Statistical Inference for Stochastic Processes (2025) 28:8
https://doi.org/10.1007/s11203-025-09326-9

Guided simulation of conditioned chemical reaction networks

Marc Corstanje[1] · **Frank van der Meulen**[1]

Received: 7 December 2023 / Accepted: 13 April 2025

Abstract
Let X be a chemical reaction process, modeled as a multi-dimensional continuous-time jump process. Assume that at given times $0 < t_1 < \cdots < t_n$, linear combinations $v_i = L_i X(t_i)$, $i = 1, \ldots, n$ are observed for given matrices L_i. We show how the process that is conditioned on hitting the states $v_1, \ldots, v_n$ is obtained by a change of measure on the law of the unconditioned process. This results in an algorithm for obtaining weighted samples from the conditioned process. Our results are illustrated by numerical simulations.

Keywords Chemical reaction processes · Doob's h-transform · Exponential change of measure · Guided process

Mathematics Subject Classification 60J27 · 60J28 · 60J74

1 Introduction

Chemical reaction networks are used to study a wide class of biological, physical and chemical processes that evolve over time. For instance, one can think of the transcription of genes to mRNA and then the translation to protein, the kinetics of a virus or the dynamics of chemical components reacting with each other. The forward evolution of such processes can be described in different ways: *(i)* a system of ordinary differential equations, see e.g. Feinberg (2019), *(ii)* a system of stochastic differential equations, see e.g. Fuchs (2013) or *(iii)* continuous-time Markov jump processes, as in Anderson (2007). It is the third option that we consider in this paper.

1.1 Chemical reaction networks

Chemical reactions are described as linear combinations of chemical components merging into each other. Typically, one denotes a reaction in which components A and B are merged into C and D by

$$A + B \to C + D. \tag{1.1}$$

✉ Marc Corstanje
M.A.Corstanje@vu.nl

[1] Department of Mathematics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Published online: 17 May 2025

More generally, a *chemical reaction network* consists of:

- a *Species* set $\mathcal{S} = \{S_1, \ldots, S_d\}$ which consists of the chemical components whose counts we model;
- a *Reaction* set $\mathcal{R}$. A reaction $\ell \in \mathcal{R}$ is modeled as $\sum_k \nu_{k\ell} S_k \to \sum_k \nu'_{k\ell} S_k$. We characterize ℓ by the change of counts of the species $\xi_\ell = \left(\nu'_{k\ell} - \nu_{k\ell}\right)_{k=1}^d$.

For example, in (1.1), the species set is $\{A, B, C, D\}$ and ξ is given by $(-1, -1, 1, 1)$. Let $\mathbb{S} \subseteq \mathbb{Z}^d_{\geq 0}$ denote the state space of the reaction system. A vector in $\mathbb{S}$ is of the form $(x_k)_{k=1}^d$, where x_k denotes the species count of type S_k. We study a Markov process $X = (X(t))_{t\geq 0}$ on $\mathbb{S}$ that models the evolution of species counts over time. That is, at time t, the k-th element of $X(t)$ is given by $\#S_k(t)$. We assume that the initial state $X(0) = x_0$ of the process is known. At a given time t, the time until a reaction of type ℓ takes place is given by τ_ℓ. When reaction $\hat{\ell} = \underset{\ell\in\mathcal{R}}{\operatorname{argmin}}\, \tau_\ell$ occurs, the process jumps at time $t + \tau_{\hat{\ell}}$ to $X\left(t + \tau_{\hat{\ell}}\right) = X(t) + \xi_{\hat{\ell}}$. The reactions are assumed to occur according to an inhomogeneous Poisson process with intensity function that we refer to as the *reaction rate*. A more detailed description of the stochastic model is given in Sect. 2. The process X evolving on the chemical reaction network is referred to as the *chemical reaction process*.

1.2 Statistical problem

Suppose at fixed times $0 < t_1 < \cdots < t_n$, we observe $v_1, \ldots v_n$, where $v_k = L_k X(t_k)$ with $L_k \in \mathbb{R}^{m_k \times d}$ and $m_k \leq d, k = 1, \ldots n$. Not assuming L to be the identity matrix is for example important in applications where the measuring device cannot distinguish two or more species, so that only sums of their counts are observed. We will assume the rows of each L_k to be linearly independent. Typically reaction rates are unknown and we wish to infer those from the data. Suppose the reaction rate depend on a parameter vector θ. Likelihood-based inference for θ is hampered by the lack of closed-form expressions for the transition probabilities of X. However, if the process were observed continuously over time, the problem would be easier. Therefore, it is natural to employ a data-augmentation scheme where we iteratively sample X on $[0, t_n]$ conditional on $v_1, \ldots, v_n$ and θ and then update θ conditional on X. In this paper, we focus on the first step, sampling from $(X \mid L_k X(t_k) = v_k,\ k = 1, \ldots n)$. It is a key objective of this paper to show rigorously how this can be done efficiently. Note that a simple rejection sampling scheme where we discard all paths contradicting the observations is valid but very inefficient in most settings.

1.3 Approach: conditioning by guiding

Our approach builds on earlier work in Corstanje et al. (2023) for general Markov processes in case of a single observation. Let us highlight the main points. The law of the process X, conditioned to be in a given state at fixed times, is obtained through Doob's h-transform. That is, there is a function $h\colon [0, T] \times \mathbb{S} \to \mathbb{R}_+$ that depends on the transition probabilities of X which induces a change of measure. Under the new measure, $\mathbb{P}^h$, the process is conditioned to hit the observed states at times of observation. Since h is typically unknown, we replace it by a fully tractable function $g\colon [0, T] \times \mathbb{S} \to \mathbb{R}_+$ that itself induces a change of measure to a measure $\mathbb{P}^g$. Under certain conditions, $\mathbb{P}^h$ is absolutely continuous with respect to $\mathbb{P}^g$ and

$$\frac{\mathrm{d}\mathbb{P}^h}{\mathrm{d}\mathbb{P}^g}(X) = \frac{1}{h(0, x_0)} F(X). \tag{1.2}$$

Here, F is known in closed form, depends on g but does not depend on h. Weighted samples of the conditioned process can therefore be obtained by sampling under $\mathbb{P}^g$. The argument for making the above precise is not too hard in the case where g is bounded and bounded away from zero. However, some natural choices we discuss and have been proposed in the literature require a more delicate argument.

1.4 Related literature

Statistical inference for chemical reaction processes has received considerable attention over the past decade. In this section we summarise related work, while in the next section we highlight contributions of this paper. Rathinam and Yu (2021) consider the setting where one observes a subset of the species counts *continuously* over time and wants to filter the latent species counts. Reeves and Bhat (2022) parametrise model transition rates by neural networks, while assuming all trajectories are *fully continuously* observed. Parameter estimation is then done by gradient ascent to maximise the log likelihood.

In this paper, we consider *discrete-time* observations and therefore the works below are more closely related to our work. Warne et al. (2019) give an introduction to chemical reaction processes and consider estimation for discrete-time partial observations with Gaussian noise focussing on Approximate Bayesian Computation. Fearnhead (2008) and Golightly and Sherlock (2019) are probably closest related to our approach. The common starting points of these works is a slightly informal computation that reveals how the reaction rate of the chemical reaction processes changes upon conditioning the process on a future observation (we present this argument at the start of Sect. 3). While the reaction rate for the conditioned process is intractable, it can be approximated and this simply boils down to choosing g as above. Fearnhead (2008) approximates g using Euler discretisation of the Chemical Langevin Equation (CLE) assuming the process is fully observed without error. Golightly and Sherlock (2019) consider conditioning on a partial observation corrupted by Gaussian noise. Their choice of g is based on the linear noise approximation to the CLE. This is shown to outperform the approach of Fearnhead (2008) and earlier work in Golightly and Wilkinson (2015).

Georgoulas et al. (2017) construct an unbiased estimator for the likelihood using random truncations and computation of matrix exponentials. This in turn is used to exploit the pseudo-marginal MCMC algorithm (Andrieu and Roberts 2009) for parameter estimation. Building upon this work Sherlock and Golightly (2023) introduce the minimal extended statespace algorithm and the nearly minimal extended statespace algorithm to alleviate the problem of choosing a proposal distribution for the truncation level, as required in Georgoulas et al. (2017).

Alt and Koeppl (2023) consider the same setting as we do and derive approximations to the filtering and smoothing distributions using expectation propagation.

1.5 Contribution

We provide sufficient conditions on g such that (1.2) holds true. We extend the result in Corstanje et al. (2023) for a single complete observation to multiple partial observations in the context of chemical reaction processes. Moreover, we discuss a variation of the *next reaction* algorithm by Gillespie (1976) for sampling from a class of reaction networks with unbounded time-dependent reaction intensities.

The proposed methods fit within the framework of Backward Filtering Forward Guiding (van der Meulen and Schauer 2020), drawing strongly on techniques for exponential changes

of measure as outlined in Palmowski and Rolski (2002). This enables us to construct a *guided* process that at any time takes into account *all* future conditionings.

Compared to Fearnhead (2008) and Golightly and Sherlock (2019), we consider the setting of multiple future conditionings (rather than one), without imposing extrinsic noise on the observations. Moreover, we derive the conditioned process and likelihood ratio in (1.2) on path space. Sufficient conditions on g to guarantee absolute continuity are given in Theorem 3.4. It turns out that the choices for g in Fearnhead (2008) and Golightly and Sherlock (2019) satisfy the assumptions of this theorem. In numerical examples we show that flexibility in choosing g is particularly beneficial in cases where some of the components of the chemical reaction process have counts that vary monotonically over time.

1.6 Outline

We introduce stochastic chemical reaction processes in Sect. 2 and discuss examples that we will study. In Sect. 3, we describe the approach of conditioning chemical reaction processes by guiding and present conditions on g such that $\mathbb{P}^h$ is absolutely continuous with respect to $\mathbb{P}^g$. In Sect. 4 we consider various choices for g. Conditions for equivalence of $\mathbb{P}^h$ and $\mathbb{P}^g$ are discussed in Sect. 5. In Sect. 6 we present methods for simulation of conditioned chemical reaction processes together with numerical illustrations. We end with a discussion section. The appendix contains various proofs.

1.7 Frequently used notation

Throughout, we assume that we have an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For a stochastic process X, we use the notation $X^t = \{X(s) : s \leq t\}$. Given $L \in \mathbb{R}^{m\times d}$ with $m \leq d$ and $v \in \mathbb{R}^m$, we denote the inverse image of v under L by $L^{-1}v = \{x \in \mathbb{R}^d : Lx = v\}$. For functions $f_1(t,x)$ and $f_2(t,x)$ of time and space, we say that $f_1 \propto f_2$ if there exists a differentiable function κ of time such that for all t, x, $f_1(t,x) = \kappa(t) f_2(t,x)$. Derivatives with respect to a variable representing t, say $\partial/\partial t$ are denoted by ∂_t. We denote by

$$A_n = \{L_k X(t_k) = v_k,\ k = 1, \ldots, n\}, \tag{1.3}$$

a set of conditionings. If a measure μ is absolutely continuous with respect to ν, we write $\mu \ll \nu$.

2 Chemical reaction processes

We construct a stochastic process to model the dynamics of the chemical reaction network described in Sect. 1.1 following Chapter 1 of Anderson and Kurtz (2015). Let X be a Markov process on $\mathbb{S}$ such that the i-th component of $X(t)$, $X_i(t)$, represents the frequency of species S_i at time t. A reaction ℓ is represented through a difference vector $\xi_\ell \in \mathbb{S}$ and an intensity $\lambda_\ell : [0,\infty) \times \mathbb{S} \to [0,\infty)$. We assume

$$\mathbb{P}\left(X(t+\Delta) - X(t) = \xi_\ell \mid \mathcal{F}_t^X\right) = \lambda_\ell(t, X(t))\Delta + o(\Delta), \qquad \Delta \downarrow 0, \tag{2.1}$$

where $\mathcal{F}_t^X = \sigma(X^t)$. The jump probabilities specified in (2.1) correspond to a process with jumps $(\xi_\ell)_{\ell\in\mathcal{R}}$ and jump rate functions $(\lambda_\ell)_{\ell\in\mathcal{R}}$. Throughout, we impose the following assumptions on the network.

Assumption 2.1 $(\lambda_\ell)_{\ell\in\mathcal{R}}$ and $(\xi_\ell)_{\ell\in\mathcal{R}}$ are such that

(2.1a) $\lambda_\ell(t,x) \geq 0$ for all $t \geq 0$, $x \in \mathbb{S}$ and $\ell \in \mathcal{R}$.
(2.1b) For all $\ell \in \mathcal{R}$, $\xi_\ell \in \mathbb{Z}^d$ is such $\lambda_\ell(t,x) > 0$ implies $x + \xi_\ell \in \mathbb{S}$ for all $t \geq 0$ and $x \in \mathbb{S}$.
(2.1c) For all $t \geq 0$ and $x \in \mathbb{S}$, $\int_0^t \sum_{\ell\in\mathcal{R}} \lambda_\ell(s,x)\,\mathrm{d}s < \infty$.

For a stochastic process $X(t)$, let $T_K = \inf\{t : |X(t)| \geq K\}$ and $T_\infty = \lim_{K\to\infty} T_K$. Let X be the jump process with jumps $(\xi_\ell)_{\ell\in\mathcal{R}}$ and jump rate functions $(\lambda_\ell)_{\ell\in\mathcal{R}}$ satisfying

$$X(t) = x_0 + \sum_{\ell\in\mathcal{R}} \xi_\ell Y_\ell\left(\int_0^t \lambda_\ell(s, X(s))\,\mathrm{d}s\right), \qquad 0 \leq t < T_\infty, \tag{2.2}$$

where $(Y_\ell)_{\ell\in\mathcal{R}}$ are independent, unit rate Poisson processes. We assume the process to be non-explosive: $\mathbb{P}(T_\infty < \infty) = 0$. For $f : [0,T]\times\mathbb{S} \to \mathbb{R}$ with $f(t,\cdot)$ finitely supported for every t define

$$(\mathcal{L}f)(t,x) = \sum_{\ell\in\mathcal{R}} \lambda_\ell(t,x)\,[f(t,x+\xi_\ell) - f(t,x)], \tag{2.3}$$

By Theorem 1.22 in Anderson and Kurtz (2015), there exists a filtration $(\mathcal{F}_t)_{t\geq 0}$ such that for all such f

$$D^f(t) := f(t,X(t)) - \int_0^t (\partial_s + \mathcal{L})f(s,X(s))\,\mathrm{d}s$$

is an $(\mathcal{F}_t)$-martingale. That is, X is the unique solution to the martingale problem for $\mathcal{A} := \partial_t + \mathcal{L}$.

2.1 Distribution of reaction times

To the ℓ-th reaction, we associate a reaction time τ_ℓ, with distribution specified by

$$\mathbb{P}(\tau_\ell > \Delta \mid X(t) = x) = \exp\left(-\int_t^{t+\Delta} \lambda_\ell(s,x)\,\mathrm{d}s\right), \qquad t \geq 0, \Delta > 0, x \in \mathbb{S}. \tag{2.4}$$

If λ_ℓ is constant in time, it follows from (2.4) that $\tau_\ell \mid X(s) = x \sim \mathrm{Exp}(\lambda_\ell(x))$. This implies that the time $\tau = \min_{\ell\in\mathcal{R}} \tau_\ell$ that the first reaction occurs satisfies $\tau \mid X(s) = x \sim \mathrm{Exp}\left(\sum_{\ell\in\mathcal{R}} \lambda_\ell(x)\right)$.

If the reaction times are inhomogeneous in time and Assumption 2.1 is satisfied, similarly the time for the first reaction to occur has distribution (2.4) with intensity function $\sum_{\ell\in\mathcal{R}} \lambda_\ell$. Moreover, the probability distribution of the first reaction in any subset $R \subseteq \mathcal{R}$ also satisfies (2.4) with intensity function $\sum_{\ell\in R} \lambda_\ell$.

