部分观测马尔可夫过程的显式可解连续时间推断

Daniel Chen ，Alexander G. Strang、Andrew W. Eckford*IEEE*高级会员、Peter J. Thoma s

arXiv:2301.00843v1 [eess.SP] 2 Jan 2023

摘要——**许多自然系统和工程系统都可以建模为离散状态马尔可夫过程。通常，只有部分状态可以直接观测。根据部分观测推断系统处于某个特定隐状态的条件概率，是一个应用广泛的问题。本文提出和积算法的一种连续时间表述。和积算法是一种广为人知的离散时间方法，用于根据一组有限的离散时间观测求出隐状态的条件概率。基于这一新表述，我们可以根据有限时间窗内的转移速率和观测，显式求出系统处于任意状态的条件概率。我们将该算法应用于囊性纤维化跨膜传导调节因子（CFTR）蛋白的一个真实模型，根据有限时间序列的部分观测，精确推断其条件占据概率。**

I. 引言

马尔可夫过程——未来行为仅取决于当前状态的动态过程——可近似描述多种自然系统和工程系统。尽管高通量数据采集和数据处理技术迅速发展，许多值得关注的系统仍包含无法直接观测的重要自由度。在给定现有观测的情况下，推断此类部分观测马尔可夫过程 处于特定隐状态的条件概率，是科学与工程领域普遍存在的问题。相关例子见于机器人学 [1], 、生态学 [2],、神经科学 [3],和算法文本分析 [4]。

本文的研究动机来自生物学示例。可兴奋膜中的离子通道，例如霍奇金和赫胥黎用于描述动作电位产生的定量模型中的钠（Na^+）通道和钾（K^+）通道 [5], [6], [7], ，提供了一个早期示例。基于霍奇金和赫胥黎钾（K^+）通道的离散状态马尔可夫模型包含五种状态，其中只有一种能够传导离子电流；其他状态都是“静默”的，无法通过直接电生理观测加以区分。同样，钠（Na^+）通道有八种状态：七种电导为零，一种电导非零。Colquhoun 和 Hawkes 引入了最大似然方法，用于推断表示烟碱型乙酰胆碱受体 [8], [9], [10], [11], 的部分观测马尔可夫过程的速率常数，但没有探讨如何根据可观测的电导时间序列推断微观状态占据。近年来，囊性纤维化（CF）的分子生物学研究聚焦于囊性纤维化跨膜电导调节因子（CFTR），它可建模为一个包含两种导通状态和五种非导通状态的七状态系统[12] （详见下文第五节）。除这些生物学示例外，从不完整观测中推断或估计隐状态的问题也在信号处理领域得到了广泛研究 [13]。

文献中已有多种方法用于近似部分观测马尔可夫过程的隐状态行为。采样是近似推断的一种常见技术 [14]。例如，Fang等人近期的研究展示了一种高效算法，利用粒子滤波器模拟具有多个分离时间尺度的随机反应网络 [15]。一般而言，马尔可夫链蒙特卡罗是一种广泛使用的采样技术，可用于推断隐状态 [16], [17], ，也已应用于离子通道 [18]。对于以离散时间运行的部分观测贝叶斯网络，基于因子图的消息传递算法提供了一种高效且精确的推断方法[19]。因子图形式具有很高的灵活性。基于消息传递概念的算法已拓展应用于定位 [20], 、压缩感知 [21], 和决策融合 [22]。因子图并不局限于含有离散数量变量的模型。例如，针对具有离散时间观测的连续时间模型，线性高斯模型中的高斯消息传递（例如卡尔曼滤波和平滑）已经得到发展 [23], [24], [25]。然而，据我们所知，文献中尚未涉及连续时间有限状态隐马尔可夫模型的状态重构问题。

在这项工作中，我们扩展了消息传递算法，以便根据离散采样的时间序列，对连续时间系统的状态占据概率进行解析插值。也就是说，我们展示了如何根据有限时间窗内的一组部分观测，推断连续时间离散状态齐次马尔可夫过程的潜在状态随时间变化的条件概率。我们推导出和积算法在连续时间下的一种等价形式，可用于求得状态占据概率的显式解析解。

本研究部分获得美国国立卫生研究院脑计划资助（项目编号 R01 NS118606）以及美国国家科学基金会资助（项目编号 DMS-2052109），资助对象为 P. J. Thomas；此外还获得欧柏林学院图书馆的研究支持，以及加拿大自然科学与工程研究委员会向 A. W. Eckford 提供的探索资助。D. Chen 和 P. J. Thomas 任职于凯斯西储大学数学、应用数学与统计学系，电气、控制与系统工程系，计算机与数据科学系及生物系，地址为美国俄亥俄州克利夫兰市 44106（电子邮箱：txc461/pjthomas@case.edu）。A. G. Strang 任职于芝加哥大学统计学系，地址为美国伊利诺伊州芝加哥市 60637（电子邮箱：alexstrang@uchicago.edu）。A. W. Eckford 任职于约克大学电气工程与计算机科学系，地址为加拿大安大略省多伦多市 M3J 1P3（电子邮箱：aeck-ford@yorku.ca）。

显式解可降低计算成本，而且与基于采样的近似方法不同，它不会牺牲准确性。此外，连续时间形式体系使解析解得以优雅地简化。对于某些系统——例如图 2 所示的三状态系统——条件概率满足一个二阶非齐次线性常微分方程。最后，我们使用模拟数据，通过 CFTR 的七状态模型展示了该算法的实际功能。

本文结构安排如下。第二节回顾具有二元观测的马尔可夫过程的消息传递算法。第三节展示本文的主要成果：和积算法的连续时间表述。我们介绍如何使用连续时间下的消息传递方法推导条件概率，以及相应的和积算法。第四节进一步讨论连续时间表述的意义，并通过小型系统的示例展示如何求得解析解。第五节则展示该算法在更大型、更贴近现实的系统中的价值。

II. 和积算法理论

和积算法通常可用于对能够表示为因子图 [19] 的概率模型进行推断。和积算法有许多变体，分别适用于不同的任务。本文将重点介绍用于隐马尔可夫模型推断的前向/后向算法。

我们考虑有限状态空间 Ω 上的连续时间、离散状态齐次马尔可夫过程。给定离散且均匀间隔的采样间隔，该连续时间过程会诱导出一个离散时间马尔可夫过程，由某个随时间保持不变的列随机转移矩阵 P 指定。设 $\mathcal{S} \subset \Omega$ 为状态的一个子集，并设 $S_t \in \Omega$ 为系统在时间 t 时所处的状态。我们假设观测者只能看到 S_t 是否属于 $\mathcal{S}$。相应地，令 $\mathcal{Y}_t = m(S_t)$ 表示观测量，其中 $m(s) : \Omega \to \{0, 1\}$ 是集合 $\mathcal{S}$ 的指示函数。该算法的目标是根据二元观测推断系统处于特定状态 $i \in \Omega$ 的条件概率。

