. These assumptions result in the following rate matrix:

$$W = \begin{pmatrix} -w_{21} & w_{12} & 0 \\ w_{21} & -(w_{12} + w_{32}) & w_{23} \\ 0 & w_{32} & -w_{23} \end{pmatrix}. \quad (21)$$

In this case, let $\mathcal{S} = \{3\}$, the singleton set of state 3. When $\mathcal{Y}(t) = 1$ the inference problem is trivial since the system takes state 3 with probability one. Thus, we emphasize the intervals when $\mathcal{Y}(t) = 0$.

First, consider the forward message given by the following system of differential equations

$$\frac{d\alpha}{dt} = \begin{pmatrix} -w_{21} & w_{12} \\ w_{21} & -(w_{12} + w_{32}) \end{pmatrix} \alpha. \quad (22)$$

Note that the matrix defining the system of equations corresponds to the upper left block of W . To simplify notation, let $w_{21} = w_{12} = a$ and $w_{32} = b$. Then, the submatrix reduces to the following form:

$$\frac{d\alpha}{dt} = \begin{pmatrix} -a & a \\ a & -a - b \end{pmatrix} \alpha. \quad (23)$$

This is a linear system of differential equations that can be solved exactly. The sub-matrix is real-symmetric, so is diagonalizable. Therefore, the solution will be of the form:

$$\alpha(t) = Ae^{\lambda_1 t} \mathbf{v}_1 + Be^{\lambda_2 t} \mathbf{v}_2. \quad (24)$$

where λ_i is an eigenvalue of the rate submatrix and $\mathbf{v}_i$ is the corresponding eigenvector. The eigenvalues and vectors are:

$$\lambda_{1/2} = a(-1 \pm \sqrt{1 + \gamma^2}) - \gamma, \quad (25)$$

$$\mathbf{v}_{1/2} = \begin{pmatrix} \gamma \pm \sqrt{1 + \gamma^2} \\ 1 \end{pmatrix}, \quad (26)$$

where $\gamma = b/2a$. Constants A and B are found through the initial condition given below. In the previous section, the initial forward message was set to the equilibrium distribution. However, the chain structure leaves no ambiguity in the state occupied when the observation changes from 1 to 0. So, the initial forward message is the delta distribution on state 2:

$$\alpha(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (27)$$

After some time T , the system re-enters the visible state, namely, state 3 again. By the same reasoning, we also have an unambiguous boundary condition for the backward message:

$$\beta(T) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (28)$$

Also, by the symmetry in the rates, the backward message evolves according to the same equations as the forward message, but backward in time. Thus, $\beta(t) = \alpha(T - t)$,

$$\beta(t) = Ae^{\lambda_1(T-t)} \mathbf{v}_1 + Be^{\lambda_2(T-t)} \mathbf{v}_2 \quad (29)$$

where λ_i 's, $\mathbf{v}_i$'s, constants A and B all remain the same due to symmetry.

The conditional probability is proportional to the component-wise product $\rho = \alpha \odot \beta$, which yields:

$$\rho(t) = \mathcal{A}[e^{\lambda_1(T-t)+\lambda_2 t} + e^{\lambda_1 t+\lambda_2(T-t)}] + \mathcal{B}, \quad (30)$$

where

$$\mathcal{A} = AB(\mathbf{v}_1 \odot \mathbf{v}_2) = (-AB \quad AB)^\top, \quad (31)$$

$$\mathcal{B} = A^2 e^{\lambda_1 T}(\mathbf{v}_1 \odot \mathbf{v}_1) + B^2 e^{\lambda_2 T}(\mathbf{v}_2 \odot \mathbf{v}_2). \quad (32)$$

To recover the conditional probability, we must normalize such that the component sum evaluates to 1. In this case, since the submatrix in Equation 22 is symmetric, the eigenvectors of the submatrix are orthogonal to each other, which means the sum of the components of $\mathbf{v}_1 \odot \mathbf{v}_2$, or the inner-product between $\mathbf{v}_1$ and $\mathbf{v}_2$, evaluates to 0. Then, the component-wise sum of $\mathcal{A}$ is zero, so the normalizing constant equals the component-wise sum of $\mathcal{B}$.