2.2 Chemical master equation and chemical Langevin equation

Denote the transition probabilities of (2.2) by p. That is $p(s,x;t,y) = \mathbb{P}(X(t) = y \mid X(s) = x)$. The Kolmogorov forward equation yields the *chemical master equation* given by

$$\partial_t p(s,x;t,y) = \sum_{\ell\in\mathcal{R}} \lambda_\ell(t, y-\xi_\ell)p(s,x;t,y-\xi_\ell) - \sum_{\ell\in\mathcal{R}} \lambda_\ell(t,y)p(s,x;t,y),$$

with initial condition $p(s,x;s,y) = 1\{y = x\}$. It is well-known, see e.g. Li (2020) that the chemical reaction process can be approximated by solutions to the *Chemical Langevin Equation (CLE)*, which is the SDE

$$\mathrm{d}Y(t) = b_{\mathrm{CLE}}(t, Y(t))\,\mathrm{d}t + \sigma_{\mathrm{CLE}}(t, Y(t))\,\mathrm{d}W_\ell(t), \qquad Y(0) = x_0 \tag{2.5}$$

where (assuming the reactions to be numbered $1, \ldots, B$)

$$\begin{aligned} b_{\mathrm{CLE}}(t,x) &= \sum_{\ell=1}^{B} \lambda_\ell(t,x)\xi_\ell \\ \sigma_{\mathrm{CLE}}(t,x) &= \left[\xi_1 \ \ldots \ \xi_B\right] \sqrt{\mathrm{diag}(\lambda_1(t,x), \ldots, \lambda_B(t,x))} \end{aligned} \tag{2.6}$$

and W is an independent $\mathbb{R}^B$-valued Brownian motion.

2.3 Examples

Example 2.2 (Pure death process) Our simplest example models a population of initial size x_0 in which an individual dies in a time interval $(t, t+\Delta)$ with probability $c\Delta + o(\Delta)$ for some constant $c > 0$ and Δ small. Such a process is modelled as chemical reaction process with just one specie and one reaction, namely $\lambda_1 \colon (t,x) \mapsto cx$ with $\xi_1 = -1$.

Example 2.3 (Gene Transcription and Translation *(GTT)*) A stochastic model for the process in which information is encoded in DNA and transferred to mRNA is described in Section 2.1.1 of Anderson and Kurtz (2015). The basic model considers the three species *Gene (G)*, *mRNA (M)* and *Protein (P)* in the set $\mathcal{S} = \{G, M, P\}$. We consider four reactions.

(1) **Transcription:** $G \to G + M$ with rate constant κ_1.
(2) **Translation:** $M \to M + P$ with rate constant $\kappa_2 > 0$.
(3) **Degradation of mRNA:** $M \to \emptyset$ with rate constant $d_M > 0$.
(4) **Degradation of protein:** $P \to \emptyset$ with rate constant $d_P > 0$.

Let $X(t)$ be the count vector at time t of species counts (G, M, P). In this example (2.2) translates to

$$\begin{aligned} X(t) = x_0 &+ Y_1\left(\int_0^t \kappa_1 X_1(s)\,\mathrm{d}s\right)\begin{pmatrix}0\\1\\0\end{pmatrix} + Y_2\left(\int_0^t \kappa_2 X_2(s)\,\mathrm{d}s\right)\begin{pmatrix}0\\0\\1\end{pmatrix} \\ &+ Y_3\left(\int_0^t d_M X_2(s)\,\mathrm{d}s\right)\begin{pmatrix}0\\-1\\0\end{pmatrix} + Y_4\left(\int_0^t d_P X_3(s)\,\mathrm{d}s\right)\begin{pmatrix}0\\0\\-1\end{pmatrix}, \end{aligned} \tag{2.7}$$

where Y_1, Y_2, Y_3, Y_4 are independent unit rate Poisson processes. A realization of this process can be found in Fig. 1.

Example 2.4 (Enzyme kinetics) The standard model for describing enzyme kinetics, see e.g. Bersani et al. (2008), where a substrate binds an enzyme reversibly to form an enzyme-substrate complex, which can in turn deteriorate into an enzyme and a product. We thus model *Substrate (S)*, *Enzyme (E)*, *Enzyme-substrate (SE)* and *Product (P)* in the species set $\mathcal{S} = \{S, E, SE, P\}$ and consider of the following reactions.

$$S + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} SE \xrightarrow{\kappa_3} P + E \tag{2.8}$$

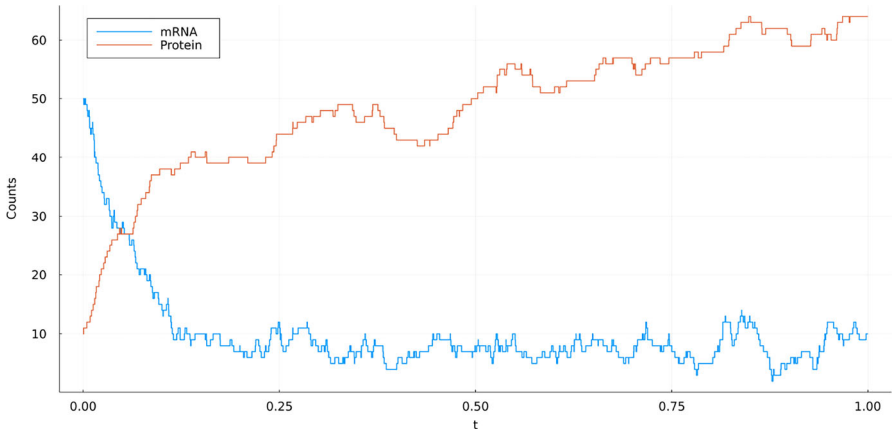

Fig. 1 Realization of (2.7) using $\kappa_1 = 200$, $\kappa_2 = 10$, $d_M = 25$, $d_P = 1$ and initial position $x_0 = (1, 50, 10)$. Note that the gene count is constant in this process. Therefore it was omitted from the figure

Equivalently:

(1) $S + E \to SE$ with rate constant κ_1.
(2) $SE \to S + E$ with rate constant κ_2.
(3) $SE \to P + E$ with rate constant κ_3.

Then the reaction rates corresponding to the above listed three reactions are given by

(1) $\lambda_1(x) = \kappa_1 x_1 x_2$ and $\xi_1 = (-1, -1, 1, 0)$.
(2) $\lambda_2(x) = \kappa_2 x_3$ and $\xi_2 = (1, 1, -1, 0)$.
(3) $\lambda_3(x) = \kappa_3 x_3$ and $\xi_3 = (0, 1, -1, 1)$.

Here (x_1, x_2, x_3, x_4) refer to species counts of (S, E, SE, P). Interesting aspects of this example are firstly that the fourth component (P) only appears in reaction (3) where 1 is added and therefore is monotonically increasing and secondly that there are absorbing states such as $x = (0, x_2, 0, x_4)$ where the process is killed as all reaction rates are zero. A realization of this process can be found in Fig. 1.

3 Guided Markov processes

3.1 First ideas

We first provide some intuition on how the dynamics of a chemical reaction network change upon conditioning on a future event. Let $\mathcal{E}$ denote some event later than time t, for example $\{X(T) = v\}$. For $t \in [0, T)$, $\Delta > 0$ and $x \in \mathbb{S}$ and $y = x + \xi_\ell$,

$$\begin{aligned}\mathbb{P}\left(X(t+\Delta) = y \mid X(t) = x, \mathcal{E}\right) &= \frac{\mathbb{P}\left(X(t+\Delta) = y,\ \mathcal{E} \mid X(t) = x\right)}{\mathbb{P}\left(\mathcal{E} \mid X(t) = x\right)} \\ &= \frac{\mathbb{P}\left(\mathcal{E} \mid X(t+\Delta) = y,\ X(t) = x\right)\mathbb{P}\left(X(t+\Delta) = y \mid X(t) = x\right)}{\mathbb{P}\left(\mathcal{E} \mid X(t) = x\right)} \\ &= \mathbb{P}\left(X(t+\Delta) = y \mid X(t) = x\right)\frac{\mathbb{P}\left(\mathcal{E} \mid X(t+\Delta) = y\right)}{\mathbb{P}\left(\mathcal{E} \mid X(t) = x\right)}\end{aligned}$$

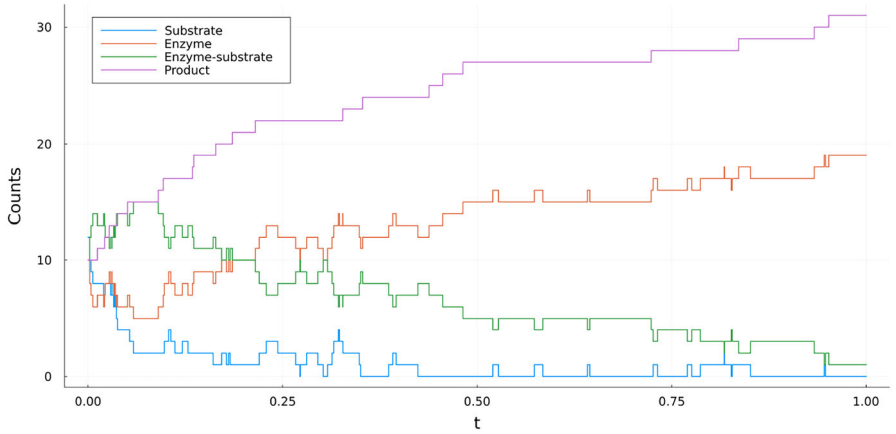

Fig. 2 Realization of the forward process of Example 2.4 using $\kappa_1 = 5$, $\kappa_2 = 5$, $\kappa_3 = 3$ and initial position $x_0 = (12, 10, 10, 10)$

Taking the limit $\Delta \downarrow 0$, this suggests that if X is conditioned on the event $\mathcal{E}$ then it is a chemical reaction process with adjusted intensities

$$\lambda_\ell^h(t, x) = \lambda_\ell(t, x) \frac{h(t, x + \xi_\ell)}{h(t, x)}, \tag{3.1}$$

where $h(t, x) = \mathbb{P}(\mathcal{E} \mid X(t) = x)$. The process with intensities $\lambda_\ell^h(t, x)$ has generator

$$\mathcal{L}_t^h f(x) = \sum_{\ell \in \mathcal{R}} \lambda_\ell^h(t, x) \left(f(x + \xi_\ell) - f(x) \right). \tag{3.2}$$

To sample the conditioned process, the function h is required which is rarely available in closed form. The general idea behind the construction of what we call a guided chemical reaction process is to replace h by a suitable tractable substitute g. This gives a process with rates λ_t^g. Discrepancies between the true conditioned process and the guided process can be accounted for by evaluating the likelihood ratio of their measures on path space. In this section we will establish sufficient conditions on g for this approach to be valid.

As shown below, conditioning the process corresponds to a change of measure. To see this connection, a direct computation shows that $\mathcal{L}_t^h f$ as defined in (3.2) can be expressed in terms of $\mathcal{L}_t$:

$$\mathcal{L}_t^h f(x) = \frac{1}{h(t, x)} \left[\mathcal{L}_t(fh)(t, x) - f(x) \mathcal{L}_t h(t, x) \right]. \tag{3.3}$$

This operator is strongly connected to exponential changes of measure studied in Palmowski and Rolski (2002) and Corstanje et al. (2023) for a wider class of Markov processes.

3.2 Guiding by a change of measure

We assume x_0 to be known and that $\mathcal{E} = \{LX(T) = v\}$ is observed for some known $v \in \mathbb{R}^m$ and a matrix $L \in \mathbb{R}^{m \times d}$ of full row rank with $m \leq d$. In case $m = d$, we assume without loss of generality that $L = I$. Additionally, we assume $\mathbb{P}(LX(T) = v \mid X(0) = x_0) > 0$. The extension to multiple observations will follow in a straightforward way in later sections. Let

$\mathcal{A}$ denote the infinitesimal generator of the space-time process $(t, X(t))$. That is $(\mathcal{A}g)(s,x) = \lim_{\Delta \downarrow 0} \Delta^{-1}\mathbb{E}\left[g(s+\Delta, X(s+\Delta)) - g(s, X(s)) \mid X(s) = x\right]$, for those g (that map to $\mathbb{R}$) for which the limit exists. Such g are said to be in the domain of the generator, denoted by $\mathcal{D}(\mathcal{A})$. While $\mathcal{D}(\mathcal{A})$ is implicitly defined, as part of the definition of $\mathcal{A}$, we can find a more explicit expression on a smaller set of functions. For those functions $\mathcal{A} = \partial_t + \mathcal{L}$, where $\mathcal{L}$ is defined in (2.3). From this, one can see that if there are finitely many reactions, bounded functions that are differentiable with respect to their time-argument are in the domain. For suitable $g\colon [0,T) \times \mathbb{S} \to \mathbb{R}$ define

$$E^g(t) = \frac{g(t, X(t))}{g(0, x_0)} \exp\left(-\int_0^t \frac{\mathcal{A}g}{g}(s, X(s))\,\mathrm{d}s\right), \qquad t \in [0,T). \tag{3.4}$$

Definition 3.1 We write $g \in \mathcal{G}$ if $g \in \mathcal{D}(\mathcal{A})$ is a strictly positive function and there exists a filtration $(\mathcal{F}_t,\ t \geq 0)$ such that $(E^g(t),\ t \in [0,T))$ is a martingale adapted to $(\mathcal{F}_t,\ t \geq 0)$.

As $\mathcal{F}_t^X = \sigma(X^t)$ is a sub-σ-algebra of $\mathcal{F}_t$ for all $t \geq 0$, $(E^g(t),\ t \in [0,T))$ is a martingale with respect to $(\mathcal{F}_t^X,\ t \geq 0)$ as well whenever $g \in \mathcal{G}$. In Lemma 3.3, we give an explicit condition under which $g \in \mathcal{G}$.

Let $g \in \mathcal{G}$ and let $\mathbb{P}_t$ denote the law of the process X restricted to $\mathcal{F}_t^X$ for $t \in [0,T)$. Since $(E^g(t),\ t \in [0,T))$ is a martingale with respect to $(\mathcal{F}_t^X,\ t \geq 0)$, we can define a new family of consistent probability measures measures $\{\mathbb{P}_t^g : t \in [0,T)\}$ by

$$\mathrm{d}\mathbb{P}_t^g = E^g(t)\,\mathrm{d}\mathbb{P}_t, \qquad t \in [0,T). \tag{3.5}$$

Since $\{\mathbb{P}_t^g : t \in [0,T)\}$ is consistent, there exists a measure $\mathbb{P}_T^g$ such that for all t, we have $\mathbb{P}_T^g\big|_{\mathcal{F}_t^X} = \mathbb{P}_t^g$ and, by Theorem 4.2 of Palmowski and Rolski (2002), the process $t \mapsto f(t, X(t)) - \int_0^t \left(\partial_t f + \mathcal{L}_t^g f\right)(s, X(s))\,\mathrm{d}s$ is a martingale under $\mathbb{P}_T^g$ for any function $f\colon [0,T) \times \mathbb{S} \to \mathbb{R}$ finitely supported in x and differentiable in t. Hence, under $\mathbb{P}_T^g$, X is a chemical reaction process with the same jumps as under $\mathbb{P}$ but with adjusted intensities $\lambda_\ell^g(t,x) = \lambda(t,x)g(t, x+\xi_\ell)/g(t,x)$. For the reader's convenience we have summarised some of the main arguments of Palmowski and Rolski (2002) for establishing this connection in Appendix D.

Definition 3.2 The *guided* process induced by g on $[0,t]$ is defined as the process X under $\mathbb{P}_T^g$.

Some authors refer to this process as the *twisted* process (see e.g. Moral (2017)). Upon denoting the transition probabilities of X by p, i.e. $\mathbb{P}(X(T) \in A \mid X(t) = x) = \sum_{y \in A} p(t,x;T,y)$ for $A \subseteq \mathbb{S}$, it follows from Example 2.4 of Corstanje et al. (2023) that the conditioned process $(X \mid LX(T) = v)$ is obtained by taking $g = h$ with

$$h(t,x) = \sum_{\zeta \in L^{-1}v} p(t,x;T,\zeta), \tag{3.6}$$

where $L^{-1}v = \{y \in \mathbb{S} : Ly = v\}$. We recall our assumption $h(0,x_0) > 0$, as otherwise conditioned paths have probability zero. Note that h is bounded and satisfies Kolmogorov's backward equation: $\mathcal{A}h = 0$. By Proposition 3.2 in Palmowski and Rolski (2002), $h \in \mathcal{G}$ and thus we can define the probability measure $\mathbb{P}_t^h$ by

$$\mathrm{d}\mathbb{P}_t^h = \frac{h(t, X(t))}{h(0,x_0)}\,\mathrm{d}\mathbb{P}_t, \qquad t \leq T.$$

Intuitively, $\mathbb{P}_t^h$ gives more mass to paths where $h(t, X(t))$ is large.