该算法涉及三个向量值量：前向消息 $\boldsymbol{\alpha}_t$、后向消息 $\boldsymbol{\beta}_t$ 和观测消息 $\boldsymbol{\chi}_t$。可以将前向消息理解为从时间 0 到时间 t 到达某一状态的概率；将后向消息理解为在测量 T 结束时处于给定目标状态或给定目标状态集合的条件下，在时间 t 占据某一状态的似然。另一方面，观测消息 $\boldsymbol{\chi}_t$ 是给定观测时可能状态的指示函数。例如，如果在时间 t 对整个系统的观测结果为 $\mathcal{Y}_t = 1$，那么 $\boldsymbol{\chi}_t$ 将是一个向量：观测集合 $\mathcal{S}$ 中的状态对应的值为 1，其余状态对应的值为 0。我们用上标表示向量索引，即 也就是 $\boldsymbol{v}^{(i)}$ 表示向量 $\boldsymbol{v}$ 的第 i 个元素，并在算法1中给出和积算法的伪代码。请注意，算法1第10行的归一化常数 $Z = \sum_k \alpha_t^{(k)} \beta_t^{(k)}$ 不随时间变化 [26]。

算法 1 前向/后向算法

输入： 转移矩阵 $P \in \mathbb{R}^{n \times n}$，观测消息 $\boldsymbol{\chi}_t$，对应 $t \in \{1, \ldots, T\}$.
输出： 推断概率 $\boldsymbol{p}_t$，对应 $t \in \{1, \ldots, T\}$。

1: 初始化前向消息 $\boldsymbol{\alpha}_0 = \boldsymbol{\pi}$，即平稳 P 的分布
2: 初始化后向消息 $\beta_T^{(i)} = 1$，对于所有 i
3: **对于**t，从 2 开始 $, \ldots, T$ **do**
4: $\quad \boldsymbol{\alpha}_t = \text{diag}(\boldsymbol{\chi}_t) P \boldsymbol{\alpha}_{t-1}$
5: **结束循环**
6: **对于** s，从 $T-1, \ldots, 1$ **do**
7: $\quad \boldsymbol{\beta}_s = P^{\mathsf{T}} \text{diag}(\boldsymbol{\chi}_{s+1}) \boldsymbol{\beta}_{s+1}$
8: **结束循环**
9: **对**t 从 1 $, \ldots, T$ **do**
10: $\quad \boldsymbol{p}_t = \frac{\boldsymbol{\alpha}_t \odot \boldsymbol{\beta}_t}{\sum_k \alpha_t^{(k)} \beta_t^{(k)}}$
11: **结束循环**

图1展示了和积算法在离散采样观测时间序列上的应用。考虑一个具有对称转移率的三状态链，如图1左上角所示。如果系统处于状态1或2，则观测值为“0”；否则观测值为“1”。使用和积算法，我们可以根据离散时间观测求出系统处于各个状态的概率。如底行所示，随着采样时间步长减小，在观测值固定（“0”或“1”）的每个区间内，条件概率似乎都收敛到一条平滑曲线。直观而言，部分观测过程的条件概率应存在一种连续的、或许是分段可微的表示。我们将在下文形式化这一直觉。

III. 连续时间消息传递

本节中，我们给出主要结果，即连续时间消息传递算法的推导。在这一新形式体系中，消息以线性微分方程的形式，在给定可观测系统的可能状态上进行传递。为了保证连续时间和积算法的存在，我们假设满足以下条件。

假设：
A1 连续时间过程 $\{S(t);\ t \in [0, T]\}$ 取值于有限状态空间 $\Omega = \{1, 2, \ldots, N\}$。A2 $S(t)$具有马尔可夫性质，其等待时间服从指数分布，参数由速率矩阵W确定，其中 w_{ji} 指定从状态 i 到状态 j的转移速率。请注意， W在区间 $[0, T]$内为常数。A3 存在一个特殊子集 $\mathcal{S} \subset \Omega$，使得可观测过程 $\mathcal{Y}(t)$满足

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

更多细节见附录 A。

在这些假设下，我们通过执行以下步骤（将在下文定理1的证明中严格论证）得到和积算法的连续时间版本。将

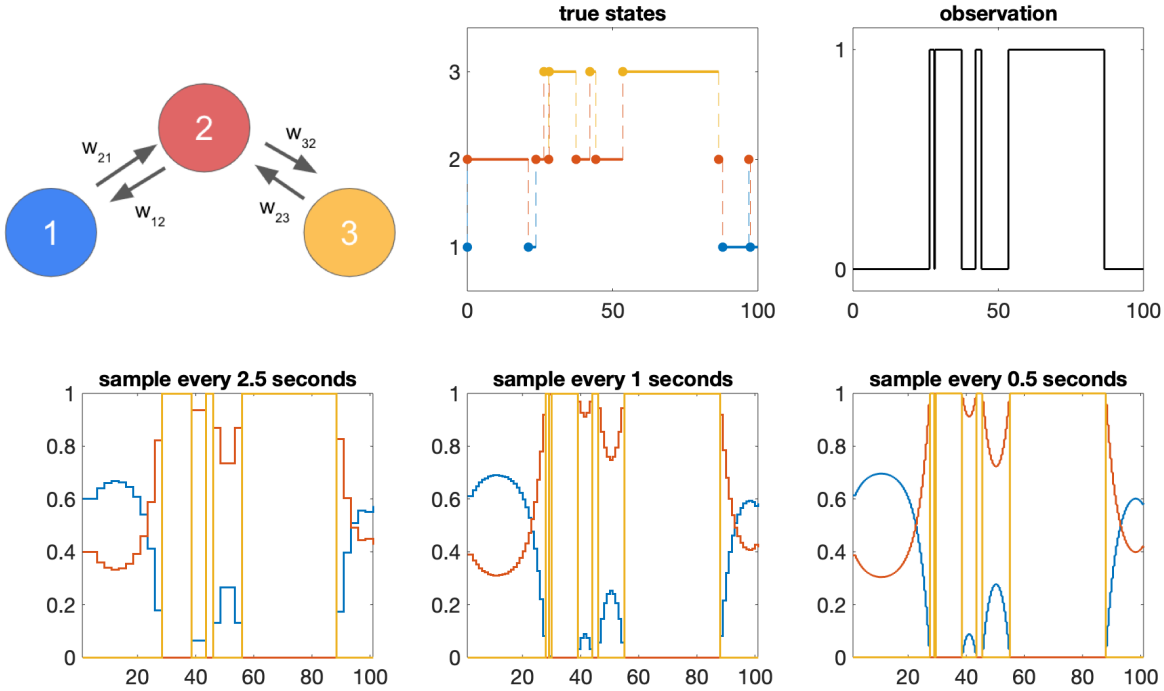

图1：展示三状态模型中条件状态占用概率如何收敛到可微函数。（**左上**）状态图。（**上中**）真实模拟状态。（**右上**）根据真实状态得到的二元观测。（**底行**）使用和积算法推断隐状态，时间步长分别为 2.5 秒（**左下**）、1.0 秒（**中下**）和 0.5 秒（**右下**）。

离散时间算法中的矩阵乘法逐元素展开。聚焦于观测保持不变的一个驻留期。在该时间区间内，令时间步长趋于零，从而推导出条件概率的连续时间动力学。通过在每个驻留期转换处施加适当的边界条件，将解扩展到整个时间区间。主要结果陈述如下定理。

定理1。假设过程 $S(t)$ 及其相关过程 $\mathcal{Y}(t)$ 满足上述假设*A1*、*A2*和*A3*。那么，给定过程 $\mathcal{Y}(t)$的一次实现，条件概率 $\boldsymbol{p}(t) = \Pr[S(t)|\mathcal{Y}(t)]$ 存在，且分段光滑（C^∞），并且在 $\mathcal{Y}(t)$ 为常数的所有区间上均为 C^∞ 。特别地， $\boldsymbol{p}(t)$ 具有如下形式

