Conditioning continuous-time Markov processes by guiding

(article starts on next page)

Conditioning continuous-time Markov processes by guiding

Marc Corstanje, Frank van der Meulen & Moritz Schauer

Published online: 05 Dec 2022.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

Conditioning continuous-time Markov processes by guiding

Marc Corstanje [iD][a], Frank van der Meulen [iD][a] and Moritz Schauer [iD][b]

[a]Department of Mathematics, Vrije Universiteit, Amsterdam, Netherlands; [b]Department of Mathematical Sciences, University of Göteburg and Chalmers Technical University, Göteborg, Sweden

ABSTRACT

A continuous-time Markov process X can be conditioned to be in a given state at a fixed time $T > 0$ using Doob's h-transform. This transform requires the typically intractable transition density of X. The effect of the h-transform can be described as introducing a guiding force on the process. Replacing this force with an approximation defines the wider class of guided processes. For certain approximations the law of a guided process approximates – and is equivalent to – the actual conditional distribution, with tractable likelihood-ratio. The main contribution of this paper is to prove that the principle of a guided process, introduced in [M. Schauer, F. van der Meulen, and H. van Zanten, *Guided proposals for simulating multi-dimensional diffusion bridges*, Bernoulli 23 (2017a), pp. 2917–2950. doi:10.3150/16-BEJ833] for stochastic differential equations, can be extended to a more general class of Markov processes. In particular we apply the guiding technique to jump processes in discrete state spaces. The Markov process perspective enables us to improve upon existing results for hypo-elliptic diffusions.

1. Introduction

1.1. Problem description and motivation

Continuous-time Markov processes are widely used for modelling phenomena that evolve over time. Examples include Brownian motion, the Poisson process, Lévy processes in general, and diffusion processes generated by a stochastic differential equation. Many applications require sampling corresponding bridge processes, that is sampling the Markov process X conditional on the value of its trajectory at some time T. For example in the statistical context, with X_t the state of the process at time t, the process is typically only observed at times $t_0 < t_1 < \ldots < t_n$. Based on these observations one may be interested in estimating a parameter θ appearing in the forward description of the process X. In this setting, likelihood based inference is difficult if the transition densities of the process are intractable. However, if the process were observed continuously rather than discretely, then likelihood computations for the continuously observed process on $[0, t_n]$ would typically be easier. This observation has led many authors to employ a data-augmentation

scheme, where one samples iteratively *(i)* $(X_t, t \in [0, t_n])$ conditional on $\{X_{t_i}\}_{i=0}^n$ and θ; *(ii)* θ conditional on $(X_t, t \in [0, t_n])$.

Clearly, step *(i)* requires a way to sample $(X_t, t \in (t_{i-1}, t_i))$ conditional on $X_{t_{i-1}}$ and X_{t_i}, i.e. a *bridge* process. More generally, both data augmentation approaches and approaches based on particle samplers naturally lead to the problem of sampling a bridge process. Whereas for Brownian motion the corresponding bridge process, the Brownian bridge, is fully tractable, for general continuous-time Markov processes this is not the case. It is the aim of this paper to provide a general framework for this.

Without loss of generality, we can state the problem as simulating $(X_t, t \in [0, T])$ conditional on $(X_0, X_T) = (x_0, x_T)$. When feasible, we sometimes consider generalizations of this setting, such as conditioning on $(X_0, LX_T) = (x_0, v)$ for a given matrix L.

1.2. Approach: conditioning by guiding

Our approach builds upon earlier work in the specific setting where X is a diffusion process generated by a stochastic differential equation (SDE). That is, X satisfies the equation

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \quad X_0 = x_0.$$

In this case the problem of bridge sampling has attracted much attention over the past two decades. The approach that we adopt here consists of *guiding*, the terminology originating from [13], the underlying ideas going back to [7,8]. Guiding refers to adjusting the dynamics of the process X to ensure that it hits x_T at time T. This can be done in multiple ways. Following [7], [8] proposed to superimpose the drift of a Brownian bridge to the original drift, leading to the process

$$dX_t^\circ = b(t, X_t^\circ)\, dt + \frac{x_T - X_t^\circ}{T - t}\, dt + \sigma(t, X_t^\circ)\, dW_t, \quad X_0 = x_0.$$

If $\sigma\sigma'$ is strictly positive definite, then indeed $X_T^\circ = x_T$, provided certain smoothness and boundedness conditions on b and σ are satisfied. Equally importantly, they derived an expression for the Radon–Nikodym derivative of the law of the conditioned process, $\mathbb{P}^\star$, with respect to the law of $X^\circ = (X_t^\circ, t \in [0, T])$, which we denote by $\mathbb{P}^\circ$. If $\mathbb{P}_t^\star$ and $\mathbb{P}_t^\circ$ denote the restrictions of $\mathbb{P}^\star$ and $\mathbb{P}^\circ$ to $[0, t]$, respectively, then proving $\mathbb{P}_t^\star \ll \mathbb{P}_t^\circ$ is relatively easy, but the limiting operation $t \uparrow T$ requires careful arguments.

[5,16] considered different, more flexible, guiding terms which can also handle hypoelliptic diffusions. A major effort in these papers consists of formulating sufficient conditions that justify taking the limit $t \uparrow T$. For diffusions on manifolds, introducing guiding terms has recently been introduced in [2,6]. While these works contain numerically convincing results of absolute continuity, no proof is given. Recently [10] proved that the approach of [8] can be extended to simulating Brownian Bridges on Riemannian manifolds.

1.3. Contribution

A first contribution of this paper is to define guided processes by means of an exponential change of measure, rigorously studied in [12]. The beauty of this approach is that it is

not restricted to diffusion processes, but applies generally to continuous-time Markov processes, including for example jump processes. Our main result, Theorem 3.3, gives a simple expression for $\mathrm{d}\mathbb{P}^\star / \mathrm{d}\mathbb{P}^\circ$ and states sufficient conditions to justify the aforementioned limiting operation $t \uparrow T$. These conditions are designed to facilitate this operation in specific examples. We first illustrate the power of our approach to non-homogeneous jump processes and a continuous-time process evolving over a Delaunay triangulation. Secondly, we apply this to diffusion processes, thereby lifting restrictions on the dynamics of the process from [5] which then enable us to fully theoretically justify the bridge simulations in [3]. Some technical proofs are gathered in the appendix.

1.4. Outline

We start this paper by stating the general setting and briefly describing Doob's h-transform. We will also use this section to introduce guided processes that are similar to the guided proposals presented in [5,16,17] for SDEs. In Section 3, we formulate conditions and prove equivalence between the tractable guided process and the intractable true conditional process. In Sections 4 and 5, we apply the theory to Markov processes in a discrete state space and Markov processes that arise as solutions to SDEs.

1.5. Frequently used notation

The transpose of a matrix A is denoted by A'. We denote the smallest and largest eigenvalue of a square matrix A by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. For matrices, we use the spectral norm, which equals the largest singular value of the matrix and is denoted by $\|\cdot\|$. We will also use that for a symmetric, positive definite matrix A, $\|A\| = \sqrt{\lambda_{\max}(A'A)} = \lambda_{\max}(A)$. The determinant and trace of the matrix A are denoted by $|A|$ and $\mathrm{tr}\,(A)$, respectively.

For stochastic differential equations with diffusion coefficient σ, we denote $a = \sigma\sigma'$.

2. General setting, Doob's h-transform and guided processes

Throughout we assume existence of an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $X = \{X_t\}_{t \in [0,T]}$ be a Markov process on a Polish space $\mathcal{S}$ equipped with a σ-algebra $\mathcal{B}$ and define the filtration $\mathcal{F}_t = \sigma(\{X_s : s \le t\})$. Let $x_0 \in \mathcal{S}$. Throughout this paper, we denote $\mathbb{P}_t = \mathbb{P}|_{\mathcal{F}_t}$, $\mathbb{E}_t = \mathbb{E}|_{\mathcal{F}_t}$ and assume all probabilities and expectations are taken conditional on $X_0 = x_0$. For a $\mathbb{P}$-Markov process starting at $x_0 \in \mathcal{S}$ and generated by a family of operators $\mathcal{L} = \{\mathcal{L}_t\}_{t \in [0,T]}$ defined on the same domain $\mathcal{D}(\mathcal{L})$, it holds that

$$M_t^f := f(X_t) - f(x_0) - \int_0^t \mathcal{L}_s f(X_s)\, \mathrm{d}s, \quad t \in (0, T),$$

is a local martingale for any function $f \in \mathcal{D}(\mathcal{L})$. We denote the space-time generator of the space-time process $\{(t, X_t)\}_{t \in [0,T]}$ by $\mathcal{A}f(t, x) = \frac{\partial f}{\partial t}(t, x) + \mathcal{L}_t f(t, x)$ and we extend $\mathcal{D}(\mathcal{L})$ to the domain $\mathcal{D}(\mathcal{A})$ of $\mathcal{A}$. For simplicity, we omit the subscript and write $\mathcal{L}f$ for $f \in \mathcal{D}(\mathcal{L})$. Observe that for $g \in \mathcal{D}(\mathcal{A})$, the process

$$M_t^g := g(t, X_t) - g(0, x_0) - \int_0^t \mathcal{A}g(s, X_s)\, \mathrm{d}s, \quad t \in (0, T), \tag{1}$$

defines a local martingale as well.

Definition 2.1 (*S*-good function): Let $S \leq T$. We call $h \in \mathcal{D}(\mathcal{A})$ an *S*-good function if h is positive and

$$D_t^h = \frac{h(t, X_t)}{h(0, x_0)} \exp\left(-\int_0^t \frac{\mathcal{A}h}{h}(s, X_s)\, ds\right), \quad 0 \leq t \leq S, \tag{2}$$

is a martingale adapted to the filtration $\{\mathcal{F}_t\}_{t \in [0, S]}$.

The following proposition is an adaptation of Proposition 3.2 of [12] and provides conditions for verifying that h is an *S*-good function

Proposition 2.2 (Adaptation of Proposition 3.2 of [12]): *Suppose that $h \in \mathcal{D}(\mathcal{A})$ is such that h and $\frac{\mathcal{A}h}{h}$ are bounded and measurable on $[0, S] \times \mathcal{S}$. Then h is an S-good function.*

Definition 2.3 (Conditioned process): Let $S \leq T$ and h be an *S*-good function. Define the change of measure

$$d\mathbb{P}_t^\star = D_t^h\, d\mathbb{P}_t, \quad t \in [0, S]. \tag{3}$$

We refer to the new measure $\mathbb{P}_t^\star$ as the conditioned measure induced by h and the process X under $\mathbb{P}^\star$ is referred to as the conditioned process induced by h. We denote expectations with respect to $\mathbb{P}_t^\star$ by $\mathbb{E}_t^\star$.

The transformation of measures is known as Doob's h-transform. The function h is typically chosen such that the process X has particular properties under $\mathbb{P}_t^\star$, which it does not possess under $\mathbb{P}_t$. The following example is a key example to illustrate this.

Example 2.4: Suppose the transition kernel of X admits a transition density with respect to a measure ν on $\mathcal{S}$, i.e. $\mathbb{P}(X_t \in A \mid X_s = x) = \int_A p(s, x; t, y)\, d\nu(y)$ for $s \leq t$ and $A \in \mathcal{B}$. For $T > 0$, let the measure $\mathbb{P}^\star$ be the measure induced by $h(t, x) = \int p(t, x; T, y)\, d\mu(y)$ for a probability measure μ under which p is also measurable. Since $\mathcal{A}h = 0$, we have for measurable f

$$\mathbb{E}^\star f(X_t) = \mathbb{E}\left(f(X_t) \frac{h(t, X_t)}{h(0, x_0)}\right)$$

$$= \int f(x) p(0, x_0; t, x) \frac{\int p(t, x; T, y)\, d\mu(y)}{\int p(0, x_0; T, y')\, d\mu(y')}\, d\nu(x)$$

$$= \int \left(\int f(x) \frac{p(0, x_0; t, x) p(t, x; T, y)}{p(0, x_0; T, y)}\, d\nu(x)\right) \frac{p(0, x_0; T, y)\, d\mu(y)}{\int p(0, x_0; T, y')\, d\mu(y')}$$

$$= \int \mathbb{E}\left(f(X_t) \mid X_T = y\right)\, d\xi(y),$$

where

$$d\xi(y) = \frac{p(0, x_0; T, y)\, d\mu(y)}{\int p(0, x_0; T, y')\, d\mu(y')}.$$

In particular, for $x_T \in \mathcal{S}$, the measure $\mu = \delta_{x_T}$ induces the process $X^\star = (X \mid X_T = x_T)$. Measurement error on the value at the endpoint can be incorporated using $d\mu(y) =$

$q_{x_T}(y) \, d\bar{\mu}(y)$, for a a probability density function q_{x_T} and dominating measure $\bar{\mu}$. Lastly, this approach can also be used when forcing diffusions to have a certain distribution at time T, such as seen in [4]. Here one can choose $d\mu(y) = \frac{q(y)}{p(0,x_0;T,y)} \, d\nu(y)$

Unfortunately, many interesting choices of h require the transition density p of the Markov process to be tractable, which is usually not the case. Instead, one can try to use a tractable approximation $\tilde{h}$ to h. This leads to the following definition.

Definition 2.5 (Guided process): Suppose $\tilde{h} \in \mathcal{D}(\mathcal{A})$ is an S-good function for $S \leq T$ and define the change of measure

$$d\mathbb{P}_t^\circ = D_t^{\tilde{h}} \, d\mathbb{P}_t, \quad t \in [0, S].$$

The process X under the law $\mathbb{P}^\circ$ is denoted by X° and is referred to as the guided process induced by $\tilde{h}$. We denote expectations with respect to $\mathbb{P}_t^\circ$ by $\mathbb{E}_t^\circ$.

By formula (1.2) in [12] the extended generators of $X^\star$ and X° are given by

$$\mathcal{A}^\star f = \frac{1}{h} \left[\mathcal{A}(fh) - f \mathcal{A}h \right] \quad \text{and} \quad \mathcal{A}^\circ f = \frac{1}{\tilde{h}} \left[\mathcal{A}(f\tilde{h}) - f \mathcal{A}\tilde{h} \right], \quad (4)$$

which characterizes the dynamics of $X^\star$ and X°, respectively.

Proposition 2.6: *Suppose h and $\tilde{h}$ are S-good functions for some $S \leq T$ and that h is space-time harmonic for $\mathcal{A}$, i.e. $\mathcal{A}h = 0$. Assume $D_t^{\tilde{h}} > 0$ on $[0, S]$. Then, for all $t \leq S$, $\mathbb{P}_t^\star \sim \mathbb{P}_t^\circ$ and*

$$\frac{d\mathbb{P}_t^\star}{d\mathbb{P}_t^\circ}(X) = \frac{h(t, X_t)}{\tilde{h}(t, X_t)} \frac{\tilde{h}(0, x_0)}{h(0, x_0)} \Psi_t(X), \quad t \in [0, S], \quad (5)$$

where

$$\Psi_t(X) = \exp\left(\int_0^t \frac{\mathcal{A}\tilde{h}}{\tilde{h}}(s, X_s) \, ds \right). \quad (6)$$

Proof: Note that, since h and $D_t^{\tilde{h}}$ are positive, both $\mathbb{P}_t^\star$ and $\mathbb{P}_t^\circ$ are equivalent to $\mathbb{P}_t$. Moreover, $d\mathbb{P}_t = (D_t^{\tilde{h}})^{-1} \, d\mathbb{P}_t^\circ$ and thus

$$\frac{d\mathbb{P}_t^\star}{d\mathbb{P}_t^\circ} = \frac{d\mathbb{P}_t^\star}{d\mathbb{P}_t} \frac{d\mathbb{P}_t}{d\mathbb{P}_t^\circ} = \frac{D_t^h}{D_t^{\tilde{h}}}.$$

The proof now follows from substituting (2) and using $\mathcal{A}h = 0$. ∎

In applications, we typically obtain a candidate for $\tilde{h}$ via an auxiliary process.