Unfortunately, h is intractable as the transition probabilities p are only known in closed-form in very specific cases including Example 2.2. To resolve this, consider another function $g \in \mathcal{G}$ that acts as a tractable substitute for h. Then, as a consequence, the process X is tractable and can be simulated under the measure $\mathbb{P}_T^g$. Moreover, for $t < T$,

$$\frac{\mathrm{d}\mathbb{P}_t^h}{\mathrm{d}\mathbb{P}_t^g}(X) = \frac{E_t^h}{E_t^g} = \frac{h(t, X(t))}{g(t, X(t))} \frac{g(0, x_0)}{h(0, x_0)} \Psi_t^g(X), \tag{3.7}$$

where

$$\Psi_t^g(X) = \exp\left(\int_0^t \frac{\mathcal{A}g}{g}(s, X(s))\, \mathrm{d}s \right). \tag{3.8}$$

To evaluate Ψ^g, a direct computation yields

$$\frac{\mathcal{A}g}{g}(s, x) = \partial_s \log g(s, x) + \sum_{\ell \in \mathcal{R}} \left(\lambda_\ell^g(s, x) - \lambda_\ell(s, x) \right). \tag{3.9}$$

Eq. 3.7 becomes particularly useful if it can be evaluated in $t = T$ as well. As $h(T, x) = \mathbf{1}_{\{Lx = v\}}$, this only depends on the choice for g. We present a class of functions that can subsequently be used for this purpose in Theorem 3.4.

Lemma 3.3 *Suppose $g \in \mathcal{D}(\mathcal{A})$ is a strictly positive function such that for some positive constant C*

$$\int_0^T \sum_{\ell \in \mathcal{R}} \lambda_\ell(s, X(s)) \left(\frac{g(s, X(s) + \xi_\ell)}{g(s, X(s))} - 1 \right)^2 \mathrm{d}s < C,$$

$\mathbb{P}$-almost surely, then $g \in \mathcal{G}$.

Proof The proof is inspired by the proof given in Example 15.2.10 in Brémaud (2020) and Lemma 19.6 in Liptser and Shiryaev (2013).

It suffices to show that $\mathbb{E}\left[E^g(T)\right] = 1$. By Lemma 3.1 of Palmowski and Rolski (2002), E^g is a local martingale. Let $\{\sigma_n\}_n$ be a localizing sequence for E^g. Then, if $(E^g(t),\ t \in [0, T])$ were uniformly integrable, then

$$\mathbb{E}\left[E^g(T)\right] = \lim_{n \to \infty} \mathbb{E}\left[E^g(T \wedge \sigma_n)\right] = \mathbb{E}\left[E^g(0)\right] = 1$$

We proceed to show uniform integrability. Given $t \in [0, T]$, suppose a trajectory X has jumps at times $t_1, \ldots, t_N$. Set $t_0 = 0$ and $t_{N+1} = t$. Then

$$\begin{aligned}
E^g(t) &= \frac{g(t, X(t))}{g(0, x_0)} \exp\left(- \sum_{j=0}^{N} \int_{t_j}^{t_{j+1}} \partial_s \log g(s, X(t_j))\, \mathrm{d}s - \int_0^t \frac{\mathcal{L}g(s, X(s))}{g(s, X(s))}\, \mathrm{d}s \right) \\
&= \frac{g(t, X(t))}{g(0, x_0)} \prod_{j=0}^{N} \frac{g(t_j, X(t_j))}{g(t_{j+1}, X(t_j))} \exp\left(- \int_0^t \frac{\mathcal{L}g(s, X(s))}{g(s, X(s))}\, \mathrm{d}s \right) \\
&= \prod_{j=0}^{N-1} \frac{g(t_{j+1}, X(t_{j+1}))}{g(t_{j+1}, X(t_j))} \exp\left(- \int_0^t \frac{\mathcal{L}g(s, X(s))}{g(s, X(s))}\, \mathrm{d}s \right) \\
&= \prod_{j=0}^{N-1} \frac{g(t_{j+1}, X(t_{j+1}))}{g(t_{j+1}, X(t_j))} \exp\left(- \int_0^t \sum_{\ell \in \mathcal{R}} \lambda_\ell(s, X(s)) \left(\frac{g(s, X(s) + \xi_\ell)}{g(s, X(s))} - 1 \right) \mathrm{d}s \right)
\end{aligned} \tag{3.10}$$

From (3.10), it is easy to verify that

$$E^g(t)^2 = E^{g^2}(t) \exp\left(\int_0^t \sum_{\ell \in \mathcal{R}} \lambda_\ell(s, X(s)) \left(\frac{g(s, X(s) + \xi_\ell)}{g(s, X(s))} - 1 \right)^2 \mathrm{d}s \right)$$

By Lemma 3.1 of Palmowski and Rolski (2002), E^{g^2} is a local martingale and, since it is bounded from below, it is a supermartingale with $\mathbb{E}\left[E^{g^2}(t)\right] \leq 1$. Hence, for all n,

$$\begin{aligned} &\mathbb{E}\left[E^g(T \wedge \sigma_n)^2\right] \\ &\leq \mathbb{E}\left[E^{g^2}(T \wedge \sigma_n) \exp\left(\int_0^T \sum_{\ell \in \mathcal{R}} \lambda_\ell(s, X(s)) \left(\frac{g(s, X(s) + \xi_\ell)}{g(s, X(s))} - 1 \right)^2 \mathrm{d}s \right) \right] \leq e^C \end{aligned}$$

Therefore $\sup_n \mathbb{E}\left[E^g(T \wedge \sigma_n)^2\right] < \infty$. □

Theorem 3.4 *Define h is as in (3.6) and let $g\colon [0, T] \times \mathbb{S} \to \mathbb{R}$ be such that the condition of Lemma 3.3 is satisfied. Then*

$$\frac{\mathrm{d}\mathbb{P}_T^h}{\mathrm{d}\mathbb{P}_T^g} = \frac{g(0, x_0)}{h(0, x_0)} \frac{\Psi_T^g(X)}{g(T, X(T))} \mathbf{1}_{\{LX(T) = v\}}. \tag{3.11}$$

Proof The form of the Radon-Nikodym derivative follows from (3.7) upon noting that for $t \uparrow T$,

$$h(t, X(t)) = \mathbb{E}\left[\mathbf{1}\{LX(T) = v\} \mid \mathcal{F}_t\right] \to \mathbb{E}\left[\mathbf{1}\{LX(T) = v\} \mid \mathcal{F}_T\right] = \mathbf{1}\{LX(T) = v\}.$$

□

The following proposition is often helpful in establishing that the condition of Lemma 3.3 is satisfied for a given g.

Proposition 3.5 *Suppose $g\colon [0, T] \times \mathbb{S} \to \mathbb{R}$ is such that for all x the map $t \mapsto g(t, x)$ is bounded from above and bounded away from zero. Then the condition of Lemma 3.3 is satisfied.*

Proof Since we assume throughout X to be nonexplosive, we have that almost surely the maps $t \mapsto g(t, X(t) + \xi_\ell)/g(t, X(t))$ are almost surely bounded on $[0, T]$ for all $\ell \in \mathcal{R}$. The result now follows upon noting that $\int_0^T \sum_{\ell \in \mathcal{R}} \lambda_\ell(s, X(s))\, \mathrm{d}s < \infty$ by Assumption (2.1c). □

Corollary 3.6 *The likelihood of x_0 based on the observation v is given by*

$$h(0, x_0) = g(0, x_0) \mathbb{E}_T^g \left[\frac{\Psi_T^g(X)}{g(T, X(T))} \mathbf{1}_{\{LX(T) = v\}} \right]. \tag{3.12}$$

Proof This follows immediately upon integrating (3.11) with respect to $\mathbb{P}_T^g$. □

Note that $g \equiv 1$ satisfies the condition in Lemma 3.3. This choice simply yields the original forward process. Sampling the conditioned process in this way is however very inefficient when $\mathbb{P}(LX(T) = v)$ is low.

The lemma below shows that the law of the guided process does not change when g is multiplied by a function only depending on time.

Lemma 3.7 *(Invariance under time scaling) Suppose $g \in \mathcal{G}$ and let $c\colon [0,T] \to \mathbb{R}_+$ be a differentiable function. Then $E^{cg}(t) = E^g(t)$.*

Proof Since c only depends on time

$$\begin{aligned}\int_0^t \frac{\mathcal{A}(cg)}{cg}(s, X(s))\,\mathrm{d}s &= \int_0^t \left[\partial_s \log\left(c(s)g(s, X(s))\right) + \frac{\mathcal{L}_s(cg)}{cg}(s, X(s))\right]\mathrm{d}s \\ &= \int_0^t \partial_s \log c(s)\,\mathrm{d}s + \int_0^t \left[\partial_s \log g(s, X(s)) + \frac{\mathcal{L}_s g}{g}(s, X(s))\right]\mathrm{d}s. \\ &= \log c(t) - \log c(0) + \int_0^t \frac{\mathcal{A}g}{g}(s, X(s))\,\mathrm{d}s.\end{aligned}$$

From this, we get $E^{cg}(t)$ by negating the right-hand-side, taking the exponent and subsequently multiplying by $c(t)g(t, X(t))/(c(0)g(0, x_0))$. It is easily seen that the terms with c cancel out. □

The following proposition is useful for numerical evaluation of the likelihood of a sampled guided process.

Proposition 3.8 *Suppose $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$, $x_0, \ldots, x_N \in \mathbb{S}$ and*

$$x(t) = \sum_{j=0}^{N-1} x_j \mathbf{1}_{[t_j, t_{j+1})}(t).$$

Then, with $\alpha_j(s) := g(s, x_{j+1})/g(s, x_j)$

$$\begin{aligned}&\frac{g(0, x_0)}{g(T, x(T))} \exp\left(\int_0^T \frac{\mathcal{A}g}{g}(s, x(s))\,\mathrm{d}s\right) \\ &\quad = \left(\prod_{j=0}^{N-2} \frac{1}{\alpha_j(t_{j+1})}\right) \exp\left(\sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \sum_{\ell \in \mathcal{R}} \lambda_\ell(s, x_j)\left[\alpha_j(s) - 1\right]\mathrm{d}s\right)\end{aligned}$$

Proof This is a consequence of (3.10). □

4 Choices for *g*

It remains to specify the maps $(t, x) \mapsto g(t, x)$ satisfying the assumption of Lemma 3.3. Moreover, to be of any use, the event $\{LX(T) = v\}$ needs to get positive probability under $\mathbb{P}_T^g$.

In this section we show the following:

- Sect. 4.1: The choices for g in Fearnhead (2008) and Golightly and Sherlock (2019), both based on the chemical Langevin equation, lead to absolute continuity.
- Sect. 4.2: A guiding function based on the transition density of a scaled Brownian motion yields absolute continuity while being computationally very efficient. It satisfies $\mathbb{P}_T^g(LX(T) = v) > 0$. In subsection 4.2.1 we show how g can be extended to the case of multiple partial observations.
- Sect. 4.3: For processes with monotone components a specific choice for g can overcome certain problems arising in the aforementioned choices.

In the following, C denotes a positive definite matrix.

4.1 Choices based on the chemical Langevin equation

Consider the CLE given in (2.5). Set $a_{\mathrm{CLE}} = \sigma_{\mathrm{CLE}}\sigma_{\mathrm{CLE}}'$. Fearnhead (2008) proposed to choose g based on the Euler discretization of the CLE. This gives

$$g_{\mathrm{F}}(t,x) = \mathcal{N}\left(v;\, L(x + b_{\mathrm{CLE}}(t,x)(T-t)),\, La_{\mathrm{CLE}}(t,x)L'(T-t) + C\right), \tag{4.1}$$

As the assumption of Proposition 3.5 is satisfied we obtain $\mathbb{P}_T^h \ll \mathbb{P}_T^{g_{\mathrm{F}}}$.

Golightly and Sherlock (2019) propose to infer g by using the Linear Noise Approximation (LNA) to the CLE. Here, we discuss their specific choice called "LNA with restart". For notational convenience, write $b = b_{\mathrm{CLE}}$ and $\sigma = \sigma_{\mathrm{CLE}}$ and again set $a = \sigma\sigma'$. Given $t < T$ and $x \in \mathbb{S}$, we denote by $z_{T|(t,x)}$ and $V_{T|(t,x)}$ the solutions at time T of the system of ordinary differential equations

$$\begin{aligned}
\mathrm{d}z_{s|(t,x)} &= b(s, z_{s|(t,x)})\,\mathrm{d}s \\
\mathrm{d}V_{s|(t,x)} &= \left(V_{s|(t,x)}\left(J_b(s, z_{s|(t,x)})\right)' + J_b\left(s, z_{s|(t,x)}\right)V_{s|(t,x)} + a\left(s, z_{s|(t,x)}\right)\right)\mathrm{d}s
\end{aligned}$$

where $s \in [t,T]$, subject to the initial conditions are given by $z_{t|(t,x)} = x$ and $V_{t|(t,x)} = 0$. Here, J_b denotes the Jacobian matrix of b (having component (i,j) given by $\partial b_i/\partial z_j$. The guiding term then is given by

$$g_{\mathrm{LNAR}}(t,x) = \mathcal{N}\left(v;\, Lz_{T|(t,x)},\, LV_{T|(t,x)}L' + C\right). \tag{4.2}$$

Absolute continuity $\mathbb{P}_T^h \ll \mathbb{P}_T^{g_{\mathrm{LNAR}}}$ follows by the same argument used for g_{F}.

4.2 Choosing *g* using the transition density of a scaled Brownian motion

For fixed $\sigma \in \mathbb{R}^{d\times d}$ define the process $\tilde{X}$ by $\mathrm{d}\tilde{X}(t) = \sigma\,\mathrm{d}W(t)$, where W is a standard Brownian motion. Denote $a = \sigma\sigma'$ and assume that a is such that LaL' is strictly positive definite. We derive g from backward filtering the process $\tilde{X}$ with observation $V \mid \tilde{X}(T) \sim \mathcal{N}(L\tilde{X}(T), C)$. Let q denote the density of the $\mathcal{N}(0,C)$-distribution. The density of V, conditional on $\tilde{X}(t) = x$ is given by $u(t,x) := \int \tilde{p}(t,x;T,y)q(v-Ly)\,\mathrm{d}y$. It follows from the results in Mider et al. (2021) that $u(t,x) \propto g(t,x)$ where

$$g(t,x) = \exp\left(-\frac{1}{2}x'H(t)x + F(t)'x\right). \tag{4.3}$$

Here, for $t \le T$, $H(t)$ and $F(t)$ satisfy the ordinary differential equations

$$\begin{aligned}
\mathrm{d}H(t) &= H(t)aH(t)\,\mathrm{d}t, \qquad & H(T) &= L'C^{-1}L \\
\mathrm{d}F(t) &= H(t)aF(t)\,\mathrm{d}t, \qquad & F(T) &= L'C^{-1}v
\end{aligned}. \tag{4.4}$$

These differential equations can be solved in closed-form:

$$H(t) = z(t)H(T) \quad \text{and} \quad F(t) = z(t)F(T), \tag{4.5}$$

with $z(t) = (I + H(T)a(T-t))^{-1}$

Note that we have not yet specified σ. Depending on its choice, g may be radically different from h. Nevertheless, it can be used as a change of measure to condition paths of X on the event $\{LX(T) = v\}$.

Theorem 4.1 *Let g be defined by (4.3) and (4.4). Then $\mathbb{P}_T^h \ll \mathbb{P}_T^g$ with*

$$\frac{\mathrm{d}\mathbb{P}_T^h}{\mathrm{d}\mathbb{P}_T^g} = \frac{g(0,x_0)}{h(0,x_0)} \exp\left(\int_0^T \frac{\mathcal{A}g}{g}(s, X(s))\,\mathrm{d}s\right) \exp\left(-\frac{1}{2}v'C^{-1}v\right) \mathbf{1}_{\{LX(T)=v\}}. \tag{4.6}$$

Proof It follows from theorems 2.4 and 2.5 in Mider et al. (2021) that if we define

$$M(t) := \left(C + LaL'(T-t)\right)^{-1},$$

then $H(t) = L'M(t)L$ and $F(t) = L'M(t)v$ and therefore

$$g(t,x) \propto \exp\left(-\frac{1}{2}(v-Lx)'M(t)(v-Lx)\right). \tag{4.7}$$

The proportionality is up to a differentiable time-dependent function and hence does not affect the change of measure by Lemma 3.7. By Lemma C.1, g is bounded from above and bounded away from 0. Since g is smooth in t and well-defined in T, we clearly have continuous differentiability of the maps $t \mapsto g(t,x)$ for all $x \in \mathbb{S}$ on $[0,T]$. The result thus follows from Proposition 3.5 and Theorem 3.4. □

Theorem 4.1 is only of interest when the guided process has a positive probability of hitting v at time T. This is ensured by Theorem 4.2 for which the proof is deferred to Appendix A.