Write $\mathbf{p}(t)$ to represent the conditional probability. We write $Z = \sum_k \mathcal{B}^{(k)}$ to represent the normalizing constant. The final equation describing the conditional probability can thus be rewritten in the following form:

$$\mathbf{p}(t) = \frac{1}{Z} \left(\mathcal{A} e^{(\lambda_1 - \lambda_2)t} e^{\lambda_2 T} + \mathcal{A} e^{(\lambda_2 - \lambda_1)t} e^{\lambda_1 T} + \mathcal{B} \right). \quad (33)$$

The components of $\mathbf{p}(t)$ may also be expressed as the solutions of a second-order ordinary differential equation, cf. (51)-(52).

B. Irreversible 3-State Loop

Light-gated Channelrhodopsin-2 (ChR2) receptors can be modeled with a 3-state chain where each vertex has out-degree 1 and forms a cycle, as depicted in Figure 2, right panel [28], [29]. Let state 1 be open, and states 2 and 3 be closed. That is, $\mathcal{S} = \{1\}$. The open/closed status of the channel is observed through voltage recordings: high conductance indicates that the channel is in the open state, and low conductance implies the closed state. The rate matrix can be written as the following:

$$W = \begin{pmatrix} -w_{21} & 0 & w_{13} \\ w_{21} & -w_{32} & 0 \\ 0 & w_{32} & -w_{13} \end{pmatrix} \quad (34)$$

where w_{ji} is the transition rate from state i to state j . Conditioning on the channel being closed, the message-passing algorithm takes the lower right 2×2 block as the transition matrix.

When transitioning from $\mathcal{S}$ to a state not in $\mathcal{S}$, the system must enter state 2 first and exit through state 3. Thus the boundary conditions for the forward and backward message are

$$\alpha(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \beta(T) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (35)$$

where the first component corresponds to state 2 and second state 3. Suppose $w_{32} \neq w_{13}$ and let $\gamma = \frac{w_{32}}{w_{13} - w_{32}}$. Then, the solution of the message-passing differential equations with the initial condition enforced satisfies:

$$\alpha(t) = e^{-w_{32}t} \begin{pmatrix} 1 \\ \gamma \end{pmatrix} - \gamma e^{-w_{31}t} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (36)$$

$$\beta(t) = e^{-w_{32}(T-t)} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{1}{\gamma} e^{-w_{13}(T-t)} \begin{pmatrix} -\gamma \\ 1 \end{pmatrix}. \quad (37)$$

To get the conditional probability, first take the component-wise product between the forward and backward message at time t .

$$\rho(t) = e^{-w_{32}t - w_{13}(T-t)} \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \begin{pmatrix} e^{-w_{32}T} \\ -e^{-w_{13}T} \end{pmatrix} \quad (38)$$

As in the previous example, the normalizing constant is invariant in time. In this case, $Z = e^{-w_{32}T} - e^{-w_{13}T}$. Thus, the time evolution of the conditional probability can be written

$$\mathbf{p}(t) = \rho(t)/Z. \quad (39)$$

Next, we consider a case in which the submatrix is not diagonalizable. We set $w = w_{32} = w_{13}$, so that the corresponding submatrix:

$$U = \begin{pmatrix} -w & 0 \\ w & -w \end{pmatrix} \quad (40)$$

admits the following Jordan normal form:

$$U = \begin{pmatrix} 0 & w \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -w & 1 \\ 0 & -w \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \frac{1}{w} & 0 \end{pmatrix}. \quad (41)$$

Using the same initial condition as Equation (35) to solve the system of differential equations yields the following forward message.

$$\alpha(t) = \begin{pmatrix} e^{-wt} \\ te^{-wt} \end{pmatrix} \quad (42)$$

We solve for the backward message by undergoing a similar procedure and yield

$$\beta(t) = \begin{pmatrix} (T-t)e^{-w(T-t)} \\ e^{-w(T-t)} \end{pmatrix}. \quad (43)$$

Upon taking the component-wise product, we **observe** that the evolution of conditional probability is independent of time:

$$\rho(t) = e^{-wT} \begin{pmatrix} T-t \\ t \end{pmatrix}, \quad Z = \frac{1}{Te^{-wT}}. \quad (44)$$

Since both the time entering and exiting the hidden states are fixed, we can see that the probability flows linearly from one state to the other at equal rates.