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

其中， $\odot$ 表示逐元素乘积，量$\boldsymbol{\alpha}(t)$ 和 $\boldsymbol{\beta}(t)$ 是时间的函数，遵循线性常微分方程

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

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

证明。 不失一般性，我们聚焦于$\mathcal{Y}(0) = \mathcal{Y}(T) = 0$且 $\mathcal{Y}(t) = 1$的情形，其中 $0 < t < T$。我们用$q(t)$表示在区间 $[0,T]$上随连续时间演化的量 q，并用 q_t 表示在离散时刻采样得到的同一过程。

将时间区间 $[0,T]$ 按步长 $\Delta t = T/n$ 离散化，其中 $n \gg 1$ 为某个整数。在每个时间步长，采样 $S(t)$。然后，可以使用和积算法求解 $\boldsymbol{p}_t$。展开矩阵乘法可得离散时间下前向和后向消息的以下方程组：

$$\alpha^{(i)}_{t+\Delta t} = \sum_k \Pr[s_{t+\Delta t} = i|s_t = k]\alpha^{(k)}_t \chi^{(i)}_t, \tag{4}$$

$$\beta^{(i)}_t = \sum_k \Pr[s_{t+\Delta t} = k|s_t = i]\beta^{(k)}_{t+\Delta t}\chi^{(k)}_{t+\Delta t}. \tag{5}$$

我们忽略不属于 $\mathcal{S}$ 的状态，因为我们只关注以观测为条件的概率。接着，对于任意状态 $i \in \mathcal{S}$，我们论证，相应的前向消息 $\boldsymbol{\alpha}^{(i)}(t)$ 和连续时间下的后向消息 $\boldsymbol{\beta}^{(i)}(t)$，在取 $\Delta t \to 0$ 的极限后，可以表示为微分方程组的解。为简化符号，对于

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

$$\frac{d\alpha^{(i)}(t)}{dt}$$
$$= \lim_{\Delta t \to 0} \frac{\alpha^{(i)}(t+\Delta t) - \alpha^{(i)}(t)}{\Delta t} \tag{6}$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\sum_k \mathcal{P}_{t+\Delta t,t}^{(i,k)} \alpha^{(k)}(t) \chi^{(i)}(t) - \alpha^{(i)}(t) \right] \tag{7}$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Bigg[\sum_{k \in \mathcal{S} \setminus \{i\}} (w_{ki}\Delta t + o(\Delta t)) \alpha^{(k)}(t) \dots - \sum_{l \neq i} (w_{il}\Delta t + o(\Delta t)) \alpha^{(i)}(t) \Bigg] \tag{8}$$
$$= \sum_{k \in \mathcal{S} \setminus \{i\}} w_{ki} \alpha^{(k)}(t) - \sum_{l \neq i} w_{il} \alpha^{(i)}(t) \tag{9}$$

$$\frac{d\beta^{(i)}(t)}{dt}$$
$$= \lim_{\Delta t \to 0} \frac{\beta^{(i)}(t+\Delta t) - \beta^{(i)}(t)}{\Delta t} \tag{10}$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Bigg[\beta^{(i)}(t+\Delta t) \dots - \sum_k \mathcal{P}_{t+\Delta t,t}^{(k,i)} \beta^{(k)}(t+\Delta t) \chi^{(k)}(t+\Delta t) \Bigg] \tag{11}$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Bigg[- \sum_{k \in \mathcal{S} \setminus \{i\}} (w_{ik}\Delta t + o(\Delta t)) \beta^{(k)}(t+\Delta t) + \sum_{l \neq i} (w_{il}\Delta t + o(t)) \beta^{(i)}(t+\Delta t) \Bigg] \tag{12}$$
$$= - \sum_{k \in \mathcal{S} \setminus \{i\}} w_{ik} \beta^{(k)}(t) + \sum_{l \neq i} w_{il} \beta^{(i)}(t) \tag{13}$$

其中 $\lim_{\Delta t \to 0} o(\Delta t)/\Delta t = 0$。关于马尔可夫跳跃过程的转移概率与转移率之间的关系，读者可参阅附录 A，或查阅现有文献，例如 [27] 。

条件概率可通过求解微分方程来求得：对每个 $t \in [0, T]$，计算 $\boldsymbol{\alpha}(t)$ 和 $\boldsymbol{\beta}(t)$ 的逐分量乘积，然后进行归一化，从而得到有效的概率分布。

微分方程形式仅适用于观测 $\mathcal{Y}(t)$保持不变的时间区间。当观测量 $\mathcal{Y}(t)$发生变化时（即系统从 $\mathcal{S}$ 中的某个状态转移到不在 $\mathcal{S}$中的某个状态），相对于时间的概率可能不可微；在某些情况下，它甚至不连续。因此，我们必须指定边界条件，以连接从一个驻留期到下一个驻留期的概率。观测量可能因系统进入 S 或离开 S而发生变化。假设在时间 $(t_* - \Delta t, t_*]$内发生了一次转移，使得对于 $t < t_* - \Delta t$， $S(t) \in \mathcal{S}$，且对于 $t \geq t_*$， $S(t) \notin \mathcal{S}$ 。将该事件称为 E。于是，我们得到前向消息的如下转移规则。

$$\Pr[S(t_*) = j | E]$$
$$= \frac{\sum_{i \in \mathcal{S}} \mathcal{P}_{t_*,t_*-\Delta t}^{j,i} \Pr[S(t_* - \Delta t) = i]}{\sum_{k \notin \mathcal{S}} \sum_{i \in \mathcal{S}} \mathcal{P}_{t_*,t_*-\Delta t}^{k,i} \Pr[S(t_* - \Delta t) = i]} \tag{14}$$
$$= \frac{\sum_{i \in \mathcal{S}} (w_{ji}\Delta t + o(\Delta t)) \Pr[S(t_* - \Delta t) = i]}{\sum_{k \notin \mathcal{S}} \sum_{i \in \mathcal{S}} (w_{ki}\Delta t + o(\Delta t)) \Pr[S(t_* - \Delta t) = i]} \tag{15}$$
$$\to \frac{\sum_{i \in \mathcal{S}} w_{ji} \Pr[S(t_*^-) = i]}{\sum_{k \notin \mathcal{S}} \sum_{i \in \mathcal{S}} w_{ki} \Pr[S(t_*^-) = i]} \tag{16}$$