Definition 2.7 (Auxiliary process): Let $\tilde{X}$ be a Markov process with generator $\tilde{\mathcal{L}}$ and space-time generator $\tilde{\mathcal{A}}$. When we consider a guided process induced by a tractable $\tilde{h}$ that satisfies $\tilde{\mathcal{A}}\tilde{h} = 0$, we refer to $\tilde{X}$ as the auxiliary process. Conditions for absolute continuity can then be stated as properties of $\tilde{X}$.

If $\tilde{h}$ is obtained from an auxiliary process then Ψ takes the form

$$\Psi_t(X) = \exp\left(\int_0^t \frac{\left(\mathcal{L} - \tilde{\mathcal{L}}\right)\tilde{h}}{\tilde{h}}(s, X_s)\, \mathrm{d}s\right).\tag{7}$$

In the setting of Example 2.4 it is often not too hard to find S-good functions h and $\tilde{h}$ where $S < T$. We would like to strengthen this to $S \le T$, i.e. to take the limit $t \uparrow T$ in (5). Sufficient conditions are given in Theorem 3.3, the main result of this paper.

If we assume $\mathrm{d}\mu(y) = q_{x_T}(y)\, \mathrm{d}\bar{\mu}(y)$ in Example 2.4, with q_{x_T} strictly positive, then it is natural to take $\tilde{h}(t, x) = \int q_{x_T}(y)\tilde{p}(t, x; T, y)\, \mathrm{d}\bar{\mu}(y)$. This simplifies showing that $\tilde{h}$ is a T-good function, as under mild conditions Proposition 2.2 can be applied.

For ease of reference, we summarize some of the introduced notation. In the table below, the third column gives the measure on the path space, the fourth column gives the corresponding expectation with respect to the measure, while the rightmost column gives the infinitesimal space-time generator of the process.

X	original, unconditioned Markov process	$\mathbb{P}_t$	$\mathbb{E}_t$	$\mathcal{A}$
$X^\star$	corresponding conditioned processes induced by h	$\mathbb{P}_t^\star$	$\mathbb{E}_t^\star$	$\mathcal{A}^\star$
X°	guided process induced by $\tilde{h}$	$\mathbb{P}_t^\circ$	$\mathbb{E}_t^\circ$	$\mathcal{A}^\circ$

3. Conditions and proof of absolute continuity of $X^\star$ and X° on $[0, T]$

Assume h and $\tilde{h}$ are S-good functions for all $S < T$. By Proposition 2.6, the processes $X^\star$ and X° are absolutely continuous on $\mathcal{F}_t$ for $t < T$. In this section, we provide conditions to ensure that absolute continuity also holds in the limit $t \uparrow T$.

To prove this, we fix a constant $t_0 \in [0, T)$ and impose the following assumptions.

Assumption 3.1: There exists a positive continuous scaling function $\kappa: [t_0, T) \to (0, \infty)$ and a family of $\mathcal{F}_t$-measurable events $\{A_k(t)\}_k$ for each $t \in [t_0, T)$ so that the following assumptions hold

(3.1a) For all k and $t_0 \le s \le t \le T$, $A_k(t) \subseteq A_{k+1}(t)$ and $A_k(t) \subseteq A_k(s)$.

(3.1b) The transition kernel of X admits a transition density p under $\mathbb{P}$ with respect to a dominating measure ν, that is $\mathbb{P}(X_t \in A \mid X_s = x) = \int_A p(s, x; t, y)\, \mathrm{d}\nu(y)$ for $0 \le s \le t \le T$, $A \in \mathcal{B}$, and $x \in \mathcal{S}$. Moreover, for all $s < T$ and $y \in \mathcal{S}$,

$$\lim_{t \uparrow T} h^\sharp(t; s, x) = h(s, x),$$

where

$$h^\sharp(t; s, x) = \int \kappa(t)\tilde{h}(t, y)p(s, x; t, y)\, \mathrm{d}\nu(y).\tag{8}$$

(3.1c) For all $t_0 \le s < t < T$, the random variable $h^\sharp(t; s, X_s)$ is $\mathbb{P}$-almost surely bounded.

(3.1d) κ is such that

$$\lim_{k\to\infty}\lim_{t\uparrow T}\mathbb{E}_t^\star\left(\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\kappa(t)\mathbf{1}_{A_k(t)}\right) = \mathbb{P}_T^\star(A(T)),$$

where $A_k(T) = \cap_{t<T}A_k(t)$, $A(t) = \bigcup_k A_k(t)$ and $A(T) = \cap_{t<T}A(t)$.

(3.1e) For all fixed k, $\Psi_t(X)\kappa(t)\mathbf{1}_{A_k(t)}$ is $\mathbb{P}^\circ$-almost surely uniformly bounded in t.

Lemma 3.2: *Under Assumption 3.1, there exists a random variable $\Psi_T^\kappa(X)$ so that for all k, $\Psi_T^\kappa(X)\mathbf{1}_{A_k(T)} = \lim_{t\uparrow T}\Psi_t(X)\kappa(t)\mathbf{1}_{A_k(t)}$*

Proof: Note that $\log\Psi$ is an integral and κ is continuous, and thus, as $A_k(T) \subseteq A_k(t)$ for all t and by (3.1e), the map $t \mapsto \Psi_t(X)\kappa(t)$ is continuous and bounded on $A_k(T)$ under $\mathbb{P}^\circ$. Hence $\lim_{t\uparrow T}\Psi_t(X)\kappa(t)\mathbf{1}_{A_k(T)}$ exists in the $\mathbb{P}^\circ$-almost sure sense. Clearly, a random variable $\Psi_T^\kappa(X)$ also exists so that $\lim_{t\uparrow T}\Psi_t(X)\kappa(t)\mathbf{1}_{A_k(T)} = \Psi_T^\kappa(X)\mathbf{1}_{A_k(T)}$. Moreover,

$$\left|\Psi_t(X)\kappa(t)\mathbf{1}_{A_k(t)} - \Psi_T^\kappa(X)\mathbf{1}_{A_k(T)}\right| \leq \left|\Psi_t(X)\kappa(t) - \Psi_T^\kappa(X)\right|\mathbf{1}_{A_k(T)}$$
$$+ \left|\Psi_t(X)\kappa(t)\right|\mathbf{1}_{A_k(t)\setminus A_k(T)}$$

It follows from the preceding that the first term tends to 0 while the second term is bounded by (3.1e) and the indicator tends to 0. ∎

Theorem 3.3: *Suppose Assumption 3.1 is satisfied. Then for any measurable function f,*

$$\mathbb{E}_T^\star\left(f(X)\mathbf{1}_{A(T)}\right) = \mathbb{E}_T^\circ\left(f(X)\frac{\tilde{h}(0,x_0)}{h(0,x_0)}\Psi_T^\kappa(X)\mathbf{1}_{A(T)}\right) \tag{9}$$

In particular, if $\mathbb{P}_T^\circ(A(T)) = 1$, the measures are equivalent with

$$\frac{d\mathbb{P}_T^\star}{d\mathbb{P}_T^\circ}(X) = \frac{\tilde{h}(0,x_0)}{h(0,x_0)}\Psi_T^\kappa(X).$$

Proof of Theorem 3.3: For simplicity, we denote $\bar{h} = \tilde{h}(0,x_0)/h(0,x_0)$. The proof is structured as follows. First we show that $\mathbb{E}_T^\circ(\bar{h}\Psi_T^\kappa(X)\mathbf{1}_{A(T)}) = \mathbb{P}_T^\star(A(T))$, then we show that $\mathbb{E}_t^\circ(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)}) \to \mathbb{P}_T^\star(A(T))$ as $t \uparrow T$. Finally, we finish the proof via Scheffé's lemma.

First note that for any fixed k, it follows from dominated convergence, combined with (3.1e), Lemma 3.2 and Proposition 2.6 that

$$\mathbb{E}_T^\circ\left(\bar{h}\Psi_T^\kappa(X)\mathbf{1}_{A_k(T)}\right) = \lim_{t\uparrow T}\mathbb{E}_t^\circ\left(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A_k(t)}\right) = \lim_{t\uparrow T}\mathbb{E}_t^\star\left(\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\kappa(t)\mathbf{1}_{A_k(t)}\right).$$

We now send $k \to \infty$ on both sides and find that, by (3.1a) combined with monotone convergence, the left hand side tends to $\mathbb{E}_T^\circ(\bar{h}\Psi_T^\kappa(X)\mathbf{1}_{A(T)})$ while the right hand side tends to $\mathbb{P}_T^\star(A(T))$ by (3.1d). Hence $\mathbb{E}_T^\circ(\bar{h}\Psi_T^\kappa(X)\mathbf{1}_{A(T)}) = \mathbb{P}_T^\star(A(T))$

Now note that for any fixed m, it follows from Proposition 2.6 that

$$\lim_{t\uparrow T}\mathbb{E}^{\circ}_t\left(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)}\right)\geq\lim_{t\uparrow T}\mathbb{E}^{\circ}_t\left(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A_m(t)}\right)=\lim_{t\uparrow T}\mathbb{E}^{\star}_t\left(\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\kappa(t)\mathbf{1}_{A_m(t)}\right).$$

Upon sending $m\to\infty$, we have, by (3.1d), $\lim_{t\uparrow T}\mathbb{E}^{\circ}_t(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)})\geq\mathbb{P}^{\star}_T(A(T))$. For the other inequality, we note that for any $s<T$, by Proposition 2.6,

$$\lim_{t\uparrow T}\mathbb{E}^{\circ}_t\left(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)}\right)\leq\lim_{t\uparrow T}\mathbb{E}^{\circ}_t\left(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A(s)}\right)\tag{10}$$

$$=\lim_{t\uparrow T}\mathbb{E}^{\star}_t\left(\mathbf{1}_{A(s)}\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\kappa(t)\right)=\mathbb{P}^{\star}_s(A(s))\tag{11}$$

where the last equality follows upon taking $g_s=\mathbf{1}_{A(s)}$ in Lemma A.1. Upon sending $s\uparrow T$, we find by monotonicity of measures, $\lim_{t\uparrow T}\mathbb{E}^{\circ}_t(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)})\leq\mathbb{P}^{\star}_T(A(T))$.

We thus conclude that $\mathbb{E}^{\circ}_t(\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)})\to\mathbb{E}^{\circ}_T(\bar{h}\Psi^{\kappa}_T(X)\mathbf{1}_{A(T)})$. Now note that upon taking $\lim_{k\to\infty}$ as well in Lemma 3.2 and interchanging $\lim_{k\to\infty}$ and $\lim_{t\uparrow T}$, we have that $\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)}\to\Psi^{\kappa}_T(X)\mathbf{1}_{A(T)}$ in the $\mathbb{P}^{\circ}$-almost sure sense. Here the interchange of limits is allowed as the sequence is monotone in k. Hence, by Scheffé's lemma, $\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)}\to\Psi^{\kappa}_T(X)\mathbf{1}_{A(T)}$ in $L^1(\mathbb{P}^{\circ})$.

Now let $s<T$ and let f_s be any bounded positive $\mathcal{F}_s$-measurable function such that the support of $f_s(X)$ is contained in $A(T)$. Then it follows from L^1 convergence and Proposition 2.6 that

$$\mathbb{E}^{\circ}_T\left(f_s(X)\bar{h}\Psi^{\kappa}_T(X)\mathbf{1}_{A(T)}\right)=\lim_{t\uparrow T}\mathbb{E}^{\circ}_t\left(f_s(X)\bar{h}\Psi_t(X)\kappa(t)\mathbf{1}_{A(t)}\right)$$

$$=\lim_{t\uparrow T}\mathbb{E}^{\star}_t\left(f_s(X)\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\kappa(t)\mathbf{1}_{A(t)}\right).$$

Since $f_s(X)$ has support only on $A(T)$ and for $t\in[s,T)$, $A(T)\subseteq A(t)\subseteq A(s)$, we have that $f_s(X)\mathbf{1}_{A(s)}=f_s(X)\mathbf{1}_{A(t)}=f_s(X)\mathbf{1}_{A(T)}$. Hence, upon applying Lemma A.1 to $f_s\mathbf{1}_{A(s)}$,

$$\lim_{t\uparrow T}\mathbb{E}^{\star}_t\left(f_s(X)\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\kappa(t)\mathbf{1}_{A(t)}\right)=\lim_{t\uparrow T}\mathbb{E}^{\star}_t\left(f_s(X)\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\kappa(t)\mathbf{1}_{A(s)}\right)$$

$$=\mathbb{E}^{\star}_s\left(f_s(X)\mathbf{1}_{A(s)}\right)=\mathbb{E}^{\star}_T\left(f_s(X)\mathbf{1}_{A(T)}\right).$$

For the equivalence, it suffices to show that $\mathbb{P}^{\circ}_T(A(T))=1\implies\mathbb{P}^{\star}_T(A(T))=1$. Note that by monotonicity $\mathbb{P}^{\star}(A(T))=\lim_{t\uparrow T}\mathbb{P}^{\star}(A(t))$. Now for all $t<T$, $\mathbb{P}^{\circ}(A(t))\geq\mathbb{P}^{\circ}(A(T))$ and thus the result follows from Proposition 2.6. ∎

3.1. Discussion of Assumption 3.1

The order of verifying the assumptions in Assumption 3.1 is often as follows. We first choose the function κ so that (3.1d) holds, then we compute $\Psi_t(X)\kappa(t)$ and find V so that $\Psi_t(X)\kappa(t)$ is bounded whenever V is bounded to ensure (3.1e) holds. We then have (3.1a) via Lemma. 3.4 and, if possible, eliminate the indicator in the expression for the

Radon–Nikodym derivative and thereby proof equivalence by showing that V is almost surely bounded.

The function κ is incorporated to enable compensating for a possible difference of smoothness in $\tilde{X}$ and X. In Section 4.2, we consider an example where X is a process in a discrete state space, while $\tilde{X}$ is a continuous process. In such a setting, incorporating a nonconstant κ is essential. If for example both X and $\tilde{X}$ are solutions to SDEs, we can always take $\kappa = 1$.

Proposition 3.4: *If the events $A_k(t)$ are defined by*

$$A_k(t) = \left\{ \sup_{t_0 \le s < t} V(x, X_s) \le k \right\} \tag{12}$$

for a nonnegative function $V : [t_0, T) \times \mathcal{S} \to [0, \infty)$, then (3.1a) is satisfied. If also $\sup_t V(t, X_t)$ is $\mathbb{P}^\circ$-almost surely bounded, then the Radon–Nikodym derivative in Theorem 3.3 can be simplified since $\mathbb{P}^\circ_T(A(T)) = 1$ and thus the measures are equivalent.