Time-invariant normalization is a fundamental property of the sum-product algorithm [26]. In Appendix B we give an elementary demonstration of this property for continuous-time systems with time-homogeneous transition rates, as illustrated by the two cases presented above.

C. Generalization

We show that under certain circumstances, the conditional probability follows a second-order nonhomogeneous linear ordinary differential equation.

Corollary 1. *For 3-state systems where the truncated submatrix U has distinct eigenvalues, the conditional probability can be written as the solution of a nonhomogeneous second-order linear differential equation with constant coefficients.*

Proof. Let $(\lambda_i, \mathbf{v}_i)$ be eigenpairs for U and let $(\lambda_i, \mathbf{w}_i)$ be eigenpairs for $U^\top$. Then the forward and backward messages are given by the expressions:

$$\alpha(t) = Ae^{\lambda_1 t} \mathbf{v}_1 + Be^{\lambda_2 t} \mathbf{v}_2, \quad (45)$$

$$\beta(t) = Ce^{\lambda_1(T-t)} \mathbf{w}_1 + De^{\lambda_2(T-t)} \mathbf{w}_2. \quad (46)$$

Taking the component-wise product yields the form:

$$\rho(t) = \mathcal{A}[e^{\lambda_1 t + \lambda_2(T-t)}] + \mathcal{B}[e^{\lambda_1(T-t) + \lambda_2 t}] + \mathcal{C}, \quad (47)$$

$$= \mathcal{A}[e^{(\lambda_1 - \lambda_2)t} e^{\lambda_2 T}] + \mathcal{B}[e^{(\lambda_2 - \lambda_1)t} e^{\lambda_1 T}] + \mathcal{C}, \quad (48)$$

where

$$\mathcal{A} = AD(\mathbf{v}_1 \odot \mathbf{w}_2), \quad \mathcal{B} = BC(\mathbf{v}_2 \odot \mathbf{w}_1), \quad (49)$$

$$\mathcal{C} = ACe^{\lambda_1 T}(\mathbf{v}_1 \odot \mathbf{w}_1) + BDe^{\lambda_2 T}(\mathbf{v}_2 \odot \mathbf{w}_2). \quad (50)$$

Since the left and right eigenvectors corresponding to different eigenvalues are orthogonal, all time-dependent terms cancel when normalizing. **Thus the normalization constant** $Z = \sum_k \mathcal{C}^{(k)}$. Upon differentiating $\mathbf{p}(t) = \rho(t)/Z$ twice with respect to time, we arrive at the second-order linear differential equation that describes the time evolution of the condition probability:

$$\frac{d^2 \mathbf{p}}{dt^2} = \frac{(\lambda_1 - \lambda_2)^2}{Z} [\mathcal{A}e^{(\lambda_1 - \lambda_2)t} e^{\lambda_2 T} + \mathcal{B}e^{(\lambda_2 - \lambda_1)t} e^{\lambda_1 T}] \quad (51)$$

$$= \frac{(\lambda_1 - \lambda_2)^2}{Z} (\mathbf{p} - \mathcal{C}). \quad (52)$$

This second order equation comes equipped with two boundary conditions set by fixing $\alpha(0)$ and $\beta(T)$. $\square$

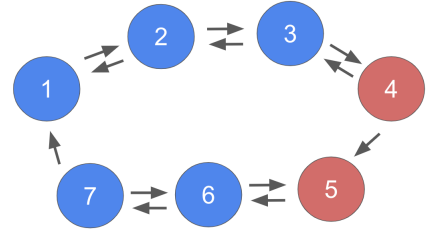


Fig. 3: State diagram for CFTR. When the protein enters one of the red states, the ion channels will open and conduct a current. On the other hand, when CFTR is in one of the blue states, then the ion channels are closed and conducts no current. For this system $\mathcal{S} = \{4, 5\}$.

