Signal Processing 225 (2024) 109613

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Sampling and filtering with Markov chains

Michael A. Kouritzin

Department of Mathematical and Statistical Sciences, University of Alberta, Canada

ARTICLE INFO

MSC:
primary 62M05
62M20
secondary 60J22
65C35

Keywords:
Continuous-time hidden Markov model
Filtering equations
Importance sampling
Measure change
Rejection sampling
Stochastic analysis
Disease forecasting

ABSTRACT

A new continuous-time Markov chain rate change formula is proven. This theorem is used to derive existence and uniqueness of novel filtering equations akin to the Duncan–Mortensen–Zakai equation and the Fujisaki–Kallianpur–Kunita equation but for Markov signals with general continuous-time Markov chain observations. The equations in this second theorem have the unique feature of being driven by both the observations and the process counting the observation transitions. A direct method of solving these filtering equations is also derived. Most results apply as special cases to the continuous-time Hidden Markov Models (CTHMM), which have become important in applications like disease progression tracking, The corresponding CTHMM results are stated as corollaries. Finally, application of our general theorems to Markov chain importance sampling, rejection sampling and branching particle filtering algorithms is also explained and these are illustrated by way of disease tracking simulations.

1. Introduction

Importance sampling (IS), introduced by Kloek and van Dijk [1], is an important statistical variance reduction technique that is vital in problems like Monte Carlo simulation (see e.g. [2,3]). From an alternative viewpoint, it is a general technique for estimating properties of a particular distribution, while only having samples generated from a different distribution. Indeed, IS can also be important in rejection resampling and in developing filtering equations, which will be further demonstrated herein. A method to extend IS to Markov chains (with possibly time and/or hidden-state dependent rates) will be introduced in this paper. Consider a process Y (which will later be the observations) with finite or countable state space O that is Markov. Let it have a potentially time-dependent rate $\gamma_{i\to j}(s)$ of going from $i \in O$ to $j \in O$ at time s. Then, its generator is

$$\mathbb{L}_s g(i) = \sum_{j\neq i, j\in O} \gamma_{i\to j}(s)[g(j) - g(i)],$$

and Y solves a well-posed martingale problem with $\mathbb{L}_s$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ such that $\mathcal{F}_t^Y \doteq \sigma\{Y_s, s\leq t\} \subset \mathcal{F}_t$ for all $t \geq 0$ under sufficient regularity. But, rather than simulating this chain directly, one samples Y with respect to a reference measure Q from a simpler chain with generator:

$$\overline{\mathbb{L}} g(i) = \sum_{j\neq i, j\in O} \overline{\gamma}_{i\to j}[g(j) - g(i)], \tag{1}$$

where the rates $\{\overline{\gamma}_{i\to j}\}_{i\neq j}$ may, for example, not depend upon time or be more balanced. Let N count Y's state changes and $\overline{\gamma}_{i\to} = \sum_{j\neq i} \overline{\gamma}_{i\to j}$, $\gamma_{i\to}(s) = \sum_{j\neq i} \gamma_{i\to j}(s)$ be the rates of leaving state i. Then, we will prove, under regularity conditions, that the likelihood ratio process

$$A_t = \exp\left(\int_0^t \overline{\gamma}_{Y_{s-}\to} - \gamma_{Y_{s-}\to}(s)\, ds\right) \prod_{0<s\leq t} \left[1 + \left(\frac{\gamma_{Y_{s-}\to Y_s}(s)}{\overline{\gamma}_{Y_{s-}\to Y_s}} - 1\right)\Delta N_s\right] \tag{2}$$

is a $\{\mathcal{F}_t^Y\}$-martingale under Q. Moreover, under a Girsanov-type measure change

$$\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t} = A_t,$$

E-mail address: michaelk@ualberta.ca.

https://doi.org/10.1016/j.sigpro.2024.109613
Received 16 February 2024; Received in revised form 6 June 2024; Accepted 7 July 2024
Available online 26 July 2024

Y has the desired generator $\mathbb{L}_s$, hence desired process distribution, under P. This importance sampling result has applications in simulation, model selection/verification, particle filters, parameter estimation, filtering theory and expanding the theory of Hidden Markov models to Markov chain observations. For example, we can simulate a Markov chain with (simple) proposal rates $\{\overline{\gamma}_{i\to j}\}$ and use rejection sampling to create a Markov chain with (elaborate, possibly time-dependent) target rates $\{\gamma_{i\to j}(s)\}$.

Hidden Markov models (HMMs) were introduced in a series of papers by Baum and collaborators [4,5]. Traditional HMMs have enjoyed tremendous success in applications like computational finance [6], single-molecule kinetic analysis [7], speech recognition, and protein folding [8]. In such applications, the unobservable hidden states X are a discrete-time or continuous-time Markov chain and the observations process Y is some distorted, corrupted partial measurement of the current state of X satisfying the condition

$$P\left(Y_t \in A \middle| X_s, s \le t\right) = P\left(Y_t \in A \middle| X_t\right)$$

at update or observation times t. These probabilities $P\left(Y_t \in A \middle| X_t\right)$ are called the *emission probabilities*. In the case of a Continuous Time HMM (CTHMM), the times $\{t_i\}$ that the observations change are also random. We consider the observations to be held constant in a cadlag manner until a new update occurs. In this way, our results herein apply to CTHMM as corollaries. CTHMM have been used successfully in applications like network performance evaluation (see [9]) and disease progression (see [10]).

CTHMM's observations are independent given the hidden state but this is limiting. Consider observing weather conditions Y based upon hidden climate state X. Wind, precipitation and temperature are not independent random samples but rather time evolutions, depending upon the climate state. In other words, they depend upon their past values in addition to the current hidden state contradicting the Hidden Markov Model assumption. A simple correction to the CTHMM is to allow observation dependence only on the immediate past as well as the hidden state and insist that the measurements only take a finite or countable number of possibilities. Then, the observations become a Markov chain, whose transition probabilities depend upon the hidden states. For generality of the hidden process, we can just assume:

(**C0**) the hidden state X is a strong Markov process on a separable metric space E that is the unique solution to a (L, μ)-martingale problem (m.p.):

$$M_t^f = f\left(X_t\right) - \int_0^t Lf\left(X_s\right) ds$$

is a martingale for each $f \in D_L$ and $\mathcal{L}\left(X_0\right) = \mu$, where L is a linear operator with domain $D_L \subset \overline{C}(E)$ of bounded, continuous $f : E \to \mathbb{R}$ such that $Lf \in \overline{C}(E)$.

$\mathcal{L}\left(X_0\right)$ denotes the law of X_0. Now, let $\widehat{D}_L = D_L \times \overline{C}(O)$ for convenience. When X is a Markov chain, the generator L has the form $Lf(x) = \sum_{j\neq x} \lambda_{x\to j}\,[f(j) - f(x)]$ for $x \in E$, where $\lambda_{x\to j}$ is the rate of X going from state x to j.

Remark 1.1. See Ethier & Kurtz [11], especially Theorem 4.4.2, for motivation, information about the m.p. and Markov processes. We avoid most of this theory but note $L1 = 0$ for Markov generators and well-posedness of m.p.s is a standard way of defining processes.

The observations are modelled naturally as a Markov chain depending upon the hidden state. (In reality they may evolve as a Markov process, fine approximated by a Markov chain.) For example, given the hidden state of the climate, the observed wind, temperature etc. evolve as a Markov chain depending on climate. Hence, the observation rates depend on the hidden state and the observations satisfy a m.p. with an operator that depends on the signal:

$$m_t^g = g\left(Y_t\right) - \int_0^t \mathbb{L}\left(X_s\right) g\left(Y_s\right) ds \tag{3}$$

is a martingale for all $g \in \overline{C}(O)$. Here, $\mathbb{L}(x)g(y) = \sum_{j\neq y} \gamma_{y\to j}(x)\,[g(j) - g(y)]$, where $\{\gamma_{i\to j}(x)\}_{i,j\in O, i\neq j}$ are the rates from state i to state j when the hidden state is x. (This setting not only includes the nonlinear filtering setup but also things like common stochastic volatility models, both in the general Markov chain setting.) If the observation noise is independent of the hidden state's, then the role of independence gives us the combined m.p. for the hidden state and observations:

$$M_t^P = M_t^P(f,g) = f\left(X_t\right) g\left(Y_t\right) - \int_0^t g\left(Y_s\right) Lf\left(X_s\right) ds - \int_0^t f\left(X_s\right) \mathbb{L}\left(X_s\right) g\left(Y_s\right) ds \tag{4}$$

is a martingale for all $(f,g) \in \widehat{D}_L$. For clarity, we will refer to the model of a hidden Markov signal X (whether it is a chain or a general process) observed through a continuous-time Markov chain Y as a continuous Markov observation model (CMOM). It subsumes the popular CTHMM and can be studied through joint m.p. (4).

A key observation for learning about the hidden state is that we can construct this model from a reference probability Q using importance sampling. Motivated by the development of the Duncan–Mortensen–Zakai equation (see [12,13]), we can first consider simple, fake observations Y with some reference probability Q that do not depend upon the hidden state at all. In particular, they could have rates $\{\overline{\gamma}_{i\to j}\}_{i\neq j}$ (that do not depend upon x) as above and be constructed to have the combined m.p.:

$$M_t^Q = f\left(X_t\right) g\left(Y_t\right) - \int_0^t g\left(Y_s\right) Lf\left(X_s\right) ds - \int_0^t f\left(X_s\right) \overline{\mathbb{L}} g\left(Y_s\right) ds \tag{5}$$

is a martingale for all $(f,g) \in \widehat{D}_L$, where $\overline{\mathbb{L}}g(y) = \sum_{j\neq y} \overline{\gamma}_{y\to j}\,[g(j) - g(y)]$. Then, adjusting likelihood ratio A from (2) to account for the hidden state, one has that

$$A_t = \exp\left(\int_0^t \overline{\gamma}_{Y_{s-}} - \gamma_{Y_{s-}}\left(X_s\right) ds\right) \prod_{0<s\le t} \left[1 + \left(\frac{\gamma_{Y_{s-}\to Y_s}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to Y_s}} - 1\right) \Delta N_s\right] \tag{6}$$

is a $\left\{\mathcal{F}_t^Y\right\}$-martingale under Q that converts Q into a new probability P, where the hidden state and observations (X,Y) solve the desired joint m.p. (4).

The *filter* $\pi_t(B) = P\left(X_t \in B \middle| \mathcal{F}_t^Y\right)$, for Borel subsets B of E, provides information on the hidden state based upon the model and the back observations for both CTHMMs and CMOMs. $\{\pi_t,\ t \geq 0\}$ is a probability measure-valued process. The *unnormalized filter* $\sigma_t(B) = E^Q\left(A_t 1_{X_t \in B} \middle| \mathcal{F}_t^Y\right)$, with A as defined in (6), is a (finite, not-necessarily-probability) measure-valued process that provides the filter through Bayes rule

$$\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)} \tag{7}$$

for (at least) all bounded, measurable functions f, where $\pi_t(f) = \int_E f\, d\pi_t$. However, the unnormalized filter provides more than a means to compute the filter. Rather as explained in [14], $\sigma_t(1) = \sigma_t(E)$ also provides the *model rating* Bayes factor (i.e. integrated Likelihood) of the model under consideration over the reference model. This is an overall rating of both overall structure and model parameters, including those in the hidden state component of the model. If we have two sets of parameters, then we can form two models M^1, M^2 and produce two Bayes factors $\sigma_t^1(1), \sigma_t^2(1)$. We can compare these models on real data by evaluating $B_{1|2}(t) = \frac{\sigma_t^1(1)}{\sigma_t^2(1)}$. These models' signals can be singular to each other, even of different dimensions, so Bayes' factor methods are very general and effective. One could even test if there is value in having a hidden state. Hence, it is often more useful to produce a *direct* or *particle filter* approximation to the unnormalized filter than the (normalized) filter.

Particle filters that give Bayes factor information can be considered *model rating* particle filters. Let $\left\{X^i\right\}_{i=1}^{\infty}$ be independent copies of the signal, called particles, recall the likelihood A that was used to convert Q into P, let

$$A_t^i = \exp\left(\int_0^t \overline{\gamma}_{Y_s \to} - \gamma_{Y_s \to}\left(X_s^i\right) ds\right) \prod_{0<s\leq t}\left[1 + \left(\frac{\gamma_{Y_{s-}\to Y_s}\left(X_s^i\right)}{\overline{\gamma}_{Y_{s-}\to Y_s}} - 1\right)\Delta N_s\right],$$

which is like A in (6) except the hidden state is replaced with the particle, and form

$$\sigma_t^N(f) = \frac{1}{N}\sum_{i=1}^N A_t^i f\left(X_t^i\right).$$

Then, by the independence of Y and $\{X^i\}_{i=1}^{\infty}$ under Q, we can fix the path Y and find that

$$\frac{1}{N}\sum_{i=1}^N A_t^i f\left(X_t^i\right) \to E^Q\left[A_t f\left(X_t\right)\middle|\mathcal{F}_t^Y\right] \tag{8}$$

i.e. $\sigma_t^N(f) \to \sigma_t(f)$ a.s. $[Q]$ for each f by the strong law of large numbers. By selecting a countable collection of f that is closed under multiplication and strongly separate points, one can show a.s. convergence as measures i.e. $\sigma_t^N(\cdot) \Rightarrow \sigma_t(\cdot)$ a.s. $[Q]$ (see Lemma 2 of [15] and Lemma 7 of [16]). We refer to σ_t^N as the weighted particle filter. It approximates the unnormalized filter. Notice that the real observations are used with each particle X^i in A^i.