The following proposition gives a condition for verifying $\mathbb{P}^\circ$-almost sure boundedness of V and thus showing $\mathbb{P}^\circ_T(A(T)) = 1$.

Proposition 3.5: *Suppose $\mathcal{A}^\circ V \le 0$, then there exists a random variable C such that $\sup_{t_0 \le t < T} V(t, X_t) \le C$ and $\mathbb{P}^\circ(C \ge \lambda) \le \lambda^{-1} \mathbb{E}_{\mathbb{P}^\circ} V(t_0, X_{t_0})$ for any $\lambda > 0$.*

Proof: We first show that $\{V(t, X_t)\}_{t_0 \le t < T}$ is a supermartingale. Note that by (1), a $\mathbb{P}^\circ$-local martingale M^V exists so that under $\mathbb{P}^\circ$

$$V(t, X_t) = V(t_0, X_{t_0}) + M_t^V + \int_{t_0}^t \mathcal{A}^\circ V(s, X_s) \, ds.$$

Since V is nonnegative and $\mathcal{A}^\circ V \le 0$, it follows that M^V is bounded from below by $-V(t_0, X_{t_0})$. Hence, by Lemma A.4, M^V is a supermartingale. Moreover, it follows that $V(t, X_t)$ is integrable for all t since

$$\mathbb{E}_t^\circ |V(t, X_t)| = \mathbb{E}_t^\circ V(t, X_t) \le \mathbb{E}_{t_0}^\circ V(t_0, X_{t_0}) + \mathbb{E}_t^\circ M_t^V$$

The supermartingale property follows as for $t_0 \le s \le t < T$,

$$\mathbb{E}_t^\circ (V(t, X_t) \mid \mathcal{F}_s) - V(s, X_s) = \mathbb{E}_t^\circ (M_t^V \mid \mathcal{F}_s) - M_s^V + \mathbb{E}_t^\circ \int_s^t \mathcal{A}^\circ V(u, X_u) \, du \le 0$$

Doob's supermartingale inequality, see e.g. Theorem 1.3.6 of [11], now states that for all $\lambda > 0$,

$$\lambda \mathbb{P}^\circ \left(\sup_{t_0 \le t < T} V(t, X_t) \ge \lambda \right) \le \mathbb{E}_{t_0}^\circ V(t_0, X_{t_0}).$$

Hence, a random variable C exists so that $\mathbb{P}^\circ$-almost surely $\sup_{t_0 \le t < T} V(t, X_t) \le C$ where $\mathbb{P}^\circ(C \ge \lambda) \le \lambda^{-1} \mathbb{E}_{\mathbb{P}^\circ} V(t_0, X_{t_0})$. ∎

4. Application 1: discrete state-space processes

Here we discuss two Markov processes that take values in a discrete state space.

4.1. Inhomogeneous poisson process

Let X be an inhomogeneous Poisson process with state-dependent rate λ, that is

$$\mathbb{P}\left(X_{t+\Delta} = x + 1 \mid X_t = x\right) = \lambda(x)\Delta + o(\Delta). \tag{13}$$

The infinitesimal generator of X is given by

$$\mathcal{L}f(x) = \lambda(x)\left(f(x+1) - f(x)\right). \tag{14}$$

We assume $X_0 = x_0$, fix $x_T > x_0$ and we consider the process $X^\star = (X \mid X_T = x_T)$. Moreover, we assume that λ takes finite values on the set $\{x_0, x_0 + 1, \ldots, x_T\}$. $X^\star$ is obtained from Doob's h-transform with $h(t, x) = \mathbb{P}(X_T = x_T \mid X_t = x)$. It follows from (4) that $X^\star$ is an inhomogeneous Poisson process with rate $\lambda^\star(t, x) = \lambda(x)\frac{h(t,x+1)}{h(t,x)}$ on $\{x_0, \ldots, x_T - 1\}$ and 0 at $x = x_T$. Since λ is state-dependent, the transition probabilities for X are intractable and thus the exact form of the process $X^\star$ cannot be determined. We thus simulate a guided process with a homogeneous Poisson process with rate $\tilde{\lambda}$ as auxiliary process, i.e.

$$\tilde{h}(t, x) = \begin{cases} \frac{\left(\tilde{\lambda}(T-t)\right)^{x_T - x}}{(x_T - x)!} e^{-\tilde{\lambda}(T-t)}, & (t, x) \in [0, T) \times \{x_0, \ldots, x_T\} \\ 0 & \text{elsewhere} \end{cases}. \tag{15}$$

Proposition 4.1: *Both h and $\tilde{h}$ are S-good functions for all $S < T$.*

Proof: Let $S < T$. Since h is a non-zero probability and $\mathcal{A}h = 0$, it immediately follows from Proposition 2.2 that h is a good function. It follows from a direct computation that

$$\frac{\mathcal{A}\tilde{h}}{\tilde{h}}(t, x) = \left(\lambda(x) - \tilde{\lambda}\right)\left(\frac{x_T - x}{\tilde{\lambda}(T - t)} - 1\right), \quad (t, x) \in [0, S] \times \{x_0, \ldots, x_T\}$$

Hence, since $S < T$, both $\tilde{h}$ and $\frac{\mathcal{A}\tilde{h}}{\tilde{h}}$ are bounded measurable on $[0, S] \times \mathcal{S}$ and $\tilde{h}$ is an S-good function by Proposition 2.2. ∎

Theorem 4.2: *Let X° be the guided process induced by (15) and suppose that $\tilde{\lambda} \leq \min\{\lambda(x) \colon x \in \{x_0, \ldots, x_T\}\}$. Then the laws of $X^\star$ and X° are equivalent on $[0, T]$. Moreover,*

$$\frac{d\mathbb{P}_T^\star}{d\mathbb{P}_T^\circ}(X) = \frac{\tilde{h}(0, x_0)}{h(0, x_0)} \exp\left(\int_0^T \left[\lambda(X_s) - \tilde{\lambda}\right]\left[\frac{x_T - X_s}{\tilde{\lambda}(T - t)} - 1\right] ds\right). \tag{16}$$

Proof: Set $\kappa(t) = 1$, $t_0 = 0$ and define $\{A_k(t)\}_{k,t}$ as in (12) with

$$V(t, x) = \frac{x_T - x}{\tilde{\lambda}(T - t)}, \quad (t, x) \in [0, T) \times \{x_0, \ldots, x_T\}, \tag{17}$$

so that (3.1a) is satisfied. The result follows follows from an application of Theorem 3.3, where the assumptions from Assumption 3.1 are satisfied via Lemmas 4.3– 4.5. The form

of the Radon–Nikodym derivative is obtained via (7) and upon noting that $\mathbb{P}^\circ_T(A(T)) = 1$ by Proposition 3.5 and Lemma 4.6. The latter also implies equivalence. ∎

In the remainder of this section, we prove the results used in the proof of Theorem 4.2 and we thus assume the conditions stated in this theorem are satisfied and take κ, t_0, V and $\{A_k(t)\}_{k,t}$ as stated in the proof.

Lemma 4.3: *Assumptions (3.1b) and (3.1c) are satisfied.*

Proof: First note that X admits a transition density p with respect to the counting measure. Let $s \in [0, T)$ and $y \in \{x_0, \ldots, x_T\}$. Then

$$
h^\sharp(t; s, x) = \int \kappa(t)\tilde{h}(t, y)p(s, x; t, y)\,d\nu(y)
$$

$$
= \sum_{y=x_0}^{\nu} \frac{\left(\tilde{\lambda}(T-t)\right)^{x_T-y}}{(x_T-y)!}e^{-\tilde{\lambda}(T-t)}p(s, x; t, y)
$$

$$
= e^{-\tilde{\lambda}(T-t)}p(s, x; t, x_T) + \sum_{y=x_0}^{x_T-1} \frac{\left(\tilde{\lambda}(T-t)\right)^{x_T-y}}{(x_T-y)!}e^{-\tilde{\lambda}(T-t)}p(s, x; t, y).
$$

As $t \uparrow T$, the first term on the right hand side tends to $p(s, x; T, x_T) = h(s, x)$, while the second term vanishes as p is bounded by 1. It can also be observed that $h^\sharp$ is uniformly bounded by 1 and thus $h^\sharp(t; s, X_s)$ is clearly bounded. ∎

Lemma 4.4:

$$
\lim_{k\to\infty}\lim_{t\uparrow T}\mathbb{E}^\star_t\left(\frac{\tilde{h}(t, X_t)}{h(t, X_t)}\mathbf{1}_{A_k(t)}\right) = \mathbb{P}^\star_T(A(T)).
$$

Proof: We first show that for all k, $\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\mathbf{1}_{A_k(t)} \to \mathbf{1}_{A_k(T)}$ in the $\mathbb{P}^\star$-almost sure sense as $t \uparrow T$. By (12) and (17),

$$
A_k(t) = \left\{\sup_{0\le s<t}\frac{x_T - X_s}{\tilde{\lambda}(T-s)} \le k\right\}
$$

Hence, for all trajectories $\omega \in A_k(T)$, we must have an $\varepsilon(\omega) > 0$ such that $X^\star_t(\omega) = x_T$ for $t \in (T - \varepsilon(\omega), T]$.

By (13) for t sufficiently close to T,

$$
h(t, x) = \begin{cases} \lambda(x)(T-t) + o(T-t), & x < x_T \\ 1 - \lambda(x)(T-t) + o(T-t), & x = x_T \end{cases}.
$$

It thus follows that $\lim_{t\uparrow T} h(t, X^\star_t(\omega)) = 1$ for all $\omega \in A_k(T)$. Similarly, we deduce from (15) that also $\lim_{t\uparrow T} \tilde{h}(t, X^\star_t(\omega)) = 1$ for all $\omega \in A_k(T)$. Now

$$
\frac{\tilde{h}(t, X_t)}{h(t, X_t)}\mathbf{1}_{A_k(t)} = \frac{\tilde{h}(t, X_t)}{h(t, X_t)}\mathbf{1}_{A_k(T)} + \frac{\tilde{h}(t, X_t)}{h(t, X_t)}\mathbf{1}_{A_k(t)\setminus A_k(T)}
$$

We proceed to show that $\frac{\tilde{h}(t,X_t)}{h(t,X_t)}$ is bounded on $A_k(t) \setminus A_k(T)$. Note

$$A_k(t) \setminus A_k(T) = \left\{ \sup_{t \leq s < T} \frac{x_T - X_s}{\tilde{\lambda}(T - s)} > k \right\} \tag{18}$$

Hence, on $A_k(t) \setminus A_k(T)$, we must have $X_t < x_T$. Now

$$\frac{\tilde{h}(t, X_t)}{h(t, X_t)} = \frac{\tilde{\lambda}^{x_T - X_t}(T - t)^{x_T - X_t - 1}}{\left(\lambda(X_t) + \frac{o(T-t)}{T-t}\right)(x_T - X_t)!} e^{-\tilde{\lambda}(T-t)}$$

Since, $x_T - X_t \geq 1$, this term is clearly bounded in t. By the preceding, under $\mathbb{P}^\star$, the first term tends to $\mathbf{1}_{A_k(T)}$ as $t \uparrow T$, while the second term in (18) tends to 0. The preceding also implies that $\frac{\tilde{h}(t,X_t)}{h(t,X_t)} \mathbf{1}_{A_k(t)}$ is bounded in t, and thus by dominated convergence,

$$\lim_{k\to\infty} \lim_{t\uparrow T} \mathbb{E}^\star_t \left(\frac{\tilde{h}(t, X_t)}{h(t, X_t)} \mathbf{1}_{A_k(t)} \right) = \lim_{k\to\infty} \mathbb{E}^\star_T \mathbf{1}_{A_k(T)} = \mathbb{P}^\star_T(A(T)),$$

where the last equality follows from monotone convergence ∎

Lemma 4.5: *For all k, $\Psi_t(X)\mathbf{1}_{A_k(t)}$ is $\mathbb{P}^\circ$-almost surely uniformly bounded in t.*

Proof: Observe that for $x \in \{x_0, \ldots, x_T - 1\}$,

$$V(t, x) = \frac{\tilde{h}(t, x + 1)}{\tilde{h}(t, x)}$$

Now note that, by (7),

$$\begin{aligned}
\log \Psi_t(X) &= \int_0^t \frac{\left(\mathcal{L} - \tilde{\mathcal{L}}\right)\tilde{h}}{\tilde{h}}(s, X_s)\, \mathrm{d}s \\
&= \int_0^t \frac{\lambda(X_s) - \tilde{\lambda}}{\tilde{h}(s, X_s)} \left(\tilde{h}(s, X_s + 1) - \tilde{h}(s, X_s)\right)\, \mathrm{d}s \\
&= \int_0^t \left(\lambda(X_s) - \tilde{\lambda}\right)(V(s, X_s) - 1)\, \mathrm{d}s.
\end{aligned}$$

Now since λ is nonnegative and $\tilde{\lambda} \leq \lambda(x)$ for all $x \in \{x_0, \ldots, x_T\}$, we have that for all $t \in [0, T)$,

$$\Psi_t(X)\mathbf{1}_{A_k(t)} \leq \exp\left((k - 1) \int_0^t \left(\lambda(X_s) - \tilde{\lambda}\right)\, \mathrm{d}s\right) \leq \exp\left((k - 1) \int_0^T \left(\lambda(X_s) - \tilde{\lambda}\right)\, \mathrm{d}s\right),$$

where the last term is $\mathbb{P}^\circ$-almost surely integrable. ∎

Lemma 4.6: $\mathcal{A}^\circ V(t, x) \leq 0$ for all $t \in [0, T]$ and $x \in \{x_0, \ldots, x_T\}$.
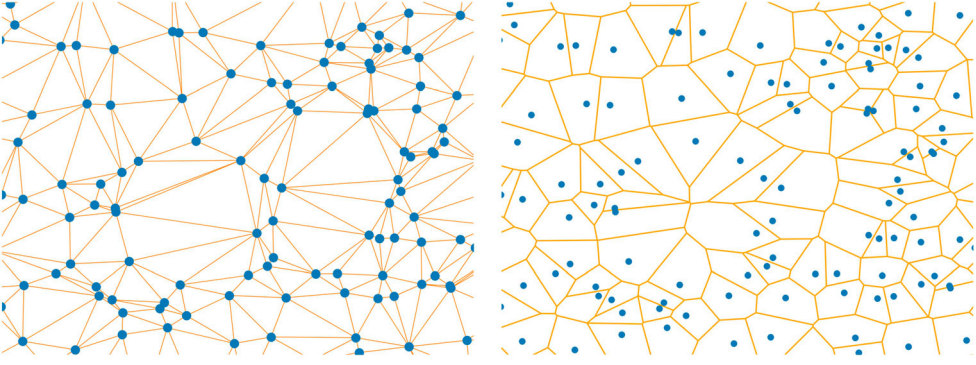
Figure 1. The blue points indicate a realization of a unit rate planar Poisson process with (left) the Delaunay triangulation and (right) the Voronoi diagram of the set.