Corollary 2. *For a system of arbitrary dimension, for which the truncated submatrix U has exactly two distinct eigenvalues and is diagonalizable, the conditional probability can be written as a second-order linear differential equation of the same form as Equation 52.*

Proof. Let λ_1 and λ_2 be the two distinct eigenvalues. Let $\mathbf{v}_i$ be the sum of all right eigenvectors corresponding to eigenvalue λ_i , and let $\mathbf{w}_i$ be the sum of all left eigenvectors corresponding to eigenvalue value λ_i . Then by diagonalizability, the proof for this corollary follows the proof of Corollary 1. $\square$

This result extends Corollary 1 to higher dimensional systems, under special conditions. These two corollaries match intuition: we are concerned with the state occupancy probability conditioned on both the entrance and exit times of an observable state, so the probability flow obeys a second-order differential equation, which also requires specifying two boundary conditions. We note, however, that if the submatrix has *three or more* distinct eigenvalues, we do not expect a similar result to hold. Moreover, generic matrices typically have as many distinct eigenvalues as their dimension, so we do not expect this scenario to occur outside of special cases when the rate submatrix is highly regular. In addition, it would be difficult to check if a given rate submatrix satisfies these conditions through numerical solvers, as equality of repeated eigenvalues might be obscured by floating point arithmetic. Therefore, as much as the second-order differential equations interpretation is intuitively appealing, it is likely not a generic phenomenon that we should expect for arbitrary rate matrices.

V. BIOLOGICAL EXAMPLE: CFTR

Here, we show that our algorithm applies to a higher dimensional system, namely the 7-state model of the cystic fibrosis transmembrane conductance regulator (CFTR) protein. CF is a common life-threatening genetic disorder. CFTR is an important protein that regulates the opening and closing of ion channels. Loss of CFTR function causes pancreatic insufficiency as well as airway infection due to excessive mucus, which in turn can cause a variety of complications such as impaired innate immunity and respiratory failure [30].

Mathematically, the behavior of CFTR can be captured with a 7-state hidden Markov process. Figure 3 illustrates