The weighted particle filter problem is particles drift away from the signal (except when living on small compact sets) and do not contribute to the conditional distributional. To maintain enough effective particles, we develop a branching method analogous to that of [17].

The Fujisaki–Kallianpur–Kunita (FKK) [18] and Duncan–Mortensen–Zakai (DMZ) [12] equations were major breakthroughs in the classical nonlinear filtering problem yielding the evolution of the filter and unnormalized filter. They are now the bases of many computational methods of solving filters on real problems. However, they are only known for the classical observation setting as well as certain specific signal-dependent noise structures (see [19]) and not for general Markov chain observations. As a final theoretical contribution we prove the FKK and DMZ equations for the CMOM filtering problem, i.e. for Markov signals and continuous-time Markov chain observations. In particular, for the setting described above, we show that the unnormalized filter is the unique strong $D_{\mathcal{M}_f(E)}[0,\infty)$-valued solution to:

$$\begin{aligned}\sigma_t(f(\cdot)) = \sigma_0(f(\cdot)) &+ \int_0^t \sigma_s(Lf(\cdot))ds + \int_0^t \sigma_s(f(\cdot)(\overline{\gamma}_{Y_s\to} - \gamma_{Y_s\to}(\cdot))\,ds \\ &+ \int_0^t \sigma_{s-}\left(\left[f(\cdot)\frac{\gamma_{Y_{s-}\to Y_s}(\cdot)}{\overline{\gamma}_{Y_{s-}\to Y_s}} - f(\cdot)\right]\right)dN_s, \quad s.t.\ \sigma_0 = \mathcal{L}(X_0).\end{aligned} \tag{9}$$

for all $f \in D_L$, where N counts the state transitions of Y. More generally, we establish:

$$\begin{aligned}&\sigma_t(f(\cdot, Y_t)) \\ &= \sigma_0(f(\cdot, Y_0)) + \int_0^t \sigma_s(Lf(\cdot, Y_s))ds + \int_0^t \sigma_s(f(\cdot, Y_s)(\overline{\gamma}_{Y_s\to} - \gamma_{Y_s\to}(\cdot))\,ds \\ &\quad + \int_0^t \sigma_{s-}\left(\left[f(\cdot, Y_s)\frac{\gamma_{Y_{s-}\to Y_s}(\cdot)}{\overline{\gamma}_{Y_{s-}\to Y_s}} - f(\cdot, Y_{s-})\right]\right)dN_s\end{aligned} \tag{10}$$

for all $f \in \widehat{D}_L$. Moreover, the probability measure-valued filter π solves

$$\begin{aligned}\pi_t(f(\cdot, Y_t)) = \pi_0(f(\cdot, Y_0)) &+ \int_0^t \pi_s(Lf(\cdot, Y_s))ds \\ &- \int_0^t \pi_s(f(\cdot, Y_s)\gamma_{Y_s\to}(\cdot)) - \pi_s(f(\cdot, Y_s))\pi_s(\gamma_{Y_s\to}(\cdot))ds \\ &+ \int_0^t \frac{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot) f(\cdot, Y_s)\right) - \pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot)\right)\pi_{s-}(f(\cdot, Y_{s-}))}{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot)\right)}dN_s,\end{aligned} \tag{11}$$

for all $f \in \widehat{D}_L$ subject to $\pi_0 = \mathcal{L}(X_0)$. In the CTHMM case, the observation rates do not depend the current observation. Instead, we have a rate for updates $\gamma(x)$ that can depend upon the hidden state and an emission probability $q_{y'}(x) = P(Y_t = y'|X_t = x, W_n = t)$ for all $t > 0$, where W_n is the time of the nth emission so $\overline{\gamma}_{i\to j}$, $\gamma_{i\to j}(x)$ become $\overline{\gamma}\,\overline{q}_j$, $\gamma(x)q_j(x)$, where $\overline{\gamma}$ and $\overline{q}_j$ are some canonical update rate and emission probability mass function that do not depend upon the hidden state x. In this CTHMM case, (5), (4) and (9) become

$$M_t^Q = f\left(X_t, Y_t\right) - \int_0^t Lf\left(X_s, Y_s\right) ds - \int_0^t \sum_j \overline{\gamma}\,\overline{q}_j [f\left(X_s, j\right) - f\left(X_s, Y_s\right)] ds, \tag{12}$$

$$M_t^P = f\left(X_t, Y_t\right) - \int_0^t Lf\left(X_s, Y_s\right) ds - \int_0^t \sum_j \gamma(X_s)\, q_j(X_s) [f(X_s, j) - f(X_s, Y_s)] ds \tag{13}$$

for all $f \in \widehat{D}_L$, where L operates only on the first variable, and

$$\sigma_t(f(\cdot)) = \sigma_0(f(\cdot)) + \int_0^t \sigma_s(Lf(\cdot)) ds + \int_0^t \sigma_s(f(\cdot)(\overline{\gamma} - \gamma(\cdot))\, ds \tag{14}$$
$$+ \int_0^t \sigma_{s-}\left(\left[f(\cdot)\frac{\gamma(\cdot)q_{Y_s}(\cdot)}{\overline{\gamma}\,\overline{q}_{Y_s}} - f(\cdot)\right]\right) dN_s, \quad s.t.\ \sigma_0 = \mathcal{L}(X_0).$$

for all $f \in D_L$. (10), (11) also simplify accordingly (see below). (14) can be used for filtering, model selection and as a basis for parameter estimation in CTHMM.

Finally, we develop a direct solution approach solving (9) (or (14)) that is being used in current work to estimate trends and volatility in financial models based upon tick-by-tick data. Naturally, there are many other potential applications for our results.

1.1. Layout

The next section is a proof of our rate-change Girsanov's theorem. Section 3 has our applications to resampling Markov chains, Monte Carlo simulation and particle filtering with Markov chain observations. A simulation study demonstrating the applicability of our results to data generation and to disease transmission is spread over this section. Section 4 contains a new development for filtering problems with Markov chain observations, establishing FKK and DMZ type equations for these filtering problems. It also contains our direct solutions to these filtering equations. Finally, our conclusions and highlights are in Section 5.

1.2. Notation

Let $B(S)$, $C(S)$ and $\overline{C}(S)$ be the bounded, continuous and continuous bounded real functionals respectively and $M(S, S)$ be the measurable functions from S to S with the topology of pointwise convergence on any Polish space S. Further, we let $\mathcal{M}_f(S)$ be the space of finite (non-negative) Borel measures with the topology of weak convergence i.e. the notion that $\nu^n \Rightarrow \nu$ if and only if $\nu^n(f) \to \nu(f)$ for all $f \in \overline{C}(S)$, and $\mathcal{P}(S) \subset \mathcal{M}_f(S)$ be the probability measures. Finally, we let $D_S[0,\infty)$ denote the space cadlag path from $[0,\infty)$ to S equipped with the Skorokhod metric.

For $z > 0$, $\lfloor z \rfloor$ is the greatest integer not more than z and $\{z\} = z - \lfloor z \rfloor$.

$\lambda_{i\to j}$ denotes the rate of a hidden Markov chain going from state i to state j and $\lambda_{i\to} = \sum_{j\neq i} \lambda_{i\to j}$ denotes the rate of leaving state i. Similarly, $\gamma_{i\to j}$ denotes the rate of an observed Markov chain going from state i to state j and $\gamma_{i\to} = \sum_{j\neq i} \gamma_{i\to j}$.

$a_{i,k} \overset{i}{\ll} b_{i,k}$ means $\forall k$, $\exists c_k > 0$ not depending on i s.t. $|a_{i,k}| \le c_k |b_{i,k}|$ $\forall i, k$.

L will be used for a generator, a linear operator $\overline{C}(S) \to \overline{C}(S)$. Here, S will either be the hidden state space E or the observation state space O.

A, A^i will be used for likelihood, particle weight processes respectively.

$\mathcal{L}(Z)$ will be used to denote the law or distribution of random variable Z.

$[W, Z]$ will denote quadratic covariation of two stochastic processes W and Z.

δ_x will be used for Dirac delta measure at point x.

$\mathcal{B}(S)$ will denote the Borel sets on metric space S.

2. Rate change formula

In this section, we derive a rate-change Girsanov-type theorem for Markov chains and hidden Markov signals with Markov chain observations. We use the notation of the later application as one can just take $X_s = s$ and $E = [0,T]$ for the former.

The following conditions will be imposed in our first main result to follow:

(C1) The observation state space O is a finite or countable metric space and $\sup_{i\in O} \overline{\gamma}_{i\to} < \infty$.
(C2) $\sup_{x\in E, i\in O} \frac{\gamma_{i\to}(x)}{\overline{\gamma}_{i\to}} < \infty$.
(C3) There are no cemetery states, meaning $\gamma_{i\to}(x), \overline{\gamma}_{i\to} > 0$ for all $i \in O$, $x \in E$.
(C4) ψ is a probability measure on $(O, \mathcal{B}(O))$.

Operator $\overline{\mathbb{L}}$, defined in (1), is bounded by (C1) and there is a unique solution to the m.p.:

$$g(Y_t) - \int_0^t \overline{\mathbb{L}} g(Y_s) ds$$

is a martingale for all $g \in \overline{C}(O)$ and $\mathcal{L}(Y_0) = \psi$. The solution can be constructed and simulated in the following manner. On our reference measure Q, let $\{T_n^{i\to}\}_{n\in\mathbb{N}, i\in O}$ be independent exponential random variables with rates $\{\overline{\gamma}_{i\to}\}_{i\in O}$, let $\{\xi_n^i, n \in \mathbb{N}_0\}_{i\in O}$ be independent discrete random variables such that $P\left(\xi_n^i = j\right) = \frac{\overline{\lambda}_{i\to j}}{\overline{\lambda}_{i\to}}$ for $j \neq i \in O$ and let θ_0 be a random sample from $\mathcal{L}(Y_0)$. Take all these to be independent of each other and of the hidden state X. Then, Y could be constructed under Q in three steps starting from some $Y_0 = \theta_0$, $W_0 = 0$:

(1) $\theta_n = \xi_{n-1}^{\theta_{n-1}}$ for $n \in \mathbb{N}$ (gets the transitions as a discrete chain)
(2) $W_n = W_{n-1} + T_n^{\theta_{n-1}\rightarrow}$ for $n \in \mathbb{N}$ (get the transition times)
(3) $Y_t = \theta_n$ for $t \in [W_n, W_{n+1})$ $n \in \mathbb{N}_0$ (create continuous time chain).

Notice that the quadratic variation $[f(X), g(Y)] = 0$ for all f, g by the X, Y independence under Q so $[f(X), N] = 0$, where N counts the transitions in Y, for all f.

Now, we have our first main result, which is a Girsanov change-of-measure result changing Markov chains by weighting.

Theorem 2.1. *Suppose (C1, C2, C3) hold and (X, Y) satisfies the (5) martingale problem starting from some initial law $\mathcal{L}(X_0, Y_0) = \nu$ under Q. Then,*

$$A_t = \exp\left(\int_0^t \overline{\gamma}_{Y_s\rightarrow} - \gamma_{Y_s\rightarrow}(X_s)\, ds\right) \prod_{0<s\le t}\left[1 + \left(\frac{\gamma_{Y_{s-}\rightarrow Y_s}(X_s)}{\overline{\gamma}_{Y_{s-}\rightarrow Y_s}} - 1\right)\Delta N_s\right] \tag{15}$$

is a $\{\mathcal{F}_t^Y\}$-martingale under Q, where N counts the transitions of Y. Moreover, if we define a new probability measure via $\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t} = A_t$, $\forall t \ge 0$, then (X, Y) satisfies the (4) m.p. starting from $\mathcal{L}(X_0, Y_0) = \nu$ under P.

Note: $\left[1 + \left(\frac{\gamma_{Y_{s-}\rightarrow Y_s}(X_s)}{\overline{\gamma}_{Y_{s-}\rightarrow Y_s}} - 1\right)\Delta N_s\right]$ is $\frac{\gamma_{Y_{s-}\rightarrow Y_s}(X_s)}{\overline{\gamma}_{Y_{s-}\rightarrow Y_s}}$ at transition times s or else 1.