Proof: The result follows from a direct computation. $A^{\circ}(t, x_T) = 0$ and for $x \in \{x_0, \ldots, x_T - 1\}$,

$$
\begin{aligned}
\mathcal{A}^{\circ} V(t, x) &= \frac{\partial V}{\partial t}(t, x) + \lambda^{\circ}(t, x) \left[V(t, x+1) - V(t, x)\right] \\
&= \frac{\partial V}{\partial t}(t, x) + \lambda(x) \frac{\tilde{h}(t, x+1)}{\tilde{h}(t, x)} \left[V(t, x+1) - V(t, x)\right] \\
&= \frac{x_T - x}{\tilde{\lambda}(T - t)^2} - \lambda(x) \frac{x_T - x}{\left[\tilde{\lambda}(T - t)\right]^2} = \frac{x_T - x}{\tilde{\lambda}(T - t)} \left(1 - \frac{\lambda(x)}{\tilde{\lambda}}\right) \leq 0. \qquad \blacksquare
\end{aligned}
$$

4.2. Jump process on a delaunay triangulation

Here, we consider a toy example to model electric flow through a city. In [9], electricity is modelled as a random walk on a discrete grid and here we slightly alter this model by considering jump processes on a triangulation of a random set of points in the plane $\mathbb{R}^2$. The network is constructed by first sampling points in the plane according to a planar Poisson process, followed by adding connections according to a Delaunay triangulation (we recap its definition below). We assume that at a given location electricity moves to a randomly chosen neighbour. We condition the model by a starting point and a final point on the grid and model the flow of electricity between the two points.

Definition 4.7 (Voronoi diagram and Delaunay triangulation): Let P be a set of points in $\mathbb{R}^d$. The Voronoi diagram associated with P is the collection of Voronoi cells

$$
\mathcal{V}_P(x) := \left\{y \in \mathbb{R}^d \colon \left|y - x\right| \leq \left|y - z\right|, \text{ for all } z \in P\right\}, \quad x \in P.
$$

The Delaunay triangulation $D(P)$ of P is the dual graph of the Voronoi diagram. It has P as vertex set and there is an edge between x and y in $D(P)$ if $\mathcal{V}_P(x)$ and $\mathcal{V}_P(y)$ share a $(d-1)$-dimensional face. An example of a Delaunay traingulation and a Voronoi diagram can be found in Figure 1.

In this section, we consider a realization ξ of a unit rate planar Poisson process that satisfies the following properties:

- for all compact $A \subseteq \mathbb{R}^2$, the set $A \cap \xi$ is finite.
- There is a constant K such that for any pair x, y of vertices in $D(\xi)$ satisfies $|y - x| \leq K$.

Note that properties are events of probability 1.

Denote the Delaunay triangulation of ξ by $D(\xi)$. Consider a unit rate jump process $X = \{X_t\}_{t \geq 0}$ on $D(\xi)$ defined through the generator

$$\mathcal{L}f(x) = \sum_{y \in N_\xi(x)} (f(y) - f(x)), \quad x \in \xi, \tag{19}$$

where $N_\xi(x)$ denotes the set of neighbours to $x \in \xi$ in $D(\xi)$. For fixed $x_T \in \xi$ and $T > 0$, we consider the process conditioned $X^\star = (X \mid X_T = x_T)$ induced by $h(t, x) = \mathbb{P}(X_T = x_T \mid X_t = x)$. By Theorem 1 of [15], the time-changed scaled process $X^\varepsilon = \{\varepsilon X_{t/\varepsilon^2}\}_{t \geq 0}$ converges in law to a scaled Brownian motion as $\varepsilon \downarrow 0$. Moreover, the scale parameter $\tilde{\sigma} > 0$ of the Brownian motion does not depend on the realization of ξ, but it does depend on the rate of ξ. We thus have an immediate candidate for $\tilde{h}$ in

$$\tilde{h}(t, x) = \eta(t) \exp\left(-H(t, x)\right), \quad \text{with}$$

$$\eta(t) = \frac{1}{2\pi \tilde{a}(T - t)} \quad \text{and} \quad H(t, x) = \frac{|x_T - x|^2}{2\tilde{a}(T - t)}, \tag{20}$$

with $\tilde{a} = \tilde{\sigma}^2$. By (4), the guided process X° induced by $\tilde{h}$ is generated by

$$\mathcal{L}^\circ f(x) = \frac{1}{\tilde{h}(t, x)} \left[\mathcal{L}\left(\tilde{h}f\right)(t, x) - f(x)\mathcal{L}\tilde{h}(t, x) \right] = \sum_{y \in N_\xi(x)} \frac{\tilde{h}(t, y)}{\tilde{h}(t, x)}(f(y) - f(x)). \tag{21}$$

Hence the guided process is a jump process on $D(\xi)$ with state-dependent jump rates. The rate of jumping from x to $y \in N_\xi(x)$ is given by

$$\frac{\tilde{h}(t, y)}{\tilde{h}(t, x)} = \exp\left(-\frac{|x_T - y|^2 - |x_T - x|^2}{2\tilde{a}(T - t)}\right) \tag{22}$$

Proposition 4.8: *Both h and $\tilde{h}$ are S-good functions for all $S < T$.*

Proof: Let $S < T$. h is defined as a non-zero probability and $\mathcal{A}h = 0$ and thus h is S-good by Proposition 2.2. Since $\mathcal{A}$ solves the martingale problem for the process X under $\mathbb{P}$, it follows from Lemma 3.1 of [12] that $D_t^{\tilde{h}}$ is a $\mathbb{P}$-local martingale. We proceed to show that $D_t^{\tilde{h}}$ is $\mathbb{P}$-almost surely bounded on $[0, S]$, so that the result follows from Theorem 47 of [14].

direct computation yields

$$\frac{\mathcal{A}\tilde{h}}{\tilde{h}}(t,x) = \frac{\mathcal{L}\tilde{h}}{\tilde{h}}(t,x) + \frac{d\log\eta}{dt} - \frac{\partial H}{\partial t} = \sum_{y\in N_\xi(x)} \left(\frac{\tilde{h}(t,y)}{\tilde{h}(t,x)} - 1\right) + \frac{d\log\eta}{dt} - \frac{\partial H}{\partial t} \quad (23)$$

Using (23), we obtain

$$D_t^{\tilde{h}} = \exp\left(\int_0^t \frac{\partial H}{\partial s}(s,X_s)\,ds - H(t,X_t) + H(0,x_0) - \int_0^t \sum_{y\in N_\xi(X_s)} \left(\frac{\tilde{h}(s,y)}{\tilde{h}(s,X_s)} - 1\right)ds\right)$$

Notice that

$$\int_0^t \frac{\partial H}{\partial s}(s,X_s)\,ds - H(t,X_t) = \int_0^t \frac{|x_T - X_s|^2}{2\tilde{a}(T-s)^2}\,ds - \frac{|x_T - X_t|^2}{2\tilde{a}(T-t)}$$

is $\mathbb{P}$-almost surely bounded on $[0,S]$. It remains to be shown that

$$\sum_{y\in N_\xi(x)} \frac{\tilde{h}(t,y)}{\tilde{h}(t,x)} = \exp\left(-\frac{1}{2\tilde{a}(T-t)}\left[|x_T - y|^2 - |x_T - x|^2\right]\right)$$

is bounded in x. For any $x \in \xi$ and $y \in N_\xi(x)$, $|x_T - y|^2 - |x_T - x|^2 = \langle 2x_T, x - y\rangle + |y|^2 - |x|^2$. Since, $y \in N_\xi(x)$, if follows from Cauchy-Schwartz that $\langle 2x_T, x - y\rangle \leq 2|x_T|K$. Moreover, since the function $(a,b) \mapsto b^2 - a^2$ is bounded on any compact disc and $||y| - |x|| \leq |y - x| \leq K$, $|y|^2 - |x|^2$ is bounded as well. ∎

Theorem 4.9: *The guided process induced by (20) is equivalent to $X^\star = (X \mid X_T = x_T)$ on $[0,T]$. Moreover,*

$$\frac{d\mathbb{P}_T^\star}{d\mathbb{P}_T^\circ}(X) = 2\pi\tilde{a}T\frac{\tilde{h}(0,x_0)}{h(0,x_0)} \exp\left(\int_0^T \left[\sum_{y\in N_\xi(X_s)} \left(\frac{\tilde{h}(s,y)}{\tilde{h}(s,X_s)} - 1\right) - \frac{|x_T - X_s|^2}{2\tilde{a}(T-s)^2}\right]ds\right).$$

Example 4.10: Figure 2 demonstrates an application of Theorem 4.9. We simulated a Poisson-Delaunay grid with intensity 5000 and chose as starting point the vertex closest to coordinates (1.25, 1.25). We considered the problem of conditioning on ending in the vertex with coordinates approximately equal to (1.75, 1.75) at time $T = 30$. Although the scale $\tilde{a}$ is not affecting the validity of Theorem 4.9, a good choice can help to improve the quality of the samples. That is, a wrong choice can lead to samples arriving at the endpoint too early or too late and, if a Metropolis-Hastings sampling algorithm is used to sample from the Radon–Nikodym derivative in Theorem 4.9, the acceptance rate suffers. Here, we estimated $\tilde{a} \approx 0.13$ by generating a long trajectory of the jump process and computing its quadratic variation.

For the proof of Theorem 4.9, we need the following property of $D(\xi)$

Proposition 4.11: *Let $x, v \in \xi$ with $x \neq x_T$. Then there exists an $y \in N_\xi(x)$ so that $|x_T - y| < |x_T - x|$.*
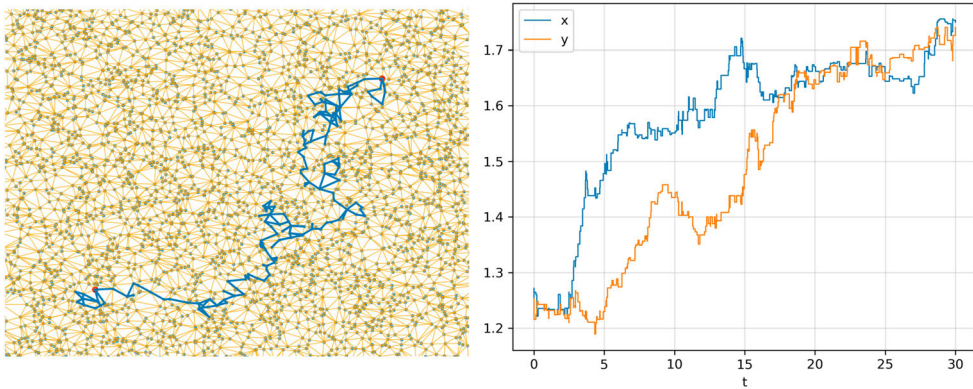
Figure 2. Illustration for Example 4.10. Left: jump process bridge between two vertices in a random Poisson-Delaunay graph – simulated using a guided process derived from the Brownian motion scaling limit of the random walk. Right: coordinate processes versus time corresponding to the same bridge realization.

Proof: If $x_T \in N_\xi(x)$, set $y = x_T$. If $x_T \notin N_\xi(x)$ we denote the Voronoi cell of x by $\mathcal{V}_\xi(x)$. Since x is not on the boundary of $\mathcal{V}_\xi(x)$, there is an edge in the boundary of $\mathcal{V}_\xi(x)$ that intersects the line segment between x_T and x. If two edges in $\mathcal{V}_\xi(x)$ meet on this line segment, we may choose either one of them. We choose y as the point in the Voronoi cell adjacent to this boundary edge of $\mathcal{V}_\xi(x)$. By construction, $y \in N_\xi(x)$ and by the triangle inequality $|x_T - y| < |x_T - x|$.

Note that the inequality is strict, as two points x, y with $|x_T - x| = |x_T - y|$ cannot have the Voronoi edge in between them also intersecting the prescribed line segment between x and x_T as the triangle with vertices x, y, x_T is an isosceles triangle. ∎

Proof of Theorem 4.9: We prove Theorem 4.9 via Theorem 3.3. Set $t_0 = 0$, $\kappa(t) = 1/\eta(t)$ and let

$$V(t, x) = \sum_{y \in N_\xi(x)} \frac{\tilde{h}(t, y)}{\tilde{h}(t, x)} \tag{24}$$

As in Lemma 3.4, define the events

$$A_k(t) = \left\{ \sup_{0 \le s < t} V(s, X_s) \le k \right\}, \tag{25}$$

We first note that (3.1a) is satisfied by construction. The remaining assumptions of Assumption 3.1 are satisfied by Lemmas 4.12, 4.4 and 4.5. Finally, the equivalence and the form of the Radon–Nikodym derivative follows from a direct computation and upon noting that $\mathbb{P}^\circ_T(A(T)) = 1$ by Corollary 4.16. ∎

In the remainder of this section, we prove the lemmas needed for Theorem 4.9. We thus assume that the conditions are satisfied and set $t_0 = 0$ and $\kappa(t) = 1/\eta(t)$ and choose the events $\{A_k(t)\}_{k,t}$ as in (25) with V as in (24).

Lemma 4.12: *Assumptions* (3.1b) *and* (3.1c) *are satisfied.*

Proof: The proof is similar to the proof of Theorem 4.3. First note that $\kappa(t)\tilde{h}(t,x) = \exp(-H(t,x))$. Now X admits a transition density p with respect to the counting measure and thus for $s \in [0, T)$ and $y \in \xi$,

$$
\begin{aligned}
h^{\sharp}(t; s, x) &= \int \kappa(t)\tilde{h}(t, y) p(s, x; t, y) \, d\nu(y) \\
&= \sum_{y \in \xi} \exp\left(-H(t, y)\right) p(s, x; t, y) \\
&= p(s, x; t, x_T) + \sum_{y \in \xi \setminus \{x_T\}} \exp\left(-\frac{|x_T - y|^2}{2\tilde{a}(T - t)}\right) p(s, x; t, y).
\end{aligned}
$$

Clearly all terms are uniformly bounded in s, t and x and, as $t \uparrow T$, the first term tends to $h(s, x)$, while the second term tends to 0 as p stays between 0 and 1. ∎

Lemma 4.13:

$$
\lim_{k \to \infty} \lim_{t \uparrow T} \mathbb{E}_t^{\star}\left(\kappa(t)\frac{\tilde{h}(t, X_t)}{h(t, X_t)}\mathbf{1}_{A_k(t)}\right) = \mathbb{P}_T^{\star}(A(T)).
$$

Proof: Through the same reasoning as in the proof of Lemma 4.4, it can be deduced that

$$
\lim_{k \to \infty} \lim_{t \uparrow T} \mathbb{E}_t^{\star}\left(\kappa(t)\frac{\tilde{h}(t, X_t)}{h(t, X_t)}\mathbf{1}_{A_k(t)}\right) = \lim_{k \to \infty} \mathbb{E}_T^{\star}\mathbf{1}_{A_k(T)} = \mathbb{P}_T^{\star}(A(T)). \qquad \blacksquare
$$

Lemma 4.14: *For all k, $\Psi_t(X)\kappa(t)\mathbf{1}_{A_k(t)}$ is uniformly bounded in t with $\mathbb{P}^{\circ}$-probability 1.*

Proof: A direct computation yields

$$
\Psi_t(X)\kappa(t)
$$

$$
= \exp\left(\log \kappa(t) + \int_0^t \frac{\mathcal{A}\tilde{h}}{\tilde{h}}(s, X_s) \, ds\right)
$$

$$
= \exp\left(-\log \eta(t) + \int_0^t \frac{d}{ds}\log \eta(s) \, ds - \int_0^t \frac{\partial H}{\partial s}(s, X_s) \, ds + \int_0^t \frac{\mathcal{L}\tilde{h}}{\tilde{h}}(s, X_s) \, ds\right)
$$

$$
\leq \exp\left(-\log \eta(0) + \int_0^t \sum_{y \in N_\xi(X_s)}\left(\frac{\tilde{h}(s, y)}{\tilde{h}(s, X_s)} - 1\right) ds\right).
$$

Here the final inequality is obtained as $\frac{\partial H}{\partial s}(s, x) \geq 0$ for any pair (s, x). Note that under $A_k(t)$, the final term is bounded by $\exp\{kt - \log \eta(0)\}$ and thus $\sup_t \Psi_t(X)\kappa(t)\mathbf{1}_{A_k(t)}$ is uniformly bounded by $\exp\{kT - \log \eta(0)\}$. ∎

Finally, we show that $\mathbb{P}^{\circ}(A(T)) = 1$. Proposition 4.11 demonstrates that if $X_t^{\circ} \neq x_T$ for $t < T$, then $\min_{y \in N_\xi(X_t^{\circ})} |x_T - y| < |x_T - X_t^{\circ}|$, i.e. there is always an $y \in N_\xi(X_t^{\circ})$ that

takes X_t° closer to x_T. Since X° is piecewise constant, it follows from the preceding and (24) that $\sup_t V(t, X_t^\circ)$ is bounded on the event $\{X_T^\circ = x_T\}$ and infinite otherwise since $\min_{y \in N_\xi(X_t^\circ)} |x_T - y| - |x_T - X_t^\circ| \geq 0$ if and only if $X_t^\circ = x_T$. This leads to the following theorem.