如 $\Delta t \to 0$。其中，$\Pr[S(t_*^-) = i]$ 表示转移发生前一刻处于状态 i 的概率，可通过求解上述微分方程来求得。

对于后向消息，我们以类似方式处理状态转移处的边界条件。定义如上所述的 E，并令 s^* 为某个特定的目标状态。则:

$$\Pr[S(T) = s^* | S(t_* - \Delta t) = j, E]$$
$$= \sum_{i \notin \mathcal{S}} \Pr[S(T) = s^*, S(t_*) = i | S(t_* - \Delta t) = j, E]$$
$$= \frac{\sum_{i \notin \mathcal{S}} \Pr[S(T) = s^* | S(t_*) = i] \mathcal{P}_{t_*,t_*-\Delta t}^{i,j}}{\sum_{k \in \mathcal{S}} \sum_{i \notin \mathcal{S}} \mathcal{P}_{T,t_*}^{s^*,i} \mathcal{P}_{t_*,t_*-\Delta t}^{i,k}} \tag{17}$$
$$= \frac{\sum_{i \notin \mathcal{S}} \Pr[S(T) = s^* | S(t_*) = i] (w_{ij}\Delta t + o(\Delta t))}{\sum_{k \in \mathcal{S}} \sum_{i \notin \mathcal{S}} \mathcal{P}_{T,t_*}^{s^*,i} (w_{ik}\Delta t + o(\Delta t))} \tag{18}$$
$$= \frac{\sum_{i \notin \mathcal{S}} \Pr[S(T) = s^* | S(t_*) = i] (w_{ij} + \frac{o(\Delta t)}{\Delta t})}{\sum_{k \in \mathcal{S}} \sum_{i \notin \mathcal{S}} \mathcal{P}_{T,t_*}^{s^*,i} (w_{ik} + \frac{o(\Delta t)}{\Delta t})} \tag{19}$$
$$\to \frac{\sum_{i \notin \mathcal{S}} w_{ij} \Pr[S(T) = s^* | S(t_*) = i]}{\sum_{k \in \mathcal{S}} \sum_{i \notin \mathcal{S}} w_{ik} \Pr[S(T) = s^* | S(t_*) = i]} \tag{20}$$

如 $\Delta t \to 0$。其中，$\Pr[S(T) = s^* | S(t_*) = i]$ 表示与时间t_*处的后向消息相关的命中概率，可通过求解后向消息微分方程来求得。这些边界条件与微分方程 (2)-(3) 一起，给出了任意有限长度观测下条件概率的连续时间演化。

请注意，通过令$\mathcal{Y}(0) = \mathcal{Y}(T) = 1$ 且 $\mathcal{Y}(t) = 0$ （对 $0 < t < T$ 而言），并将$\mathcal{S} \leftarrow$ 视为$\Omega \setminus \mathcal{S}$，方程(2)–(3)可扩展到这种情形。因此，给定时间序列观测 $\mathcal{Y}(t)$，其中观测（即 $\mathcal{Y}(t)$的值）在时刻 $0 < t_1 < t_2 < .. t_m$发生变化，我们可以利用方程(2)和(3)，求出观测保持不变的任意区间内的解析解（$(t_i, t_{i+1}]$ ）。随后，按照(16)和(20)中的规定，利用结果计算下一个区间的初始条件——即前向消息的$(t_{i+1}, t_{i+2}]$)，以及后向消息的$(t_{i-1}, t_i]$)。因此，定理1的陈述成立。 □

虽然可以得到条件概率的一般表达式，但它对大多数系统并没有太大用处。不过，也有一些特殊情形能够得出简洁优雅的解；我们将在第四节介绍几个例子。

连续时间和积算法可直接由上述推导得出，并在算法 2 中概述。

离散算法通过截断转移矩阵的矩阵乘法传递信息；连续时间算法则通过使用截断速率矩阵求解线性微分方程组来实现相同功能。由于最终的条件概率仅为分段可微，因此每当发生进出可观测集合 $\mathcal{S}$ 的转移时，都必须应用边界条件。

算法 2 连续时间前向/后向算法 输入：速率转移矩阵 $W \in \mathbb{R}^{n\times n}$，观测过程 $\mathcal{Y}(t)$。输出：$t \in [0,T]$ 的推断概率 $\boldsymbol{p}(t)$。1：令 $[t_1, t_2, \ldots, t_m]$ 为发生转移的时间列表 2：$\tau \leftarrow 0$ 3：$\boldsymbol{\alpha}^* \leftarrow \boldsymbol{\pi}$，平稳分布 4：对于 j 从 1，...，$m$，执行 5：$\boldsymbol{\alpha}_j \leftarrow$ 从 τ 到 t_j 、初始条件为 $\boldsymbol{\alpha}^*$ 的前向消息微分方程（方程 2）的解 6：$\tau \leftarrow t_j$ 7：$\boldsymbol{\alpha}^* \leftarrow$ 按照方程 16 指定的分布 8：结束循环 9：$\tau \leftarrow T$ 10：$\boldsymbol{\beta}^* \leftarrow$ 均匀分布 11：对于 j 从 m，...，1，执行 12：$\boldsymbol{\beta}_j \leftarrow$ 从 τ 到 t_j 、初始条件为 β^* 的后向消息微分方程（方程 3 ）的解 13：$\tau \leftarrow t_j$ 14：$\boldsymbol{\beta}^* \leftarrow$ 按照方程 20 指定的分布 15：结束循环 16：$\boldsymbol{\alpha}(t)\ \boldsymbol{\beta}(t) \leftarrow$ 拼接 $\boldsymbol{\alpha}_j$ 和 $\boldsymbol{\beta}_j$ 17：计算 $\boldsymbol{\rho}(t) = \boldsymbol{\alpha}(t) \odot \boldsymbol{\beta}(t)$，即

$$\boldsymbol{\alpha}\ t \qquad \boldsymbol{\beta}\ t \qquad t\ (\ ) \text{ 与 } \boldsymbol{\rho}\ t$$

18 $(\ )$ 关于 t 的逐点逐分量乘积：计算条件概率 $\boldsymbol{p}(\ )\ (\ ) = \boldsymbol{\rho}\ t\ (\ ) / \sum_k \rho^{(k)}(t)$

从实践角度来看，通过微分方程精确求解条件概率的能力，大幅降低了前向/后向算法的计算成本。传统的离散时间算法会在每个时间步长通过矩阵运算传播前向和后向消息。对于较长的时间序列和/或高维系统，这在计算上难以承受。借助我们的连续时间形式体系，我们对微分方程进行解析求解；这项运算与时间序列的长度无关，因此可以求得任意时间点的前向或后向消息。这一区别实际上将渐近尺度从 $\mathcal{O}(\Delta t^{-1})$ 降至 $\mathcal{O}(1)$，后者的尺度仅取决于可观测转移的数量。在确定适当边界条件需要反复求解前向和后向消息的迭代过程时，我们的连续时间方法应比传统离散时间方法高效得多。我们将在第五节进一步讨论连续时间消息传递算法的性能

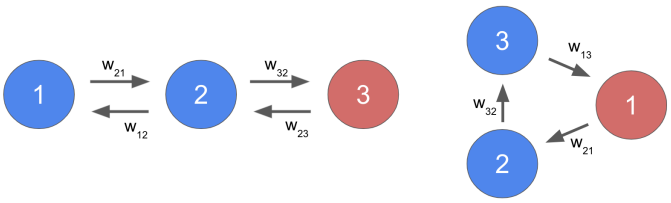

图2：连续时间消息传递算法在解析上可简化的两个系统的状态图。（左）具有对称速率的三态链$w_{12} = w_{21}$。（右）不可逆三态环。红色标记的状态返回 $\mathcal{Y}(t) = 1$，蓝色标记的状态返回 $\mathcal{Y}(t) = 0$。