Proof. By our construction, N counts the transitions of Y and the $\{\xi_i^k\}$ determine the actual transitions. Let $\mathcal{F}_t \doteq \sigma\{N_s, \xi_i^k, X_u : i \le N_s, k \in O, s \le t, u \in [0,\infty)\}$, $\overline{q}_{i\rightarrow j} = \frac{\overline{\lambda}_{i\rightarrow j}}{\overline{\lambda}_{i\rightarrow}}$ and $q_{i\rightarrow j}(x) = \frac{\lambda_{i\rightarrow j}(x)}{\lambda_{i\rightarrow}(x)}$. Under Q, the combined chain (Y, N) satisfies the m.p.

$$g(Y_t, N_t) - \int_0^t \overline{\mathbb{L}}^N g(Y_s, N_s)\, ds \tag{16}$$

is a $\{\mathcal{F}_t\}$-martingale, where $\overline{\mathbb{L}}^N g(y,n) = \sum_{j\ne y} \overline{\gamma}_{y\rightarrow j}[g(j, n+1) - g(y,n)]$. In particular, taking $g(y,n) = n$

$$M_t^N = N_t - \int_0^t \overline{\gamma}_{Y_s\rightarrow}\, ds \tag{17}$$

and so

$$\begin{aligned} &\int_0^t A_{s-}^1 \left(\frac{\gamma_{Y_{s-}\rightarrow}(X_s)}{\overline{\gamma}_{Y_{s-}\rightarrow}} - 1\right) dM_s^N \\ &= \int_0^t A_{s-}^1 \left(\frac{\gamma_{Y_{s-}\rightarrow}(X_s)}{\overline{\gamma}_{Y_{s-}\rightarrow}} - 1\right) dN_s - \int_0^t A_{s-}^1 \left(\frac{\gamma_{Y_{s-}\rightarrow}(X_s)}{\overline{\gamma}_{Y_{s-}\rightarrow}} - 1\right) \overline{\gamma}_{Y_s\rightarrow}\, ds, \end{aligned} \tag{18}$$

with A_t^1 defined below, are Q-local martingale. We break the weight A from (15) into two factors A^1 and A^2 defined by $A_0^1 = A_0^2 = 1$ and

$$dA_t^1 = A_{t-}^1 \left[\left(\frac{\gamma_{Y_{t-}\rightarrow}(X_t)}{\overline{\gamma}_{Y_{t-}\rightarrow}} - 1\right) dN_t + \left[\overline{\gamma}_{Y_{t\rightarrow}} - \gamma_{Y_{t\rightarrow}}(X_t)\right] dt\right] \tag{19}$$

$$dA_t^2 = A_{t-}^2 \left(\frac{q_{Y_{t-}\rightarrow Y_t}(X_t)}{\overline{q}_{Y_{t-}\rightarrow Y_t}} - 1\right) dN_t. \tag{20}$$

To show this, we use integration by parts with $\alpha = A^1 A^2$

$$\begin{aligned} d\alpha_t &= \alpha_{t-}\left[\left(\frac{\gamma_{Y_{t-}\rightarrow}(X_t)}{\overline{\gamma}_{Y_{t-}\rightarrow}} - 1\right) dN_t + \left[\overline{\gamma}_{Y_{t\rightarrow}} - \gamma_{Y_{t\rightarrow}}(X_t)\right] dt\right] \\ &+ \alpha_{t-}\left(\frac{q_{Y_{t-}\rightarrow Y_t}(X_t)}{\overline{q}_{Y_{t-}\rightarrow Y_t}} - 1\right) dN_t + \alpha_{t-}\left(\frac{\gamma_{Y_{t-}\rightarrow}(X_t)}{\overline{\gamma}_{Y_{t-}\rightarrow}} - 1\right)\left(\frac{q_{Y_{t-}\rightarrow Y_t}(X_t)}{\overline{q}_{Y_{t-}\rightarrow Y_t}} - 1\right) dN_t \\ &= \alpha_{t-}\left(\frac{\gamma_{Y_{t-}\rightarrow}(X_t)}{\overline{\gamma}_{Y_{t-}\rightarrow}}\frac{q_{Y_{t-}\rightarrow Y_t}(X_t)}{\overline{q}_{Y_{t-}\rightarrow Y_t}} - 1\right) dN_t + \alpha_{t-}\left[\overline{\gamma}_{Y_{t\rightarrow}} - \gamma_{Y_{t\rightarrow}}(X_t)\right] dt \\ &= \alpha_{t-}\left(\frac{\gamma_{Y_{t-}\rightarrow Y_t}(X_t)}{\overline{\gamma}_{Y_{t-}\rightarrow Y_t}} - 1\right) dN_t + \alpha_{t-}\left[\overline{\gamma}_{Y_{t\rightarrow}} - \gamma_{Y_{t\rightarrow}}(X_t)\right] dt. \end{aligned} \tag{21}$$

But, this stochastic exponential equation has unique solution (see Protter [20, Theorem II.36])

$$\alpha_t = \exp\left(\int_0^t \overline{\gamma}_{Y_{s\rightarrow}} - \gamma_{Y_{s\rightarrow}}(X_s)\, ds\right) \prod_{0<s\le t}\left[1 + \left(\frac{\gamma_{Y_{s-}\rightarrow Y_s}(X_s)}{\overline{\gamma}_{Y_{s-}\rightarrow Y_s}} - 1\right)\Delta N_s\right] = A_t$$

so $A_t = A_t^1 A_t^2$. Moreover, by Ito's formula (see Theorem II.36 of Protter) again

$$A_t^2 = \prod_{0<s\le t}\left[1 + \left(\frac{q_{Y_{s-}\rightarrow Y_s}(X_s)}{\overline{q}_{Y_{s-}\rightarrow Y_s}} - 1\right)\Delta N_s\right] \tag{22}$$

and

$$A_t^1 = \exp\left(\int_0^t \overline{\gamma}_{Y_{s\rightarrow}} - \gamma_{Y_{s\rightarrow}}(X_s)\, ds\right) \prod_{0<s\le t}\left[1 + \left(\frac{\gamma_{Y_{s-}\rightarrow}(X_s)}{\overline{\gamma}_{Y_{s-}\rightarrow}} - 1\right)\Delta N_s\right] \tag{23}$$

$$= \exp\left(\int_0^t \left[\overline{\gamma}_{Y_s\to} - \gamma_{Y_s\to}\left(X_s\right)ds + \ln\left(\frac{\gamma_{Y_{s-}\to}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to}}\right)dN_s\right]\right).$$

(*Show A^1 is a martingale*) By (18), (19) A^1 is a Q local martingale. Under (C1, C2), it is a martingale by (23). Let P^1 be defined by

$$\left.\frac{dP^1}{dQ}\right|_{\mathcal{F}_t} = A_t^1,\ \forall t \geq 0.$$

(*Apply Girsanov–Meyer*) It follows by (5) with $f \equiv 1$ and Theorem III.20 of Protter that

$$\widetilde{M}_t = g(Y_t) - g(Y_0) - \int_0^t \overline{\mathbb{L}}g(Y_s)ds - \int_0^t \frac{1}{A_s^1}d[A^1, M^Q]_s \tag{24}$$

is a local martingale under P^1. However, by (5) with $f \equiv 1$ again, (19), (23)

$$\left[A^1, M^Q\right]_t = \left[A^1, g(Y)\right]_t \tag{25}$$

$$\int_0^t \frac{1}{A_s^1}d[A^1, M^Q]_s = \int_0^t \frac{A_{s-}^1}{A_s^1}\left(\frac{\gamma_{Y_{s-}\to}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to}} - 1\right)\left[g\left(\xi_{N_{s-}}^{Y_{s-}}\right) - g\left(Y_{s-}\right)\right]dN_s \tag{26}$$

$$= \int_0^t \frac{\overline{\gamma}_{Y_{s-}\to}}{\gamma_{Y_{s-}\to}\left(X_s\right)}\left(\frac{\gamma_{Y_{s-}\to}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to}} - 1\right)\left[g\left(\xi_{N_{s-}}^{Y_{s-}}\right) - g\left(Y_{s-}\right)\right]dN_s.$$

Putting (24), (26) together, we have that

$$\widetilde{M}_t = \int_0^t \frac{\overline{\gamma}_{Y_{s-}\to}}{\gamma_{Y_{s-}\to}\left(X_s\right)}\left[g\left(\xi_{N_{s-}}^{Y_{s-}}\right) - g\left(Y_{s-}\right)\right]dN_s - \int_0^t \overline{\mathbb{L}}g(Y_s)ds$$

and so

$$\int_0^t \frac{\gamma_{Y_{s-}\to}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to}}d\widetilde{M}_s = g\left(Y_t\right) - g\left(Y_0\right) - \int_0^t \frac{\gamma_{Y_{s-}\to}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to}}\overline{\mathbb{L}}g(Y_s)ds \tag{27}$$

is a P^1-local martingale.

(*New Y m.p.*) Rewriting (27) and using (C1-C3) as well as $g \in \overline{C}(O)$, we have that

$$g\left(Y_t\right) - g\left(Y_0\right) - \int_0^t \gamma_{Y_s\to}\left(X_s\right)\sum_{j\neq Y_s}\overline{q}_{Y_s\to j}[g(j) - g(Y_s)]ds \tag{28}$$

is a P^1 martingale and following (16), (17) that

$$N_t - \int_0^t \gamma_{Y_{s-}\to}\left(X_s\right)ds \tag{29}$$

is a P^1-local martingale.

(*Show A^2 is a P^1-martingale*) Under (C1, C2), $\mathfrak{L}(x)g(y) \doteq \gamma_{y\to}(x)\sum_{j\neq y}\overline{q}_{y\to j}[g(j) - g(y)]$ are uniformly (in x) bounded operators and for a given $\{X_t, t \geq 0\}$, the $\mathfrak{L}(X_t)$ m.p. is well posed. The solution can be constructed using the same 3 step procedure given above except in (2) the distribution of T_n is determined by $P(T_n^{\theta_{n-1}} > t) = e^{-\int_{W_n}^{W_n+t}\gamma_{\theta_{n-1}\to}(X_s)ds}$ for $t \geq 0$. Hence, we still use independent $\{\xi_n^k\}$ such that $P^1\left(\xi_n^k = j\right) = Q\left(\xi_n^k = j\right) = \overline{q}_{k\to j}$ and

$$Y_s = \begin{cases}\xi_{N_{s-}}^{Y_{s-}} & \Delta N_s = 1\\ Y_{s-} & \Delta N_s = 0\end{cases}.$$

Therefore, by independence and our representation

$$E^{P^1}\left[\prod_{u<s\leq t}\left[1 + \left(\frac{q_{Y_{s-}\to Y_s}\left(X_s\right)}{\overline{q}_{Y_{s-}\to Y_s}} - 1\right)\Delta N_s\right]\middle|\mathcal{F}_u\right] \tag{30}$$

$$= E^{P^1}\left[\prod_{u<s\leq t}\left[1 + \left(\sum_j q_{Y_{s-}\to j}\left(X_s\right) - 1\right)\Delta N_s\right]\middle|\mathcal{F}_u\right] = 1$$

and A^2 is a P^1-martingale.

(*Use new m.p.*) (28) is our intermediate m.p. We know that

$$\begin{aligned} g\left(Y_t\right) - g\left(Y_0\right) &= \int_0^t \left[g\left(Y_s\right) - g\left(Y_{s-}\right)\right]dN_s \\ &= \int_0^t \left[g\left(\xi_{N_{s-}}^{Y_{s-}}\right) - g\left(Y_{s-}\right)\right]dN_s \\ &= \int_0^t \left[g\left(\xi_{N_{s-}}^{Y_{s-}}\right) - g\left(Y_{s-}\right)\right]\gamma_{Y_{s-}\to}\left(X_s\right)ds + \mathcal{M}_t \end{aligned} \tag{31}$$

and $\mathcal{M}$ is a local P^1-martingale since it is the sum of the following two terms, the first

$$\int_0^t \left[\sum_j \overline{q}_{Y_{s-}\to j}\,g(j) - g\left(Y_{s-}\right)\right][dN_s - \gamma_{Y_{s-}\to}\left(X_s\right)ds]$$

would clearly be a local martingale by (29) and the second

$$E^{P^1}\left[\int_u^t \left[g\left(\xi^{Y_{s-}}_{N_{s-}}\right) - \sum_j \overline{q}_{Y_{s-}\to j} g(j)\right][dN_s - \gamma_{Y_{s-}\to}(X_s)\,ds]\Big|\mathcal{F}_u\right] = 0$$

by the independence (from everything) and distribution of the $\{\xi^k_n\}$ under P^1.

Now, by (29), (31)

$$\int_0^t \frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}}(g(Y_s)-g(Y_{s-}))\,dN_s - \int_0^t \frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}}\left[g\left(\xi^{Y_{s-}}_{N_{s-}}\right)-g(Y_{s-})\right]\gamma_{Y_{s-}\to}(X_s)\,ds$$
$$= \widehat{M}^N_t$$

is a P^1-local martingale. Furthermore, if $\mathcal{G}_s = \sigma\{Y_v : v \le s\} \vee \mathcal{F}_u$ for $s > u$, then for $t > u$

$$E\left[\int_u^t \frac{\gamma_{Y_{s-}\to\xi^{Y_{s-}}_{N_{s-}}}(X_s)}{\overline{q}_{Y_{s-}\to\xi^{Y_{s-}}_{N_{s-}}}}\left[g\left(\xi^{Y_{s-}}_{N_{s-}}\right)-g(Y_{s-})\right] - \sum_j [g(j)-g(Y_{s-})]\,\gamma_{Y_{s-}\to j}(X_s)\,ds\Big|\mathcal{F}_u\right]$$
$$=\int_u^t E\left[E\left[\frac{\gamma_{Y_{s-}\to\xi^{Y_{s-}}_{N_{s-}}}(X_s)}{\overline{q}_{Y_{s-}\to\xi^{Y_{s-}}_{N_{s-}}}}\left[g\left(\xi^{Y_{s-}}_{N_{s-}}\right)-g(Y_{s-})\right] - \sum_j [g(j)-g(Y_{s-})]\,\gamma_{Y_{s-}\to j}(X_s)\Big|\mathcal{G}_{s-}\right]\Big|\mathcal{F}_u\right]ds$$
$$= 0.$$

Consequently,

$$\widetilde{M}^N_t = \int_0^t \frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}}(g(Y_s)-g(Y_{s-}))\,dN_s - \int_0^t \sum_j [g(j)-g(Y_{s-})]\,\gamma_{Y_{s-}\to j}(X_s)\,ds \tag{32}$$

is a P^1-local martingale.