Theorem 4.15: $\mathbb{P}^\circ(X_T = x_T) = 1$.

Proof: Given $X_t^\circ = x$, X° jumps to $y \in N_\xi(x)$ with rate $\frac{\tilde{h}(t,y)}{\tilde{h}(t,x)}$. Hence, as $\varepsilon \downarrow 0$,

$$\mathbb{P}^\circ\left(X_{t+\varepsilon} = y \mid X_t = x\right) = \frac{\tilde{h}(t, y)}{\tilde{h}(t, x)} \varepsilon + o(\varepsilon). \tag{26}$$

Define the first time the process is closer to x_T as τ_+. That is,

$$\tau_+ = \inf\left\{s \geq t \colon |x_T - X_s| < |x_T - X_t|\right\}$$

It can be quickly derived that the distribution of τ_+ satisfies

$$\mathbb{P}^\circ(\tau_+ > s \mid X_t = x) \leq \exp\left(-\int_t^s V_+(u, x)\,\mathrm{d}u\right), \quad x \neq x_T, \tag{27}$$

where

$$V_+(u, x) = \sum_{\substack{y \in N_\xi(x) \\ |x_T - y| < |x_T - x|}} \frac{\tilde{h}(u, y)}{\tilde{h}(u, x)} = \sum_{\substack{y \in N_\xi(x) \\ |x_T - y| < |x_T - x|}} \exp\left(-\frac{|x_T - y|^2 - |x_T - x|^2}{2\tilde{a}(T - u)}\right)$$

Upon plugging in $s = T$ in (27), we deduce that for any $t < T$ and $x \neq x_T$, the probability, conditional on $X_t = x$, of reaching a point closer to x_T before time T equals 1. That is, for all $t < T$ and $x \in \xi \setminus \{x_T\}$,

$$\mathbb{P}^\circ(\exists s \in [t, T) \text{ such that } |x_T - X_s| < |x_T - x| \mid X_t = x) = 1. \tag{28}$$

Since (28) holds for any $t < T$, X° keeps reaching points close to x_T with probability 1. Since there are only finitely many points in all compact sets around around x_T,

$$\mathbb{P}(\exists s \in [t, T) \text{ such that } X_s = x_T \mid X_t = x) = 1$$

for all $x \in \xi$. Hence, by the law of total probability, for all $t < T$,

$$\mathbb{P}^\circ(\exists s \in [t, T) \text{ such that } X_s = x_T) = 1 \tag{29}$$

Define the collection of events $B_t = \bigcup_{t \leq s < T}\{X_s = x_T\}$. By (29), $\mathbb{P}^\circ(B_t) = 1$ for all $t < T$. By monotonicity of probability measures

$$\mathbb{P}^\circ(X_T = x_T) = \mathbb{P}^\circ\left(\bigcap_{t < T} B_t\right) = \lim_{t \uparrow T} \mathbb{P}^\circ(B_t) = 1 \qquad \blacksquare$$

Corollary 4.16: $\mathbb{P}^\circ(\sup_{0 \leq t < T} V(t, X_t) < \infty) = 1$.

Proof: Clearly, for each $\omega \in \{X_T^\circ = x_T\}$, we have that $\sup_{0 \leq t < T} V(t, X_t(\omega)) < \infty$, which finishes the proof as $\{X_T^\circ = x_T\}$ is an event with probability 1 by Theorem 4.15. $\qquad \blacksquare$

5. Application 2: conditional stochastic differential equations

In this section, we focus on Markov processes that arise as the solution to Euclidean SDEs and verify Assumption 3.1 under certain conditions on the coefficients.

5.1. Setting, assumptions and main result

We assume that $b\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times d'}$ are such that X, solving the SDE

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \quad X_0 = x_0, \tag{30}$$

uniquely exists (in the strong sense). Here, W is a d'-dimensional Wiener process with all components independent. A Doob h-transform then yields a process $X^\star$ that solves

$$dX_t^\star = b^\star(t, X_t^\star)\, dt + \sigma(t, X_t^\star)\, dW_t, \quad X_0^\star = x_0, \tag{31}$$

where, with $a = \sigma \sigma'$,

$$b^\star(t, x) = b(t, x) + a(t, x)\nabla_x \log h(t, x).$$

The guided process induced by a function $\tilde{h}$ is found as the solution to

$$dX_t^\circ = b^\circ(t, X_t^\circ)\, dt + \sigma(t, X_t^\circ)\, dW_t, \quad X_0^\circ = x_0, \tag{32}$$

where

$$b^\circ(t, x) = b(t, x) + a(t, x)\nabla_x \log \tilde{h}(t, x).$$

This can be derived using (4). This setting has initially been considered in [16], followed by generalization in [5]. However, especially when the diffusion is hypo-elliptic with state-dependent diffusivity, the results in these papers are insufficient to obtain absolute continuity of $\mathbb{P}^\star$ with respect to $\mathbb{P}^\circ$. The following example exemplifies the setting that we wish to study that is not covered by theoretical results in [5].

Example 5.1 (Integrated diffusion): Suppose we study the movement of a particle and $X_{t,1}$ denotes its position and $X_{t,2}$ its velocity at time t. Assume the velocity is driven by a Wiener process, so that we have the system of SDEs given by

$$dX_{t,1} = X_{t,2}\, dt$$
$$dX_{t,2} = \beta(t, X_t)\, dt + \gamma(t, X_{t,1})\, dW_t.$$

This example was studied in [5] in the special case where γ is a constant. Computational results in there suggest that absolute continuity may also hold if γ is state-dependent, but the assumptions imposed in the main result are too strong to be verified for this example.

As in [5], we will consider the process X, conditioned on $LX_T = v$, where L is an $m \times d$ matrix with $m \leq d$ and $v \in \mathbb{R}^m$. Without loss of generality, we assume rank $(L) = m$. The conditional process $X^\star = (X \mid LX_T = v)$ arises from a Doob h-transform. In Section 1.3.2

of [5], it is shown that it suffices to choose $h(t, x)$ as the density of the measure $\mathbb{P}(LX_T \in \cdot \mid X_t = x)$, i.e.

$$h(t, x) = \begin{cases} p(t, x; T, \nu) & \text{if } m = d \\ \displaystyle\int_{\mathbb{R}^{d-m}} p\left(t, x; T, \sum_{i=1}^{d} \xi_i f_i\right) \, \mathrm{d}\xi_{m+1} \cdots \mathrm{d}\xi_d & \text{if } m < d \end{cases}. \tag{33}$$

Here, p denotes the transition density of X and we assume without loss of generality that $L = I$ in the case where rank $(L) = d$. Moreover, $f_1, \ldots, f_m$ form an orthonormal basis of $\mathrm{Col}(L')$ and $f_{m+1}, \ldots, f_d$ form an orthonormal basis of $\ker L$. $\xi_1, \ldots, \xi_d$ are so that $L \sum_{i=1}^{d} \xi_i f_i = \nu$, for which $\xi_1, \ldots, \xi_m$ are uniquely determined.

5.1.1. Choice of appropriate Doob's h-transform $\tilde{h}$

Since h, as defined in (33) is generally intractable, we define a guided process induced by $\tilde{h}$, which is derived similarly to (33), but with the transition density $\tilde{p}$ from an auxiliary process $\tilde{X}$ that satisfies the SDE

$$\mathrm{d}\tilde{X}_t = \tilde{b}\left(t, \tilde{X}_t\right) \, \mathrm{d}t + \tilde{\sigma}(t) \, \mathrm{d}W_t,$$

where

$$\tilde{b}(t, x) = \tilde{B}(t)x + \tilde{\beta}(t).$$

Here $\tilde{B} \colon [0, T] \to \mathbb{R}^{d \times d}, \tilde{\beta} \colon [0, T] \to \mathbb{R}^d$ and $\tilde{\sigma} \colon [0, T] \to \mathbb{R}^{d \times d'}$ should be so that $\tilde{X}$ exists uniquely and possesses smooth transition densities. We introduce the following notation, assuming $t < T$

$$L(t) = L \exp\left\{\int_t^T \tilde{B}(s) \, \mathrm{d}s\right\}$$

$$\mu(t) = \int_t^T L(s)\tilde{\beta}(s) \, \mathrm{d}s$$

$$M(t) = \left(\int_t^T L(s)\tilde{a}(s)L(s)' \, \mathrm{d}s\right)^{-1} \tag{34}$$

$$\zeta(t, x) = \nu - \mu(t) - L(t)x$$

Here, we implicitly assume the inverse appearing in the definition of $M(t)$ exists and set $\tilde{a} = \tilde{\sigma}\tilde{\sigma}'$. By Lemma 2.5 of [5],

$$\tilde{h}(t, x) = \eta(t) \exp\{-H(t, x)\}, \tag{35}$$

where

$$\eta(t) = (2\pi)^{-m/2} |M(t)|^{1/2} \quad \text{and} \quad H(t, x) = \frac{1}{2}\zeta(t, x)'M(t)\zeta(t, x). \tag{36}$$

5.1.2. Main result

Let $\Delta(t)$ be an invertible $m \times m$ diagonal matrix valued measurable function on $[0, T)$ with $t \mapsto \Delta_{ii}(t)$ nondecreasing for $i \in \{1, \dots, m\}$. Define

$$L_\Delta(t) = \Delta(t)L(t)$$
$$M_\Delta(t) = \Delta(t)^{-1}M(t)\Delta(t)^{-1} \tag{37}$$
$$\zeta_\Delta(t, x) = \Delta(t)\zeta(t, x)$$

Note that for $t \approx T$, we have $L(t) \approx L$ and $\mu(t) \approx 0$ and thus $\zeta_\Delta(t, X_t^\circ) \approx \Delta(t)(v - LX_t^\circ)$. For a uniformly elliptic diffusion we will always take $\Delta(t) = I$. In the hypo-elliptic setting, the matrix $\Delta(t)$ is included to account for potential differences in smoothness of components X_t°.

Assumption 5.2: Let $t_0 \in [0, T)$. The scaling matrix $\Delta(t)$ and the coefficients of the process X and $\tilde{X}$ are such that

(5.2a)   There exist positive constants $\underline{c}, \overline{c}$ so that for all $t \in [t_0, T)$

$$\underline{c}(T - t)^{-1} \leq \lambda_{\min}(M_\Delta(t)) \leq \lambda_{\max}(M_\Delta(t)) \leq \overline{c}(T - t)^{-1}.$$

(5.2b)   There exists a positive constant c_1 so that for all $t \in [t_0, T)$,

$$\left| L_\Delta(t) \left(\tilde{b}(t, X_t^\circ) - b(t, X_t^\circ) \right) \right| \leq c_1.$$

(5.2c)   There exists a positive constant c_2 so that for all $t \in [t_0, T)$,

$$\mathrm{tr}\left(L_\Delta(t) a(t, X_t^\circ) L_\Delta(t)' \right) \leq c_2.$$

(5.2d)   There exists a positive constant c_3 and a function $\theta \colon [t_0, T) \times \mathbb{R}^d \to [0, \infty)$ so that for all $t \in [t_0, T)$,

$$\left\| L_\Delta(t) \left(\tilde{a}(t) - a(t, X_t^\circ) \right) L_\Delta(t)' \right\| \leq c_3 \left| \zeta_\Delta(t, X_t^\circ) \right| + \theta(t, X_t^\circ),$$

where there exist positive c_4, α so that for all x,

$$\theta(t, x) \leq c_4(T - t)^\alpha.$$

Assumption 5.2 is similar to Assumption 2.7 of [5]. However, instead of (5.2d) a stronger assumption is formulated in there, which excludes diffusions with state-dependent diffusivity apart from a few exceptional cases. The present assumption is less restrictive and therefore provides a solution to a conjecture posed in section 7.2 of [5].

We require one final assumption on the transition density of the process under $\mathbb{P}$ for the proof of Theorem 5.6.

Assumption 5.3: For $K > 1$ and $u \geq 0$, define $g_K(u) = \max(1/K, 1 - Ku)$. There exist constants $\Lambda, C > 0, K > 1$ and a function $\mu_t(s, x) \colon \{s, t \colon 0 \leq s \leq t \leq T\} \times \mathbb{R}^d \to \mathbb{R}^d$ with

$|\mu_t(s,x) - x| < K(t-s)|x|$ and $|\mu_t(s,x)|^2 \geq g_K(t-s)|x|^2$, so that for all $s < t \leq T$ and $x, y \in \mathbb{R}^d$,

$$p(s,y;t,x) \leq C(t-s)^{-d/2} \exp\left(-\Lambda \frac{|x - \mu_t(s,y)|^2}{t-s}\right)$$

Remark 5.4 (On Assumption 5.3): This assumption is the same as Assumption 2 in [16] and is implied by the stronger Aronson inequality. However, Assumption 5.3 is also satisfied for example for linear diffusion processes that have unbounded drift.

Proposition 5.5: *Let $S < T$. Then h and $\tilde{h}$, defined in (33) and (35), respectively, are S-good functions.*

Proof: Note that by Itô's formula,

$$\mathrm{d}h(t, X_t) = \sigma(t, X_t)' \nabla_x h(t, X_t)\, \mathrm{d}W_t + \mathcal{A}h(t, X_t)\, \mathrm{d}t.$$

Hence, since $\mathcal{A}h = 0$, $D_t^h = \frac{h(t,X_t)}{h(0,x_0)}$ is a $\mathbb{P}$-local martingale. Since also $\mathbb{E}\sup_{s \leq t} h(s, X_s) < \infty$ for all $t \in [0, S]$, h is an S-good function by Theorem 47 of [14].