在第五节进一步介绍传递算法。

IV. 解析解

上一节中的定理1表明，对前向和后向消息的逐分量乘积进行归一化后，始终可以解析地得到条件概率，即 $\boldsymbol{p}(t) = \boldsymbol{\rho}(t)/Z$ 其中 $Z = \sum_k \rho^{(k)}(t)$。与离散时间情形一样，在连续时间情形下，归一化项 Z 也与时间无关。因此，在某些情况下，条件概率可以用一种特别简洁的形式表示，即作为一个线性非齐次二阶微分方程的解。本节首先讨论两个示例，然后探讨向高维情形的扩展。

A. 对称三状态链

考虑图 2 左侧面板所示的三态链，其中状态 1 和 2 是隐藏的。假设隐藏块内的单位个体转移速率是对称的，即 $w_{12} = w_{21} > 0$，并假设 $w_{13} = w_{31} = 0$。速率 $w_{23} > 0$ 和 $w_{32} > 0$ 可以是任意值。这些假设得到以下速率矩阵：

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

在此情况下，令 $\mathcal{S} = \{3\}$ 为状态 3 的单元素集合。当 $\mathcal{Y}(t) = 1$ 时，推断问题很简单，因为系统以概率 1 处于状态 3。因此，我们重点关注 $\mathcal{Y}(t) = 0$ 的区间。

首先，考虑由以下微分方程组给出的前向消息

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

注意，定义该方程组的矩阵对应于 W 的左上块。为简化记号，令$w_{21} = w_{12} = a$ 和 $w_{32} = b$。那么，该子矩阵可化为以下形式：

$$\frac{d\boldsymbol{\alpha}}{dt} = \begin{pmatrix} -a & a \\ a & -a-b \end{pmatrix} \boldsymbol{\alpha}\,. \tag{23}$$

这是一个可以精确求解的线性微分方程组。该子矩阵是实对称矩阵，因此可对角化。所以，解的形式为：

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

其中，λ_i 是速率子矩阵的一个特征值，而 $\boldsymbol{v}_i$ 是对应的特征向量。特征值和特征向量如下：

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

$$\boldsymbol{v}_{1/2} = \begin{pmatrix} \gamma \pm \sqrt{1+\gamma^2} \\ 1 \end{pmatrix}, \tag{26}$$

其中，$\gamma = b/2a$。常数 A 和 B 通过下面给出的初始条件确定。在上一节中，初始前向消息被设为平衡分布。然而，链结构使得观测从 1 变为 0 时所处的状态毫无歧义。因此，初始前向消息是在状态 2 上的狄拉克分布：

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

经过一段时间 T后，系统再次进入可见状态，即再次进入状态 3。根据同样的推理，我们也可以为后向消息得到明确的边界条件：

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

此外，由于速率对称性，后向消息遵循与前向消息相同的方程，但沿时间反向演化。因此，$\boldsymbol{\beta}(t) = \boldsymbol{\alpha}(T-t)$，

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

由于对称性，λ_i、$\boldsymbol{v}_i$以及常数 A 和 B 都保持不变。

条件概率正比于 the 逐分量乘积 $\boldsymbol{\rho} = \boldsymbol{\alpha} \odot \boldsymbol{\beta}$，由此得到：

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

其中

$$\mathcal{A} = AB(\boldsymbol{v}_1 \odot \boldsymbol{v}_2) = \begin{pmatrix} -AB & AB \end{pmatrix}^\intercal, \tag{31}$$

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

为了求出条件概率，我们必须进行归一化，使分量和为 1。在这种情况下，由于方程 22 中的子矩阵是对称的，该子矩阵的特征向量彼此正交，这意味着 $\boldsymbol{v}_1 \odot \boldsymbol{v}_2$ 的分量和，或 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 的内积，为 0。因此，$\mathcal{A}$ 的分量和为零，所以归一化常数等于 $\mathcal{B}$ 的分量和。

用 $\boldsymbol{p}(t)$ 表示条件概率。用 $Z = \sum_k \mathcal{B}^{(k)}$ 表示归一化常数。因此，描述条件概率的最终方程可以改写为以下形式：

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

$\boldsymbol{p}(t)$ 的分量也可以表示为二阶常微分方程的解，参见 (51)-(52)。

B. 不可逆三态环

光控通道视紫红质-2（ChR2）受体可用一个三态链建模，其中每个顶点的出度均为 1，并构成一个环，如图 2 右图所示 [28],[29]。设状态 1 为开放，状态 2 和状态 3 为关闭。也就是说，$\mathcal{S} = \{1\}$。通道的开放/关闭状态通过电压记录进行观测：高电导表示通道处于开放状态，低电导则表示通道处于关闭状态。速率矩阵可写为：

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

其中，w_{ji} 是从状态 i 到状态 j的转移速率。在通道关闭的条件下，消息传递算法将右下角的 2×2 块作为转移矩阵。

当从 $\mathcal{S}$ 转移到不属于 $\mathcal{S}$的状态时，系统必须先进入状态 2，再经由状态 3 离开。因此，前向消息和后向消息的边界条件为

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

其中，第一个分量对应状态 2，第二个分量对应状态 3。假设 $w_{32} \neq w_{13}$，并令 $\gamma = \frac{w_{32}}{w_{13}-w_{32}}$。那么，满足初始条件的消息传递微分方程的解为：

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

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

要得到条件概率，首先计算时间 t处前向消息和后向消息的分量逐元素乘积。

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

与前一个示例一样，归一化常数不随时间变化。在这种情况下，$Z = e^{-w_{32}T} - e^{-w_{13}T}$。因此，条件概率的时间演化可写为

$$\boldsymbol{p}(t) = \boldsymbol{\rho}(t)/Z. \tag{39}$$

接下来，我们考虑子矩阵不可对角化的情形。令 $w = w_{32} = w_{13}$，使得相应的子矩阵：

$$U = \begin{pmatrix} -w & 0 \\ w & -w \end{pmatrix} \tag{40}$$

具有如下约当标准形：

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

使用与方程（35）相同的初始条件求解微分方程组，可得到以下前向消息。

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

通过类似的过程求解后向消息，可得

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

进行逐分量乘积后，我们观察到条件概率的演化与时间无关：

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

由于进入和离开隐状态的时间均已固定，我们可以看到，概率以相同的速率在线性地从一个状态流向另一个状态。

时间不变归一化是和积算法 [26]的一项基本性质。在附录 B 中，我们针对具有时间齐次转移率的连续时间系统，对这一性质进行了基础说明，正如上述两个案例所示。

C. 广义化

我们将证明，在某些情况下，条件概率满足一个二阶非齐次线性常微分方程。

推论 1. 对于三状态系统，若截断子矩阵 U 的特征值互不相同，则条件概率可表示为一个常系数非齐次二阶线性微分方程的解。

证明. 设 $(\lambda_i, \boldsymbol{v}_i)$ 是 U的特征对，且 $(\lambda_i, \boldsymbol{w}_i)$ 是 U 的特征对。那么，前向和后向消息可表示为:

$$\boldsymbol{\alpha}(t) = Ae^{\lambda_1 t}\boldsymbol{v}_1 + Be^{\lambda_2 t}\boldsymbol{v}_2, \tag{45}$$
$$\boldsymbol{\beta}(t) = Ce^{\lambda_1 (T-t)}\boldsymbol{w}_1 + De^{\lambda_2 (T-t)}\boldsymbol{w}_2. \tag{46}$$