(*Apply Girsanov–Meyer*) By (32), Theorem III.20 of Protter and the fact that A^2 is pure jump, one has that

$$\check{M}^N_t = \int_0^t \frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}}(g(Y_s)-g(Y_{s-}))\,dN_s - \int_0^t \sum_j [g(j)-g(Y_{s-})]\,\gamma_{Y_{s-}\to j}(X_s)\,ds \tag{33}$$
$$- \int_0^t \frac{1}{A^2_s}\,d[A^2, \int_0^\cdot \frac{q_{Y_{u-}\to Y_u}(X_u)}{\overline{q}_{Y_{u-}\to Y_u}}(g(Y_u)-g(Y_{u-}))\,dN_u]_s$$

is a local martingale under P. However, by (20) and (31)

$$\left[A^2, \int \frac{q_{Y_{u-}\to Y_u}(X_u)}{\overline{q}_{Y_{u-}\to Y_u}}(g(Y_u)-g(Y_{u-}))\,dN_u\right]_t \tag{34}$$
$$= \sum_{0<s\le t}\left[A^2_{s-}\left(\frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}} - 1\right)\frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}}(g(Y_s)-g(Y_{s-}))\right]\Delta N_s$$

so by (22)

$$\int_0^t \frac{1}{A^2_s}\,d\left[A^2, \int \frac{q_{Y_{u-}\to Y_u}(X_u)}{\overline{q}_{Y_{u-}\to Y_u}}(g(Y_u)-g(Y_{u-}))\,dN_u\right]_s \tag{35}$$
$$= \int_0^t \frac{A^2_{s-}}{A^2_s}\left(\frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}} - 1\right)\frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}}(g(Y_s)-g(Y_{s-}))\,dN_s$$
$$= \int_0^t \frac{\overline{q}_{Y_{s-}\to Y_s}}{q_{Y_{s-}\to Y_s}(X_s)}\left(\frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}} - 1\right)\frac{q_{Y_{s-}\to Y_s}(X_s)}{\overline{q}_{Y_{s-}\to Y_s}}(g(Y_s)-g(Y_{s-}))\,dN_s.$$

Substituting (35) into (33) and using bounded g, one has that

$$\check{M}^N_t = \int_0^t (g(Y_s)-g(Y_{s-}))\,dN_s - \int_0^t \sum_j [g(j)-g(Y_{s-})]\,\gamma_{Y_{s-}\to j}(X_s)\,ds \tag{36}$$
$$= g(Y_t) - g(Y_0) - \int_0^t \sum_j [g(j)-g(Y_{s-})]\,\gamma_{Y_{s-}\to j}(X_s)\,ds$$

is a martingale under P.

(*Work on X*) A third application of the Girsanov–Meyer theorem with $[f(X), N] = 0$ shows

$$m^f_t = f(X_t) - \int_0^t Lf(X_s)\,ds \tag{37}$$

is also a P local martingale. One now obtains from boundedness, integration by parts, (36), the fact $\left[m^f, N\right] = 0$ and (37) that

$$f(X_t)\,g(Y_t) - f(X_0)\,g(Y_0) - \int_0^t g(Y_s)Lf(X_s) + f(X_s)\mathbb{L}(X_s)g(Y_s)ds \tag{38}$$

is a P martingale for continuous, bounded g and $f \in D(L)$. □

As mentioned previously, CTHMM is a special case of our CMOM model. (C1-C3) are modified for the CTHMM model as follow:

(A1) The observation state space O is a finite or countable space.
(A2) $\sup_{x\in E}\gamma(x)<\infty$.
(A3) $\gamma(x)>0$ for all $x\in E$.

Now, the following result is an immediate corollary of Theorem 2.1.

Corollary 2.2. *Suppose (A1, A2, A3) hold and (X,Y) satisfies the (12) m.p. starting from some initial law $\mathcal{L}(X_0,Y_0)=\nu$ under Q. Then,*

$$A_t=\exp\left(\int_0^t \overline{\gamma}-\gamma\left(X_s\right)ds\right)\prod_{0<s\le t}\left[1+\left(\frac{\gamma(X_s)q_{Y_s}(X_s)}{\overline{\gamma}\,\overline{q}_{Y_s}}-1\right)\Delta N_s\right] \tag{39}$$

is a $\{\mathcal{F}_t^Y\}$-martingale under Q, where N counts the transitions of Y. Moreover, if P satisfies $\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t}=A_t,\ \ \forall t\ge 0$, then (X,Y) solves the (13) m.p. starting from $\mathcal{L}(X_0,Y_0)=\nu$ under P.

3. Simulation and model testing

3.1. Rejection sampling

Theorem 2.1 was partially motivated by the desire to change the rates of Markov chains. We start with a (simple Markov chain say) m.p. under the reference probability measure Q as the proposal process Y satisfying:

$$m_t^Q=g\left(Y_t\right)-\int_0^t\overline{\mathbb{L}}g\left(Y_s\right)ds \tag{40}$$

is a martingale for all g, with $\overline{\mathbb{L}}$ defined in (1). This can be expanded to the m.p.

$$M_t^Q=f(t)g\left(Y_t\right)-\int_0^t g\left(Y_s\right)\frac{d}{ds}f(s)\,ds-\int_0^t f(s)\overline{\mathbb{L}}g\left(Y_s\right)ds, \tag{41}$$

is a martingale for all f,g by integration by parts, which is just (5) with $X_s=s$ and $E=[0,\infty)$. Take D_L to be the bounded, one-time continuously differentiable functions with bounded derivative, where the derivative at 0 is from the right. (41) is then turned into target m.p. (4) by Theorem 2.1 with $X_s=s$:

$$M_t^P=M_t^P(f,g)=f(t)g\left(Y_t\right)-\int_0^t g\left(Y_s\right)\frac{d}{ds}f(s)\,ds-\int_0^t f(s)\mathbb{L}_s g\left(Y_s\right)ds \tag{42}$$

is a martingale for all $(f,g)\in\widehat{D}_L$ under the new measure P, where

$$\mathbb{L}_s g(i)=\sum_{j\neq i,j\in O}\gamma_{i\to j}(s)[g(j)-g(i)]. \tag{43}$$

The likelihood ratio martingale for this change is then given by (2) by Theorem 2.1.

In summary, we

1. Simulate Markov chain Y under the (simple) proposal process distribution. We think of this as being done on the reference probability space $(\Omega,\mathcal{F},Q)$.
2. Reweight the simulation by A so that the combined effect is like it came from the target distribution with a different probability P.

It is natural to wonder if there is some way to stay on the simulation space (with Q) and get rid of the weight. Notice we use sample Y dependence of the likelihood ratio weight:

$$A=A_T(Y)=\exp\left(\int_0^T\overline{\gamma}_{Y_s\to}-\gamma_{Y_s\to}(s)ds\right)\prod_{0<s\le T}\left[1+\left(\frac{\gamma_{Y_{s-}\to Y_s}(s)}{\overline{\gamma}_{Y_{s-}\to Y_s}}-1\right)\Delta N_s\right], \tag{44}$$

where N counts the transitions of Y, to convert proposal Markov chain simulations as target ones. In fact, the weight A tells us how good a proposal sample would be as a sample from the target m.p. Now, von Neumann's Acceptance–Rejection algorithm:
(Step 1) Simulate Y^Q with proposal m.p. and a $[0,C]$-Uniform U independent of Y^Q.
(Step 2) If $U\le A_T\left(Y^Q\right)$, then accept by setting $Y=Y^Q$ and quitting the algorithm. Otherwise, reject by returning to Step 1.
(with independence implied between iterations of the algorithm) will allow us to create a target sample without requiring the likelihood ratio weight. We constrain A to be bounded (for now) in order to show the algorithm works.
The following conditions will be imposed in our simulation result:

(C2′) $\sup_{s\in[0,T],i\in O}\frac{\gamma_{i\to}(s)}{\overline{\gamma}_{i\to}}<\infty$.
(C3′) There are no cemetery states, meaning $\gamma_{i\to}(s),\overline{\gamma}_{i\to}>0$ for all $i\in O,\ s\in[0,T]$.
(Good A) There is a $C>1$ such that $A=A_T(Y)\in[0,C]$ for all Y and $E^Q[A]=1$.

Proposition 3.1. *Suppose (C1, C2', C3', C4, Good A) hold, A is defined in (44) and $\{(Y_n^Q,U_n)\}_{n=1}^\infty$ are the independent samples produced by Step 1 of the rejection algorithm. Then, the rejection algorithm output Y has the target distribution on Q.*

For clarity, the theorem says one can simulate on a computer with rates $\{\overline{\gamma}_{i\to j}\}$ apply rejection resampling and get a Markov chain with rates $\{\gamma_{i\to j}(s)\}$ at times s.

Proof. Let $f : \mathbb{R}^T \to \mathbb{R}$ be continuous and bounded. By the algorithm's acceptance criterion and independence one has (with $A^n = A_T(Y_n^Q)$) that

$$\begin{aligned} E^Q[f(Y)] &= \sum_{n=1}^{\infty} E^Q\left[f\left(Y_n^Q\right)1_{U_1>A^1},\dots,1_{U_{n-1}>A^{n-1}},1_{U_n\le A^n}\right] \\ &= \sum_{n=1}^{\infty} E^Q\left[f\left(Y_n^Q\right),1_{U_n\le A^n}\right]Q\left(U_1>A^1,\dots,U_{n-1}>A^{n-1}\right) \\ &= \sum_{n=1}^{\infty} \frac{E^Q\left[f\left(Y_n^Q\right)A^n\right]}{C}Q\left(U_1>A^1\right)\cdots Q\left(U_{n-1}>A^{n-1}\right) \\ &= \sum_{n=1}^{\infty} \frac{E^Q\left[f\left(Y_1^Q\right)A^1\right]}{C}Q\left(U_1>A^1\right)^{n-1} \\ &= E^Q\left[f\left(Y_1^Q\right)A^1\right]\frac{1}{CQ\left(U_1\le A^1\right)} \\ &= E^P\left[f\left(Y_1^Q\right)\right]\frac{1}{CQ\left(U_1\le A^1\right)}. \end{aligned} \tag{45}$$

Substituting $f=1$, one finds that $1=\frac{1}{CQ(U_1\le A^1)}$ so $Q\left(U_1\le A^1\right)=\frac{1}{C}$. □

The acceptance rate decreases as C increases. Unfortunately, there is no absolute bound on the number of transitions N. If $\{\gamma_{i\to j}\}$ are chosen so that $\frac{\gamma_{Y_{s-}\to Y_s}(s)}{\overline{\gamma}_{Y_{s-}\to Y_s}}\le 1$ for all samples Y and all s, then we can use the rejection method as above without stopping times since

$$A=\exp\left(\int_0^T \overline{\gamma}_{Y_s\to}-\gamma_{Y_s\to}(s)\,ds\right)\prod_{0<s\le T}\left[1+\left(\frac{\gamma_{Y_{s-}\to Y_s}(s)}{\overline{\gamma}_{Y_{s-}\to Y_s}}-1\right)\Delta N_s\right]$$

can be bounded. (This can be either impossible or else force a very inefficient simulation.) Otherwise, we use the algorithm multiple times with stopping times.

Let W_n be the time of the n^{th} jump of N, which is the nth transition of Y and assume (Bounded transitions) Suppose there is a $c>0$ such that $\sup_i\left|\overline{\gamma}_{i\to}\right|\le c$ and $\sup_{i,j}\left|\frac{\gamma_{i\to j}(s)}{\overline{\gamma}_{i\to j}}\right|\le c$.

Now, the following lemma is trivial:

Lemma 3.2. *Suppose (Bounded transitions) holds, $T>0$ and $n\in\mathbb{N}$. Then, $A_{T\wedge W_n}$ satisfies (good A).*

Indeed, it can be a good idea to simulate a fixed number of transitions at a time based upon the following formula. Y only changes when N jumps and N only jumps by 1 and we can consider our explicit formula along the jump times $\{W_n\}$ of N. In which case we get:

$$A_{W_n}=A_{W_{n-1}}\exp\left(\int_{W_{n-1}}^{W_n}\left(\overline{\gamma}_{Y_{W_{n-1}}\to}-\gamma_{Y_{W_{n-1}}\to}(s)\right)ds\right)\left(\frac{\gamma_{Y_{W_{n-1}}\to Y_{W_n}}(W_n)}{\overline{\gamma}_{Y_{W_{n-1}}\to Y_{W_n}}}\right)$$

which shows how the explicit solution weight updates at jump times. Here, we have that $W_0=0$ and $A_0=1$. In this way, we can resample a few jump times at a time under (Bounded transitions), provided we do not exceed some fixed time $T>0$.

We begin our simulation study of disease spread by explaining how rejection sampling can be used to reduce the computation cost of generating samples.

Example. Tracking and forecasting infectious disease spread using mathematical models to assist policy creation has become increasingly important since the COVID-19 pandemic. Traditionally, deterministic, non-spatial models based on compartmental differential equations or cellular automata were used. More recently, stochastic models (see e.g. [21]) and methods to handle spatial interactions (see e.g. [22], [23]) have been employed. However, these references do not use filtering to estimate and track actual infections from point-of-care tests. We will explain how this could be done while illustrating some of our results in our three examples.