Theorem 4.2 $\mathbb{P}^g(LX(T) = v) > 0$

A convenient choice for the matrix C has computational advantages.

Definition 4.3 Let $\varepsilon > 0$ and set $C = \varepsilon LaL'$. The corresponding g in (4.3) is denoted by g_ε.

Lemma 4.4 *The induced measure $\mathbb{P}_T^{g_\varepsilon}$ is the law of a chemical reaction process with intensities $\lambda_\ell^{g_\varepsilon}(t,x) = \alpha_\ell^{g_\varepsilon}(t,x)\lambda_\ell(t,x)$, where*

$$\alpha_\ell^{g_\varepsilon}(t,x) = \exp\left(-\frac{d(v, L(x+\xi_\ell))^2 - d(v,Lx)^2}{2(\varepsilon + T - t)}\right), \tag{4.8}$$

with d the metric given by

$$d(x,y)^2 = (y-x)'(LaL')^{-1}(y-x). \tag{4.9}$$

Proof This result follows by substituting the expression for g in (4.7) into the definition of $\alpha_\ell^g(t,x)$ and rewriting in terms of the metric d. □

This implies that only one matrix inverse, $(LaL')^{-1}$, needs to be calculated. By Eq. 4.8, the intensity of the guided process either becomes very large or small depending on whether $L(x+\xi_\ell)$ is closer or further away from v than Lx with respect to the metric d, which has intuitive appeal. Also note that ε appears only in the denominator at $T+\varepsilon$, which implies that the choice $C = \varepsilon LaL'$ imposes the conditioning at at time $T+\varepsilon$ instead of T, thereby precluding explosive behaviour in (4.7) as $t \uparrow T$.

Remark 4.5 Ideally, we would like to have $\mathbb{P}^g(LX(T) = v) = 1$. For $g = g_\varepsilon$ as in Lemma 4.4 this is not guaranteed. Essentially, the choice $C = \varepsilon LaL'$ yields a process where $LX(t)$ hits v at time $T+\varepsilon$. Since the intensities are bounded from below and above, there is a positive probability of at least one reaction occurring in $(T, T+\varepsilon)$ resulting in $LX(T) \neq v$. In Sect. 5 we discuss the case where $\varepsilon = 0$, which does yield $\mathbb{P}^g(LX(T) = v) = 1$.

Remark 4.6 Comparing g_{F}, g_{LNAR} and g_ε, it is likely that g_ε will deviate most from h. In particular when the drift and diffusion coefficient of the CLE are nonlinear, g_ε will likely deviate from h the most leading to the guided paths deviating from the conditioned paths. However, from a computational point of view g_ε is most attractive, as can be seen from (4.8). Moreover, in Sect. 6 it turns out that simulation of the guided process is simplified in case of g_ε.

Remark 4.7 In dimension 1, we derive a better intuition for the choice for a from (4.8) and (4.9). A small choice leads to a higher guided intensity, and thus a process arriving around v quickly and likely to stay near v, while a large value of a leads to trajectories that remain unaffected by $\alpha_\ell^{g_\varepsilon}$ until t approaches T. Ideally, we mimic trajectories under $\mathbb{P}^h$ and thus a good choice of a ensures $g(0,x_0)\Psi_T^g(X)\mathbf{1}\{LX(T)=v\}/g(T,X(T))$ is large (with $\Psi_t^g(X)$ defined in Equation (3.8)). This can be verified using Monte-Carlo simulation sampling X under $\mathbb{P}^{g_\varepsilon}$. Alternatively, a simple choice for a is $a_{\mathrm{CLE}}(0,x_0)$ or, if we have a complete observation, $a = a_{\mathrm{CLE}}(T,v)$.

4.2.1 Extension to multiple observations

We now extend Theorem 4.1 to a result for multiple observations. Consider observations $v_i = L_i X(t_i)$, $i = 1,\dots,n$ where $0 = t_0 < t_1 < \cdots < t_n$ and assume without loss of generality that $L_i \in \mathbb{R}^{m_i\times d}$ are of full column rank with $m_i \le d$ and $L_i = I$ when $m_i = d$.

Proposition 4.8 *Let p denote the transition probabilities of X and define for $t \in [t_{k-1}, t_k)$ and $x \in \mathbb{S}$*

$$\begin{aligned} h(t,x) &= \mathbb{P}\left(L_i X(t_i) = v_i,\ i = k,\dots,n \mid X(t) = x\right) \\ &= \sum_{\zeta_k \in L_k^{-1}v_k} \cdots \sum_{\zeta_n \in L_n^{-1}v_n} p(t,x;t_k,\zeta_k) \prod_{i=k}^{n-1} p(t_i,\zeta_i;t_{i+1},\zeta_{i+1}). \end{aligned} \tag{4.10}$$

Then $h \in \mathcal{G}$ and the change of measure (3.5) induces $(X \mid L_k X(t_k) = v_k,\ k = 1,\dots,n)$.

Proof This result is obtained upon following Example 2.4 of Corstanje et al. (2023) using the delta-dirac distribution $\mu(\zeta_k) = \delta(v_k - L_k\zeta_k)$. □

We deduce the form of g from Mider et al. (2021) in a similar way compared to a single observation. We consider an auxiliary process $\tilde{X}$ that solves the SDE

$$\mathrm{d}\tilde{X}(t) = \sigma(t)\,\mathrm{d}W(t), \qquad X(0) = x_0,\ t \in [0,t_n), \tag{4.11}$$

where $\sigma(t) = \sum_{k=1}^n \sigma_k \mathbf{1}_{[t_{k-1},t_k)}(t)$ and $a_k = \sigma_k\sigma_k'$ are positive definite $d\times d$ matrices for $k = 1,\dots,n$.

For each observation k, we consider $V_k \mid X(t_k) \sim \mathcal{N}(0,C_k)$ where C_k is an $m_k \times m_k$ covariance matrix. Suppose q_k denotes the density of the $\mathcal{N}(0,C_k)$-distribution. Then the transition density of $\tilde{X}$ satisfies for $t \in [t_{k-1},t_k)$,

$$\begin{aligned} &\int \tilde{p}(t,x;t_k,\zeta_k) q_k(v_k - L_k\zeta_k) \prod_{i=k}^{n-1} \tilde{p}(t_i,\zeta_i;t_{i+1},\zeta_{i+1}) q_{i+1}(v_{i+1} - L_{i+1}\zeta_{i+1})\,\mathrm{d}\zeta_k \cdots \mathrm{d}\zeta_n \\ &\qquad \propto \exp\left(-\frac{1}{2}x'H(t)x + F(t)'x\right) =: g(t,x). \end{aligned} \tag{4.12}$$

Here, the expressions for H and F can be found by backward solving a system of equations given by

$$H(t_n) = L_n' C_n^{-1} L_n$$
$$F(t_n) = L_n' C_n^{-1} v_n$$

and for $t \in (t_{k-1}, t_k)$, $k = 1, \dots, n$

$$\begin{aligned} \mathrm{d}H(t) &= H(t) a_k H(t)\,\mathrm{d}t, \qquad & H(t_k) &= H_k := L_k' C_k^{-1} L_k + H(t_k+) \\ \mathrm{d}F(t) &= H(t) a_k F(t)\,\mathrm{d}t, \qquad & F(t_k) &= F_k := L_k' C_k^{-1} v_k + F(t_k+) \end{aligned} \tag{4.13}$$

Finally, we let H and F be a right continuous modification of the solution to (4.13), i.e. setting $H(t_k) = H(t_k+)$ and $F(t_k) = F(t_k+)$, $k = 1, \dots, n-1$. This system can be solved in closed form: $H(t) = z_k(t) H_k$, $F(t) = z_k(t) F_k$, where, for $k = 1, \dots, n$, $z_k(t) = (I + H_k a_k (t_k - t))^{-1}$.

Remark 4.9 Alternatively, following the computations in Section 2 of Mider et al. (2021), g can be expressed as

$$g(t, x) \propto \exp\left(-\frac{1}{2} \left(v(t) - L(t)x\right)' M(t) \left(v(t) - L(t)x\right) \right), \tag{4.14}$$

where, the proportionality is up to a time-dependent function and for $t \in [t_{k-1}, t_k)$

$$L(t) = \begin{pmatrix} L_k \\ \vdots \\ L_n \end{pmatrix} \qquad \text{and} \qquad v(t) = \begin{pmatrix} v_k \\ \vdots \\ v_n \end{pmatrix}$$

and $M(t) = M^\dagger(t)^{-1}$ with $M^\dagger(t)$ a block matrix with entries

$$M^\dagger(t) = \left(C_i \mathbf{1}\{i = j\} + \sum_{l=k}^{i \wedge j - 1} L_i a_{l+1} L_j' (t_{l+1} - t_l) + L_i a_k L_j' (t_k - t) \right)_{i,j=k}^{n},$$

$t \in [t_{k-1}, t_k)$. While this representation of g is useful in most proofs, it is computationally more demanding as the matrix dimensions of $H(t)$ and $F(t)$ are $d \times d$, while $M(t)$, $L(t)$ and $v(t)$ have dimensions that increase with the amount of observations.

Using Lemma C.2, the proof of Theorem 4.1 can be repeated for a result like Theorem 4.1 for each observation k to find that $\mathbb{P}^h_{t_n} \ll \mathbb{P}^g_{t_n}$ with

$$\frac{\mathrm{d}\mathbb{P}^h_{t_n}}{\mathrm{d}\mathbb{P}^g_{t_n}} = \frac{g(0, x_0)}{h(0, x_0)} \exp\left(\int_0^{t_n} \frac{\mathcal{A}g}{g}(s, X(s))\,\mathrm{d}s - \frac{1}{2} \sum_{k=1}^{n} v_k' C_k^{-1} v_k \right) \mathbf{1}_{A_n}. \tag{4.15}$$

The term $\exp\left(-\frac{1}{2} \sum_{k=1}^n v_k' C_k^{-1} v_k \right)$ results from evaluating each fraction $g(t_k, x)/g(t_k-, x)$ on the set $\{L_k x = v_k\}$. We can repeat Theorem 4.2 for each observation to obtain $\mathbb{P}^g(A_n) > 0$. However, similar to Remark 4.5, we generally do not have $\mathbb{P}^g(A_n) = 1$.

4.3 Choosing *g* to guide a process with monotone components

Theorem 4.1 shows that choosing g using the transition density of a scaled Brownian motion yields a process that is absolutely continuous with respect to the conditioned process. However, depending on the network, this might not always be a desirable choice.

Consider for instance Example 2.4. Here, the fourth component of the process (P) is monotonically increasing as it only appears in reaction 3, where 1 is added, which can also be seen in Fig. 2. Let us for simplicity assume a complete observation at time T, that is $X(T) = v_T$. Clearly, if $X_4(t) = v_{T,4}$ for some $t < T$, reaction 3 cannot occur anymore for the process to satisfy the conditioning. However, $\lambda_3^g(t, X(t)) \neq 0$ for the choices of g discussed so far. This can lead to trajectories that don't satisfy the conditioning, which in turn means a low acceptance ratio when sampling.

Alternatively, we choose an auxiliary process $\tilde{X} = (\tilde{Z}, \tilde{Y})$, where $\tilde{Z}$ is the $\mathbb{R}^3$-valued process that solves $\mathrm{d}\tilde{Z}(t) = \sigma\, \mathrm{d}W_t$ and $\tilde{Y}$ is a homogeneous Poisson process with intensity $\tilde{\theta}$. We stick with the notation $x = (z, y) \in \mathbb{R}^4$ with $z \in \mathbb{R}^3$ and $y \in \mathbb{R}$, and denote $v_T = (z_T, y_T)$. Similar to earlier computations, we deduce that

$$g(t, (z, y)) = \exp\left(-\frac{d(z_T, z)^2}{2(T + \epsilon - t)}\right) \frac{\left(\tilde{\theta}(T-t)\right)^{y_T - y}}{(y_T - y)!} \exp\left(-\tilde{\theta}(T-t)\right), \tag{4.16}$$

where $\epsilon > 0$ and

$$d(z_T, z) = \sqrt{(z_T - z)' a^{-1} (z_T - z)}.$$

This choice satisfies the assumption of Lemma 3.3.

5 Choosing g so that $\mathbb{P}^h$ and $\mathbb{P}^g$ are equivalent

By (4.15), the measures $\mathbb{P}^h$ and $\mathbb{P}^g$ are equivalent when $\mathbb{P}^g(A_n) = 1$, as all other terms in the Radon-Nikodym derivative are nonzero. In this section we propose a choice for g which yields equivalence of the measures $\mathbb{P}^h$ and $\mathbb{P}^g$ under an additional assumption on the network.

By Remark 4.5, the choice $C = \varepsilon L a L'$ can be interpreted as imposing a condition of hiting v at time $T + \varepsilon$. Therefore, intuitively, equivalence can be achieved upon artificially setting all elements in the matrices C_k equal to 0. In this case, $g(t_k, x)$ becomes ill-defined whenever $L_k x \neq v_k$ and boundedness of g is lost, rendering the earlier proofs invalid. To utilize this choice for g and show equivalence, we build on earlier work in Corstanje et al. (2023). Proofs of results in this section are given in Appendix B.

5.1 Absolute continuity

The initial conditions for the differential equations (4.13) for the functions H and F appearing in (4.12) are not defined as the matrices $C_1, \ldots, C_n$ are not invertible, but we can still obtain an equivalent of (4.14). Observe that for $t \in [t_{k-1}, t_k)$,

$$\begin{aligned} &\int \tilde{p}(t, x; t_k, \zeta_k) \delta_{v_k}(L_k \zeta_k) \prod_{i=k}^{n-1} \tilde{p}(t_i, \zeta_i; t_{i+1}, \zeta_{i+1}) \delta_{v_{i+1}}(L_{i+1}\zeta_{i+1})\, \mathrm{d}\zeta_k \cdots \mathrm{d}\zeta_n \\ &\qquad \propto \exp\left(-\frac{1}{2}(v(t) - L(t)x)' M(t) (v(t) - L(t)x)\right) =: g(t, x), \end{aligned} \tag{5.1}$$

where L, v and M are defined in Remark 4.9 but with $C_1, \ldots, C_n = 0$ and we define $g(t_k, x) = g(t_k+, x)$ for $x \in L_k^{-1} v_k$.

Theorem 5.1 *Let g be defined by (5.1). Then $\mathbb{P}^h \ll \mathbb{P}^g$ with*

$$\frac{\mathrm{d}\mathbb{P}^h_{t_n}}{\mathrm{d}\mathbb{P}^g_{t_n}} = \frac{g(0,x_0)}{h(0,x_0)} \exp\left(\int_0^{t_n} \frac{\mathcal{A}g}{g}(s, X(s))\,\mathrm{d}s\right) \mathbf{1}_{A[X^{t_n}]}, \tag{5.2}$$

where

$$A[X^{t_n}] = \left\{ \sup_{0\le s<t_n} \sum_{\ell\in\mathcal{R}} \lambda^g_\ell(s, X(s)) < \infty \right\}.$$

To show that $\mathbb{P}^g_{t_n}(A[X^{t_n}]) > 0$, we require Proposition 5.2, combined with the observation that the proof of Theorem 4.2 can be repeated with this choice for g upon observing that $\sum_{\ell\in\mathcal{R}} \lambda^g_\ell$ stays finite on the sets $\{L_k x = v_k\}$.

Proposition 5.2 *$A_n \subseteq A[X^{t_n}]$ with A_n as defined in (1.3).*

Remark 5.3 To see that the reverse inclusion generally does not hold, consider a process conditioned to hit v at time T. If $LX(t) = u \neq v$ for $t \in (T-\varepsilon, T]$ where u is such that no reactions exist such that $d(v, L(u+\xi_\ell)) < d(v, Lu)$, it can be shown that $t \mapsto \frac{g(t,u+\xi_\ell)}{g(t,u)}$ is bounded and therefore such trajectories are included in $A[X^T]$.