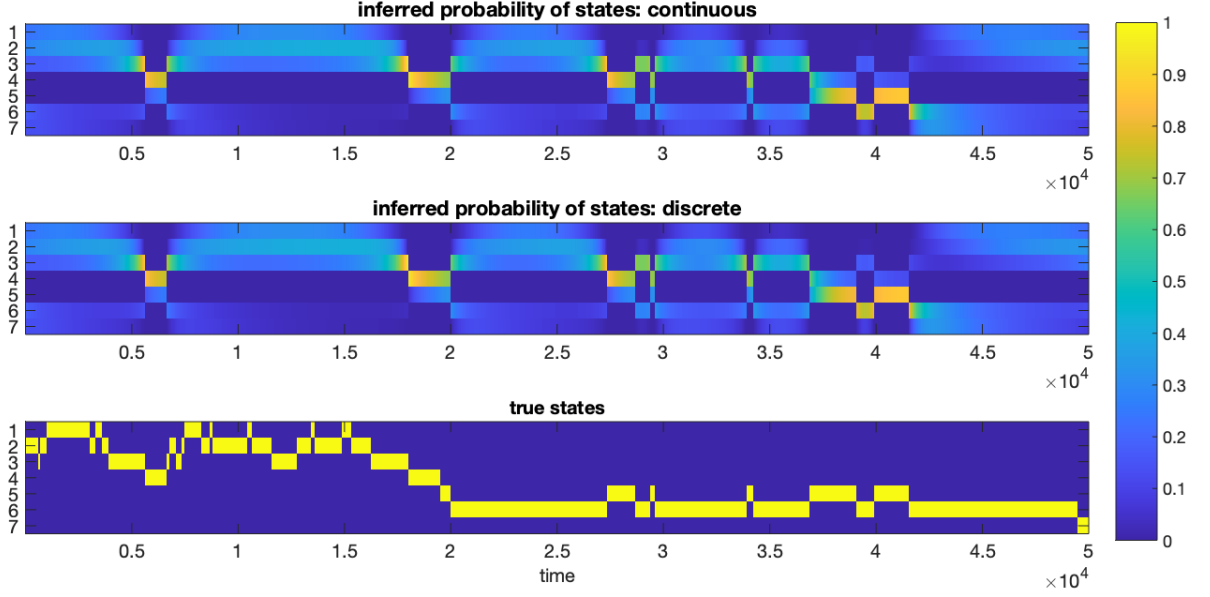


Fig. 4: The inference result of the continuous-time algorithm (**top**) and the discrete-time algorithm (**middle**) for the true state occupancy simulated from the Gillespie algorithm (**bottom**). Sampling time step $\Delta t = 10^{-4}$.

the state diagram [12]. The ion channel opens (conducts an ionic current) when CFTR is in state four or five (marked in red). When in the nonconducting states (marked in blue) the ion channel is closed. Transmembrane conductance recordings report whether CFTR is in a conducting or a nonconducting state, but do not directly indicate which of the possible states is occupied. Using the numbering shown in Figure 3, the rate matrix is defined as follows:

$$W = \begin{pmatrix} -9.9 & 5.0 & 0 & 0 & 0 & 0 & 1.7 \\ 9.9 & -12.7 & 5.8 & 0 & 0 & 0 & 0 \\ 0 & 7.7 & -10.7 & 10.0 & 0 & 0 & 0 \\ 0 & 0 & 4.9 & -17.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7.1 & -3.0 & 7.0 & 0 \\ 0 & 0 & 0 & 0 & 3.0 & -13.0 & 12.8 \\ 0 & 0 & 0 & 0 & 0 & 6.0 & -14.5 \end{pmatrix}. \quad (53)$$

We used Gillespie’s exact stochastic simulation algorithm [31], [32], [33] to generate sample traces of the ion channel states and recordings. The simulation produces discrete state, continuous time trajectories. We introduced a finite sampling time step to discretize the simulated data along the time axis, consistent with data obtained through experimental recordings. Then, we place the generated data into discretized time bins where the size of each bin (one can think of this as the sampling time step) is a parameter that can be altered. We solved the differential equations giving the forward and backward messages exactly via function handles in Matlab. Figure 4 compares the traces produced by the classical discrete-time algorithm, our continuous-time algorithm, and the true states, and shows excellent agreement among the respective curves.

With a sufficiently small time step ($\Delta t = 10^{-4}$), the continuous and discrete-time algorithms show no visible discrepancy, as expected. Figure 5 shows that the maximum discrepancy

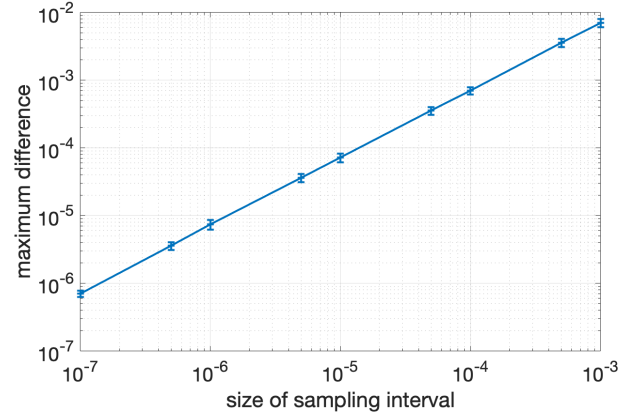


Fig. 5: First-order convergence of the message passing algorithms as a function of sampling time step. The mean maximum difference was plotted along with one standard deviation about the mean as the error bar for an ensemble of forty trajectories.

(ℓ_∞ norm) between the conditional probabilities generated by the continuous time and discrete time algorithms decreases linearly with the sampling time step Δt . The figure shows the results from an ensemble of forty independent repeated trials for each sample size. Again, we emphasize that while the discrete time algorithm requires iterating over all samples, which scales $\mathcal{O}(\Delta t^{-1})$, the continuous time algorithm outputs a *description* of the conditional probability — that is, a function that returns a value at an exact time point queried — in $\mathcal{O}(1)$ time, scaling only with the number of transitions.