We consider a *contact process* disease spread model over $S=\{1,\dots,M\}^2$ with $M=20$ up to time $T=10$ days, which is meant to represent spread within individuals in some isolated town. In particular, we include the outskirts in $\overline{S}=\{0,1,\dots,M,M+1\}^2$ but insist there are no people hence disease in $\overline{S}\backslash S$ and model the disease spread by a contact process on S with radius $r=1$, infection rate b and recovery rate d. We need to break with the general notation of this paper momentarily and consider what will later be the hidden state X observable so we can apply rejection sampling and create possible hidden state and particle simulations. Each $X_t(\omega)$ maps $\overline{S}$ to $\{0,1\}$, with 1 representing infected and 0 representing healthy. However, the boundary conditions are set as follows:

$$X_t((0,j),\omega)=X_t((i,0),\omega)=X_t((M+1,j),\omega)=X_t((i,M+1),\omega)=0 \tag{46}$$

for all $i,j\in\{0,\dots,M+1\}$ and t,ω and

$$\{X_0(i,j)\}_{i,j=1}^M \quad \text{are independent } p=\frac{1}{20}\text{-Bernoulli.}$$

Then, X evolves as a Markov chain with rates at each site $(i,j)\in S$

$$\lambda_{1\to 0}=d \quad\text{and}\quad \lambda_{0\to 1}=b\{x(i-1,j)+x(i+1,j)+x(i,j-1)+x(i,j+1)\}.$$

Hence, there is a constant rate of an individual becoming healthy and the rate she gets infected is proportional to how many of her direct (since $r=1$) neighbours are infected. In this example, we assume there are two choices for each of d and b, $d=0.0999$ or $d=0.1001$ and $b=0.0499$

Table 1
Data reuse by rejection sampling.

b	d	$C_{b,d}$	Number accepted 1000 samples
0.0499	0.0999	1.075	957
0.0499	0.1001	1.075	948
0.0501	0.0999	1.075	908
0.0501	0.1001	1.075	891

or $b = 0.0501$ to test. Having the goal of saving simulation time, we just produce 1000 simulations with midpoint $(\bar{b}, \bar{d}) = (0.05, 0.1)$ and then use acceptance rejection to accept these simulations as $(b, d) = (0.0499, 0.0999)$, $(0.0501, 0.0999)$, $(0.0499, 0.1001)$ and $(0.0501, 0.1001)$. The details are as follows: The likelihood weights became

$$\begin{aligned} A \; &= \exp\left(\int_0^1 (0.3 - d - 4b)\overline{X}_s - (0.05 - b)\widehat{X}_s \; ds\right) \\ &\times \prod_{0<s\leq 1} \left[1 + \left(\frac{d}{0.1} 1_{\overline{X}_s < \overline{X}_{s-}} + \frac{b}{0.05} 1_{\overline{X}_s > \overline{X}_{s-}} - 1\right) \Delta \overline{N}_s\right], \end{aligned} \tag{47}$$

for each of our choices $(b, d) = (0.0499, 0.0999)$, $(0.0501, 0.0999)$, $(0.0499, 0.1001)$ and $(0.0501, 0.1001)$, where $\overline{N}$ counted the jumps of $\overline{X}$,

$$\overline{X}_s = \sum_{i,j=1}^{M} X_s(i,j) \quad \text{and} \quad \widehat{X}_s = \sum_{k=1}^{M} \{X_s(k,1) + X_s(k,M) + X_s(1,k) + X_s(M,k)\}.$$

Based upon these specifics, we decided to take $C = 1.075$ regardless of b, d for simplicity and just test that A never exceeded C in our simulations. It did not. This value of C translates into an acceptance probability of 0.9302, meaning we expect 930.2 samples for each of the four possibilities, a total of 3721, to be accepted. Indeed, we ran von Neumann's rejection algorithm (with four independent $[0, C]$-uniforms for the different b, d combinations), and reported the results Table 1. Since the total accepted samples far exceeds the simulated ones we reduced the simulation cost of producing samples. Clearly, some samples were used in different parameter groups. If the goal was to produce 800 samples of each type, then we would be done. Otherwise, we could simulate more. While this particular example is modest, it is easy imagine situations where the method in this example could be a huge savings.

3.2. Monte Carlo simulation

Very often, we are not after just one sample but rather an ensemble of samples to form a distribution. We are really after some probabilities, expectations or conditional expectations, which we can approximate by independent proposal particles on the reference probability space $(\Omega, \mathcal{F}, Q)$ weighted by likelihood martingale weights. In particular, suppose $\{Y^m\}_{m=1}^M$ are independent proposal particles on $(\Omega, \mathcal{F}, Q)$ with rates $\{\overline{\gamma}_{i\to j}\}$ say and we weight the mth particle with likelihood weight

$$A^m = A_T^m = \exp\left(\int_0^T \overline{\gamma}_{Y_s^m \to} - \gamma_{Y_s^m \to}(s)\, ds\right) \prod_{0<s\leq T} \left[1 + \left(\frac{\gamma_{Y_{s-}^m \to Y_s^m}(s)}{\overline{\gamma}_{Y_{s-}^m \to Y_s^m}} - 1\right) \Delta N_s^m\right],$$

where N^m counts Y^m's jumps. Then, it follows by the strong law of large numbers that

$$\frac{1}{M} \sum_{m=1}^{M} A_T^m f\left(Y_s^m, s \in [0,T]\right) \to E^Q\left[A_T f\left(Y_s, s \in [0,T]\right)\right] = E^P\left[f\left(Y_s, s \in [0,T]\right)\right].$$

Hence, *any* target distribution expectation can be estimated without (rejection) resampling.

We continue our disease spread simulation by using Monte Carlo to estimate probabilities.

Example. Consider the *contact process* disease spread model of the prior example with $M = 20$ and final time $T = 500$ days. Using the same four choices of b and d but still only simulating the average one where $b = 0.05$ and $d = 0.1$, we will estimate the probability $P(\overline{X}_T \leq 10)$. We will now use 10,000 samples but, instead of resampling, we just weight. The particle locations are independent copies $\{X^m\}$ of the proposal $b = 0.05$, $d = 0.1$ contact process X and adapting the likelihood from (47), we have four different likelihood weights

$$\begin{aligned} A_t^m(b,d) \; &= \exp\left(\int_0^t (0.3 - d - 4b)\overline{X}_s^m - (0.05 - b)\widehat{X}_s^m \; ds\right) \\ &\times \prod_{0<s\leq t} \left[1 + \left(\frac{d}{0.1} 1_{\overline{X}_s^m < \overline{X}_{s-}^m} + \frac{b}{0.05} 1_{\overline{X}_s^m > \overline{X}_{s-}^m} - 1\right) \Delta \overline{N}_s^m\right], \end{aligned} \tag{48}$$

for each of our choices $(b, d) = (0.0499, 0.0999)$, $(0.0501, 0.0999)$, $(0.0499, 0.1001)$ and $(0.0501, 0.1001)$, where $\overline{N}^m$ counts the jumps of $\overline{X}^m$,

$$\overline{X}_s^m = \sum_{i,j=1}^{M} X_s^m(i,j) \quad \text{and} \quad \widehat{X}_s^m = \sum_{k=1}^{M} \{X_s^m(k,1) + X_s^m(k,M) + X_s^m(1,k) + X_s^m(M,k)\}.$$

The results, recorded in Table 2, are computed by

$$\frac{1}{10000} \sum_{m=1}^{10000} A_t^m(b,d) 1_{\overline{X}_T^m \leq 10} \approx P(\overline{X}_T \leq 10 \mid b, d).$$

Table 2
Probability estimation by Monte Carlo.

b	d	$P(\overline{X}_T \le 10)$
0.0499	0.0999	0.0082
0.0499	0.1001	0.0081
0.0501	0.0999	0.0036
0.0501	0.1001	0.0035

3.3. Particle filtering and model selection

Particle filters use Monte Carlo methods to approximate filters. Branching particle filters are often among the best performers and have the advantages of readily providing Bayes factor and facilitating maximum likelihood estimation (see [24]). The idea of these particle filters is to start with the weighted particle filter (8) but then, as mentioned in the introduction, branch in an unbiased way so as to keep the weights somewhat uniform while still keeping the number of particles relatively constant. The following algorithm is an adaptation to our setting of the simple residual algorithm introduced and explained in [17]. The other algorithms given in [17] could also be employed to improve performance.

Algorithm Setting: Let $\{t_n\}_{n=1}^{\infty}$ be the random transition times of the observations and set $t_0 = 0$. Let $r \in (1, \infty)$ and $\left\{V_n^m\right\}_{n,m=1}^{\infty}$, $\left\{U_n^m\right\}_{n,m=1}^{\infty}$ be independent $[-0.1, 0.1]$-uniform, $[0, 1]$-uniform random variables. r is a branching parameter. $\left\{V_n^m\right\}$ are smoothing random variable's to ensure we can show convergence. They can be tighter than $[-0.1, 0.1]$.

All particles evolve independently of each other between observations and interact weakly (i.e. through their empirical measure) at observation times.

Let: $\left\{X_0^m\right\}_{m=1}^{N}$ be independent, $\mathcal{L}(X_0^m) = \mu$; $N_0^m = A_0^m = 1$, $m = 1, \dots, N$; $N_n = 0$, $n \in \mathbb{N}$.

Repeat: for $n = 1, 2, \dots$ do

Repeat: for $m = 1, \dots, N_{n-1}$ do

1. Evolve Particle Behaviour Independently: Simulate each X^m on $(t_{n-1}, t_n]$ independently according to the signal's generator. Call final point $\widetilde{X}_{t_n}^m$.
2. Get observation Y_{t_n}, which also gives the value of Y_s on $[t_n, t_{n+1})$.
3. Weight Particles: $\widehat{A}_{t_n}^m = A_{t_{n-1}}^m \exp\left(\int_{t_{n-1}}^{t_n} \overline{\gamma}_{Y_{s\to}} - \gamma_{Y_{s\to}}\left(X_s^m\right) ds\right) \frac{\gamma_{Y_{t_{n-1}} \to Y_{t_n}}\left(X_{t_n}^m\right)}{\overline{\gamma}_{Y_{t_{n-1}} \to Y_{t_n}}}$
4. Resample Decision: Let $\overline{A}_{t_n} = \frac{1}{N} \sum_{m=1}^{N_n} \widehat{A}_{t_n}^m$. If $\widehat{A}_{t_n}^m + V_n^m \notin \left(\frac{\overline{A}_{t_n}}{r}, r\overline{A}_{t_n}\right)$ then the Offspring Number, Weight are: $N_n^i = \left\lfloor \frac{\widehat{A}_{t_n}^m}{\overline{A}_{t_n}} \right\rfloor + 1_{U_n^m \le \left\{\frac{\widehat{A}_{t_n}^m}{\overline{A}_{t_n}}\right\}}$, $\widetilde{A}_{t_n}^m = \overline{A}_{t_n}$; or else the Offspring Number, Weight are: $N_n^m = 1$, $\widetilde{A}_{t_n}^m = \widehat{A}_{t_n}^m$.
5. Resample: $A_{t_n}^{N_n+j} = \widetilde{A}_{t_n}^m$, $X_{t_n}^{N_n+j} = \widetilde{X}_{t_n}^m$ for $j = 1, \dots, N_n^m$
6. Add Offspring Number: $N_n = N_n + N_n^m$

The branching particle filter approximations of σ_{t_n} are then

$$S_{t_n}^N(f) = \frac{1}{N} \sum_{m=1}^{N_n} A_{t_n}^m f\left(X_{t_n}^m\right).$$

Notice compared to regular Monte Carlo we only have one observable process Y but many hidden processes $\{X^m\}_{m=1}^{N}$, called particles. Compared to the weighted particle filter, the particles are adjusted so they are more effective and a better approximation is achieved with the same number of particles. Branching particle filters continue to be an active area of research (see e.g. [25,26]). Based upon [27], the following results are expected.

Conjecture 3.3. *Under general regularity conditions, for any $n \in \mathbb{N}$, the above Residual Branching particle filter satisfies:*

Slln $S_{t_n}^N \Rightarrow \sigma_{t_n}$ *(i.e. weak convergence) as $N \to \infty$ a.s. $[Q^Y]$;*

Mlln $\left|S_{t_n}^N(f) - \sigma_{t_n}(f)\right| \overset{N}{\ll} N^{-\beta}$ *a.s. $[Q^Y]$ $\forall f \in \overline{C}(E)_+$, $0 \le \beta < \frac{1}{2}$.*

For related background in nonlinear filtering and sequential Monte Carlo, the reader is referred to the books [28,29] as well as the vast literature, including [12,13,18,30–35] and their references and citations.

Remark 3.4. For the CTHMM particle filter, we need only change Step 3 to:

Weight Particles: $\widehat{A}_{t_n}^m = A_{t_{n-1}}^m \exp\left(\int_{t_{n-1}}^{t_n} \overline{\gamma} - \gamma\left(X_s^m\right) ds\right) \frac{\gamma(X_{t_n}^m) q_{Y_{t_n}}(X_{t_n}^m)}{\overline{\gamma}\, \overline{q}_{Y_{t_n}}}$.

The rest of the algorithm is unchanged.

We finish our simulation study of disease spread by adding observation testing. We refer the reader to [36] as well as its citations and references for the importance and difficulty of inferring general population health from *point of care* testing.