5.2 Equivalence

Lemma 5.1 shows that $\mathbb{P}^h$ is absolutely continuous with respect to $\mathbb{P}^g$ on the set $A[X^{t_n}]$. For simulation purposes, equivalence would be preferable. It follows from (5.2) that this is indeed the case if $\mathbb{P}^g\left(A[X^{t_n}]\right) = 1$. This can be shown if the network also satisfies a greedy property.

Assumption 5.4 For $k = 1, \ldots, n$, define the metric d_k on $\mathbb{R}^{m_k}$ through (B.6) and suppose that $d_1, \ldots, d_n$ are such that for all k, $t \in [t_{k-1}, t_k)$ $x \in \mathbb{S} \setminus L_k^{-1} v_k$, there is a reaction $\ell \in \mathcal{R}$ such that $\lambda_\ell(t,x) > 0$ and $d_k(v_k, L_k(x+\xi_\ell)) < d_k(v_k, L_k x)$.

Intuitively, Assumption 5.4 is satisfied when there is always a reaction available that takes the process closer to the first desired conditioning given being in state x at time t. The choice for g will then guarantee that eventually, the path will jump to this point closer to the conditioned state and will therefore hit the observation in finitely many reactions. This argument is formalized in Theorem 5.5.

Theorem 5.5 *Suppose Assumption 5.4 is satisfied. Then $\mathbb{P}^g_{t_n}(A_n) = 1$.*

Corollary 5.6 *It follows from Proposition 5.2 and Theorem 5.5 that, under Assumption 5.4, $\mathbb{P}^g_{t_n}(A[X^{t_n}]) = 1$.*

The choice for g described in this section has two advantages. The first being that the set $A[X^{t_n}]$ on which absolute continuity is obtained is larger than A_n and the second being that an additional assumption yields equivalence. However, a representation such as (4.12) is not available in this case. When evaluating (4.12), the matrix products and inversions are computed for matrices of size $d \times d$, while (4.14) requires matrix computations where the size increases with the amount of observations. We thus conclude that from a theoretical point of view we may prefer the construction in this section, but for computational purposes (4.12) should be preferred, especially in case n is large.

6 Simulation methods

In this section we discuss methods for simulating the guided process on $[0, T]$. This is a Markov jump process with time-dependent intensity. In case $\|LX(t) - v\|$ is large for t close to T, this intensity may blow up.

Simulating a Markov jump process is easy when intensities do not depend on time. In this case, given a state x at time t, we simply simulate reaction times $\tau_\ell \sim \text{Exp}(\lambda_\ell(x))$ for each reaction, set $\hat{\ell} = \text{argmin}_{\ell\in\mathcal{R}}\, \tau_\ell$ and move $t \leftarrow t + \tau_{\hat{\ell}}$ and $X(t + \tau_{\hat{\ell}}) \leftarrow X(t) + \xi_{\hat{\ell}}$. When $\mathcal{R}$ contains many reactions, one could alternatively use Gillespie's algorithm, see e.g. Gillespie (1976, 1977), which first samples the reaction time and subsequently the reaction that takes place at that time.

To extend this method to chemical reaction processes with time-dependent intensities, we have to sample reaction times satisfying (2.4). In general this is hard and therefore therefore we consider a Poisson thinning step. This gives Algorithm 1. The efficiency of Algorithm 1 will depend on whether sharp bounds $\bar{\lambda}_\ell$ can be derived.

Algorithm 1: Next reaction method for time-inhomogeneous rates with a Poisson thinning step

Input: $x \in \mathbb{S}$ and $t \geq 0$ and an upperbound $\bar{\lambda}_\ell$ for $\lambda_\ell(\cdot, x)$.
Result: The next reaction $\hat{\ell} \in \mathcal{R}$ and the corresponding reaction time $\tau_{\hat{\ell}}$ from the state (t, x).

```
for ℓ ∈ R do
    Set t* = t ;
    Sample τ*_ℓ ~ Exp(λ̄_ℓ) and set t* ← t* + τ*_ℓ;
    Sample U ~ Unif(0, 1) ;
    if U ≤ λ_ℓ(t*, x)/λ̄_ℓ then
        Accept, set τ_ℓ = t* − t ;
    else
        Reject and return to line 3 ;
    end
end
Set ℓ̂ = argmin_{ℓ∈R} τ_ℓ ;
return ℓ̂ and τ_ℓ̂.
```

6.1 Simulation of the guided process for underlying processes with time-homogeneous intensities

In many applications, such as the examples in Section 2.3, the intensities λ_ℓ, $\ell \in \mathcal{R}$ of the underlying process only depend on the state x and don't have a direct dependence on t. Here we consider this case and present a method for simulating the guided process in such a scenario. Note that if the underlying rates are time-dependent, but bounded, these upper bounds can easily be included in the thinning step.

6.1.1 Special case: guided process induced by g_ε

We consider the guided process with $\lambda_\ell^{g_\varepsilon} = \alpha_\ell^{g_\varepsilon}\lambda_\ell$, where $\alpha_\ell^{g_\varepsilon}$ is defined in Eq. 4.8. As the map $t \mapsto \alpha_\ell^{g_\varepsilon}(t,x)$ is monotone, given (t,x), an upper bound can be found at either (t,x) or (T,x). Unfortunately, the upper bound at T is typically far from sharp, hampering efficiency, especially when ε is chosen to be small. To resolve this issue, we employ Algorithm 2. In lines 7–18, $\lambda_\ell^{g_\varepsilon}(\cdot,x)$ is increasing in time. Thus we can choose δ such that $t+\delta < T$ and use that, on the interval $[t,t+\delta]$, $\lambda_\ell^{g_\varepsilon}(\cdot,x)$ is upper bounded by $\lambda_\ell^{g_\varepsilon}(t+\delta,x)$. Hence, on $[t,t+\delta]$, we apply the thinning property for an inhomogeneous Poisson process with rate $\lambda_\ell^{g_\varepsilon}(\cdot,x)$ and, if no reaction occurs in this interval, we move t to $t+\delta$. Typically, once $LX(t) = v$ for t near T the map $s \mapsto g(s,x+\xi_\ell)/g(s,x)$ will be decreasing for choices of g considered in this article. If at line 7 $LX(t) \neq v$ *and* $T-(t+\delta) \leq \varepsilon$ we reject the sampled path for specified small $\varepsilon > 0$.

Algorithm 2: Next reaction method with a Poisson thinning step for guided rates where the original intensity is independent of time

Input: $x \in \mathbb{S}$ and $t \geq 0$.
Result: The next reaction $\hat{\ell} \in \mathcal{R}$ and the corresponding reaction time $\tau_{\hat{\ell}}$ from the state (t,x).

1 Set $\mathcal{R}_+ = \{\ell \in \mathcal{R} \mid \lambda_\ell(x) > 0\}$ as the set of possible reactions to occur;
2 **for** $\ell \in \mathcal{R}_+$ **do**
3 **if** $s \mapsto \frac{g(s,x+\xi_\ell)}{g(s,x)}$ *is decreasing in* s *on* $[t,T)$ **then**
4 Apply lines 3-9 of Algorithm 1 with $\bar{\lambda}_\ell = \lambda_\ell^g(t,x)$;
5 **else**
6 Set $t^* = t$ and $t_{\text{start}} = t$;
7 Choose $\delta > 0$ such that $t+\delta < T$;
8 Sample $\tau_\ell^* \sim \text{Exp}\left(\lambda_\ell^g(t+\delta,x)\right)$ and set $t^* \leftarrow t^* + \tau_\ell^*$;
9 Sample $U \sim \text{Unif}(0,1)$;
10 **if** $t^* > t+\delta$ **then**
11 Set $t \leftarrow t+\delta$ and return to line 7
12 **else**
13 **if** $U \leq \lambda_\ell^g(t^*,x)/\lambda_\ell^g(t+\delta,x)$ **then**
14 Accept and set $\tau_\ell = t^* - t_{\text{start}}$;
15 **else**
16 Reject and return to line 8
17 **end**
18 **end**
19 **end**
20 **end**
21 Set $\hat{\ell} = \text{argmin}_{\ell\in\mathcal{R}_+}\tau_\ell$;
22 **return** $\hat{\ell}$ and $\tau_{\hat{\ell}}$.

6.1.2 Choice of δ

We now explain that δ can be chosen such that a minimum acceptance rate is attained.

The acceptance rate of a proposed time τ_ℓ is given by

$$\eta = \frac{\lambda_\ell^{g_\varepsilon}(t+\tau_\ell,x)}{\lambda_\ell^{g_\varepsilon}(t+\delta,x)} \tag{6.1}$$

A direct computation yields that

$$\eta = \exp\left(\frac{d(v, L(x+\xi_\ell))^2 - d(v, Lx)^2}{2(T+\varepsilon-t-\delta)} - \frac{d(v, L(x+\xi_\ell))^2 - d(v, Lx)^2}{2(T+\varepsilon-t-\tau_\ell)}\right), \quad (6.2)$$

Upon solving (6.2) for δ, we find that

$$\delta = T+\varepsilon-t-\left(\frac{2\log\eta}{d(v, L(x+\xi_\ell))^2 - d(v, Lx)^2} + \frac{1}{T+\varepsilon-t-\tau_\ell}\right)^{-1}$$

Now τ_ℓ is not known when choosing δ, but since $\tau_\ell \geq 0$, we obtain the desired acceptance rate η by choosing

$$\delta \geq T+\varepsilon-t-\left(\frac{2\log\eta}{d(v, L(x+\xi_\ell))^2 - d(v, Lx)^2} + \frac{1}{T+\varepsilon-t}\right)^{-1}. \quad (6.3)$$

6.2 Simulation studies

Julia written code for the simulation examples of this section is available in Corstanje (2023).

6.2.1 Death model and comparison to Golightly and Sherlock (2019)

We consider the pure death process of Example 2.2 and estimate the distribution of $X(T)$ conditional on $X(0) = x_0$. For convenience, we denote by g^v the guiding term g chosen for the conditioning $X(T) = v$. By Corollary 3.6, for any $v \in \mathbb{S}$

$$p(0, x_0; T, v) = g^v(0, x_0)\mathbb{E}^{g^v}\left[\frac{\Psi_T^{g^v}(X)}{g^v(T, X(T))}\mathbf{1}_{\{X(T)=v\}}\right]$$

Hence, for large N, we can estimate $p(0, x_0; T, v)$ by sampling $X_1, \ldots, X_N$ from $\mathbb{P}^{g^v}$ using Algorithm 2 and computing

$$\hat{p}(v) = \frac{1}{N}\sum_{i=1}^{N} g^v(0, x_0)\frac{\Psi_T^{g^v}(X_i)}{g^v(T, X_i(T))}\mathbf{1}_{\{X_i(T)=v\}} \quad (6.4)$$

It is a well-known result that $X(T) \sim \text{Binom}\left(x_0, e^{-cT}\right)$, so we will use this mass function for comparison. In the following experiment, we use (6.4) to estimate the mass function of the $\text{Binom}\left(x_0, e^{-cT}\right)$-distribution with $x_0 = 50$, $T = 1$ and $c = 1/2$. We consider four choices for g:

- g_F as in (4.1) with $C = 10^{-5}$ for $v < 32$ and $C = 0.3$ for $v \geq 32$ (we comment on this choice of C below).
- g_{LNA} from the LNA method with restart as in (4.2), with $C = 10^{-5}$.
- g_ε chosen using a scaled diffusion as in (4.4). We took $\varepsilon = 10^{-5}$. For a given value of v, the tuning parameter $a = \sigma^2$ appearing in g_ε was chosen as $2.5(x_0 - v)$. This was found to be roughly the optimum of the map $a \mapsto g^v(0, x_0)\mathbb{E}^{g_\varepsilon}\Psi_T^{g^v}(X)\mathbf{1}_{\{X(T)=v\}}/g^v(T, X(T))$, where the expectation was estimated using 100 forward simulated paths and v was taken to be the 1%, 50% and 99%-quantiles of $X(T)$.
- g chosen using the density of a reversed (decreasing) constant rate Poisson process with rate constant θ. We took $\theta = cv$ (this choice ensures equivalence by Theorem 4.2 of Corstanje et al. (2023)).

Table 1 Mean squared errors of the estimates from Fig. 3 for each of the methods

Method	Mean Squared Error
Fearnhead	$1.5 \cdot 10^{-5}$
LNA (with restart)	$6.5 \cdot 10^{-6}$
Diffusion guiding term	$2.1 \cdot 10^{-5}$
Poisson guiding term	$4.0 \cdot 10^{-7}$

For forward simulation under $\mathbb{P}^g$ we employed Algorithm 2 where for g_ε, δ was chosen according to (6.3). For the other choices of g such a closed form expression cannot be derived, and for $t < T$ we simply took $\delta = \frac{T-t}{2}$. This does not affect the validity of the algorithm, only the acceptance probability of a sampled reaction time.

Results: We took Monte-Carlo sample size $N = 15000$. The estimated probability mass functions are depicted in Fig. 3. This figure confirms that $\mathbb{P}^h_T \ll \mathbb{P}^g_T$ for all choices of g considered. The percentages of paths conditioned on $X(T) = v$ and actually ending in v are depicted in Fig. 4 for a range of values of v. We now comment on the choice of C for the Fearnhead guiding term. For $C = 10^{-5}$ we noticed that conditioning on high values of v caused the guided process to jump very closely to T with high probability. This resulted in numerical instability when computing the integral appearing in Ψ^g_T. For that reason, we took a larger value of C in case v is large. A side effect of that is that there is less strong forcing to hit v at time T which explains the lower fractions of paths ending in v for $v \geq 32$.

We observe that the overall error of the Poisson guiding term is lowest and the diffusion guiding term performs particularly well for low values of v for which the process has to make a lot of jumps. However, when conditioning on high values, the amount of sample paths that end up in the correct state is low, leading to a larger variance of $\hat{p}(v)$ defined in Eq. 6.4. The performance of the Poisson guiding term is explained due to the typically lower guiding term, which is exactly 0 when the conditioning is reached at a time prior to T. This property is not shared by the other choices considered.

In Table 1 we report the mean squared errors $\frac{1}{\#v} \sum_v (q(v) - \hat{p}(v))^2$, with q denoting the probability mass function of the Binom(x_0, e^{-cT})-distribution.

6.2.2 Gene transcription and translation

We consider the *GTT*-model presented in Example 2.3. Figure 5 contains realizations of the process X under $\mathbb{P}^g$ conditioned to hit $(1, 11, 56)$ at time $T = 1$. We used guiding induced by g_ε with $\varepsilon = 10^{-5}$ and $a = a_{\text{CLE}}(0, x_0)$. The plot shows 10 sampled trajectories. Out of 1000 more trajectories sampled, all of these end in the correct point $X(T) = (1, 11, 56)$.

Next, from a simulated forward path we chose 15 times at random and saved the values of randomly chosen components of the process. Taking these values as observations we show in Fig. 6 multiple realisations of the guided process. Out of 1000 simulated trajectories we found that all of those passed through all partial observations.