取逐分量乘积可得:

$$\boldsymbol{\rho}(t) = \mathcal{A}\left[e^{\lambda_1 t + \lambda_2 (T-t)}\right] + \mathcal{B}\left[e^{\lambda_1 (T-t) + \lambda_2 t}\right] + \mathcal{C}, \tag{47}$$
$$= \mathcal{A}\left[e^{(\lambda_1-\lambda_2)t}e^{\lambda_2 T}\right] + \mathcal{B}\left[e^{(\lambda_2-\lambda_1)t}e^{\lambda_1 T}\right] + \mathcal{C}, \tag{48}$$

其中

$$\mathcal{A} = AD(\boldsymbol{v}_1 \odot \boldsymbol{w}_2), \quad \mathcal{B} = BC(\boldsymbol{v}_2 \odot \boldsymbol{w}_1), \tag{49}$$
$$\mathcal{C} = ACe^{\lambda_1 T}(\boldsymbol{v}_1 \odot \boldsymbol{w}_1) + BDe^{\lambda_2 T}(\boldsymbol{v}_2 \odot \boldsymbol{w}_2). \tag{50}$$

由于对应不同特征值的左、右特征向量相互正交，归一化时所有随时间变化的项都会抵消。因此，归一化常数 $Z = \sum_k \mathcal{C}^{(k)}$。对 $\boldsymbol{p}(t) = \boldsymbol{\rho}(t)/Z$ 关于时间求两次导数后，可得描述条件概率时间演化的二阶线性微分方程:

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

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

这个二阶方程附带两个边界条件，分别由固定 $\boldsymbol{\alpha}(0)$ 和 $\boldsymbol{\beta}(T)$ 给出。 □

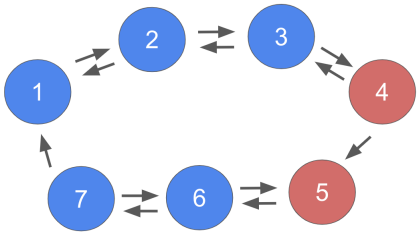

图 3：CFTR 的状态图。当蛋白质进入红色状态之一时，离子通道将开放并传导电流。另一方面，当 CFTR 处于蓝色状态之一时，离子通道关闭且不传导电流。对于该系统 $\mathcal{S} = \{4, 5\}$。

推论 2. 对于任意维度的系统，若截断子矩阵 U 恰好有两个不同的特征值且可对角化，则条件概率可以写成与方程 *52* 形式相同的二阶线性微分方程。

证明。 设 λ_1 和 λ_2 是两个不同的特征值。设 $\boldsymbol{v}_i$ 为所有对应于特征值 λ_i的右特征向量之和， $\boldsymbol{w}_i$ 为所有对应于特征值 λ_i的左特征向量之和。那么，根据可对角化性，这个推论的证明与推论 1 的证明相同。 □

这一结果在特殊条件下将推论 1 推广到更高维的系统。这两个推论符合直觉：我们关注的是在给定观测状态的进入和离开时间的条件下，该状态的占据概率，因此概率流满足一个二阶微分方程，而这也要求指定两个边界条件。然而，我们注意到，如果子矩阵有三个或更多不同的特征值，就不认为类似的结果仍然成立。此外，一般矩阵通常具有与其维数相同数量的不同特征值，因此我们不认为除速率子矩阵高度规则的特殊情形外，会出现这种情况。此外，通过数值求解器检查给定的速率子矩阵是否满足这些条件可能很困难，因为重复特征值相等这一事实可能会被浮点运算掩盖。因此，尽管二阶微分方程的解释在直觉上颇具吸引力，但它很可能并非我们应当预期其对任意速率矩阵都成立的一般现象。

V. 生物学示例：CFTR

在这里，我们展示了我们的算法适用于更高维的系统，即囊性纤维化跨膜电导调节因子（CFTR）蛋白的七状态模型。囊性纤维化是一种常见且危及生命的遗传性疾病。CFTR 是一种重要的蛋白质，负责调节离子通道的开放与关闭。CFTR 功能丧失会导致胰腺功能不全，并因黏液过多而引发气道感染，进而可能造成多种并发症，例如先天免疫受损和呼吸衰竭 [30]。

从数学上讲，可以用一个包含7个状态的隐马尔可夫过程来描述CFTR的行为。图3展示了

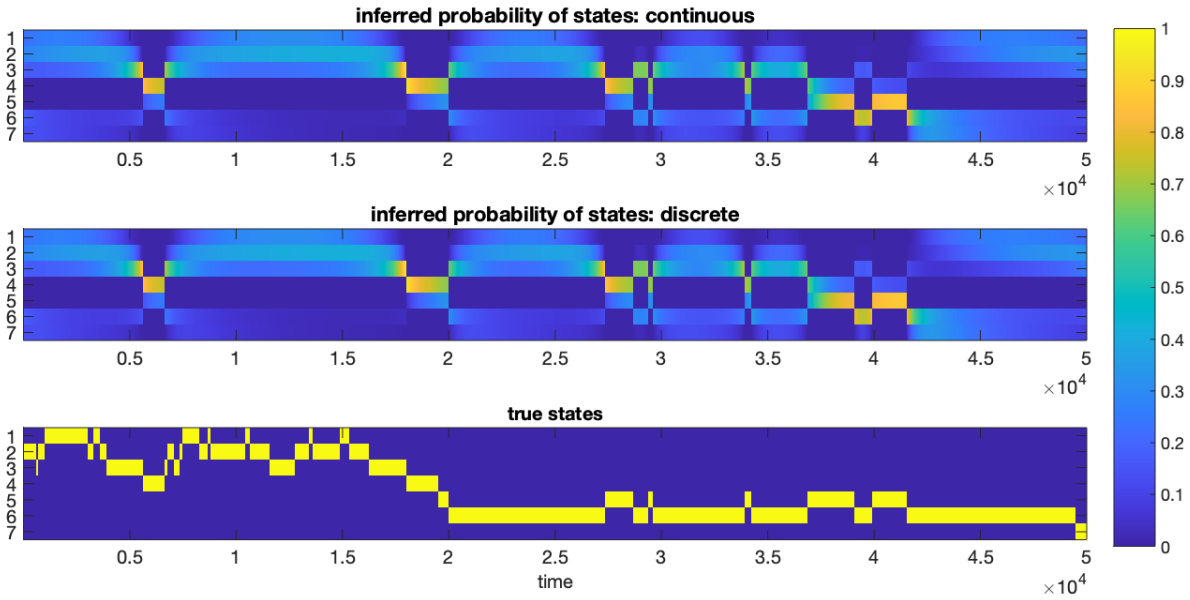

图 4：连续时间算法（上）和离散时间算法（中）对由 Gillespie 算法（下）模拟得到的真实状态占据情况的推断结果。采样时间步长为 $\Delta t = 10^{-4}$。

状态图 [12]。当CFTR处于状态4或状态5（标为红色）时，离子通道开放（传导离子电流）。当处于非导通状态（标为蓝色）时，离子通道关闭。跨膜电导记录显示CFTR处于导通状态还是非导通状态，但不能直接指出其处于哪一个可能的状态。按照图3所示的编号，速率矩阵定义如下：