Example. [37] uses filtering in some sense to track disease transmission. However, our results can make this process far easier. Consider the same contact process disease spread model as the prior examples with $M = 10$ and $T = 100$ days. Now, we just have one population X with $b = 0.05$ and $d = 0.1$ that we cannot observe but rather only see testing data constructed as follows: $t_0 = 0$ and the times between tests $\{t_k - t_{k-1}\}_{k=1}^{\infty}$ are independent 100-exponential random variables so we would expect 10,000 tests over the 100 days. At each test time t_k, $k \ge 1$ an individual

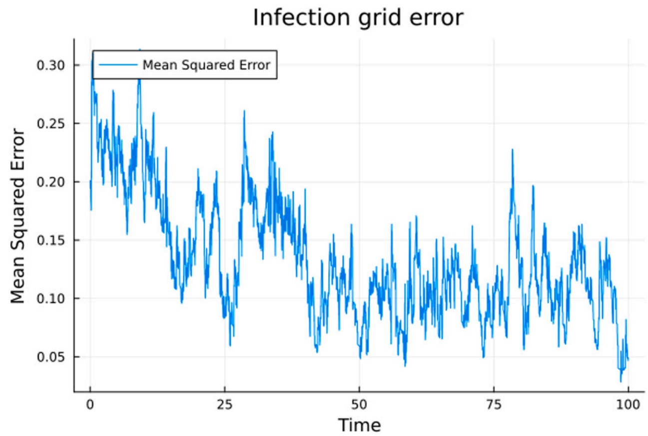

Fig. 1. Error of infections with branching particle filter (Initial Particles: 30,000, Branching Number: $r = 5$).

$(i_k, j_k) \in S^2$ is equally likely tested independent of all prior selections. Her location (i_k, j_k) is part of the observation as is her test result $Y_{t_k}(i_k, j_k)$. However, there is a chance of false positives and negatives, resulting in a confusion matrix with the real probability P: If this individual is infected, i.e. $X_{t_k}(i_k, j_k) = 1$, then $P(Y_{t_k}(i_k, j_k) = 1) = 0.80$ but, if this individual is healthy, then $P(Y_{t_k}(i_k, j_k) = 0) = 0.95$. This implies an observation rate of $\gamma(x) = 100$ and emission probabilities:

$$q_y(x(i_k, j_k)) = \begin{cases} 0.95\delta_0(y) + 0.05\delta_1(y), & x(i_k, j_k) = 0 \\ 0.8\delta_1(y) + 0.2\delta_0(y), & x(i_k, j_k) = 1 \end{cases}.$$

On the other hand, under the reference probability Q, we just take $\overline{\gamma} = 100$ and $\overline{q}_y = \frac{1}{2}\delta_0(y) + \frac{1}{2}\delta_1(y)$. Substituting these into the CTHMM weights, one finds

$$\widehat{A}^m_{t_n} = A^m_{t_{n-1}} \frac{q_{Y_{t_n}(i_n,j_n)}(X^m_{t_n}(i_n, j_n))}{\overline{q}_{Y_{t_n}(i_n,j_n)}} = A^m_{t_{n-1}} 2q_{Y_{t_n}(i_n,j_n)}(X^m_{t_n}(i_n, j_n))$$

in Step 3 of the particle filter algorithm. The branching particle algorithm was run using these A's in Step 3 and the unnormalized filter approximation

$$S^N_{t_n}(f) = \frac{1}{N} \sum_{m=1}^{N_n} A^m_{t_n} f\left(X^m_{t_n}\right)$$

was computed. This gives us a computer approximation of the optimal filter for the disease spread based only upon the observations. The whole conditional distribution of X over all of S given the back observations is available. However, due to space limitations, we only plot the error estimate of number of the infections averaged over the whole grid. We start the signal and particles with independent $\frac{1}{3}$-Bernoulli random variables (instead of $\frac{1}{20}$ in the previous examples), so the expected initial error is $\left(\frac{1}{3}\right)^2 \frac{2}{3} + \left(\frac{2}{3}\right)^2 \frac{1}{3} = 0.222$. The branching particle filter based upon our results then locates and tracks the infections throughout the population as shown in Fig. 1. Indeed, we see that the mean-square error estimate of the infections over the whole city grid continually improves from the expected start of 0.22 down to a low point of 0.03, which appears to be very good performance for a proof-of-concept type study.

4. Filtering equations

In addition to immediate, practical applications like those in the prior section, our main measure-change result for Markov chains, Theorem 2.1, can be used to establish other important theory. In particular in this section, we use this theorem to establish filtering equations akin to the Duncan–Mortensen–Zakai and the Fujisaki–Kallianpur–Kunita equations. The form and well-posedness of these new filtering equations for Markov signals with Markov chain observations represents our second main result. At the end of this section, we establish a direct solution method for these equations that can used in applications.

Our filtering equation approach uses the *reference probability* Q and related *unnormalized filter* process σ. P restricted to $\mathcal{F}_t$ is absolutely continuous with respect to Q restricted to $\mathcal{F}_t$ for $t \geq 0$ and the observation process Y is independent of the hidden state X under Q. Let $E^Q[\cdot]$ denote expectation with respect to Q, and consider the additional regularity condition:

(U) X is a Markov chain with state space $E \subset \mathbb{N}$.

Some additional regularity is required to establish uniqueness of (9). We chose to restrict X to be a Markov chain, which is immediately verifiable and built into the CTHMM model.

The observations Y are much simpler under Q so our strategy is to work under the reference probability and first derive an equation for σ_t. Then, apply Itô's formula to obtain the equation for the desired conditional distribution given by Bayes' formula

$$\pi_t(f) = \sigma_t(f)/\sigma_t(1).$$

Now, we can state our second main result, which is on the filtering equations.

Theorem 4.1. *Suppose (C0, C1, C2, C3) hold, (X,Y) satisfies the (5) martingale problem starting from some initial law $\mathcal{L}(X_0,Y_0)=\nu$ under Q and*

$$\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t}=A_t,\quad \forall t\ge 0,$$

where A is defined in (15). Then, σ, defined by $\sigma_t(B)=E^Q\left(A_t 1_{X_t\in B}\middle|\mathcal{F}_t^Y\right)$, solves (10) and π, defined by $\pi_t(B)=P\left(X_t\in B\middle|\mathcal{F}_t^Y\right)$, for Borel subsets B of E, solves (11). Moreover, if (U) also holds, then σ is the unique strong $D_{\mathcal{M}_f(E)}[0,\infty)$-valued solution to (9).

Remark 4.2. We have the ideal situation of existence in the more general setting (10) but uniqueness holding already in the narrow setting (9) for σ.

Remark 4.3. In the case that f only depends upon X, one has that

$$\begin{aligned}\pi_t(f)&=\pi_0(f)+\int_0^t\pi_s(Lf)ds-\int_0^t\pi_s(f\gamma_{Y_s\to})-\pi_s(f)\pi_s(\gamma_{Y_s\to})ds\\&+\int_0^t\frac{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}f\right)-\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}\right)\pi_{s-}(f)}{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}\right)}dN_s,\quad\forall f\in D_L.\end{aligned}\tag{49}$$

Proof σ satisfies (10). One notes by Q-independence, our representation $Y_s=\xi_{N_{s-}}^{Y_{s-}}$ and integration by parts that for $f=h\times g\in\widehat{D}_L$

$$\begin{aligned}&f(X_t,Y_t)-f(X_0,Y_0)\\&=\int_0^t Lf(X_s,Y_s)ds+m_t(f)+\int_0^t[f(X_{s-},\xi_{N_{s-}}^{Y_{s-}})-f(X_{s-},Y_{s-})]\,dN_s,\end{aligned}\tag{50}$$

where $m_t(f)=\int_0^t g(Y_{s-})dM_s^h$ and M^h is defined in (C0), so by (50) and (21) (with $\alpha=A$) as well as independence

$$[f(X,Y),A]_t=\int_0^t A_{s-}[f(X_{s-},\xi_{N_{s-}}^{Y_{s-}})-f(X_{s-},Y_{s-})]\left[\frac{\gamma_{Y_{s-}\to Y_s}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to Y_s}}-1\right]dN_s.\tag{51}$$

(L only operates on the first variable of f in (50) and below.) Utilizing integration by parts, (50), (21) and (51), one finds that

$$\begin{aligned}&f(X_t,Y_t)A_t-f(X_0,Y_0)\\&=\int_0^t A_{s-}df(X_s,Y_s)+\int_0^t f(X_{s-},Y_{s-})dA_s+[f(X,Y),A]_t\\&=\int_0^t A_sLf(X_s,Y_s)ds+\int_0^t A_{s-}dm_s(f)\\&+\int_0^t A_{s-}f(X_{s-},Y_{s-})\left(\frac{\gamma_{Y_{s-}\to Y_s}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to Y_s}}-1\right)dN_s\\&+\int_0^t A_sf(X_s,Y_s)(\overline{\gamma}_{Y_s\to}-\gamma_{Y_s\to}\left(X_s\right))ds\\&+\int_0^t A_{s-}[f(X_{s-},Y_s)-f(X_{s-},Y_{s-})]\frac{\gamma_{Y_{s-}\to Y_s}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to Y_s}}dN_s\\&=\int_0^t A_sLf(X_s,Y_s)ds+\int_0^t A_{s-}dm_s(f)+\int_0^t A_sf(X_s,Y_s)(\overline{\gamma}_{Y_s\to}-\gamma_{Y_s\to}\left(X_s\right))ds\\&+\int_0^t A_{s-}\left[f(X_{s-},Y_s)\frac{\gamma_{Y_{s-}\to Y_s}\left(X_s\right)}{\overline{\gamma}_{Y_{s-}\to Y_s}}-f(X_{s-},Y_{s-})\right]dN_s.\end{aligned}\tag{52}$$

Next, we show $E^Q[\int_0^t A_{s-}dm_s(f)|\mathcal{F}_t^Y]=0$. For each $n\in\mathbb{N}$, let $t_0^n=0$ and $\{t_i^n\}_{i=1}^\infty$ be a refining partition of stopping times that include the transition times of Y such that

$$A_{s-}^n\doteq 1_{\{0\}}(s)+\sum_{i=0}^n A_{t_i^n}1_{(t_i^n,t_{i+1}^n]}(s)$$

satisfies

$$\sup_{0\le s\le t}|A_{s-}-A_{s-}^n|\to 0\ \text{ in probability for any } t>0.$$

Then, $\int_0^t A_{s-}^n dm_s(f)\to\int_0^t A_{s-}dm_s(f)$ in probability and

$$E^Q\left|\int_0^t A_{s-}^n dm_s(f)-\int_0^t A_{s-}dm_s(f)\right|\to 0\tag{53}$$

by the boundedness of $m(f)$ and Condition (C1). Moreover, it follows by the tower property, independence and Doob's Optional Stopping that

$$E^Q\left[\int_0^t A^n_{s-}dm_s(f)\Big|\mathcal{F}^Y_t\right] = \sum_i E^Q\left[E^Q\left[A_{t^n_i}(m_{t^n_{i+1}}(f)-m_{t^n_i}(f))\Big|\mathcal{F}^X_{t^n_i}\vee\mathcal{F}^Y_t\right]\Big|\mathcal{F}^Y_t\right] \tag{54}$$
$$= \sum_i E^Q\left[A_{t^n_i}E^Q\left[(m_{t^n_{i+1}}(f)-m_{t^n_i}(f))\Big|\mathcal{F}^X_{t^n_i}\right]\Big|\mathcal{F}^Y_t\right]$$
$$= 0 \quad a.s.$$

Thus, it follows by (53), (54) and Jensen's inequality that

$$E^Q\left|E^Q\left[\int_0^t A_{s-}dm_s(f)\Big|\mathcal{F}^Y_t\right]\right| = \lim_{n\to\infty} E^Q\left|E^Q\left[\int_0^t A_{s-}dm_s(f) - \int_0^t A^n_{s-}dm_s(f)\Big|\mathcal{F}^Y_t\right]\right| \tag{55}$$
$$\le \lim_{n\to\infty} E^Q\left[\left|\int_0^t A_{s-}dm_s(f) - \int_0^t A^n_{s-}dm_s(f)\right|\right]$$
$$= 0.$$

Letting E^* denote Q-expectation with respect to X only, and setting

$$\sigma_t(f) \equiv E^Q[f(X_t,Y_t)A_t|\mathcal{F}^Y_t] = E^*[f(X_t,Y_t)L_t],$$

we find by (52) that the Zakai-type equation for $\sigma_t(f)$ becomes

$$\sigma_t(f(\cdot,Y_t)) = \sigma_0(f(\cdot,Y_0)) + \int_0^t \sigma_s(Lf(\cdot,Y_s))ds + \int_0^t \sigma_s(f(\cdot,Y_s)(\overline{\gamma}_{Y_s\to} - \gamma_{Y_s\to}(\cdot))\,ds$$
$$+ \int_0^t \sigma_{s-}\left(\left[f(\cdot,Y_s)\frac{\gamma_{Y_{s-}\to Y_s}(\cdot)}{\overline{\gamma}_{Y_{s-}\to Y_s}} - f(\cdot,Y_{s-})\right]\right)dN_s.$$