VI. DISCUSSION & CONCLUSIONS

This paper presents an algorithm for continuous-time inference on partially observable Markov processes with discrete state spaces. We show that the well-known sum-product algorithm can be extended to the continuous-time domain via two sets of differential equations and pointwise normalization. In the continuous time setting, we were able to solve the trajectory of conditional probabilities exactly given a finite time-series. Moreover, we find that the dynamics of the state occupancy probabilities can be reduced to second-order differential equations under special circumstances. These results are valuable not only for their mathematical interest, but also because they have the potential to reduce the inference problem to solving systems of linear differential equations, with a potentially significant reduction in computational complexity for long time series. Numerically, the continuous-time algorithm is consistent with the discrete-time algorithm in the limit of small time step, but executes in approximately constant time rather than linearly in the number of time steps.

Our formalism extends naturally to non-binary observations. Briefly, let $S_1, S_2, \dots, S_m$ be a partition over the sample space Ω , and suppose the observable process $\mathcal{Y}$ is given as:

$$\mathcal{Y}(t) = i \text{ if } S(t) \in S_i. \quad (54)$$

For a sojourn with observation i , we can apply the forward and backward message-passing scheme as introduced for the binary case, viewing the observation as either in S_i or $\Omega \setminus S_i$. At the boundaries, the same technique (as in the proof of Theorem 1) can be used for finding the update rule by noting the possible transitions from S_i to S_j , for all $j \neq i$. This approach can finally be extended to an arbitrary collection of subsets of Ω where the elements are not necessarily disjoint. One may accomplish this extension by expanding all unions and intersections as disjoint sets possibly with the same observable. We defer detailed investigations in this direction to future work.

Extending the continuous-time message passing to the case with inhomogeneous transition rates remains an open problem. We expect a derivation similar to the proof of the main theorem would be applicable, possibly with smoothness constraints on the transition rates. While the normalization constant will remain invariant in time, we do not expect a result such as Corollary 1 to hold beyond constant transition rates.

Finally, we note the relationship between conditional probability and a second-order differential equation is intuitively satisfying: the entry and exit times act as two boundary conditions that fix the endpoints of the evolution, whereas the unconstrained forward evolution equation, a first-order differential equation, requires only the starting condition. It is an interesting question for future work to investigate under which assumptions a similar result as Corollary 1 would hold for systems with more than two distinct eigenvalues.

APPENDIX A

NOTATION AND PRELIMINARIES

For completeness, we define a continuous-time discrete-space Markov process with exponential waiting times.

Definition 1 (Markov Process). *Let $\{S(t); t \in [0, T]\}$ be a discrete-space, continuous-time stochastic process where for each t , $S(t)$ is a random variable with state space $\Omega = \{1, 2, \dots, N\}$, $N < \infty$. Then, the process $S(t)$ has the Markov Property if for any $s_1, s_2, \dots, s_m \in \Omega$, $0 < \tau_1 < \tau_2 < \dots < \tau_{m-1} < t$,*

$$\begin{aligned} \Pr[S(t) = s_m | S(\tau_1) = s_1, S(\tau_2) = s_2, \dots, S(\tau_{m-1}) = s_{m-1}] \\ = \Pr[S(t) = s_m | S(\tau_{m-1}) = s_{m-1}]. \end{aligned} \quad (55)$$

Furthermore, the process has exponential waiting times if for any $i, j \in \Omega$, $i \neq j$, there is a constant rate $0 \leq w_{ji} < \infty$ such that

$$\begin{cases} \Pr[S(t + \Delta t) = i | S(t) = i] = 1 - \sum_{k \neq i} w_{ki} \Delta t + o(\Delta t) \\ \Pr[S(t + \Delta t) = j | S(t) = i] = w_{ji} \Delta t + o(\Delta t) \\ \Pr[S(t + \Delta t) = j, S(t + 2\Delta t) = k | S(t) = i] = o(\Delta t), \quad \forall k \end{cases}$$

for sufficiently small Δt .

The following table lists notation used in the paper.