6.2.3 Enzyme kinetics

As described in Example 2.4, this is a very interesting example as it contains a monotone component as well as absorbing states, where no more reactions can occur. For example, the state $X(t) = (0, 20, 0, 32)$ can be reached with a sequence of reactions from x_0, but no further

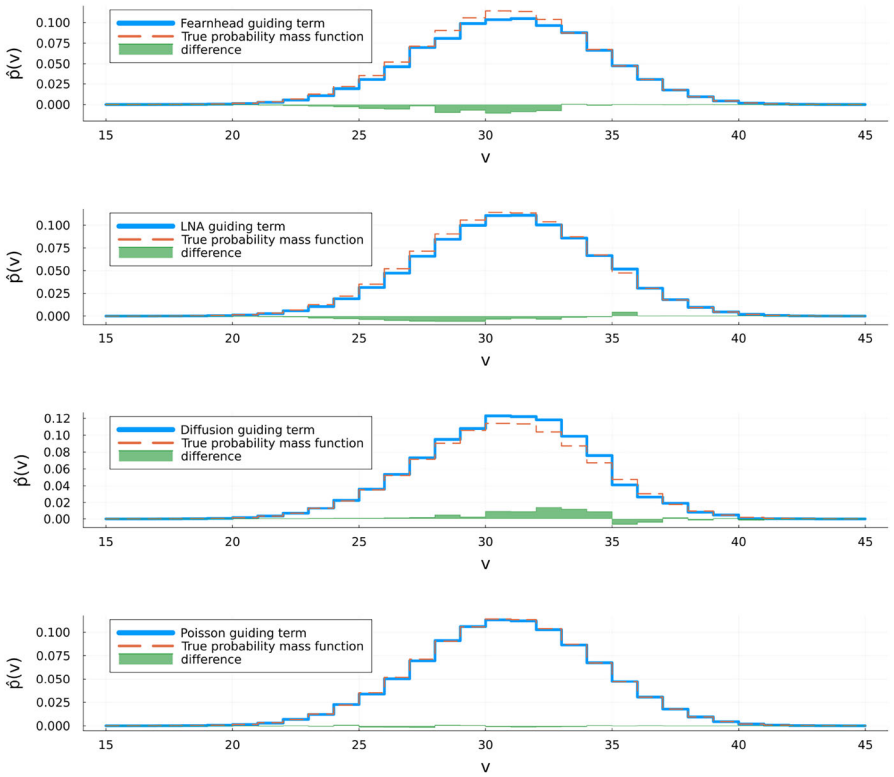

Fig. 3 Estimates of the probability mass function of $X(T) \mid X(0) = x_0$ using the guiding functions f from the LNA method, a scaled diffusion and a Poisson process. The upper barplot is the true density. For each v, we estimated using (6.4) with $N = 15000$

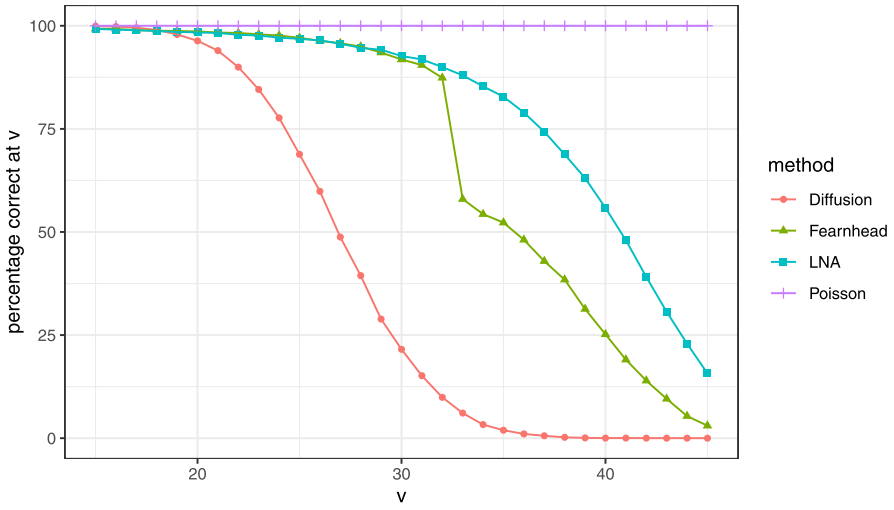

Fig. 4 The percentage of paths ending in the point of conditioning (v) versus v for the four methods considered (with the same settings as in Fig. 3)

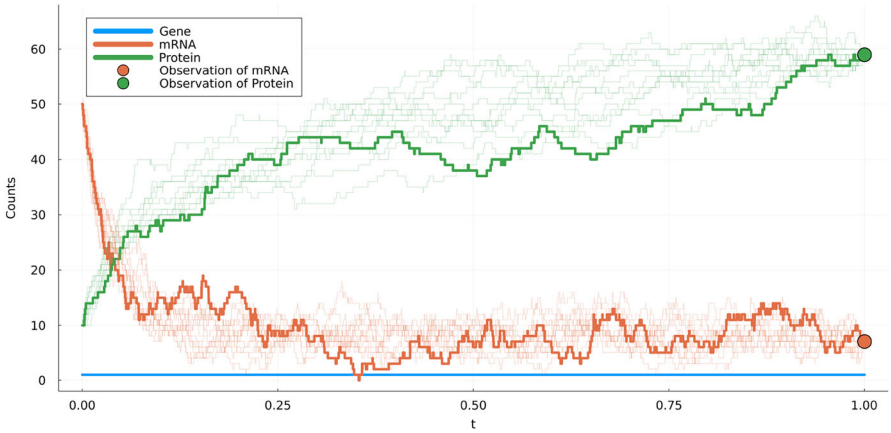

Fig. 5 Realization of a guided process starting from $x_0 = (1, 50, 10)$ conditioned to be at $(1, 10, 50)$ at time $T = 1$ of the *GTT*-model from Example 2.3 with $\kappa_1 = 100$, $\kappa_2 = 10$, $d_M = 25$ and $d_P = 1$. The thick line is the original (unconditioned) process

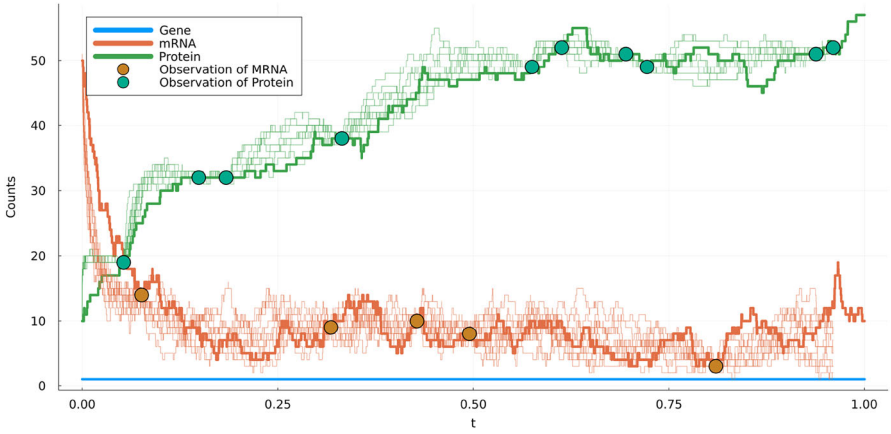

Fig. 6 Realization of a guided process through 15 arbitrary observations, with the same settings as in Fig. 5

reactions are possible in this state. We use this example to compare various guiding terms. We choose the starting point $x_0 = (12, 10, 10, 10)$ and parameters $(\kappa_1, \kappa_2, \kappa_3) = (5, 5, 3)$. We consider three scenarios in which the LNA method with restart, the guiding term obtained from a scaled diffusion presented in Sect. 6.2.3 and the guiding term in which one of the components is replaced by a Poisson guiding term as presented in Sect. 4.3.

- **Scenario A:** The process is conditioned to be at $x_{T,0.01} = (0, 15, 5, 27)$ at time $T = 1$. 27 is the 1% quantile of $X_4(T)$, determined through forward simulation.
- **Scenario B:** The process is conditioned to be at $x_{T,0.5} = (0, 19, 1, 31)$ at time $T = 1$. 31 is the 50% quantile of $X_4(T)$, determined through forward simulation.
- **Scenario C:** The process is conditioned to be at $x_{T,0.99} = (0, 20, 0, 32)$ at time $T = 1$. 32 is the 99% quantile of $X_4(T)$, determined through forward simulation and $x_{T,0.99}$ is also an absorbing state of the process.

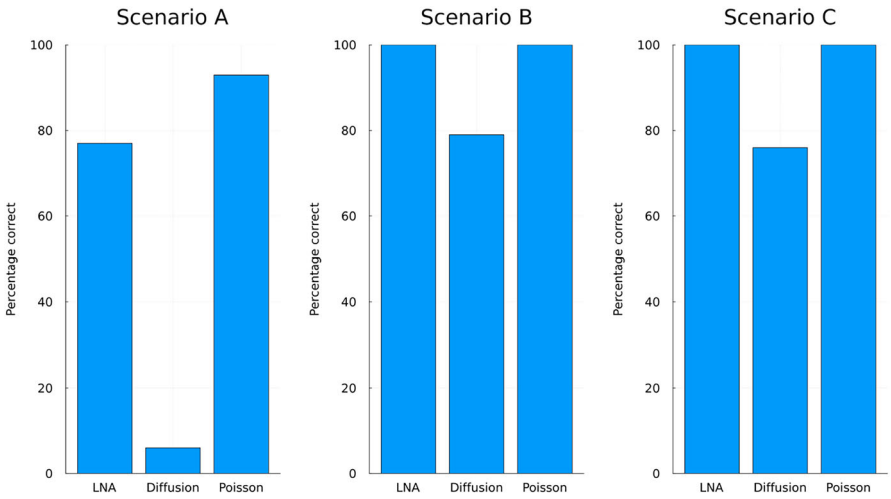

Fig. 7 Percentages out of 100 trajectories that satisfy the conditionings set in scenario's **A**, **B** and **C**

Table 2 Mean Squared Error (MSE) of the estimates from Figs. 8, 9 and 10 for each of the methods

Method	MSE scenario A	MSE scenario B	MSE scenario C
LNA (without restart)	$8.8 \cdot 10^{-6}$	$2.3 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$
Diffusion guiding term	$9.0 \cdot 10^{-6}$	$1.8 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$

First we simulate 100 trajectories of each process and check the percentage of paths that satisfy $X(T) = x_{T,q}$ for $q = 0.01, 0.5, 0.99$. We used the same scheme for simulating the guided process as for the death process. For the diffusion guiding term, we took $a = a_{\text{CLE}}(0, x_0)$. For the Poisson guiding term, the auxiliary process from which g is obtained contains a scaled diffusion in the 3 components and a Poisson processes in the third component. The diffusion is scaled by a matrix a, in which we used the first three rows and the first three columns of $a_{\text{CLE}}(0, x_0)$. The intensity of the Poisson component was taken as $\lambda_3(0, x_0)$, which is a lower bound of the reaction intensity for the reaction that induces the monotone component. The results are summarised in Fig. 7.

Similar to the death process in Fig. 3, we see that the diffusion guiding term and the LNA guiding term struggle when the monotone component is conditioned not to make many jumps.

We used the same method as for the death processes to estimate $p(0, x_0; T, v)$ for scenario's A, B and C. In each scenario, we computed $\hat{p}$ through (6.4). In each scenario, we make 200 estimates for $p(0, x_0; T, v)$ by computing $\hat{p}$ with $N = 1000$ processes and we compared the LNA method (without restart cf. Section 4.3 of Golightly and Sherlock (2019)) with the diffusion guiding term with $\varepsilon = 10^{-5}$ and $a = 100 a_{\text{CLE}}(T, x_{T,q})$ for $q = 0.01, 0.5, 0.99$. For the LNA method we assumed extrinsic noise $C = 2000I$, $C = 500I$ and $C = 200I$ for scenario's A, B and C, respectively. Histograms of the estimates are given in Figs. 8, 9 and 10. The MSEs of the estimates are summarised in Table 2.

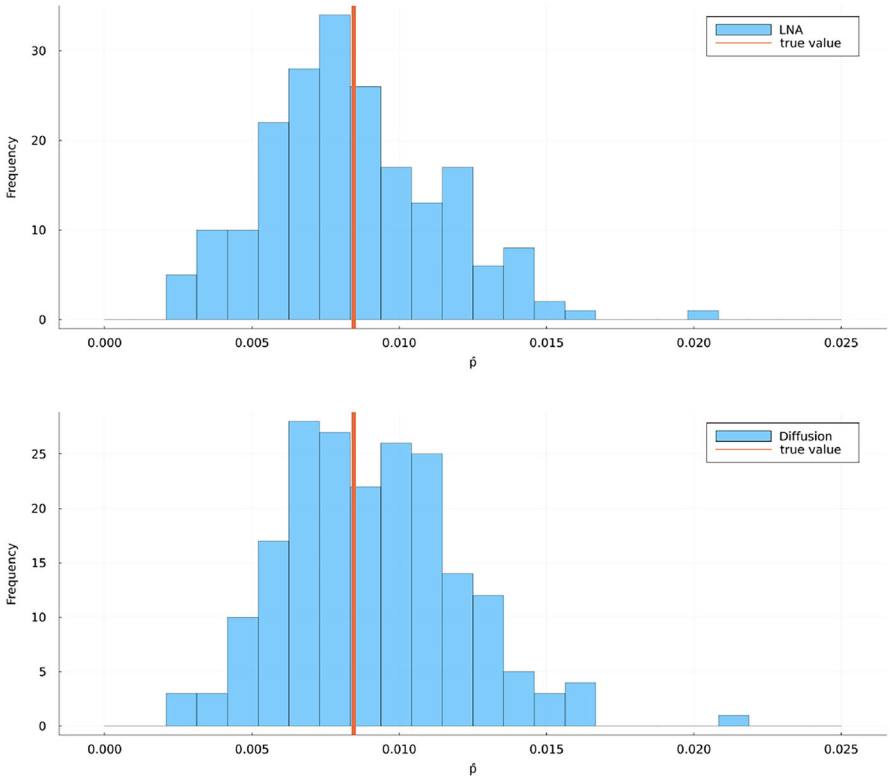

Fig. 8 Histograms of 200 estimates for $p(0, x_0; T, v)$ in scenario A using the LNA method without restart (top) and and the diffusion method (bottom). The true value was estimated using 10,000 forward simulations

7 Discussion

In this paper we have provided sufficient conditions for constructing valid guided processes, where "valid" refers to the law of the true conditioned process being absolutely continuous with respect to the law of the guided process induced by g with the laws as defined in Sect. 3. We have presented various choices of g and conclude that among those there is no best choice in terms of closeness to the true conditioned process and computational cost. When used within a sequential Monte Carlo or Markov chain Monte Carlo algorithm, a mixture of proposals may be beneficial. The inherent discreteness of chemical reaction processes makes it a hard problem, but it works to our advantage in the sense that guided processes can be constructed on $[0, T]$ where $g(T, x)$ is well defined. This is accomplished by using a guiding term derived from conditioning a diffusion process that is observed with (small) extrinsic noise. The guiding term defined in Sect. 4.3 has the additional advantage that monotonicity can be exploited for simulating paths of the guided process efficiently.

In the analogous problem of continuous-discrete smoothing for diffusion processes (see e.g. Mider et al. (2021), Beskos et al. (2008), Golightly and Wilkinson (2008)) there exists a simple random-walk like MCMC-sampler on path space to update guided processes: the pre-conditioned Crank-Nicolson scheme. Unfortunately, we are not aware of a similar construct for chemical reaction processes.

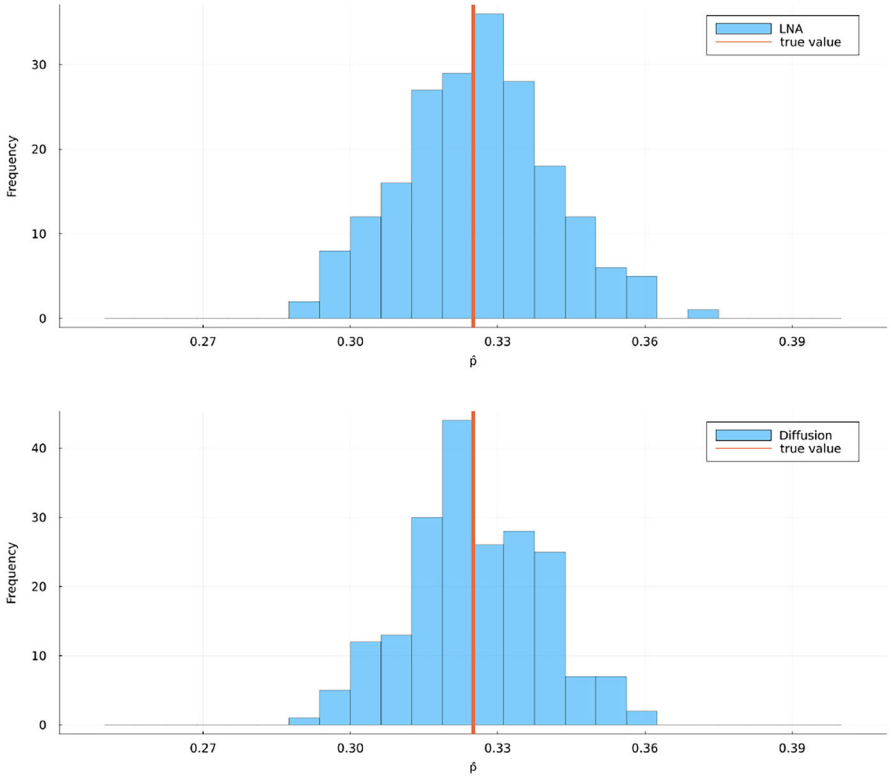

Fig. 9 Histograms of 200 estimates for $p(0, x_0; T, v)$ in scenario B using the LNA method without restart (top) and and the diffusion method (bottom). The true value was estimated using 10,000 forward simulations

Appendix A Proof of Theorem 4.2

The proof relies on the existence of a chain of reactions that result in the conditioning $LX(T) = v$. We formalize this in Definition A.1.