Proof of (9) uniqueness. For ease of notation, we take $E=\mathbb{N}$ and $t_0=0$. Let $\{t_l\}_{l=1}^\infty$ be the random transition times (in order) for Y and $Lf(i)=\sum_{j\neq i}\lambda_{i\to j}[f(j)-f(i)]$ be X's generator. Then, the adjoint operator $L^*p(j)$ satisfies

$$L^*p(j) = \sum_i\left[\lambda_{i\to j}p(i) - \lambda_{j\to i}\,p(j)\right] = [L^*]p(j), \quad [L^*] = \begin{bmatrix} -\lambda_{1\to} & \lambda_{2\to 1} & \lambda_{3\to 1} & \cdots \\ \lambda_{1\to 2} & -\lambda_{2\to} & \lambda_{3\to 2} & \cdots \\ \lambda_{1\to 3} & \lambda_{2\to 3} & -\lambda_{3\to} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Now, if we let $\sigma^i_t=\sigma_t(\delta_i)$ for $i=1,2,\ldots$, then we will discover that (9) gives us the closed system of linear differential equations parameterized by the observations:

$$d\begin{bmatrix}\sigma^1_t\\ \sigma^2_t\\ \vdots\end{bmatrix} = \begin{bmatrix} \overline{\gamma}_{Y_t\to}-\gamma_{Y_t\to}(1)-\lambda_{1\to} & \lambda_{2\to 1} & \cdots \\ \lambda_{1\to 2} & \overline{\gamma}_{Y_t\to}-\gamma_{Y_t\to}(2)-\lambda_{2\to} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}\begin{bmatrix}\sigma^1_t\\ \sigma^2_t\\ \vdots\end{bmatrix}dt \tag{56}$$
$$+ \begin{bmatrix} \frac{\gamma_{Y_{t-}\to Y_t}(1)}{\overline{\gamma}_{Y_{t-}\to Y_t}}-1 & 0 & \cdots \\ 0 & \frac{\gamma_{Y_{t-}\to Y_t}(2)}{\overline{\gamma}_{Y_{t-}\to Y_t}}-1 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}\begin{bmatrix}\sigma^1_{t-}\\ \sigma^2_{t-}\\ \vdots\end{bmatrix}dN_t.$$

Using the *mild solution* and *Trotter product*, we can write the solution explicitly between observation times. Let $P_t(i\to j)$, $i,j\in\{1,2,\ldots,m\}$ be the transition function for the hidden Markov chain X. Then, the unnormalized filter is $\sigma_t(\cdot)=\sum_{i\in E}\sigma^i_t\delta_i(\cdot)$, where

$$\begin{bmatrix}\sigma^1_t\\ \sigma^2_t\\ \vdots\end{bmatrix} = \left[T^n_{t-t_{n-1}}\right]\begin{bmatrix}\sigma^1_{t_{n-1}}\\ \sigma^2_{t_{n-1}}\\ \vdots\end{bmatrix} \tag{57}$$

for all $t\in[t_{n-1},t_n)$ with $T^n_t=\lim_{N\to\infty}\left[S^n_{\frac{t-t_{n-1}}{N}}\right]^N$ and Trotter product factor

$$S^n_t = \begin{bmatrix} P_t(1\to 1) & P_t(2\to 1) & \cdots \\ P_t(1\to 2) & P_t(2\to 2) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}\begin{bmatrix} e^{t(\overline{\gamma}_{Y_{t_{n-1}}\to}-\gamma_{Y_{t_{n-1}}\to}(1))} & 0 & \cdots \\ 0 & e^{t(\overline{\gamma}_{Y_{t_{n-1}}\to}-\gamma_{Y_{t_{n-1}}\to}(2))} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{58}$$

Each T^n behaves as a semi-group on $[0, t_n - t_{n-1})$. Then,

$$\begin{bmatrix}\sigma^1_{t_n}\\ \sigma^2_{t_n}\\ \vdots\end{bmatrix} = \begin{bmatrix}\frac{\gamma_{Y_{t_{n-1}}\to Y_{t_n}}(1)}{\overline{\gamma}_{Y_{t_{n-1}}\to Y_{t_n}}}\sigma^1_{t_n-}\\ \frac{\gamma_{Y_{t_{n-1}}\to Y_{t_n}}(2)}{\overline{\gamma}_{Y_{t_{n-1}}\to Y_{t_n}}}\sigma^2_{t_n-}\\ \vdots\end{bmatrix}, \tag{59}$$

and the equations start at $\begin{bmatrix}\sigma^1_{t_0}\\ \sigma^2_{t_0}\\ \vdots\end{bmatrix} = \begin{bmatrix}P(X_0=1)\\ P(X_0=2)\\ \vdots\end{bmatrix}$.

Now, suppose $\sigma, \widehat{\sigma}$ are $D_{\mathcal{M}_f(E)}[0,\infty)$-valued solutions to (9) and $\sigma = \widehat{\sigma}$ on $[0, t_{n-1}]$. Then,

$$\begin{bmatrix}\sigma^1_{t}\\ \sigma^2_{t}\\ \vdots\end{bmatrix} = T^n_{t-t_{n-1}}\begin{bmatrix}\sigma^1_{t_{n-1}}\\ \sigma^2_{t_{n-1}}\\ \vdots\end{bmatrix} = T^n_{t-t_{n-1}}\begin{bmatrix}\widehat{\sigma}^1_{t_{n-1}}\\ \widehat{\sigma}^2_{t_{n-1}}\\ \vdots\end{bmatrix} = \begin{bmatrix}\widehat{\sigma}^1_{t}\\ \widehat{\sigma}^2_{t}\\ \vdots\end{bmatrix}$$

for $t \in [t_{n-1}, t_n)$ so $\sigma_t(\cdot) = \sum_{i\in E}\sigma^i_t\delta_i(\cdot) = \sum_{i\in E}\widehat{\sigma}^i_t\delta_i(\cdot) = \widehat{\sigma}_t(\cdot)$ and uniqueness holds on $[0, t_n)$. Finally, (59) yields

$$\begin{bmatrix}\sigma^1_{t_n}\\ \sigma^2_{t_n}\\ \vdots\end{bmatrix} = \begin{bmatrix}\frac{\gamma_{Y_{t_{n-1}}\to Y_{t_n}}(1)}{\overline{\gamma}_{Y_{t_{n-1}}\to Y_{t_n}}}\sigma^1_{t_n-}\\ \frac{\gamma_{Y_{t_{n-1}}\to Y_{t_n}}(2)}{\overline{\gamma}_{Y_{t_{n-1}}\to Y_{t_n}}}\sigma^2_{t_n-}\\ \vdots\end{bmatrix} = \begin{bmatrix}\frac{\gamma_{Y_{t_{n-1}}\to Y_{t_n}}(1)}{\overline{\gamma}_{Y_{t_{n-1}}\to Y_{t_n}}}\widehat{\sigma}^1_{t_n-}\\ \frac{\gamma_{Y_{t_{n-1}}\to Y_{t_n}}(2)}{\overline{\gamma}_{Y_{t_{n-1}}\to Y_{t_n}}}\widehat{\sigma}^2_{t_n-}\\ \vdots\end{bmatrix} = \begin{bmatrix}\widehat{\sigma}^1_{t_n}\\ \widehat{\sigma}^2_{t_n}\\ \vdots\end{bmatrix}, \tag{60}$$

so (9) (strong) uniqueness holds on $[0, t_n]$ and on $[0,\infty)$ by induction.

Proof of (11). It follows by (10) that

$$\sigma_s(f(\cdot, Y_s)) = \sigma_{s-}(f(\cdot, Y_{s-})) + \sigma_{s-}\left(\left[f(\cdot, Y_s)\frac{\gamma_{Y_{s-}\to Y_s}(\cdot)}{\overline{\gamma}_{Y_{s-}\to Y_s}} - f(\cdot, Y_{s-})\right]\right)\Delta N_s \tag{61}$$

so

$$\frac{\sigma_s(1)}{\sigma_{s-}(1)} = 1 + \pi_{s-}\left(\frac{\gamma_{Y_{s-}\to Y_s}(\cdot)}{\overline{\gamma}_{Y_{s-}\to Y_s}} - 1\right)\Delta N_s\ . \tag{62}$$

Next, recalling $\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)}$ and using (61) twice then (62), one has that

$$\begin{aligned}
&\pi_s(f(\cdot, Y_s)) - \pi_{s-}(f(\cdot, Y_{s-}))\\
&= \frac{\sigma_s(f(\cdot, Y_s)) - \sigma_{s-}(f(\cdot, Y_{s-})) - (\sigma_s(1) - \sigma_{s-}(1))\pi_{s-}(f(\cdot, Y_{s-}))}{\sigma_s(1)}\\
&= \frac{\sigma_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot) f(\cdot, Y_s)\right) - \sigma_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot)\right)\pi_{s-}(f(\cdot, Y_{s-}))}{\sigma_s(1)\overline{\gamma}_{Y_{s-}\to Y_s}}\Delta N_s\\
&= \frac{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot) f(\cdot, Y_s)\right) - \pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot)\right)\pi_{s-}(f(\cdot, Y_{s-}))}{\pi_{s-}\left(\frac{\gamma_{Y_{s-}\to Y_s}(\cdot)}{\overline{\gamma}_{Y_{s-}\to Y_s}}\right)\overline{\gamma}_{Y_{s-}\to Y_s}}\Delta N_s\\
&= \frac{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot) f(\cdot, Y_s)\right) - \pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot)\right)\pi_{s-}(f(\cdot, Y_{s-}))}{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot)\right)}\Delta N_s.
\end{aligned} \tag{63}$$

Ito's formula on $\pi_t(f(\cdot, Y_t)) = \frac{\sigma_t(f(\cdot, Y_t))}{\sigma_t(1)}$ gives

$$d\pi_t(f(\cdot, Y_t)) = \frac{d\sigma^c_t(f(\cdot, Y_t))}{\sigma_t(1)} - \pi_t(f(\cdot, Y_t))\frac{d\sigma^c_t(1)}{\sigma_t(1)} + \pi_t(f(\cdot, Y_t)) - \pi_{t-}(f(\cdot, Y_t)), \tag{64}$$

where c indicates the continuous part. Hence, using (10), (64) and (63), one has

$$\begin{aligned}
\pi_t(f(\cdot, Y_t)) &= \pi_0(f(\cdot, Y_0)) + \int_0^t \pi_s(Lf(\cdot, Y_s))ds\\
&\quad - \int_0^t \pi_s(f(\cdot, Y_s)(\gamma_{Y_s\to}(\cdot) - \overline{\gamma}_{Y_s\to})) - \pi_s(f(\cdot, Y_s))\pi_s(\gamma_{Y_s\to}(\cdot) - \overline{\gamma}_{Y_s\to})ds\\
&\quad + \int_0^t \frac{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot) f(\cdot, Y_s)\right) - \pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot)\right)\pi_{s-}(f(\cdot, Y_{s-}))}{\pi_{s-}\left(\gamma_{Y_{s-}\to Y_s}(\cdot)\right)}dN_s\ . \quad \square
\end{aligned}$$

We state the filtering equations result for the popular CTHMM, as an immediate corollary. However, to do this we have to first give the CTHMM versions of (11) and (10), which are:

$$\sigma_t(f(\cdot,Y_t)) = \sigma_0(f(\cdot,Y_0)) + \int_0^t \sigma_s(Lf(\cdot,Y_s))ds + \int_0^t \sigma_s(f(\cdot,Y_s)(\overline{\gamma} - \gamma(\cdot))\,ds + \int_0^t \sigma_{s-}\left(\left[f(\cdot,Y_s)\frac{\gamma(\cdot)q_{Y_s}(\cdot)}{\overline{\gamma}\,\overline{q}_{Y_s}} - f(\cdot,Y_{s-})\right]\right)dN_s \tag{65}$$

for all $f \in \widehat{D}_L$ subject to $\sigma_0 = \mathcal{L}(X_0)$ and

$$\begin{aligned}\pi_t(f(\cdot,Y_t)) = {} & \pi_0(f(\cdot,Y_0)) + \int_0^t \pi_s(Lf(\cdot,Y_s))ds \\ & - \int_0^t \pi_s(f(\cdot,Y_s)\gamma(\cdot)) - \pi_s(f(\cdot,Y_s))\pi_s(\gamma(\cdot))ds \\ & + \int_0^t \frac{\pi_{s-}\left(\gamma(\cdot)q_{Y_s}(\cdot)f(\cdot,Y_s)\right) - \pi_{s-}\left(\gamma(\cdot)q_{Y_s}(\cdot)\right)\pi_{s-}(f(\cdot,Y_{s-}))}{\pi_{s-}\left(\gamma(\cdot)q_{Y_s}(\cdot)\right)}dN_s,\end{aligned} \tag{66}$$

for all $f \in \widehat{D}_L$ subject to $\pi_0 = \mathcal{L}(X_0)$. Now, the CTHMM corollary is:

Corollary 4.4. *Suppose (C0, A1, A2, A3) hold, (X,Y) satisfies the (12) martingale problem starting from some initial law $\mathcal{L}(X_0,Y_0) = \nu$ under Q and $\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t} = A_t$, $\forall t \geq 0$, where A is defined in (39). Then, σ solves (65) and π solves (66). Moreover, if (U) also holds, then σ is the unique strong $D_{\mathcal{M}_f(E)}[0,\infty)$-valued solution to (14).*

4.1. Direct solution

An effective computer workable solution to many real filtering problems can constructed from the DMZ equation based on uniqueness technique above. Consider the case where X is (or has been approximate by) a Markov chain on a finite space $E = \{1,2,\ldots,m\}$ with generator $Lf(i) = \sum_{j\neq i}\lambda_{i\to j}[f(j) - f(i)]$ for $i \in \{1,2,\ldots,m\}$. Now, if $\sigma_t^i = \sigma_t(\delta_i)$ for $i = 1,2,\ldots,m$, then (9) gives us the system of equations:

$$d\begin{bmatrix}\sigma_t^1\\ \sigma_t^2\\ \vdots\\ \sigma_t^m\end{bmatrix} = \begin{bmatrix}\overline{\gamma}_{Y_t\to} - \gamma_{Y_t\to}(1) - \lambda_{1\to} & \lambda_{2\to 1} & \cdots & \lambda_{m\to 1}\\ \lambda_{1\to 2} & \overline{\gamma}_{Y_t\to} - \gamma_{Y_t\to}(2) - \lambda_{2\to} & \cdots & \lambda_{m\to 2}\\ \vdots & \vdots & \ddots & \vdots\\ \lambda_{1\to m} & \lambda_{2\to m} & \cdots & \overline{\gamma}_{Y_t\to} - \gamma_{Y_t\to}(m) - \lambda_{m\to}\end{bmatrix}\begin{bmatrix}\sigma_t^1\\ \sigma_t^2\\ \vdots\\ \sigma_t^m\end{bmatrix}dt + \begin{bmatrix}\frac{\gamma_{Y_{t-}\to Y_t}(1)}{\overline{\gamma}_{Y_{t-}\to Y_t}} - 1 & 0 & \cdots & 0\\ 0 & \frac{\gamma_{Y_{t-}\to Y_t}(2)}{\overline{\gamma}_{Y_{t-}\to Y_t}} - 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & & \frac{\gamma_{Y_{t-}\to Y_t}(m)}{\overline{\gamma}_{Y_{t-}\to Y_t}} - 1\end{bmatrix}\begin{bmatrix}\sigma_{t-}^1\\ \sigma_{t-}^2\\ \vdots\\ \sigma_{t-}^m\end{bmatrix}dN_t.$$