Symbol	Meaning
S_t	discrete-time process with (integer) time index t
$S(t)$	continuous-time process with time index t
Ω	sample space of a process at fixed time
$\mathcal{S}$	states that give observable “1”
$\mathcal{Y}(t)$	the observed process (indicating $S(t) \in \mathcal{S}$)
P	transition matrix of a Markov chain
W	transition rate matrix of Markov process
$\alpha_t; \alpha(t)$	forward message
$\beta_t; \beta(t)$	backward message
$\chi_t; \chi(t)$	observation message
$\rho_t; \rho(t)$	unnormalized conditional state-occupancy probability
$Z(t)$	normalizing constant for $\rho(t)$
$p_t; p(t)$	conditional state-occupancy probability
Δt	sampling time step

APPENDIX B

INVARIANT NORMALIZATION IN CONTINUOUS TIME

Time-invariant invariant normalization is a fundamental property of the sum-product algorithm [26]. In the continuous time case, we observed that this property can be easily confirmed when the submatrices are diagonalizable, because the left and right eigenvectors corresponding to different eigenvalues are orthogonal. When the submatrix has nontrivial Jordan blocks, the time-dependent term cancels in less obvious ways. Here we show that time-independent normalization holds in general, using only elementary calculus without invoking the machinery of factor-graphs.

Recall that the normalizing constant is $Z(t) = \sum_i \alpha^{(i)}(t) \beta^{(i)}(t)$. Using Equation 2 and 3, we can obtain the

following series of expressions.

$$\frac{dZ(t)}{dt} = \frac{d}{dt} \sum_{i \in S} \alpha^{(i)}(t) \beta^{(i)}(t) \quad (56)$$

$$= \sum_{i \in S} \frac{d\alpha^{(i)}(t)}{dt} \beta^{(i)}(t) + \alpha^{(i)}(t) \frac{d\beta^{(i)}(t)}{dt} \quad (57)$$

$$= \sum_{i \in S} \left(\sum_{k \in S \setminus \{i\}} w_{ki} \alpha^{(k)} - \sum_{l \neq i} \alpha^{(i)} \right) \beta^{(i)} \dots \\ + \alpha^{(i)} \left(- \sum_{k \in S \setminus \{i\}} w_{ik} \beta^{(k)} + \sum_{l \neq i} \beta^{(i)} \right) \quad (58)$$

$$= \sum_{i \in S} \sum_{k \in S \setminus \{i\}} w_{ki} \alpha^{(k)} \beta^{(i)} - w_{ik} \alpha^{(i)} \beta^{(k)} \quad (59)$$

$$= 0 \quad (60)$$

Since the rate of change of $Z(t)$ is zero, the normalizing constant is invariant of time.