Definition A.1 For $x, y \in \mathbb{S}$, we say that y can be reached from x within a time interval $(a, b) \subseteq (0, T)$, denoted $x \hookrightarrow y$ in (a, b), if there is a finite collection $\{\ell_1, \ldots, \ell_I\} \subseteq \mathcal{R}$ of reactions and a partition $U_1, \ldots, U_I$ of disjoint intervals for (a, b) such that $y = x + \sum_{i=1}^{I} \xi_{\ell_i}$ and for all the set i, $\operatorname{supp} \lambda_{\ell_i} \left(x + \sum_{j<i} \xi_{\ell_j} \right) \cap U_i$ is of positive Lebesgue measure.

Since we consider a process conditioned on $LX(T) = v$, we can assume there exists an $x_T \in \mathbb{S}$ such that $Lx_T = v$ and $x_0 \hookrightarrow x_T$ in $(0, T)$.

Proof of Theorem 4.2 Notice that, given any $t < T$,

$$\mathbb{P}^g \left(X(s) = x_T \text{ for all } s \in [t, T] \mid X(t) = x_T \right) = \exp\left(- \int_t^T \sum_{\ell \in \mathcal{R}} \lambda_\ell^g(s, x_T) \, \mathrm{d}s \right).$$

Now it can be shown that $\sum_{\ell \in \mathcal{R}} \lambda_\ell^g$ is bounded through Lemma C.1 and Assumption Assumption (2.1c) that the right hand side is strictly positive and hence, it suffices to show

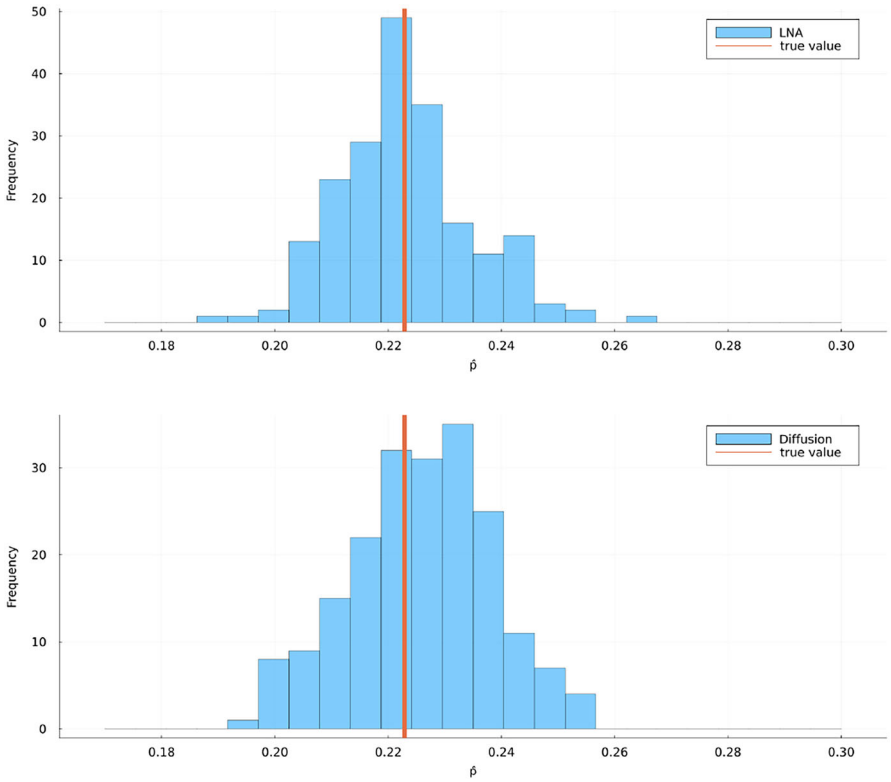

Fig. 10 Histograms of 200 estimates for $p(0, x_0; T, v)$ in scenario C using the LNA method without restart (top) and and the diffusion method (bottom). The true value was estimated using 10,000 forward simulations

that $\mathbb{P}^g(\exists t < T : X(t) = x_T) > 0$. In the remainder of this proof, we show this by showing that the probability that the chain of reactions described in Definition A.1 connection x_0 to x_T occurs is positive.

Let $\ell_1, \ldots, \ell_I$ be the collection of reactions described in Definition A.1 and let $U_1, \ldots, U_I$ be the respective partition of $(0, T)$. For the first reaction to occur, we denote by $T_\ell^{(1)}$ the reaction time for each reaction $\ell \in \mathcal{R}$ and let L_1 be the first reaction to occur. That is,

$$L_1 = \underset{\ell \in \mathcal{R}}{\operatorname{argmin}}\, T_\ell^{(1)}, \qquad \text{where} \qquad \mathbb{P}^g\left(T_\ell^{(1)} > t\right) = \exp\left(-\int_0^t \lambda_\ell^g(s, x_0)\, \mathrm{d}s\right).$$

Note that,

$$\left\{T_{\ell_1}^{(1)} \in U_1,\ T_\ell^{(1)} > \sup U_1 \text{ for all } \ell \in \mathcal{R} \setminus \{\ell_1\}\right\} \subseteq \left\{T_{\ell_1}^{(1)} \in U_1,\ L_1 = \ell_1\right\}$$

and thus

$$\begin{aligned}
&\mathbb{P}^g\left(T_{\ell_1}^{(1)} \in U_1,\ L_1 = \ell_1\right) \\
&\quad \geq \left(1 - \exp\left(-\int_{U_1} \lambda_{\ell_1}^g(s, x_0)\, \mathrm{d}s\right)\right) \exp\left(-\int_{U_1} \sum_{\ell \in \mathcal{R} \setminus \{\ell_1\}} \lambda_\ell^g(s, x_0)\, \mathrm{d}s\right).
\end{aligned}$$

Since $\lambda^g_{\ell_1}(\cdot, x_0)$ has support in U_1, the first term is nonzero, while the second term is nonzero by Assumption (2.1c).

For the second reaction, we set

$$L_2 = \operatorname*{argmin}_{\ell \in \mathcal{R}} T^{(2)}_\ell \qquad \text{where} \qquad \mathbb{P}^g\left(T^{(2)}_\ell > t \mid L_1\right) = \exp\left(-\int_{T^{(1)}_{L_1}}^t \lambda^g_\ell(s, x_0 + \xi_{L_1})\,\mathrm{d}s\right).$$

Through similar reasoning, we deduce that

$$\begin{aligned}
&\mathbb{P}^g\left(L_2 = \ell_2,\ T^{(2)}_{\ell_2} \in U_2,\ L_1 = \ell_1,\ T^{(1)}_{\ell_1} \in U_1\right) \\
&\quad = \mathbb{P}^g\left(L_2 = \ell_2,\ T^{(2)}_{\ell_2} \in U_2 \mid L_1 = \ell_1,\ T^{(1)}_{\ell_1} \in U_1\right)\mathbb{P}^g\left(L_1 = \ell_1,\ T^{(1)}_{\ell_1} \in U_1)\right) \\
&\quad \geq \left(1 - \exp\left(-\int_{U_2} \lambda^g_{\ell_2}\left(s, x_0 + \xi_{\ell_1}\right)\,\mathrm{d}s\right)\right)\exp\left(-\int_{U_2} \sum_{\ell \in \mathcal{R}\setminus\{\ell_2\}} \lambda^g_\ell(s, x_0 + \xi_{\ell_1})\,\mathrm{d}s\right) \\
&\qquad \times \left(1 - \exp\left(-\int_{U_1} \lambda^g_{\ell_1}(s, x_0)\,\mathrm{d}s\right)\right)\exp\left(-\int_{U_1} \sum_{\ell \in \mathcal{R}\setminus\{\ell_1\}} \lambda^g_\ell(s, x_0)\,\mathrm{d}s\right).
\end{aligned}$$

Upon iteratively repeating this process, we find

$$\begin{aligned}
&\mathbb{P}^g\left(L_i = \ell_i,\ T^{(i)}_{\ell_i} \in U_i,\ i = 1, \ldots, I\right) \geq \\
&\prod_{i=1}^I \left(1 - \exp\left(-\int_{U_i} \lambda^g_{\ell_i}\left(s, x_0 + \sum_{j<i} \xi_{\ell_j}\right)\,\mathrm{d}s\right)\right) \\
&\times \exp\left(-\int_{U_i} \sum_{\ell \in \mathcal{R}\setminus\{\ell_i\}} \lambda^g_{\ell_i}\left(s, x + \sum_{j<i} \xi_{\ell_j}\right)\,\mathrm{d}s\right).
\end{aligned}$$

The right hand size is strictly positive as a finite product of nonzero terms. Therefore $\mathbb{P}^g(\exists t \leq T : X(t) = x_T) > 0$, which finishes the proof. □

Appendix B Proofs for Sect. 5

Theorem B.1 *(Theorem 3.3 of Corstanje et al. (2023)]) Suppose that there exist a family of $\mathcal{F}_t$-measurable events $\{A_j[X^t]\}_j$ for each $t \in [0, t_n)$ so that the following assumptions hold.*

(B.1a) For all j $s \leq t$, $A_j[X^t] \subseteq A_{j+1}[X^t]$ and $A_j[X^t] \subseteq A_j[X^s]$.

(B.1b) For all k and $s \in [0, t_k)$ and $x \in \mathbb{S}$,

$$\lim_{t \uparrow t_k} \mathbb{E}\left[g(t, X(t)) \mid X(s) = x\right] = \mathbb{E}\left[g(t_k, X(t_k))\mathbf{1}\{L_k X(t_k) = v_k\}\right],$$

Moreover, $\mathbb{E}\left[g(t, X(t)) \mid X(s)\right]$ is $\mathbb{P}$-almost surely bounded for all t.

(B.1c) For all $k = 1, \ldots, n$,

$$\lim_{j \to \infty} \lim_{t \uparrow t_k} \mathbb{E}^h_t\left[\frac{g(t, X(t))}{h(t, X(t))}\mathbf{1}_{A_j[X^t]}\right] = \frac{\mathbb{E}\left[g(t_k, X(t_k))\mathbf{1}\{L_k X(t_k) = v_k\}\mathbf{1}_{A[X^{t_k}]}\right]}{h(0, x_0)},$$

where $A_j[X^{t_k}] = \cap_{t \in [0, t_k)} A_j[X^t]$, $A[X^t] = \bigcup_j A_j[X^t]$ and $A[X^{t_k}] = \cap_{t \in [0, t_k)} A[X^t]$.

(B.1d) For all fixed j, $\Psi_t^g(X)\mathbf{1}_{A_j[X^t]}$ is $\mathbb{P}^g$-almost surely uniformly bounded in t.

Then for any bounded measurable function f,

$$\mathbb{E}\left[f(X)\frac{\mathbb{E}\left[g(t_k, X(t_k))\mathbf{1}\{L_k X(t_k) = v_k\}\right]}{h(0,x_0)}\mathbf{1}_{A[X^{t_k}]}\right] = \mathbb{E}^g\left[f(X)\frac{g(0,x_0)}{h(0,x_0)}\Psi_{t_k}^g(X)\mathbf{1}_{A[X^{t_k}]}\right] \tag{B.1}$$

for $k = 1, \ldots, n$. In particular, for $k = n$,

$$\mathbb{E}^h\left[f(X)\mathbf{1}_{A[X^{t_n}]}\right] = \mathbb{E}^g\left[f(X)\frac{g(0,x_0)}{h(0,x_0)}\Psi_{t_n}^g(X)\mathbf{1}_{A[X^{t_n}]}\right]. \tag{B.2}$$

Proof The proof of Theorem 3.3 of Corstanje et al. (2023) can be followed with the new limits in (B.1b) and (B.1c) for each of the limits $t \uparrow t_k$. □

Proof of Theorem 5.1 The proof utilizes Theorem B.1 with

$$V(t,x) = \sum_{\ell\in\mathcal{R}} \lambda_\ell^g(t,x) \qquad \text{and} \qquad A_j[X^t] = \left\{\sup_{0\le s<t} V(s, X(s)) \le j\right\}. \tag{B.3}$$

Then (B.1a) is satisfied by construction. Lemmas B.3-B.5 prove the remaining conditions. □

Lemma B.2 *The map g, defined in (5.1), is such that $g \in \mathcal{G}$ and for all $k = 1, \ldots, n$ and $x \in \mathbb{S}$,*

$$\lim_{t\uparrow t_k} g(t,x) = \begin{cases} g(t_k, x) & \text{if } L_k x = v_k \\ 0 & \text{otherwise} \end{cases}. \tag{B.4}$$

and $\partial_t \log g(t,x) \le 0$ for all $(t,x) \in (0,t_n)\times\mathbb{S}$.

Proof Through a direct computation using the Schur complement, it can be derived that as $t \uparrow t_k$,

$$(v(t) - L(t)x)' M(t)(v(t) - L(t)x) = \frac{d_k(v_k, L_k x)^2}{(t_k - t)} + o(t_k - t), \tag{B.5}$$

where d_k is the metric on $\mathbb{R}^{m_k}$ defined as

$$d_k(x,y) = \sqrt{(y-x)'\left(L_k a_k L_k'\right)^{-1}(y-x)}, \qquad x, y \in \mathbb{R}^{m_k}. \tag{B.6}$$

which yields (B.4). Since v and L are piecewise constant and $\frac{\mathrm{d}M}{\mathrm{d}t} = M(t)\left(L_i a_k L_j'\right)_{i,j=k}^n M(t)$ is positive semidefinite, $\partial_t \log g(t,x) \le 0$. □

Lemma B.3 *For all $k = 1, \ldots, n$, $s \in [0, t_n)$ and $x \in \mathbb{S}$,*

$$\lim_{t\uparrow t_k} \mathbb{E}\left[g(t, X(t)) \mid X(s) = x\right] = \mathbb{E}\left[g(t_k, X(t_k))\mathbf{1}\{L_k X(t_k) = v_k\} \mid X(s) = x\right].$$

Proof By dominated convergence and Lemma B.2

$$\begin{aligned}
&\lim_{t\uparrow t_k} \mathbb{E}\left[g(t, X(t)) \mid X(s) = x\right] \\
&= \lim_{t\uparrow t_k} \sum_{y\in\mathbb{S}} g(t,y) p(s,x;t,y) \\
&= \sum_{y\in\mathbb{S}} g(t_k, y)\mathbf{1}\{L_k y = v_k\} p(s,x;t_k,y) \\
&= \mathbb{E}\left[g(t_k, X(t_k))\mathbf{1}\{L_k X(t_k) = v_k\} \mid X(s) = x\right].
\end{aligned}$$

□

Lemma B.4 *For all $k = 1, \ldots, n$,*

$$\lim_{j\to\infty}\lim_{t\uparrow t_k}\mathbb{E}^h\left[\frac{g(t,X(t))}{h(t,X(t))}\mathbf{1}_{A_j[X^t]}\right] = \frac{\mathbb{E}\left[g(t_k,X(t_k))\mathbf{1}\{L_kX(t_k)=v_k\}\mathbf{1}_{A[X^{t_k}]}\right]}{h(0,x_0)}.$$

Proof Upon defining $Z(t) = g(t, X(t))$, it follows from dominated convergence and Lemma B.2 that for any bounded continuous function f,

$$\begin{aligned}\lim_{t\uparrow t_k}\mathbb{E}f(Z(t)) &= \sum_{x\in\mathbb{S}}\lim_{t\uparrow t_k} f\left(g(t,x)\right)p(0,x_0;t,x)\\ &= \sum_{x\in\mathbb{S}} f\left(g(t_k,x)\mathbf{1}\{L_kx=v_k\}\right)p(0,x_0;t_k,x)\\ &= \mathbb{E}f\left(g(t_k,X(t_k))\mathbf{1}\{L_kX(t_k)=v_k\}\right).\end{aligned}$$

Hence,

$$Z(t) \rightsquigarrow g(t_k,X(t_k))\mathbf{1}\{L_kX(t_k)=v_k\} \text{ as } t\uparrow t_k. \tag{B.7}$$

Now

$$\mathbb{E}\left[Z(t)\mathbf{1}_{A_j[X^t]}\right] = \mathbb{E}\left[Z(t)\mathbf{1}_{A_j[X^{t_k}]}\right] + \mathbb{E}\left[Z(t)\mathbf{1}_{A_j[X^t]\setminus A_j[X^{t_k}]}\right]$$