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

我们使用 Gillespie 精确随机模拟算法 [31], [32], [33] 生成离子通道状态和记录的样本轨迹。该模拟生成离散状态和连续时间轨迹。我们引入有限的采样时间步长，沿时间轴对模拟数据进行离散化，以与通过实验记录获得的数据保持一致。然后，我们将生成的数据放入离散化的时间区间中，每个区间的大小（可以将其视为采样时间步长）是一个可调整的参数。我们通过 Matlab 中的函数句柄，精确求解了给出前向和后向消息的微分方程。图 4 比较了经典离散时间算法、我们的连续时间算法以及真实状态生成的轨迹，并显示各条曲线之间吻合度极高。

当时间步长足够小（$\Delta t = 10^{-4}$）时，连续时间算法与离散时间算法之间没有明显差异，这符合预期。图 5 显示，最大差异

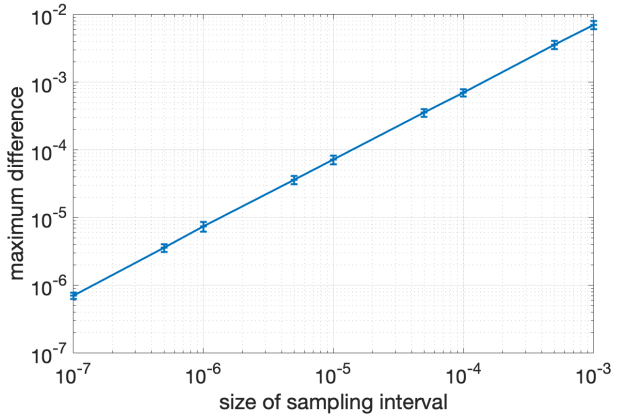

图 5：消息传递算法的一阶收敛随采样时间步长的变化。图中绘出了最大差异的均值，并以均值上下一个标准差作为误差条，数据来自由四十条轨迹构成的集合。

连续时间算法与离散时间算法生成的条件概率之间的ℓ_∞范数随采样时间步长 Δt线性减小。图中展示了每种样本量下进行四十次独立重复试验所得的结果。再次强调，离散时间算法需要遍历所有样本，其复杂度随 $\mathcal{O}(\Delta t^{-1})$增长；而连续时间算法可在 $\mathcal{O}(1)$时间内输出条件概率的描述 —— 也就是说，它能在给定的精确时间点返回一个值的函数，其复杂度仅随跃迁次数增长。

VI. 讨论与结论

本文提出了一种用于离散状态空间上的部分可观测马尔可夫过程的连续时间推断算法。我们证明，通过两组微分方程和逐点归一化，可以将著名的和积算法扩展到连续时间领域。在连续时间情形下，给定有限的时间序列，我们能够精确求解条件概率的轨迹。此外，我们发现，在特殊情况下，状态占用概率的动态变化可以化简为二阶微分方程。这些结果不仅具有数学意义，而且有望将推断问题简化为求解线性微分方程组，从而显著降低长时间序列的计算复杂度。从数值上看，当时间步长趋近于零时，连续时间算法与离散时间算法一致；但其执行时间近似为常数，而非随时间步数线性增长。

我们的形式体系可自然地扩展到非二元观测。简而言之，设 $\mathcal{S}_1$，$\mathcal{S}_2$，…，$\mathcal{S}_m$ 是样本空间 Ω 的一个划分，并假设可观测过程 $\mathcal{Y}$ 定义如下:

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

对于观测为 i 的驻留期，我们可以采用为二元情形所介绍的前向和后向消息传递方案，将观测视为属于 $\mathcal{S}_i$ 或 $\Omega \backslash \mathcal{S}_i$。在边界处，可使用相同的技术（如定理1的证明中所述），通过考察从 $\mathcal{S}_i$ 到 $\mathcal{S}_j$ 的可能转移来求得更新规则，其中所有 $j \neq i$ 均适用。最后，这种方法还可扩展到 Ω 的任意子集集合，其中各元素不一定互不相交。可通过将所有并集和交集展开为互不相交的集合（这些集合可能具有相同的观测量）来实现这一扩展。关于这一方向的详细研究留待未来开展。

将连续时间消息传递扩展到具有非齐次转移率的情形，仍是一个开放问题。我们预计，采用类似主定理证明的推导方式或许适用，但可能需要对转移率施加光滑性约束。虽然归一化常数将保持时间不变，但我们不认为推论 1 这样的结果能够适用于常数转移率以外的情形。

最后，我们指出，条件概率与二阶微分方程之间的关系在直觉上令人满意：进入和离开时间充当两个边界条件，用以确定演化的端点；而不受约束的前向演化方程是一阶微分方程，只需要初始条件。未来一个有趣的研究问题是：在什么假设下，类似推论 1 的结果能够适用于具有两个以上不同特征值的系统。

附录 A 符号与预备知识

为完整起见，我们定义一种具有指数等待时间的连续时间离散状态空间马尔可夫过程。

定义 1（马尔可夫过程）。令 $\{S(t); t \in [0,T]\}$ 为一个离散状态空间的连续时间随机过程，其中对于每个 t，$S(t)$ 都是一个状态空间为$\Omega = \{1,2,\ldots,N\}$的随机变量，$N < \infty$。那么，对于任意 s_1、s_2、…、$s_m \in \Omega$，$0 < \tau_1 < \tau_2 < \cdots < \tau_{m-1} < t$，过程 $S(t)$ 具有马尔可夫性质，

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

此外，如果对于任意 i、$j \in \Omega$ 和 $i \neq j$，存在一个常数速率 $0 \leq w_{ji} < \infty$，使得

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

当 Δt足够小时。

下表列出了本文中使用的符号。

符号	含义
S_t	时间索引为 t（整数）的离散时间过程
$S(t)$	时间索引为 t的连续时间过程
Ω	过程在固定时刻的样本空间
$\mathcal{S}$	观测量为“1”的状态
$\mathcal{Y}(t)$	观测过程（表示 $S(t) \in \mathcal{S}$）
P	马尔可夫链的转移矩阵
W	马尔可夫过程的转移率矩阵
$\boldsymbol{\alpha}_t$; $\boldsymbol{\alpha}(t)$	前向消息
$\boldsymbol{\beta}_t$; $\boldsymbol{\beta}(t)$	后向消息
$\boldsymbol{\chi}_t$; $\boldsymbol{\chi}(t)$	观测消息
$\boldsymbol{\rho}_t$; $\boldsymbol{\rho}(t)$	未归一化的条件状态占用概率
$Z(t)$	$\boldsymbol{\rho}(t)$ 的归一化常数
$\boldsymbol{p}_t$; $\boldsymbol{p}(t)$	条件状态占用概率
Δt	采样时间步长

附录 B 连续时间中的不变归一化

时不变的不变归一化是和积算法 [26] 的一项基本性质。在连续时间情形下，我们发现，当子矩阵可对角化时，很容易验证这一性质，因为对应于不同特征值的左、右特征向量彼此正交。当子矩阵具有非平凡的若尔当块时，时间相关项的抵消方式则不那么显而易见。这里我们将证明，一般情况下时间无关的归一化仍然成立，而且只需用到初等微积分，无须借助因子图的理论工具。