REFERENCES

- [1] G. Grisetti, C. Stachniss, and W. Burgard, "Improving grid-based slam with Rao-Blackwellized particle filters by adaptive proposals and selective resampling," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE, 2005, pp. 2432–2437.
- [2] L. E. Baum and J. A. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bulletin of the American Mathematical Society*, vol. 73, no. 3, pp. 360–363, 1967.
- [3] D. F. Anderson, B. Ermentrout, and P. J. Thomas, "Stochastic representations of ion channel kinetics and exact stochastic simulation of neuronal dynamics," *Journal of Computational Neuroscience*, vol. 38, no. 1, pp. 67–82, 2015.
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning research*, vol. 3, pp. 993–1022, 2003.
- [5] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, vol. 117, no. 4, p. 500, 1952.
- [6] E. Skaugen and L. Walløe, "Firing behaviour in a stochastic nerve membrane model based upon the Hodgkin–Huxley equations," *Acta Physiologica Scandinavica*, vol. 107, no. 4, pp. 343–363, 1979.
- [7] B. Hille, "Ionic channels in excitable membranes. current problems and biophysical approaches," *Biophysical Journal*, vol. 22, no. 2, pp. 283–294, 1978.
- [8] D. Colquhoun and A. G. Hawkes, "Relaxation and fluctuations of membrane currents that flow through drug-operated channels," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 199, no. 1135, pp. 231–262, 1977.
- [9] D. Colquhoun and A. Hawkes, "On the stochastic properties of single ion channels," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 211, no. 1183, pp. 205–235, 1981.
- [10] D. Colquhoun and A. G. Hawkes, "On the stochastic properties of bursts of single ion channel openings and of clusters of bursts," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, pp. 1–59, 1982.
- [11] F. Qin, A. Auerbach, and F. Sachs, "Maximum likelihood estimation of aggregated Markov processes," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 264, no. 1380, pp. 375–383, 1997.
- [12] M. D. Fuller, Z.-R. Zhang, G. Cui, and N. A. McCarty, "The block of CFTR by scorpion venom is state-dependent," *Biophysical Journal*, vol. 89, no. 6, pp. 3960–3975, 2005.
- [13] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.
- [14] D. Koller and N. Friedman, *Particle-based Approximate Inference*. MIT Press, 2012.
- [15] Z. Fang, A. Gupta, and M. Khammash, "Stochastic filtering for multiscale stochastic reaction networks based on hybrid approximations," *Journal of Computational Physics*, vol. 467, p. 111441, 2022.
- [16] O. Cappé, C. P. Robert, and T. Rydén, "Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 65, no. 3, pp. 679–700, 2003.
- [17] P. M. Djuric and J.-H. Chun, "Estimation of nonstationary hidden Markov models by MCMC sampling," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, vol. 3. IEEE, 1999, pp. 1737–1740.
- [18] R. Rosales, J. A. Stark, W. J. Fitzgerald, and S. B. Hladky, "Bayesian restoration of ion channel records using hidden Markov models," *Biophysical Journal*, vol. 80, no. 3, pp. 1088–1103, 2001.
- [19] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [20] D. Jin, F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Bayesian cooperative localization using received signal strength with unknown path loss exponent: Message passing approaches," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1120–1135, 2020.
- [21] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a Markov-tree prior," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3439–3448, 2012.
- [22] A. Abrardo, M. Barni, K. Kallas, and B. Tondi, "A message passing approach for decision fusion of hidden-Markov observations in the presence of synchronized attacks," in *Int. Conf. Advances in Multimedia (MMEDIA), Special Track on Models and Algorithms for Spatially and Temporally Correlated Data (STCD), Venice, Italy*, 2017.
- [23] L. Bolliger, *Digital Estimation of Continuous-Time Signals Using Factor Graphs*. Hartung-Gorre Verlag, Konstanz, Series in Signal and Information Processing, 2012, vol. 22.
- [24] L. Bolliger, H.-A. Loeliger, and C. Vogel, "LMMSE estimation and interpolation of continuous-time signals from discrete-time samples using factor graphs," *arXiv preprint arXiv:1301.4793*, 2013.
- [25] L. Bruderer and H.-A. Loeliger, "Estimation of sensor input signals that are neither bandlimited nor sparse," in *2014 Information Theory and Applications Workshop (ITA)*. IEEE, 2014, pp. 1–5.
- [26] G. D. Forney Jr and P. O. Vontobel, "Partition functions of normal factor graphs," *arXiv preprint arXiv:1102.0316*, 2011.
- [27] R. G. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.
- [28] A. W. Eckford and P. J. Thomas, "The channel capacity of Channel-rhodopsin and other intensity-driven signal transduction receptors," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 4, no. 1, pp. 27–38, 2018.
- [29] G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann, and E. Bamberg, "Channelrhodopsin-2, a directly light-gated cation-selective membrane channel," *Proceedings of the National Academy of Sciences*, vol. 100, no. 24, pp. 13940–13945, 2003.
- [30] D. Goetz and C. L. Ren, "Review of cystic fibrosis," *Pediatric Annals*, vol. 48, no. 4, pp. e154–e161, 2019.
- [31] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.
- [32] —, "Stochastic simulation of chemical kinetics," *Annual Review of Physical Chemistry*, vol. 58, pp. 35–55, 2007.
- [33] D. J. Wilkinson, *Stochastic Modelling for Systems Biology*. Chapman and Hall/CRC, 2018.