By (B.7), the first term tends to $\mathbb{E}\left[g(t_k,X(t_k))\mathbf{1}\{L_kX(t_k)=v_k\}\mathbf{1}_{A_j[X^{t_k}]}\right]$ as $t\uparrow t_k$. Moreover, $\mathbf{1}_{A_j[X^t]\setminus A_j[X^{t_k}]}\downarrow 0$ as $t\uparrow t_k$ and thus the second term tends to 0 by Slutsky's theorem. Hence, by (3.5)

$$\begin{aligned}\lim_{j\to\infty}\lim_{t\uparrow t_k}\mathbb{E}^h\left[\frac{Z(t)}{h(t,X(t))}\mathbf{1}_{A_j[X^t]}\right] &= \frac{1}{h(0,x_0)}\lim_{j\to\infty}\lim_{t\uparrow t_k}\mathbb{E}\left[Z(t)\mathbf{1}_{A_j[X^t]}\right]\\ &= \frac{1}{h(0,x_0)}\lim_{j\to\infty}\mathbb{E}\left[g(t_k,X(t_k))\mathbf{1}\{L_kX(t_k)=v_k\}\mathbf{1}_{A_j[X^{t_k}]}\right]\\ &= \frac{\mathbb{E}\left[g(t_k,X(t_k))\mathbf{1}\{L_kX(t_k)=v_k\}\mathbf{1}_{A[X^{t_k}]}\right]}{h(0,x_0)}.\end{aligned}$$

□

Lemma B.5 *For all j,*

$$\sup_{t<t_n}\exp\left(\int_0^t\frac{\mathcal{A}g}{g}(s,X(s))\,\mathrm{d}s\right)\mathbf{1}_{A_j[X^t]} \le \exp(jt_n).$$

Proof Since $\mathcal{A} = \partial_t + \mathcal{L}$,

$$\frac{\mathcal{A}g}{g}(s,x) = \partial_s\log g(s,x) + \frac{\mathcal{L}g}{g}(s,x).$$

Note that

$$\frac{\mathcal{L}g}{g}(s,x) = \sum_{\ell\in\mathcal{R}}\frac{\lambda_\ell(s,x)\left(g(s,x+\xi_\ell)-g(s,x)\right)}{g(s,x)} = V(s,x) - \sum_{\ell\in\mathcal{R}}\lambda_\ell(s,x) \le V(s,x).$$

By Lemma B.2, $\partial_t\log g \le 0$ and hence, for all $t < t_n$ and on $A_j[X^t]$,

$$\int_0^t\frac{\mathcal{A}g}{g}(s,X(s))\,\mathrm{d}s = \int_0^t\partial_s\log g(s,X(s))\,\mathrm{d}s + \int_0^t\frac{\mathcal{L}g}{g}(s,X(s)) \le jt \le jt_n.$$

□

Proof of Proposition 5.2 By (B.5),

$$\lim_{t\uparrow t_k}\frac{g(t,x+\xi_\ell)}{g(t,x)}=\begin{cases}\frac{g(t_k,x+\xi_\ell)}{g(t_k,x)} & \text{if } L_kx=L_k(x+\xi_\ell)=v_k\\ 0 & \text{if } d_k(v_k,L_kx)>d_k(v_k,L_k(x+\xi_\ell))\\ \infty & \text{if } d_k(v_k,L_kx)<d_k(v_k,L_k(x+\xi_\ell)\end{cases}. \tag{B.8}$$

And therefore

$$\lim_{t\uparrow t_k}\lambda_\ell^g(t,x)=\begin{cases}\lambda_\ell(t_k,x)\frac{g(t_k,x+\xi_\ell)}{g(t_k,x)} & \text{if } L_kx=L_k(x+\xi_\ell)=v_k\\ 0 & \text{if } d_k(v_k,L_k(x+\xi_\ell))>d_k(v_k,L_kx)\\ \infty & \text{if } d_k(v_k,L_k(x+\xi_\ell))<d_k(v_k,L_kx)\end{cases}. \tag{B.9}$$

Now, for $t\in[t_{k-1},t_k)$, define $B_t^k=\{L_kX(s)=v_k \text{ for all } s\in[t,t_k]\}$. Then it follows that for all k and $t\in[t_{k-1},t_k)$,

$$B_t^k\subseteq\left\{\sup_{t_{k-1}\le s<t_k}\sum_{\ell\in\mathcal{R}}\lambda_\ell^g(s,X(s))<\infty\right\}.$$

Hence

$$A_n=\bigcup_{k=1}^n\bigcap_{t_{k-1}\le t<t_k}B_t^k\subseteq\bigcup_{k=1}^n\left\{\sup_{t_{k-1}\le s<t_k}\sum_{\ell\in\mathcal{R}}\lambda_\ell^g(s,X(s))<\infty\right\}=A[X^{t_n}].$$

□

Proof of Theorem 5.5 Observe that the result follows immediately if we show that for $k=1,\dots,n$, $\mathbb{P}^g(L_k(X(t_k)=v_k)>0$. Now note The probability distribution of the next reaction time τ satisfies, for $t_{k-1}<s<t<t_k$ and $x\in\mathbb{S}$,

$$\mathbb{P}^g(\tau>t\mid X(s)=x)=\exp\left(-\int_s^t\sum_{\ell\in\mathcal{R}}\lambda_\ell^g(u,x)\,\mathrm{d}u\right).$$

Note that in (B.9), the convergence/divergence is at an exponential rate and therefore, by (B.9) and Assumption 5.4,

$$\mathbb{P}^g(\tau>t_k\mid X(s)=x)=0$$

for all $s<t_k$ whenever $L_kx\neq v_k$. Hence, if $L_kX(s)\neq v_k$ for any $s<t_k$, X jumps before time t_k with probability 1. Since this holds for any $s<t_k$, X jumps an infinite amount of times if L_kX does not hit v_k.

We now consider the jumps that take L_kX closer to v_k with respect to the metric d_k. Let

$$\tau_+=\inf\{t\ge s\colon d_k(v_k,L_kX(t))<d(v_k,L_kX(s))\}.$$

Then the distribution of τ_+ satisfies for any x such that $L_kx\neq v_k$ and $t_{k-1}<s<t<t_k$,

$$\mathbb{P}^g(\tau_+>t\mid X(s)=x)\le\exp\left(-\int_s^t\sum_{\ell\in\mathcal{R}}\lambda_\ell^g(u,x)\mathbf{1}_{\{d_k(v_k,L_k(x+\xi_\ell))<d_k(v_k,L_kx)\}}\,\mathrm{d}u\right).$$

By Assumption 5.4, there are nonzero entries in this summation that tend to ∞ exponentially as $u\uparrow t_k$ and thus

$$\mathbb{P}^g(\tau_+>t_k\mid X(s)=x)=0$$

for any $s < t_k$ and $L_k x \neq v_k$. In other words, for any $s < t_k$ and $L_k x \neq v_k$,

$$\mathbb{P}^g \left(\exists t \in [s, t_k] \text{ such that } d_k(v_k, L_k X(t)) < d(v_k, L_k x) \mid X(s) = x\right) = 1.$$

Now $\mathbb{S}$ is a discrete lattice and thus any set $\mathbb{S} \cap K$, where $K \subseteq \mathbb{R}^d$ is compact, contains only finitely many points. Hence, since $L_k X$ keeps jumping with probability 1 to points strictly closer to v_k if it does not hit v_k, it must hit v_k with probability 1 and thus

$$\mathbb{P}^g \left(\exists t \in [s, t_k] \text{ such that } L_k X(t) = v_k \mid X(s) = x\right) = 1$$

for any x with $L_k x \neq v_k$. Since this obviously also holds for $L_k x = v_k$, we have by the law of total probability

$$\mathbb{P}^g \left(\exists t \in [s, t_k] \text{ such that } L_k X(t) = v_k\right) = 1. \tag{B.10}$$

Now denote the collection of events $B_s = \bigcup_{s \leq t \leq t_k} \{L_k X(t) = v_k\}$, $s \in [t_{k-1}, t_k]$. By (B.10), $\mathbb{P}^g(B_s) = 1$ for all s. Since $B_s \downarrow \{L_k X(t_k) = v_k\}$ as $s \uparrow t_k$, we have by monotonicity of measures that $\mathbb{P}^g(L_k X(t_k) = v_k) = \lim_{s \uparrow t_k} \mathbb{P}^g(B_s) = 1$. □

Appendix C Additional Lemmas

Lemma C.1 *Let g be defined by (4.7). Then for all x, $g(t, x)$ is bounded and bounded away from zero uniformly in t.*

Proof Note

$$(v - Lx)' M(t)(v - Lx) \geq \lambda_{\min}(M(t)) \, |v - Lx|^2 = \left(\lambda_{\max}(M^{\dagger}(t))\right)^{-1} |v - Lx|^2 .$$

Since for all symmetric matrices A, B,

$$\lambda_{\min}(A) + \lambda_{\min}(B) \leq \lambda_{\min}(A + B) \leq \lambda_{\max}(A + B) \leq \lambda_{\max}(A) + \lambda_{\max}(B),$$

we have that

$$\left(\lambda_{\max}(M^{\dagger}(t))\right)^{-1} \geq \left(\lambda_{\max}(C) + \lambda_{\max}(LaL')(T - t)\right)^{-1} .$$

Now LaL' is positive definite and thus the right hand side is decreasing in t. Hence, a lower bound is given by $\left(\lambda_{\max}(C) + T\lambda_{\max}\left(LaL'\right)\right)^{-1}$. We conclude

$$g(t, x) \leq \exp\left(-\frac{1}{2} |v - Lx|^2 \left(\lambda_{\max}(C) + T\lambda_{\max}\left(LaL'\right)\right)^{-1}\right).$$

In a similar way, we can show

$$g(t, x) \geq \exp\left(-\frac{1}{2} |v - Lx|^2 \lambda_{\min}(C)^{-1}\right).$$

□

Corollary C.2 *(Multiple observations) Let g be defined by (4.14). Then for all x, $g(t, x)$ is bounded and bounded away from zero uniformly in t.*

Proof The proof of Lemma C.1 can be repeated using, for $t \in (t_{k-1}, t_k]$,

$$|v(t) - L(t)x|^2 = \sum_{j=k}^{n} \left|v_j - L_j x\right|^2 \in \left[\left|v_n - L_n x\right|^2, \sum_{j=1}^{n} \left|v_j - L_j x\right|^2\right].$$

□

Appendix D Change of generator under change of measure

Here, we briefly recap the argument given in Section 4.1 of Palmowski and Rolski (2002). Fix g and define $\{\tilde{\mathbb{P}}_t,\ t \geq 0\}$ by

$$\frac{\mathrm{d}\tilde{\mathbb{P}}_t}{\mathrm{d}\mathbb{P}_t} = E^g(t) := \frac{g(t, X(t))}{g(0, x_0)} \exp\left(-\int_0^t \frac{\mathcal{A}g}{g}(s, X(s))\,\mathrm{d}s\right),$$

where $\mathcal{A}$ is the generator of the space-time process under $\mathbb{P}$. We aim to show that under $\tilde{\mathbb{P}}$ the generator of this process is given by $\tilde{\mathcal{A}}$ where $g\tilde{\mathcal{A}}f = \mathcal{A}(fg) - f\mathcal{A}g$. For this, it suffices to prove that

$$\tilde{D}^f(t) := f(t, X(t)) - \int_0^t (\tilde{\mathcal{A}}f)(s, X(s))\,\mathrm{d}s$$

is a local martingale under $\tilde{\mathbb{P}}$. To show this, by Lemma 3.1 in Palmowski and Rolski (2002) $\tilde{D}^f(t)$ is a local $\tilde{\mathbb{P}}$-martingale if and only if $\tilde{E}^f(t)$ is a local $\tilde{\mathbb{P}}$-martingale, where $\tilde{E}^f(t)$ is defined as $E^f(t)$ but with $\mathcal{A}$ replaced by $\tilde{\mathcal{A}}$. To show that $\tilde{E}^f(t)$ is a local $\tilde{\mathbb{P}}$-martingale, it suffices to show that $\tilde{E}^f(t)E^g(t)$ is a local $\mathbb{P}$-martingale (Cf. Lemma 4.1 in Palmowski and Rolski (2002)). Straightforward calculus gives

$$\tilde{E}^f(t)E^g(t) = E^{fg}(t),$$

which is indeed a local $\mathbb{P}$ martingale.

Acknowledgements The authors would like to thank the anonymous referees for their helpful comments. We thank the associate editor for detailed comments and M. Schauer (Gothenburg University and Chalmers University of Technology, Sweden) for stimulating discussions on the simulation algorithms.

Funding This work is part of the research project "Bayes for longitudinal data on manifolds" with project number OCENW.KLEIN.218, which is financed by the Dutch Research Council (NWO).

Declarations

Conflict of interest The authors have no Conflict of interest to declare that are relevant to the content of this article.

References

Alt B, Koeppl H (2023) Entropic Matching for Expectation Propagation of Markov Jump Processes. arXiv preprint arXiv:2309.15604

Anderson DF (2007) A modified next reaction method for simulating chemical systems with time dependent propensities and delays. J Chem Phys 127(21):214107

Anderson DF, Kurtz TG (2015) Stochastic analysis of biochemical systems. Springer International Publishing, Mathematical Biosciences Institute Lecture Series

Andrieu C, Roberts GO (2009) The pseudo-marginal approach for efficient Monte Carlo computations. The Annals of Statistics 37(2):697–725
Bersani AM, Bersani E, Mastroeni L (2008) Deterministic and stochastic models of enzymatic networks–applications to pharmaceutical research. Comput Math Appl 55(5):879–888
Beskos A, Roberts G, Stuart A, Voss J (2008) MCMC Methods for diffusion bridges. Stochast Dyn 08(03):319–350
Brémaud P (2020) Probability theory and stochastic processes. Springer
Corstanje M (2023) https://github.com/macorstanje/CRMP.jl. Code examples for simulations
Corstanje M, van der Meulen F, Schauer M (2023) Conditioning continuous-time Markov processes by guiding. Stochastics 95(6):963–996
Moral P del (2017) From Applications to Theory Pierre, & Penev, Spiridon. CRC Press, Stochastic Processes
Fearnhead P (2008) Computational methods for complex stochastic systems: a review of some alternatives to MCMC. Statist Comput 18(2):151–171
Feinberg M (2019) Foundations of chemical reaction network theory. Springer, Cham
Fuchs C (2013) Inference for diffusion processes - with applications in life sciences. Springer-Verlag, Berlin Heidelberg
Georgoulas A, Hillston J, Sanguinetti G (2017) Unbiased Bayesian inference for population Markov jump processes via random truncations. Statist Comput 27(4):991–1002
Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Computat Phys 22(4):403–434
Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81(25):2340–2361
Golightly A, Wilkinson DJ (2008) Bayesian inference for nonlinear multivariate diffusion models observed with error. Comput Statist Data Anal 52(3):1674–1693
Golightly A, Sherlock C (2019) Efficient sampling of conditioned Markov jump processes. Statist Comput 29(5):1149–1163
Golightly A, Wilkinson DJ (2015) Bayesian inference for Markov jump processes with informative observations. Statist Appl Genet Mol Biol 14(2):169–188
Li T (2020) Chemical langevin equation for complex reactions. J Phys Chem A 124(5):810–816
Liptser, Robert S, Shiryaev, Albert N (2013). *Statistics of random processes II: Applications*. Vol. 6. Springer Science & Business Media
Mider M, Schauer M, van der Meulen F (2021) Continuous-discrete smoothing of diffusions. Elect J Statist 15(2):102
Palmowski Z, Rolski T (2002) A technique for exponential change of measure for Markov processes. Bernoulli 8(6):767–785
Rathinam M, Yu M (2021) State and parameter estimation from exact partial state observation in stochastic reaction networks. The Journal of Chemical Physics 154(3):034103
Reeves M, Bhat HS (2022) *Neural Continuous-Time Markov Models*
Sherlock C, Golightly A (2023) Exact Bayesian inference for discretely observed markov jump processes using finite rate matrices. J Computat Graph Statist 32(1):36–48
Van der Meulen F, Schauer M (2020) Automatic Backward Filtering Forward Guiding for Markov processes and graphical models. arXiv preprint arXiv:2010.03509
Warne DJ, Baker RE, Simpson MJ (2019) Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. J Royal Soc Int 16(151):20180943