Let $\{t_n\}_{n=1}^\infty$ be the random transition times of the observations, $t_0 = 0$, $P_t(i \to j)$, $i,j \in \{1,2,\ldots,m\}$ be the transition function for the hidden-state Markov chain X and

$$S_t^n = \begin{bmatrix}P_t(1\to 1) & P_t(2\to 1) & \cdots & P_t(m\to 1)\\ P_t(1\to 2) & P_t(2\to 2) & \cdots & P_t(m\to 2)\\ \vdots & \vdots & \ddots & \vdots\\ P_t(1\to m) & P_t(2\to m) & \cdots & P_t(m\to m)\end{bmatrix} * \begin{bmatrix}e^{t(\overline{\gamma}_{Y_{t_{n-1}}\to} - \gamma_{Y_{t_{n-1}}\to}(1))} & 0 & \cdots & 0\\ 0 & e^{t(\overline{\gamma}_{Y_{t_{n-1}}\to} - \gamma_{Y_{t_{n-1}}\to}(2))} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & e^{t(\overline{\gamma}_{Y_{t_{n-1}}\to} - \gamma_{Y_{t_{n-1}}\to}(m))}\end{bmatrix}. \tag{67}$$

Then, the solution is approximated recursively by: (i) weighted evolution

$$\begin{bmatrix}\sigma_t^1\\ \sigma_t^2\\ \vdots\\ \sigma_t^m\end{bmatrix} = \left[S^n_{\frac{t-t_{n-1}}{N}}\right]^N\begin{bmatrix}\sigma_{t_{n-1}}^1\\ \sigma_{t_{n-1}}^2\\ \vdots\\ \sigma_{t_{n-1}}^m\end{bmatrix} \tag{68}$$

for large N (Trotter product approximation) and all $t \in [t_{n-1}, t_n)$. (ii) observation update

$$\begin{bmatrix} \sigma^1_{t_n} \\ \sigma^2_{t_n} \\ \vdots \\ \sigma^m_{t_n} \end{bmatrix} = \begin{bmatrix} \frac{\gamma_{Y_{t_{n-1}} \to Y_{t_n}}(1)}{\overline{\gamma}_{Y_{t_{n-1}} \to Y_{t_n}}} \sigma^1_{t_n-} \\ \frac{\gamma_{Y_{t_{n-1}} \to Y_{t_n}}(2)}{\overline{\gamma}_{Y_{t_{n-1}} \to Y_{t_n}}} \sigma^2_{t_n-} \\ \vdots \\ \frac{\gamma_{Y_{t_{n-1}} \to Y_{t_n}}(m)}{\overline{\gamma}_{Y_{t_{n-1}} \to Y_{t_n}}} \sigma^m_{t_n-} \end{bmatrix}, \; n \geq 1; \quad \text{s.t.} \begin{bmatrix} \sigma^1_{t_0} \\ \sigma^2_{t_0} \\ \vdots \\ \sigma^m_{t_0} \end{bmatrix} = \begin{bmatrix} P(X_0 = 1) \\ P(X_0 = 2) \\ \vdots \\ P(X_0 = m) \end{bmatrix}. \tag{69}$$

The unnormalized filter is then $\sigma_t(\cdot) = \sum_{i=1}^{m} \sigma^i_t \delta_i(\cdot)$.

5. Conclusions and highlights

- We developed methods to change continuous-time Markov chains, add time dependence or even add hidden states. Importance sampling, rejection sampling, and Monte Carlo simulation methods were developed. Parameter estimation and model learning are future applications.
- We provided a Bayes' factor model selection approach to Continuous-time Hidden Markov Chains, both through particle and direct methods. Filtering equations for both the unnormalized and normalized filter were developed and solved.
- We introduced the Continuous-time Markov Observation Models (CMOM). and proved all our results in this more general context. All applications and solutions discussed apply to CMOM.
- Our explicit measure-change formula for Markov chains is new. Our introduction of the transition count processes into any measure change and filtering equations is believed to be new.
- Our branching particle filter approach to infectious disease monitoring is novel and extendable to a stochastic Susceptible–Infected–Recovered (SIR) hidden models with point-of-care observations. Our CMOM model allows the amount of testing to increase with the amount of infection.

6. Configuration

All simulations were implemented in Julia version 1.9.4 and executed on a Windows Subsystem for Linux (WSL) virtual machine on a Windows 11 PC equipped with an Intel 13th Gen i5-13400F processor and 16 GB of RAM.

Funding

NSERC Discovery Grant support is gratefully acknowledged.

CRediT authorship contribution statement

Michael A. Kouritzin: Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Kloek T., H.K. van Dijk, Bayesian estimates of equation system parameters: An application of integration by Monte Carlo, Econometrica 46 (1) (1978) 1–19, http://dx.doi.org/10.2307/1913641.
[2] H.K. van Dijk, T. Kloek, Experiments with some alternatives for simple importance sampling in Monte Carlo integration, in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith (Eds.), Bayesian Statistics, Vol. II, North Holland, Amsterdam, ISBN: 0-444-87746-0, 1984.
[3] V. Elvira, L. Martino, D. Luengo, M.F. Bugallo, Improving population Monte Carlo: Alternative weighting and resampling schemes, Signal Process. 131 (2017) 77–91, http://dx.doi.org/10.1016/j.sigpro.2016.07.012.
[4] L.E. Baum, T. Petrie, Statistical inference for probabilistic functions of finite state Markov chains, Ann. Math. Stat. 37 (6) (1966) 1554–1563, http://dx.doi.org/10.1214/aoms/1177699147.
[5] L.E. Baum, J.A. Eagon, An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology, Bull. Amer. Math. Soc. 73 (3) (1967) 360, http://dx.doi.org/10.1090/S0002-9904-1967-11751-8, Zbl 0157.11101.
[6] A. Petropoulos, S.P. Chatzis, S. Xanthopoulos, A novel corporate credit rating system based on student's-t hidden Markov models, Expert Syst. Appl. 53 (2016) 87–105, http://dx.doi.org/10.1016/j.eswa.2016.01.015.
[7] C. Nicolai, Solving ion channel kinetics with the QuB software, Biophys. Rev. Lett. 8 (3n04) (2013) 191–211, http://dx.doi.org/10.1142/S1793048013300053.
[8] J. Stigler, F. Ziegler, A. Gieseke, J.C.M. Gebhardt, M. Rief, The complex folding network of single calmodulin molecules, Science 334 (6055) (2011) 512–516, http://dx.doi.org/10.1126/science.1207598, Bibcode:2011Sci.334.512S.
[9] W. Wei, B. Wang, D. Towsley, Continuous-time hidden Markov models for network performance evaluation, Perform. Eval. 49 (2002) 129–146, http://dx.doi.org/10.1016/S0166-5316(02)00122-0.
[10] Y.Y. Liu, S. Li, F. Li, L. Song, Efficient learning of continuous-time hidden markov models for disease progression, Adv. Neural Inf. Process. Syst. 28 (2015) 3599–3607.
[11] N. Ethier Stewart, G. Kurtz Thomas, Markov Processes, Characterization and Convergence, Wiley, New York, 1986.

[12] M. Zakai, On the optimal filtering of diffusion processes, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 11 (1969) 230–243.
[13] M.A. Kouritzin, H. Long, On extending classical filtering equations, Statist. Probab. Lett. 78 (2008) 3195–3202, http://dx.doi.org/10.1016/j.spl.2008.06.005.
[14] M.A. Kouritzin, Y. Zeng, Bayesian model selection via filtering for a class of micro-movement models of asset price, Int. J. Theor. Appl. Finance 8 (2005) 97–121.
[15] D. Blount, M.A. Kouritzin, On convergence determining and separating classes of functions, Stochastic Process. Appl. 120 (2010) 1898–1907.
[16] M.A. Kouritzin, Y.-X. Ren, A strong law of large numbers for super-stable processes, Stochastic Process. Appl. 124 (2014) 505–521, http://dx.doi.org/10.1016/j.spa.2013.08.009.
[17] M.A. Kouritzin, Residual and stratified branching particle filters, Comput. Statist. Data Anal. 111 (2017) 145–165, http://dx.doi.org/10.1016/j.csda.2017.02.003.
[18] M. Fujisaki, G. Kallianpur, H. Kunita, Stochastic differential equations for the nonlinear filtering problem, Osaka J. Math. 9 (1972) 19–40.
[19] D. Crisan, M.A. Kouritzin, J. Xiong, Nonlinear filtering with signal dependent observation noise, Electron. J. Probab. 14 (2009) 1863–1883, http://dx.doi.org/10.1214/EJP.v14-687.
[20] Protter Philip, Stochastic Integration and Differential Equations: A New Approach, Springer, Berlin-Heidelberg-New York, 1990.
[21] X. Zhu, B. Gao, Y. Zhong, C. Gu, K.-S. Choi, Extended Kalman filter based on stochastic epidemiological model for COVID-19 modelling, Comput. Biol. Med. 137 (2021) http://dx.doi.org/10.1016/j.compbiomed.2021.104810, Paper 104810.
[22] M.G. Martínez, E. Pérez-Castro, Ramón Reyes-Carreto, R. Acosta-Pech, Spatial modeling in epidemiology, in: Recent Advances in Medical Statistics, IntechOpen, 2022, http://dx.doi.org/10.5772/intechopen.104693, Ch. 6.
[23] G. Davidson, A. Fahlman, E. Mereu, et al., A methodological approach for modeling the spread of disease using geographical discrete-event spatial models, Simulation 100 (1) (2024) 39–70, http://dx.doi.org/10.1177/00375497231152458.
[24] U. Picchini, A. Samson, Coupling stochastic EM and approximate Bayesian computation for parameter inference in state-space models, Comput. Statist. 33 (2018) 179–212, http://dx.doi.org/10.1007/s00180-017-0770-y.
[25] A.M.G. Cox, E. Horton, D. Villemonais, Binary branching processes with moran type interactions, Ann. Inst. Henri Poincaré Probab. Stat. (2024) in press.
[26] M.A. Kouritzin, A. Mackay, N. Vellone-Scott, New branching filters with explicit negative dependence, IEEE Access 8 (2020) 157306-157321, http://dx.doi.org/10.1109/ACCESS.2020.3019226.
[27] M.A. Kouritzin, Convergence rates for residual branching particle filters, J. Math. Anal. Appl. 449 (2017) 1053–1093, http://dx.doi.org/10.1016/j.jmaa.2016.12.046.
[28] J. Xiong, An Introduction to Stochastic Filtering Theory, in: Oxford Graduate Texts in Mathematics, vol. 18, Oxford University Press, 2008.
[29] N. Chopin, O. Papaspiliopoulos, An Introduction to Sequential Monte Carlo, Springer Nature, Switzerland AG, 2020, http://dx.doi.org/10.1007/978-3-030-47845-2.
[30] T.G. Kurtz, D.L. Ocone, Unique characterization of conditional distributions in nonlinear filtering, Ann. Probab. 16 (1988) 80–107.
[31] P. Del Moral, M.A. Kouritzin, L. Miclo, On a class of discrete generation interacting particle systems, Electron. J. Probab. 6 (2001) 26, Paper No. 16.
[32] N. Chopin, Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference, Ann. Statist. 32 (6) (2004) 2385–2411.
[33] V. Maroulas, A. Nebenführ, Tracking rapid intracellular movements: A Bayesian random set approach, Ann. Appl. Stat. 9 (2) (2015) 926–949, http://dx.doi.org/10.1214/15-AOAS819.
[34] P.J. Van Leeuwen, H.R. Künsch, L. Nerger, R. Potthast, S. Reich, Particle filters for high-dimensional geoscience applications: A review, Q. J. R. Meteorol. Soc. 145 (2019) 2335–2365, http://dx.doi.org/10.1002/qj.3551.
[35] J. Elfring, E. Torta, R. van de Molengraft, Particle filters: A hands-on tutorial, Sensors (Basel) 21 (2) (2021) 438, http://dx.doi.org/10.3390/s21020438.
[36] T.R. Kozel, A.R. Burnham-Marusich, Point-of-care testing for infectious diseases: Past, present, and future, J. Clin. Microbiol. 55 (8) (2017) 2313–2320, http://dx.doi.org/10.1128/JCM.00476-17.
[37] E.L. Ionides, K. Asfaw, J. Park, A.A. King, Bagged filters for partially observed interacting systems, J. Amer. Statist. Assoc. 118 (542) (2021) 1078–1089, http://dx.doi.org/10.1080/01621459.2021.1974867.