回顾一下，归一化常数为 $Z(t) = \sum_i \alpha^{(i)}(t)\beta^{(i)}(t)$。利用公式 2 和 3，我们可以得到

以下一系列表达式。

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

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

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

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

$$= 0 \tag{60}$$

由于 $Z(t)$的变化率为零，归一化常数不随时间变化。

参考文献

[1] G. Grisettiyz、C. Stachniss 和 W. Burgard，“通过自适应提议和选择性重采样，利用 Rao-Blackwell 化粒子滤波器改进基于网格的同时定位与地图构建”，载于*2005*年*IEEE*国际机器人与自动化会议论文集。IEEE，2005年，第2432–2437页。[2] L. E. Baum 和 J. A. Eagon，“一个不等式及其在马尔可夫过程概率函数的统计估计和生态学模型中的应用”，《美国数学会通报》，第73卷，第3期，第360–363页，1967年。[3] D. F. Anderson、B. Ermentrout 和 P. J. Thomas，“离子通道动力学的随机表示与神经元动力学的精确随机模拟”，《计算神经科学杂志》，第38卷，第1期，第67–82页，2015年。[4] D. M. Blei、A. Y. Ng 和 M. I. Jordan，“潜在狄利克雷分配”，《机器学习研究杂志》，第3卷，第993–1022页，2003年。[5] A. L. Hodgkin 和 A. F. Huxley，“对膜电流的定量描述及其在神经传导和兴奋中的应用”，《生理学杂志》，第117卷，第4期，第500页，1952年。[6] E. Skaugen 和 L. Walløe，“基于 Hodgkin–Huxley 方程的随机神经膜模型中的放电行为”，《斯堪的纳维亚生理学学报》，第107卷，第4期，第343–363页，1979年。[7] B. Hille，“可兴奋膜中的离子通道：当前问题与生物物理方法”，《生物物理学杂志》，第22卷，第2期，第283–294页，1978年。[8] D. Colquhoun 和 A. G. Hawkes，“流经药物门控通道的膜电流的弛豫与涨落”，《伦敦皇家学会会刊 *B* 辑：生物科学》，第199卷，第1135期，第231–262页，1977年。[9] D. Colquhoun 和 A. Hawkes，“单离子通道的随机特性”，《伦敦皇家学会会刊 *B* 辑：生物科学》，第211卷，第1183期，第205–235页，1981年。[10] D. Colquhoun 和 A. G. Hawkes，“单离子通道开放簇发及簇发簇的随机特性”，《伦敦皇家学会哲学汇刊 *B* 辑：生物科学》，第1–59页，1982年。[11] F. Qin、A. Auerbach 和 F. Sachs，“聚合马尔可夫过程的最大似然估计”，《伦敦皇家学会会刊 *B* 辑：生物科学》，第264卷，第383页，1997年，第1380、375页。[12] M. D. Fuller、Z.-R. Zhang、G. Cui 和 N. A. McCarty，“蝎毒对 CFTR 的阻断具有状态依赖性”，《生物物理学杂志》，第89卷，第6期，第3960–3975页，2005年。[13] H.-A. Loeliger、J. Dauwels、J. Hu、S. Korl、L. Ping 和 F. R. Kschischang，“基于因子图的模型驱动信号处理方法”，《*IEEE* 会刊》，第95卷，第6期，第1295–1322页，2007年。[14] D. Koller 和 N. Friedman，基于粒子的近似推断。麻省理工学院出版社，2012年。[15] Z. Fang、A. Gupta 和 M. Khammash，“基于混合近似的多尺度随机反应网络随机滤波”，《计算物理学杂志》，第467卷，文章编号111441，2022年。[16] O. Capp´e、C. P. Robert 和 T. Ryd´en，“可逆跳跃、生灭及更一般的连续时间马尔可夫链蒙特卡罗采样器”，英国皇家统计学会杂志：*B*辑（统计方法学），第65卷，第3期，第679–700页，2003年。[17] P. M. Djuric 和 J.-H. Chun，“通过马尔可夫链蒙特卡罗采样估计非平稳隐马尔可夫模型”，载于*1999*年*IEEE*国际声学、语音与信号处理会议论文集。*ICASSP99*（目录号 *99CH36258*），第3卷。IEEE，1999年，第1737–1740页。[18] R. Rosales、J. A. Stark、W. J. Fitzgerald 和 S. B. Hladky，“使用隐马尔可夫模型复原离子通道记录”，《生物物理学杂志》，第80卷，第3期，第1088–1103页，2001年。[19] F. R. Kschischang、B. J. Frey 和 H.-A. Loeliger，“因子图与和积算法”，《*IEEE*信息论汇刊》，第47卷，第2期，第498–519页，2001年。[20] D. Jin、F. Yin、C. Fritsche、F. Gustafsson 和 A. M. Zoubir，“路径损耗指数未知时利用接收信号强度进行贝叶斯协作定位：消息传递方法”，《*IEEE*信号处理汇刊》，第68卷，第1120–1135页，2020年。[21] S. Som 和 P. Schniter，“利用近似消息传递和马尔可夫树先验进行压缩成像”，《*IEEE*信号处理汇刊》，第60卷，第7期，第3439–3448页，2012年。[22] A. Abrardo、M. Barni、K. Kallas 和 B. Tondi，“存在同步攻击时对隐马尔可夫观测进行决策融合的消息传递方法”，载于多媒体进展国际会议（*MMEDIA*），时空相关数据的模型与算法（*STCD*）特别专题，意大利威尼斯，2017年。[23] L. Bolliger，使用因子图对连续时间信号进行数字估计，Hartung-Gorre出版社，康斯坦茨，信号与信息处理系列，第22卷，2012年。[24] L. Bolliger、H.-A. Loeliger 和 C. Vogel，“利用因子图，根据离散时间样本对连续时间信号进行线性最小均方误差估计与插值”，*arXiv*预印本 *arXiv:1301.4793*，2013年。[25] L. Bruderer 和 H.-A. Loeliger，“估计既非带限也非稀疏的传感器输入信号”，载于*2014*年信息论与应用研讨会（*ITA*）。IEEE，2014年，第1–5页。[26] G. D. Forney Jr 和 P. O. Vontobel，“正规因子图的配分函数”，*arXiv*预印本 *arXiv:1102.0316*，2011年。[27] R. G. Gallager，随机过程：应用理论。剑桥大学出版社，2013年。[28] A. W. Eckford 和 P. J. Thomas，“通道视紫红质及其他强度驱动型信号转导受体的信道容量”，《*IEEE*分子、生物与多尺度通信汇刊》，第4卷，第1期，第27–38页，2018年。[29] G. Nagel、T. Szellas、W. Huhn、S. Kateriya、N. Adeishvili、P. Berthold、D. Ollig、P. Hegemann 和 E. Bamberg，“通道视紫红质-2：一种直接由光门控的阳离子选择性膜通道”，《美国国家科学院院刊》，第100卷，第13 945页，2003年，第13 940、24页。[30] D. Goetz 和 C. L. Ren，“囊性纤维化综述”，《儿科年鉴》，第48卷，第4期，第e154–e161页，2019年。[31] D. T. Gillespie，“耦合化学反应的精确随机模拟”，《物理化学杂志》，第81卷，第25期，第2340–2361页，1977年。[32] ——，“化学动力学的随机模拟”，《物理化学年度评论》，第58卷，第35–55页，2007年。[33] D. J. Wilkinson，系统生物学随机建模。Chapman and Hall/CRC出版社，2018年。