To see that $\tilde{h}$ is an S-good function, denote $\gamma(t,x) = \sigma(t,x)' \nabla_x \log \tilde{h}(t,x)$. Following the reasoning in the proof of Theorem 3.3 in [17], we deduce that

$$D_t^{\tilde{h}} = \exp\left(\int_0^t \gamma(s, X_s)'\, \mathrm{d}W_s - \frac{1}{2}\int_0^t |\gamma(s, X_s)|^2\, \mathrm{d}s\right) \qquad \blacksquare$$

In the proof of Theorem 1 of [8], it is shown that the right hand side is a positive $\mathbb{P}$-martingale and thus $\tilde{h}$ is an S-good function.

Our main result of this section is that Assumptions 5.2 and 5.3 suffice to justify the limiting operation $t \uparrow T$ and hence absolute continuity of $\mathbb{P}^\star$ with respect to $\mathbb{P}^\circ$.

Theorem 5.6: *Let X, $X^\star$ and X° be defined through (30)–(32), respectively. Here $X^\star$ and X° are induced by (33) and (35). Suppose Assumptions 5.2 and 5.3 are satisfied and there exists a positive Δ such that $|\Delta(t)| \lesssim (T-t)^{-\delta}$. Then $\mathbb{P}_T^\star$ and $\mathbb{P}_T^\circ$ are equivalent and*

$$\frac{\mathrm{d}\mathbb{P}_T^\star}{\mathrm{d}\mathbb{P}_T^\circ}(X) = \frac{\tilde{h}(0,x_0)}{h(0,x_0)}\Psi_T(X).$$

Moreover, using (7), it can be deduced that $\Psi_t(X) = \exp(\int_0^t \mathcal{G}(s, X_s)\, \mathrm{d}s)$, where $\mathcal{G}$ denotes the tractable quantity

$$\mathcal{G}(s,x) = \left(b(s,x) - \tilde{b}(s,x)\right)' \tilde{r}(s,x)$$

$$- \frac{1}{2}\mathrm{tr}\left([a(s,x) - \tilde{a}(s)]\left[L(s)'M(s)L(s) - \tilde{r}(s,x)\tilde{r}(s,x)'\right]\right)$$

with $\tilde{r}(s,x) = \nabla_x \log \tilde{h}(s,x)$.

The proof is deferred to the next subsection. The following result can simplify verifying (5.2d).

Lemma 5.7: *Suppose*

- *there exists a map $\bar{a}\colon [t_0, T] \times \mathbb{R}^m \to \mathbb{R}^{d \times d}$ such that $\bar{a}(t, Lx) = a(t, x)$ for all $(t, x) \in [t_0, T] \times \mathbb{R}^d$.*
- *the map $t \mapsto L_\Delta(t)\tilde{a}(t)L_\Delta(t)'$ is Lipschitz and the map $(t, y) \mapsto L_\Delta(t)\bar{a}(t, y)L_\Delta(t)'$ is Lipschitz in y, uniformly in t. Moreover, $a_T = \lim_{t \uparrow T} L_\Delta(t)\tilde{a}(t)L_\Delta(t)' = \lim_{t \uparrow T} L_\Delta(t)\bar{a}(t, v) L_\Delta(t)'$ exists.*

   Then Assumption (5.2d) is satisfied.

Proof: By the Lipschitz-assumptions, there exist constants k_1 and k_2 such that

$$\left\| L_\Delta(t)\left(\tilde{a}(t) - \bar{a}(t, Lx)\right)L_\Delta(t)'\right\| = \left\| L_\Delta(t)\tilde{a}(t)L_\Delta(t)' - a_T + a_T - L_\Delta(t)\bar{a}(t, Lx)L_\Delta(t)'\right\|$$
$$\leq k_1 \left| v - Lx \right| + k_2 (T - t).$$

Now note that

$$v - Lx = v - \mu(t) - L(t)x + \mu(t) + L(t)x - Lx$$
$$= \Delta(t)^{-1}\zeta_\Delta(t, x) + \mu(t) + (L(t) - L)x.$$

Hence

$$|v - Lx| \leq \left\|\Delta(t)^{-1}\right\| |\zeta_\Delta(t, x)| + |\mu(t) + (L(t) - L)x|.$$

Recall that Δ has nondecreasing entries and therefore $\left\|\Delta(t)^{-1}\right\| \leq \left\|\Delta(t_0)^{-1}\right\|$. We can thus choose $c_3 = k_1 \left\|\Delta(t_0)^{-1}\right\|$ and $\theta(t, x) = k_2(T - t) + k_1 |\mu(t) + (L(t) - L)x|$. ∎

Example 5.8 (Example 5.1 continued): Following up on Example 5.1, note that if we measure a location v at time T, we impose the condition $LX_T = v$ where $L = \begin{pmatrix} I & 0 \end{pmatrix}$. Now σ is constant on $\{x\colon Lx = v\}$ and thus we may choose $x^v \in \{x\colon Lx = v\}$ arbitrarily and define the coefficients of the auxiliary process via

$$\tilde{B}(t) = \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix}, \quad \tilde{\beta}(t) = 0 \quad \text{and} \quad \tilde{\sigma} = \begin{pmatrix} 0 \\ \gamma(T, x^v) \end{pmatrix}.$$

Now observe that $L(t) = \begin{pmatrix} I & (T-t)I \end{pmatrix}$ and $M(t) = \frac{3}{(T-t)^3}(\gamma\gamma')^{-1}(T, x^v)$. Direct computations now show that Assumptions (5.2a), (5.2b) and (5.2c) are satisfied with $\Delta(t) = (T - t)^{-1}I$. Another direct computation now shows that

$$L_\Delta(t)\left(\tilde{a}(t) - a(t, x)\right)L_\Delta(t)' = (\gamma\gamma')(T, x^v) - (\gamma\gamma')(t, x).$$

Hence, under smoothness conditions on $\gamma\gamma'$, Assumption (5.2d) is also satisfied.

5.2. Proof of theorem 5.6

We show that Assumptions 5.2 and 5.3 imply that Assumption 3.1 holds and then the result follows by an application of Theorem 3.3 and proving $\mathbb{P}^\circ\left(\bigcup_k A_k(T)\right) = 1$.

Choose $\kappa(t) = 1$. Suppose t_0 is so that $\log(\frac{1}{T-t}) > 0$ for all $t \geq t_0$ and define

$$V(t,x) = \log^{-1}\left(\frac{1}{T-t}\right) H(t,x), \quad (t,x) \in [t_0, T) \times \mathbb{R}^d, \tag{38}$$

with H as defined in (36). We choose the events $A_k(t)$ cf. Lemma 3.4 so that (3.1a) is satisfied. Note that it is an immediate consequence of Lemma 6.3 of [5] that (3.1b) is satisfied as well. Moreover, (3.1c) is a direct consequence of Assumption 5.3 as the proof of Lemma 2 of [16] can be repeated with $\tilde{h}$ also denoting a normal density.

Theorem 5.9: *Suppose Assumption* 5.3 *holds and* $\delta > 0$ *exists so that* $|\Delta(t)| \lesssim (T-t)^{-\Delta}$. *Then (3.1d) is satisfied.*

Proof: Note that

$$\mathbb{E}_t^\star\left(\kappa(t)\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\mathbf{1}_{A_k(t)}\right) = \mathbb{E}_t^\star\left(\kappa(t)\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\right) - \mathbb{E}_t^\star\left(\kappa(t)\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\mathbf{1}_{A_k(t)^c}\right).$$

It follows from Lemma A.1 and Lemma 6.5 in [5] upon noticing that the stopping time used there coincides with $T \wedge \inf_{t_0 \leq t < T}\{V(t,X_t) \geq k\}$ that the right hand side tends to 1. By monotonicity, $\mathbb{P}^\star(A(T)) = \lim_{t \uparrow T} \mathbb{P}^\star(A(t))$. Now note that for any $t < T$, $\mathbb{P}^\circ(A(t)) \geq \mathbb{P}^\circ(A(T)) = 1$, by Theorem 5.11 and thus by Proposition 2.6, $\mathbb{P}^\star(A(T)) = \lim_{t \uparrow T} \mathbb{P}^\star(A(t)) = 1$. ∎

Theorem 5.10: *For fixed k and Ψ, defined through* (6), $\sup_{t_0 \leq t < T} \Psi_t(X)\mathbf{1}_{A_k(t)}$ *is bounded.*

Proof: This proof is located in Appendix A.2. ∎

Theorem 5.11: *Under Assumption* 5.2 $\limsup_{t \uparrow T} V(t,X_t)$ *is* $\mathbb{P}^\circ$-*almost surely bounded by a finite random variable with V as in* (38).

Proof: The proof is located in Appendix A.1. ∎

We conclude that, as stated in the first part of this section, (3.1a), (3.1b) and (3.1c) are satisfied. (3.1d) is satisfied by Theorem 5.9 and (3.1e) by Theorem 5.10. Lastly, by Theorem 5.11, $\mathbb{P}_T^\circ(A(T)) = \mathbb{P}_T^\circ(\sup_{t_0 \leq t < T} V(t,X_t) < \infty) = 1$ and thus Theorem 5.6 is proven via Theorem 3.3.

5.3. Application: stochastic landmarks registration

In this section we apply our results to a class of models that has recently appeared in shape analysis. Suppose a shape is characterized by a finite set of points, referred to as landmarks. One problem in landmarks registration consists of finding a flow of diffeomorphisms mapping an ordered set of landmarks of one shape to that of another shape, assuming both shapes are summarized by an equal number of landmarks (see for instance [18]). Whereas traditional models assume the flow to be generated by an *ordinary* differential equation, *stochastic* differential equations have been proposed more recently. Here,

we consider specifically the model proposed in [2]. Suppose $q = (q_1, \ldots, q_n)$ denotes a configuration of n distinct landmarks $q_i \in \Omega$ in a domain $\Omega \subseteq \mathbb{R}^d$. Suppose that to each position q_i a momentum vector p_i is attached. Define the Hamiltonian

$$H(q, p) = \frac{1}{2} \sum_{i,j=1}^{n} p_i^T K(q_i, q_j) p_j,$$

the kernel K typically being chosen as Gaussian. The Eulerian model proposed by [2] specifies a flow on landmark positions induced by the system of stochastic differential equations

$$dq_i = (\partial_{p_i} H)(q, p) \, dt + \sum_{l=1}^{J} \sigma_l(q_i) \circ dW_t^l,$$

$$dp_i = -(\partial_{q_i} H)(q, p) \, dt - \sum_{l=1}^{J} \left(\partial_{q_i} \langle p_i, \sigma_l(q_i) \rangle \right) (q_i, p_i) \circ dW_t^l, \qquad (39)$$

where we have surpressed dependence of q_i and p_i on t, for readability. Here, $\sigma_1, \ldots, \sigma_J$ are noise fields centred at prespecified locations $\Delta_1, \ldots, \Delta_J \in \Omega$ defined by

$$\sigma_\ell^\alpha(q_i) = \gamma_\alpha \Lambda_\tau(q_i - \delta_\ell), \quad \alpha = 1, \ldots, d \qquad (40)$$

for noise-amplitudes $\gamma \in \mathbb{R}^d$ and kernel Λ_τ with length-scale τ. For more explanation and details on the derivation of this model, we refer the reader to [2].

We illustrate the stochastic landmarks registration problem using Figure 3, where the number of landmarks equals $n = 75$. Here, the black and orange points correspond to $q(0) := (q_1(0), \ldots, q_n(0))$ and $q(1) := (q_1(1), \ldots, q_n(1))$, respectively. If initial momenta $p(0) := (p_1(0), \ldots, p_n(0))$ are specified, then the system (39) defines a flow that takes $x(0) := (q(0), p(0))$ to $x(1)$ at time 1. The landmarks registration problem corresponds to conditioning the process such that the vector of positions at time 1, "the q-part of $x(1)$", equals $q(1)$. The left-hand panel of the figure shows an unconditional forward simulation of the process; the right-hand panel shows a sample of the guided process defined below.

Conditioning the system (39) on $q(1)$ is challenging as the diffusivity coefficient in the system (39) is state-dependent with the dynamics of each q_i and p_i driven by all Wiener processes W^j. As in [3] we define a guided process based on the auxiliary process

$$d\tilde{q}_i = (\partial_{p_i} H)(q(1), \tilde{p}) \, dt + \sum_{l=1}^{J} \sigma_l(q_i(1)) \circ dW_t^l$$

$$d\tilde{p}_i = -\sum_{l=1}^{J} \left(\partial_{q_i} \langle p_i, \sigma_l(q_i) \rangle \right) (q_i(1), p_i(1)) \circ dW_t^l, \qquad (41)$$

where $q_i(1)$ are the landmarks of the shape that we condition on and $p_i(1)$ can be freely chosen.

Then it follows from an Itô-Stratonovich conversion, see Proposition 1 of [3], that $\tilde{X} = \begin{pmatrix} \tilde{q} & \tilde{p} \end{pmatrix}$ satisfies the linear stochastic differential equation $d\tilde{X}_t = (\tilde{B}\tilde{X}_t + \tilde{\beta}) \, dt + \tilde{\sigma} \, dW_t$
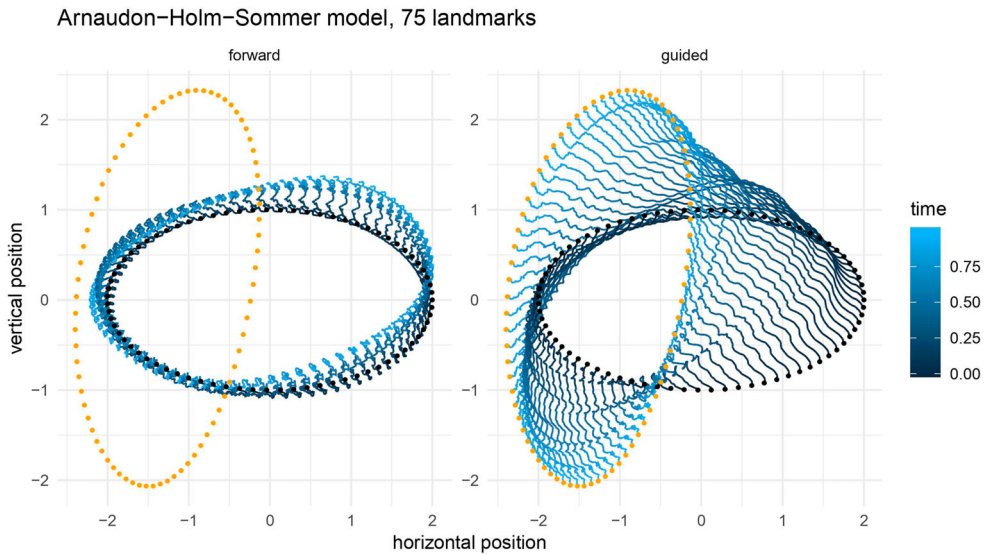
Arnaudon–Holm–Sommer model, 75 landmarks

Figure 3. Illustration of the landmarks registration problem. In both figures, the black and orange points are landmarks that represent the shape at time 0 and time 1, respectively. Left: a sample path of the unconditioned process. Right: a sample path of the guided process.

where $\tilde{B}$ is of the form

$$\tilde{B} = \begin{pmatrix} 0 & G \\ 0 & C \end{pmatrix}$$

and where G and C are known, constant matrices.

Theorem 5.12: *Let* $\tilde{\sigma} = \begin{pmatrix} \tilde{\sigma}_q \\ \tilde{\sigma}_p \end{pmatrix}$. *If*

- $\tilde{\sigma}_q \tilde{\sigma}_q'$ *is strictly positive definite;*
- K, Λ_τ *and* $\nabla \Lambda_\tau$ *are continuous on* Ω;
- *the maps* $\sigma_q \sigma_q'$, $\sigma_p \sigma_q'$ *and* $\sigma_p \sigma_p'$ *are Lipschitz in space.,*

then Assumption 5.2 is satisfied.

Proof of Theorem 5.12: Clearly, we have $L = \begin{pmatrix} I & 0 \end{pmatrix}$ and since all elements of X have the same Hölder-regularity as Brownian motion, we choose $\Delta(t) = I$. A direct computation yields $L(t) = \begin{pmatrix} I & Q(t) \end{pmatrix}$ where $Q(t) = \sum_{n=1}^{\infty} GC^{n-1} \frac{(1-t)^n}{n!}$. We have

$$\int_t^1 L(s)\tilde{a}L(s)' \, ds = (1-t)\tilde{\sigma}_q\tilde{\sigma}_q' + \int_t^1 Q(s) \, ds\tilde{\sigma}_p\tilde{\sigma}_q' + \tilde{\sigma}_q\tilde{\sigma}_p' \int_t^1 Q(s)' \, ds$$

$$+ \int_t^1 Q(s)\tilde{\sigma}_p\tilde{\sigma}_p'Q(s)' \, ds$$

$$= (1-t)\tilde{\sigma}_q\tilde{\sigma}_q' + \mathcal{O}\left((1-t)^2\right).$$

Now observe that $\lambda_{\max}(M(t)) = \lambda_{\min}(M(t)^{-1})^{-1}$ and $\lambda_{\min}(M(t)) = \lambda_{\max}(M(t)^{-1})^{-1}$ and thus (5.2a) is satisfied when $\tilde{\sigma}_q \tilde{\sigma}_q'$ is positive definite. We now note that (5.2b) is satisfied by continuity assumptions on K, Λ_τ, and $\nabla \Lambda_\tau$. We also note that

$$L(t)a(t,x)L(t)' = \sigma_q \sigma_q' + Q(t)\sigma_p \sigma_q' + \sigma_q \sigma_p Q(t)' + Q(t)\sigma_p \sigma_p' Q(t)'.$$

Since $Q(t)$ is of order $1-t$, we see that $\mathrm{tr}\left(L(t)a(t,x)L(t)'\right)$ behaves as $\mathrm{tr}\left(\sigma_q \sigma_q'\right)$ as $t \uparrow 1$ and thus (5.2c) is satisfied as well. Lastly, we note that the map $(t,x) \mapsto L(t)a(t,x)L(t)'$ is Lipschitz in space since $\sigma_q \sigma_q'$, $\sigma_p \sigma_q'$ and $\sigma_p \sigma_p'$ are. Since $\tilde{\sigma}_q \tilde{\sigma}_q'$, $\tilde{\sigma}_p \tilde{\sigma}_q'$ and $\tilde{\sigma}_p \tilde{\sigma}_p'$ are constant, $t \mapsto L(t)\tilde{a}L(t)'$ is Lipschitz. By choice of the auxiliary process, $La(t,x)L'$ and $L\tilde{a}L'$ are equal on the set $\{x: Lx = \{q_i(1)\}_{i=1}^n\}$ and thus (5.2d) is satisfied through Lemma 5.7. ∎

Remark 5.13: Note that the requirement of $\tilde{\sigma}_q \tilde{\sigma}_q'$ being positive definite implies that the number of noise fields J should satisfy $J \geq nd$. Numerical simulations have confirmed that if this assumption is not satisfied, the guided processes used in [2] behave erratically.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Marc Corstanje http://orcid.org/0000-0002-2900-8773
Frank van der Meulen http://orcid.org/0000-0001-7246-8612
Moritz Schauer http://orcid.org/0000-0003-3310-7915

References

[1] R. Agarwal, S. Deng, and W. Zhang, *Generalization of a retard gronwall-like inequality and its applications*, Appl. Math. Comput. 165 (2005), pp. 599–612.
[2] A. Arnaudon, D.D. Holm, and S. Sommer, *A geometric framework for stochastic shape analysis*, Found. Comput. Math. 16 (2019), pp. 653–701.
[3] A. Arnaudon, F.H. van der Meulen, M. Schauer, and S. Sommer, *Diffusion bridges for stochastic hamiltonian systems and shape evolutions* (2021).
[4] F. Baudoin, *Conditioned stochastic differential equations: Theory, examples and application to finance*, Stoch. Process. Their. Appl. 100 (2002), pp. 109–145. Available at https://www.sciencedirect.com/science/article/pii/S0304414902001096.
[5] J. Bierkens, F.H. van der Meulen, and M. Schauer, *Simulation of elliptic and hypo-elliptic conditional diffusions*, Adv. Appl. Probab. 52 (2020), pp. 173–212.
[6] M.N. Bui, Y. Pokern, and P. Dellaportas, *Inference for partially observed riemannian ornstein–uhlenbeck diffusions of covariance matrices*, arXiv preprint arXiv:2104.03193 (2021).

[7] J.M.C. Clark, *The simulation of pinned diffusions*, Decision and Control, Proceedings of the 29th IEEE Conference, 1990, pp. 1418–1420.

[8] B. Delyon and Y. Hu, *Simulation of conditioned diffusions and application to parameter estimation*, Stoch. Process. Their. Appl. 116 (2006), pp. 1660–1675.

[9] P.G. Doyle and J.L. Snell, *Random walks and electric networks* (2000).

[10] M.H. Jensen and S. Sommer, *Simulation of conditioned semimartingales on riemannian manifolds*, arXiv preprint arXiv:2105.13190 (2021).

[11] X. Mao, *Stochastic Differential Equations and Applications*, 2nd ed. Horwood Publishing Chichester, UK, 2008.

[12] Z. Palmowski and T. Rolski, *A technique for exponential change of measure for markov processes*, Bernoulli 8 (2002), pp. 767–785. Available at https://projecteuclid.org:443/euclid.bj/107636 4805.

[13] O. Papaspiliopoulos and G. Roberts, *Importance sampling techniques for estimation of diffusion models*, Stat. Methods for Stochastic Differ. Equ. 124 (2012), pp. 311–340.

[14] P. Protter, *Stochastic Integration and Differential Equations*, Stochastic Modelling and Applied Probability, Springer Berlin, Heidelberg, 1990.

[15] A. Rousselle, *Quenched invariance principle for random walks on delaunay triangulations*, Electron. J. Probab. 20 (2015), pp. 1–32. doi:10.1214/EJP.v20-4006

[16] M. Schauer, F. van der Meulen, and H. van Zanten, *Guided proposals for simulating multidimensional diffusion bridges*, Bernoulli 23 (2017a), pp. 2917–2950. doi:10.3150/16-BEJ833

[17] F.H. van der Meulen and M. Schauer, *Bayesian estimation of incompletely observed diffusions*, Stochastics 90 (2018), pp. 641–662. doi:10.1080/17442508.2017.1381097

[18] L. Younes, *Shapes and Diffeomorphisms*, Springer, 2010.

Appendices

Appendix 1: Proof of Theorems 5.10 and 5.11

Notation: For convenience, we denote a subscript t for evaluation of a space-time function in (t, X_t°) throughout this section.

A.1 Proof of Theorem 5.11

We start by studying $\mathcal{A}^\circ V$ and assume that t_0 is so that $\log(\frac{1}{T-t}) > 0$ for all $t \in [t_0, T]$. It follows that

$$
\begin{aligned}
\mathcal{A}^\circ V(t, x) &= \frac{\partial V}{\partial t} + \mathcal{L}^\circ V(t, x) \\
&= -\frac{H(t, x)}{(T - t) \log^2 \left(\frac{1}{T-t} \right)} + \log^{-1} \left(\frac{1}{T - t} \right) \mathcal{A}^\circ H(t, x) \\
&= \log^{-1} \left(\frac{1}{T - t} \right) \left[\mathcal{A}^\circ H(t, x) - \frac{V(t, x)}{T - t} \right].
\end{aligned}
\tag{A1}
$$

Here, the second equality follows form the product rule and the fact that $\mathcal{L}^\circ$ only acts on the space-variable x. To compute $\mathcal{A}^\circ H$, we note that by Itô's formula,

$$
dH_t = \mathcal{A}^\circ H_t \, dt + \sigma_t' \nabla_x H_t \, dX_t^\circ
$$

Hence, it follows from (32) and Lemma 5.3 of [5] that

$$
\begin{aligned}
\mathcal{A}^\circ H_t = {} & \left(\tilde{b}_t - b_t\right)' L_\Delta(t)' M_\Delta(t)\zeta_{\Delta,t} \\
& + \frac{1}{2}\zeta'_{\Delta,t} M_\Delta(t) L_\Delta(t)\left(\tilde{a}(t) - a_t\right) L_\Delta(t)' M_\Delta(t)\zeta_{\Delta,t} \\
& + \frac{1}{2}\operatorname{tr}\left(a_t L_\Delta(t)' M_\Delta(t) L_\Delta(t)\right) \\
& - \frac{1}{2}\zeta'_{\Delta,t} M_\Delta(t) L_\Delta(t) a_t L_\Delta(t)' M_\Delta(t)\zeta_{\Delta,t}.
\end{aligned}
\tag{A2}
$$

Now in order to upper bound $\limsup_{t\uparrow T} V(t, X_t^\circ)$, we start by upper bounding $\mathcal{A}^\circ H$ by applying the assumptions stated in Assumption 5.2 to each of the terms of (A2). First, we note that it is trivial to verify that $H(t,x) = \frac{1}{2}\zeta_\Delta(t,x)' M_\Delta(t)\zeta_\Delta(t,x)$ and therefore, it follows from a standard quadratic forms inequality that

$$
\frac{1}{2}\lambda_{\min}(M_\Delta(t))\,|\zeta_\Delta(t,x)|^2 \le \frac{1}{2}\zeta_\Delta(t,x)' M_\Delta(t)\zeta_\Delta(t,x) \le \frac{1}{2}\lambda_{\max}(M_\Delta(t))\,|\zeta_\Delta(t,x)|^2,
$$

and thus, by Assumption (5.2a), we have the relation

$$
\sqrt{\frac{2(T-t)H(t,x)}{\bar{c}}} \le \sqrt{\frac{2H(t,x)}{\lambda_{\max}(M_\Delta(t))}} \le |\zeta_\Delta(t,x)| \le \sqrt{\frac{2H(t,x)}{\lambda_{\min}(M_\Delta(t))}} \le \sqrt{\frac{2(T-t)H(t,x)}{\underline{c}}}. \tag{A3}
$$

For the first term of (A2), it follows from the Cauchy-Schwarz inequality, Assumption (5.2b) and (A3) that

$$
\begin{aligned}
\left|\left(\tilde{b}_t - b_t\right)' L_\Delta(t)' M_\Delta(t)\zeta_{\Delta,t}\right| & \le \lambda_{\max}(M_\Delta(t))\left|L_\Delta(t)\left(\tilde{b}_t - b_t\right)\right|\,|\zeta_{\Delta,t}| \\
& \le \bar{c}(T-t)^{-1} c_1 \sqrt{\frac{2(T-t)H_t}{\underline{c}}} \\
& = \gamma_1 (T-t)^{-1/2} H_t^{1/2},
\end{aligned}
\tag{A4}
$$

where $\gamma_1 = \bar{c} c_1 \sqrt{2/\underline{c}}$.

For the second term, we first note that

$$
\begin{aligned}
& \left|\frac{1}{2}\zeta'_{\Delta,t} M_\Delta(t) L_\Delta(t)\left(\tilde{a}(t) - a_t\right) L_\Delta(t)' M_\Delta(t)\zeta_{\Delta,t}\right| \\
& \quad \le \frac{1}{2}\zeta'_{\Delta,t} M_\Delta(t)^2 \zeta_{\Delta,t}\left\|L_\Delta(t)\left(\tilde{a}(t) - a_t\right) L_\Delta(t)'\right\|.
\end{aligned}
$$

Since $M_\Delta(t)$ is symmetric and strictly positive definite, its matrix square root exists and thus $\frac{1}{2}\zeta_\Delta(t,x)' M_\Delta(t)^2 \zeta_\Delta(t,x) \le \lambda_{\max}(M_\Delta(t))H(t,x)$. We now apply Assumption (5.2d) to upper bound the absolute value of the second term of (A2) by

$$
\begin{aligned}
\lambda_{\max}(M_\Delta(t))H_t\left[c_3\,|\zeta_{\Delta,t}| + \theta_t\right] & \le \bar{c}(T-t)^{-1}H_t\left[c_3\sqrt{\frac{2(T-t)H_t}{\underline{c}}} + \theta_t\right] \\
& \le \gamma_2(T-t)^{-1/2}H_t^{3/2} + \gamma_3(T-t)^{-1+\alpha}H_t,
\end{aligned}
\tag{A5}
$$

where $\gamma_2 = \bar{c} c_3 \sqrt{2/\underline{c}}$ and $\gamma_3 = \bar{c} c_4$.

For the third term of (A2), we note that, for positive definite $m \times m$ matrices A and C, there is a relation $|\mathrm{tr}\,(AC)| \leq \mathrm{tr}\,(A)\,\mathrm{tr}\,(C) \leq m\lambda_{\max}(A)\mathrm{tr}\,(C)$. It thus follows from Assumption (5.2c) that

$$
\left| \frac{1}{2}\mathrm{tr}\left(a_t L_\Delta(t)' M_\Delta(t) L_\Delta(t) \right) \right| \leq \frac{1}{2}m\lambda_{\max}(M_\Delta(t))\mathrm{tr}\left(L_\Delta(t) a_t L_\Delta(t)' \right)
$$

$$
\leq \gamma_4 (T-t)^{-1},
$$

(A6)

where $\gamma_4 = m\bar{c}c_3/2$.

Now by combining (A2) with (A4) –(A6), we deduce that

$$
\begin{aligned}
\left| \mathcal{A}^\circ H_t \right| &\leq \gamma_1 (T-t)^{-1/2} H_t^{1/2} \\
&\quad + \gamma_2 (T-t)^{-1/2} H_t^{3/2} + \gamma_3 (T-t)^{-1}\theta_t H_t \\
&\quad + \gamma_4 (T-t)^{-1} \\
&\quad - \frac{1}{2}\zeta_{\Delta,t}' M_\Delta(t) L_\Delta(t) a_t L_\Delta(t)' M_\Delta(t) \zeta_{\Delta,t}.
\end{aligned}
$$

(A7)

For $t \geq t_0$ we can go back to Equation (A1) substitute $V_t = \log^{-1}(\frac{1}{T-t})H_t$ in these equations to find that

$$
\begin{aligned}
\mathcal{A}^\circ V_t &\leq \ell_0(t) + \ell_1(t) V_t^{1/2} + \ell_2(t) V_t + \ell_3(t) V_t^{3/2} \\
&\quad - \frac{1}{2}\zeta_{\Delta,t}' M_\Delta(t) L_\Delta(t) a_t L_\Delta(t)' M_\Delta(t) \zeta_{\Delta,t},
\end{aligned}
$$

where

$$
\begin{aligned}
\ell_0(t) &= \frac{\gamma_4}{(T-t)\log\left(\frac{1}{T-t}\right)} \\[2mm]
\ell_1(t) &= \frac{\gamma_1}{\sqrt{(T-t)\log\left(\frac{1}{T-t}\right)}} \\[2mm]
\ell_2(t) &= \frac{\gamma_3}{(T-t)^{1-\alpha}} - \frac{1}{(T-t)\log\left(\frac{1}{T-t}\right)} \\[2mm]
\ell_3(t) &= \gamma_2 \sqrt{\frac{\log\left(\frac{1}{T-t}\right)}{T-t}}.
\end{aligned}
$$

(A8)

Now recall that a martingale M^V exists so that

$$
V_t = V_{t_0} + M_t^V + \int_{t_0}^t \mathcal{A}^\circ V_s \, \mathrm{d}s
$$

It is a consequence of Itô's formula that

$$
[M^V]_t = \zeta_{\Delta,t}' M_\Delta(t) L_\Delta(t) a_t L_\Delta(t)' M_\Delta(t) \zeta_{\Delta,t}.
$$

Hence,

$$
\begin{aligned}
V_t &\leq V_{t_0} + M_t^V - \frac{1}{2}[M^V]_t + \int_{t_0}^t \ell_0(s)\,\mathrm{d}s + \int_{t_0}^t \ell_1(s) V_s^{1/2}\,\mathrm{d}s \\
&\quad + \int_{t_0}^t \ell_2(s) V_s \,\mathrm{d}s + \int_{t_0}^t \ell_3(s) V_s^{3/2}\,\mathrm{d}s
\end{aligned}
$$

(A9)

Now by Lemma A.2, we have an almost surely finite random variable C so that $M_t^V - \frac{1}{2}[M^V]_t \le C\log(\frac{1}{T-t})$. Moreover, by Lemma A.3, we now have

$$V_t \le 4 \left[\frac{4\exp\left\{-\frac{1}{2}\int_{t_0}^t \ell_2(s)\,\mathrm{d}s\right\}}{2\sqrt{V\left(t_0, X_{t_0}^\circ\right) + C\log\left(\frac{1}{T-t}\right) + \int_{t_0}^t \ell_0(s)\,\mathrm{d}s + \int_{t_0}^t \ell_1(s)\,\mathrm{d}s}} - \int_{t_0}^t \ell_3(s)\,\mathrm{d}s \right]^{-2}.$$

Direct computations show that

$$\kappa_0(t) := \int_{t_0}^t \ell_0(s)\,\mathrm{d}s = \gamma_4 \log \frac{\log\left(\frac{1}{T-t}\right)}{\log\left(\frac{1}{T-t_0}\right)}$$

$$\kappa_1(t) := \int_{t_0}^t \ell_1(s)\,\mathrm{d}s = \gamma_1 \int_{\log\left(\frac{1}{T-t_0}\right)}^{\log\left(\frac{1}{T-t}\right)} u^{-1/2}e^{-u/2}\,\mathrm{d}u$$

$$\kappa_2(t) := \int_{t_0}^t \ell_2(s)\,\mathrm{d}s = \gamma_3 \frac{(T-t_0)^\alpha - (T-t)^\alpha}{\alpha} - \log \frac{\log\left(\frac{1}{T-t}\right)}{\log\left(\frac{1}{T-t_0}\right)}$$

$$\kappa_3(t) := \int_{t_0}^t \ell_3(s)\,\mathrm{d}s = \gamma_2 \int_{\log\left(\frac{1}{T-t_0}\right)}^{\log\left(\frac{1}{T-t}\right)} u^{1/2}e^{-u/2}\,\mathrm{d}u.$$

(A10)

Now note that κ_1 and κ_3 are bounded in the limit $t \uparrow T$. A substitution of the results in (A10) and some direct computations now yields

$$V_t \le 4 \left[\frac{4\log^{-1/2}\left(\frac{1}{T-t_0}\right)\exp\left\{-\frac{1}{2}\gamma_3\frac{(T-t_0)^\alpha-(T-t)^\alpha}{\alpha}\right\}}{\sqrt{\dfrac{V\left(t_0,X_{t_0}^\circ\right)+C\log\left(\frac{1}{T-t}\right)+\gamma_4\log\frac{\log\left(\frac{1}{T-t}\right)}{\log\left(\frac{1}{T-t_0}\right)}}{\log\left(\frac{1}{T-t}\right)} - \dfrac{\kappa_1(t)}{\sqrt{\log\left(\frac{1}{T-t}\right)}}}} - \kappa_3(t) \right]^{-2},$$

which is almost surely bounded in the limit $t \uparrow T$.

A.2 Proof of Theorem 5.10

To determine the form of Ψ, we note that direct computations show that

$$\frac{\partial H}{\partial t}(t,x) = \tilde{b}(t,x)'L_\Delta(t)'M_\Delta(t)\zeta_\Delta(t,x) + \frac{1}{2}\zeta_\Delta(t,x)'M_\Delta(t)L_\Delta(t)\tilde{a}(t)L_\Delta(t)'M_\Delta(t)\zeta_\Delta(t,x). \quad \text{(A11)}$$

Moreover,

$$\nabla_x H(t,x) = -L_\Delta(t)'M_\Delta(t)\zeta_\Delta(t,x) \quad \text{and} \quad \mathrm{Hess}_H(t,x) = L_\Delta(t)'M_\Delta(t)L_\Delta(t). \quad \text{(A12)}$$

Using the formulas derived in (A11) and (A12) for $\frac{\partial H}{\partial t}$, $\nabla_x H$ and Hess_H and the relation $\nabla_x \tilde{h}(t,x) = -\tilde{h}(t,x)\nabla_x H(t,x)$, it can be derived that

$$
\begin{aligned}
\frac{\mathcal{A}\tilde{h}}{\tilde{h}}(s,x) &= \left(b(s,x) - \tilde{b}(s,x)\right)' L_\Delta(s)' M_\Delta(s)\zeta_\Delta(s,x) \\
&\quad + \frac{1}{2}\zeta_\Delta(s,x)' M_\Delta(s)L_\Delta(s)\left(a(s,x) - \tilde{a}(s)\right)L_\Delta(s)'M_\Delta(s)\zeta_\Delta(s,x) \\
&\quad - \text{tr}\left(\left(a(s,x) - \tilde{a}(s)\right)L_\Delta(s)'M_\Delta(s)L_\Delta(s)\right).
\end{aligned}
\tag{A13}
$$

We now apply the upper bounds derived in Appendix A.1 for each of the terms. First note that under $\{\sup_{t_0 \leq s \leq t} V(t, X_t^\circ) \leq k\}$, we have that $H(s, X_s^\circ) \leq k\log(\frac{1}{T-s}) \leq k\log(\frac{1}{T-t})$ for all $s \in [t_0, t)$. It now follows from Equation (A4) that the absolute value of first term of (A13) can be upper bounded by

$$
\gamma_1 \sqrt{k\frac{\log\left(\frac{1}{T-t}\right)}{T-t}},
$$

which is integrable over $[t_0, T]$ (see the last line of (A10)). Using a similar approach, we can also combine (A5) and (A10) to see that the absolute value of the second term of (A13) is integrable. Similarly, the relation $|\text{tr}(AC)| \leq \text{tr}(A)\text{tr}(C) \leq m\lambda_{\max}(A)\text{tr}(C)$ for $m \times m$ matrices A and C, in combination with Assumption (5.2d) and the known integrals in (A10) yields that the final term is integrable.

Appendix 2: Additional lemmas

In this section, we discuss some lemmas that were used in various proofs throughout the paper.

Lemma A.1: *Suppose* (3.1b) *and* (3.1c) *hold, let $s < T$ and let g_s be a bounded $\mathcal{F}_s$-measurable function.*

$$
\lim_{t\uparrow T}\mathbb{E}_t^\star\left(g_s(X)\kappa(t)\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\right) = \mathbb{E}_s^\star g_s(X).
$$

Proof: By (3), upon noting $\mathcal{A}h = 0$,

$$
\begin{aligned}
\mathbb{E}_t^\star\left(g_s(X)\kappa(t)\frac{\tilde{h}(t,X_t)}{h(t,X_t)}\right) &= \mathbb{E}_t\left(g_s(X)\kappa(t)\frac{\tilde{h}(t,X_t)}{h(0,x_0)}\right) \\
&= \mathbb{E}_s\left(\frac{g_s(X)}{h(0,x_0)}\mathbb{E}_t\left(\kappa(t)\tilde{h}(t,X_t)\mid\mathcal{F}_s\right)\right) = \mathbb{E}_s\left(\frac{g_s(X)}{h(0,x_0)}h^\sharp(t;s,X_s)\right)
\end{aligned}
$$

We now take $\lim_{t\uparrow T}$. Note that it follows from (3.1c), dominated convergence and (3.1b) that

$$
\lim_{t\uparrow T}\mathbb{E}_s\left(\frac{g_s(X)}{h(0,x_0)}h^\sharp(t;s,X_s)\right) = \mathbb{E}_s\left(g_s(X)\frac{h(s,X_s)}{h(0,x_0)}\right)
$$

The result now follows upon changing measures using (3) again with $\mathcal{A}h = 0$. ∎

Lemma A.2 (Exponential martingale bound): *Suppose $\{N_t\}_{t\geq t_0}$ is a martingale. Then an almost surely finite random variable C exists so that $N_t - \frac{1}{2}[N]_t \leq C\log(\frac{1}{T-t})$.*

Proof: First note that, for any fixed $t < T$ and positive λ, Doob's maximal inequality yields

$$\mathbb{P}\left(\sup_{t_0 \leq s \leq t} \exp\left\{N_s - \frac{1}{2}[N]_s\right\} > e^\lambda\right) \leq e^{-\lambda}\mathbb{E}\left(\exp\left\{N_t - \frac{1}{2}[N]_t\right\}\right) = k_0 e^{-\lambda},$$

where $k_0 = \exp\{N_{t_0} - \frac{1}{2}[N]_{t_0}\}$ is a bounded random variable. It follows that

$$\mathbb{P}\left(\sup_{0 \leq s \leq t}\left\{N_s - \frac{1}{2}[N]_s\right\} > \lambda\right) \leq k_0 e^{-\lambda}.$$

Now set $t_n = T - \frac{1}{n}$ and assume n is sufficiently large so that $t_n > 0$. It follows from the preceding that we can choose an arbitrary positive sequence $\{\lambda_n\}_n$ and have that

$$\mathbb{P}\left(\sup_{0 \leq s \leq t_{n+1}}\left\{N_s - \frac{1}{2}[N]_s\right\} > \lambda_n\right) \leq k_0 e^{-\lambda_n}.$$

Upon choosing $\lambda_n = 2\log n$, one has $\sum_n e^{-\lambda_n} < \infty$ and thus, by the Borel-Cantelli lemma

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{\sup_{0 \leq s \leq t_{n+1}}\left\{N_s - \frac{1}{2}[N]_s\right\} > \lambda_n\right\}\right) = 0.$$

We can thus almost surely find a random variable $n_0(\omega)$ so that for all $n \geq n_0(\omega)$

$$\sup_{0 \leq s \leq t_{n+1}}\left\{N_s - \frac{1}{2}[N]_s\right\} \leq \lambda_n$$

Now notice that for any $t \in [t_n, t_{n+1}]$, one has $\lambda_n \leq 2\log(\frac{1}{T-t})$ and therefore

$$\sup_{0 \leq t < T} \frac{N_t - \frac{1}{2}[N]_t}{\log\left(\frac{1}{T-t}\right)} = \sup_n \sup_{t_n \leq t \leq t_{n+1}} \frac{N_t - \frac{1}{2}[N]_t}{\log\left(\frac{1}{T-t}\right)}$$

$$\leq \sup_{0 \leq t \leq t_{n_0}} \frac{N_t - \frac{1}{2}[N]_t}{\log\left(\frac{1}{T-t}\right)} \vee \sup_{n \geq n_0} \sup_{t_n \leq t \leq t_{n+1}} \frac{\lambda_n}{\log\left(\frac{1}{T-t}\right)}$$

$$\leq C \vee \sup_n \frac{2\log n}{\log\left(\frac{1}{T-t_{n+1}}\right)}$$

$$= C \vee 2\sup_n \frac{\log n}{\log(n+1)}$$

$$\leq C \vee 2,$$

where C is a random variable depending on n_0, which is finite since t_{n_0} is almost surely bounded away from T. ∎

Lemma A.3 (Application of Theorem 2.1 of [1]): *Suppose V satisfies* (A9) *with ℓ_0, ℓ_1, ℓ_2 and ℓ_3 as in* (A8). *Then*

$$V(t, X_t^\circ) \leq 4\left[\frac{4\exp\left\{-\frac{1}{2}\int_{t_0}^t \ell_2(s)\,ds\right\}}{2\sqrt{V\left(t_0, X_{t_0}^\circ\right) + C\log\left(\frac{1}{T-t}\right) + \int_{t_0}^t \ell_0(s)\,ds + \int_{t_0}^t \ell_1(s)\,ds}} - \int_{t_0}^t \ell_3(s)\,ds\right]^{-2}.$$

Proof: By Lemma A.2, we have an almost surely finite random variable C so that $M_t^V - \frac{1}{2}[M^V]_t \leq C\log(\frac{1}{T-t})$. The lemma is an application of Theorem 2.1 of [1] with, in their notation, $a(t) =$

$V(t_0, X_{t_0}^\circ) + C \log(\frac{1}{T-t}) + \int_{t_0}^t \ell_0(s)\, ds$, $f_i(t, s) = \ell_i(s)$ and $b_i(t) = t$ for $i = 1, 2, 3$, $w_1(u) = \sqrt{u}$, $w_2(u) = u$, $w_3(u) = u\sqrt{u}$. To achieve the result, we choose $u_1 = 0$, $u_2 = 1$ and $u_3 = 4$. ∎

Lemma A.4: *Suppose M is a local martingale bounded from below with $\mathbb{E}M_0 < \infty$, then M is a supermartingale.*

Proof: Without loss of generality, we assume M is bounded from below by 0. Now let $\{\tau_n\}_n$ be a sequence of stopping times such that $\tau_n \uparrow \infty$ and $\{M_{t \wedge \tau_n}\}_t$ is a martingale for all n. It follows from Fatou's lemma that

$$\mathbb{E}\, |M_t| = \mathbb{E}M_t = \mathbb{E}\left(\liminf_{n \to \infty} M_{t \wedge \tau_n}\right) \leq \liminf_{n \to \infty} \mathbb{E}M_{t \wedge \tau_n} = \mathbb{E}M_0 < \infty.$$

Hence, M is integrable. Moreover, for $s \leq t$, if also follows from Fatou's lemma that

$$\mathbb{E}\,(M_t \mid \mathcal{F}_s) = \mathbb{E}\left(\liminf_{n \to \infty} M_{t \wedge \tau_n} \mid \mathcal{F}_s\right) \leq \liminf_{n \to \infty} \mathbb{E}\left(M_{t \wedge \tau_n} \mid \mathcal{F}_s\right) = \liminf_{n \to \infty} M_{s \wedge \tau_n} = M_s \quad ∎$$