Guided proposals for simulating multi-dimensional diffusion bridges

MORITZ SCHAUER, FRANK VAN DER MEULEN and HARRY VAN ZANTEN

Mathematical Institute
Leiden University
P.O. Box 9512
2300 RA Leiden
The Netherlands
E-mail: m.r.schauer@math.leidenuniv.nl

Delft Institute of Applied Mathematics (DIAM)
Delft University of Technology
Mekelweg 4
2628 CD Delft
The Netherlands
E-mail: f.h.vandermeulen@tudelft.nl

Korteweg-de Vries Institute for Mathematics
University of Amsterdam
P.O. Box 94248
1090 GE Amsterdam
The Netherlands
E-mail: hvzanten@uva.nl

A Monte Carlo method for simulating a multi-dimensional diffusion process conditioned on hitting a fixed point at a fixed future time is developed. Proposals for such diffusion bridges are obtained by superimposing an additional guiding term to the drift of the process under consideration. The guiding term is derived via approximation of the target process by a simpler diffusion processes with known transition densities. Acceptance of a proposal can be determined by computing the likelihood ratio between the proposal and the target bridge, which is derived in closed form. We show under general conditions that the likelihood ratio is well defined and show that a class of proposals with guiding term obtained from linear approximations fall under these conditions.

Keywords: Multidimensional diffusion bridge; change of measure; data augmentation; linear processes.

Primary 60J60; secondary 65C30, 65C05.

*Research supported by the Netherlands Organization for Scientific Research (NWO)

1. Introduction

1.1. Diffusion bridges

Suppose X is a d-dimensional diffusion with time dependent drift $b\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and dispersion coefficient $\sigma\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times d'}$ governed by the stochastic differential equation (SDE)

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \quad X_0 = u, \tag{1.1}$$

where W is a standard d'-dimensional Brownian motion. When the process X is conditioned to hit a point $v \in \mathbb{R}^d$ at time $T > 0$, the resulting process $X^\star$ on $[0, T]$ is called the *diffusion bridge from u to v*. In this paper we consider the problem of simulating realizations of this bridge process. Since we are conditioning on an event of probability zero and in general no closed form expression for the transition densities of the original process X or the bridge $X^\star$ exist, this is known to be a difficult problem.

This problem arises for instance when making statistical inference for diffusion models from discrete-time, low-frequency data. In that setting the fact that the transition densities are unavailable implies that the likelihood of the data is not accessible. A successful approach initiated by Roberts and Stramer (2001) is to circumvent this problem by viewing the continuous segments between the observed data points as missing data. Computational algorithms can then be designed that augment the discrete-time data by (repeatedly) simulating the diffusion bridges between the observed data points. This statistical application of simulation algorithms for diffusion bridges was our initial motivation for this work. The present paper however focusses on the simulation problem as such and can have other applications as well.

The simulation of diffusion bridges has received much attention over the past decade, see for instance the papers Elerian et al. (2001), Eraker (2001), Roberts and Stramer (2001), Durham and Gallant (2002), Stuart et al. (2004), Beskos and Roberts (2005), Beskos et al. (2006), Beskos et al. (2008), Fearnhead (2008), Papaspiliopoulos and Roberts (2012), Lin et al. (2010), Bladt and Sørensen (2012), Bayer and Schoenmakers (2013) to mention just a few. Many of these papers employ accept-reject-type methods. The common idea is that while sampling directly from the law $\mathbb{P}^\star$ of the bridge process $X^\star$ is typically impossible, sampling from an equivalent law $\mathbb{P}^\circ$ of some proposal process X° might in fact be feasible. If this proposal is accepted with an appropriately chosen probability, depending on the Radon-Nikodym derivative $(\mathrm{d}\mathbb{P}^\star/\mathrm{d}\mathbb{P}^\circ)(\mathrm{X}^\circ)$, then either exact or approximate draws from the target distribution $\mathbb{P}^\star$ can be generated. Importance sampling and Metropolis-Hastings algorithms are the prime examples of methods of this type.

To be able to carry out these procedures in practice, simulating paths from the proposal process has to be relatively easy and, up to a normalizing constant, an expression for the derivative $(\mathrm{d}\mathbb{P}^\star/\mathrm{d}\mathbb{P}^\circ)(\mathrm{X}^\circ)$ has to be available that is easy to evaluate. The speed of the procedures greatly depends on the acceptance probability, which in turn depends on $(\mathrm{d}\mathbb{P}^\star/\mathrm{d}\mathbb{P}^\circ)(\mathrm{X}^\circ)$. This can be influenced by working with a cleverly chosen proposal process X°. A naive choice might result in a proposal process that, although its law is equivalent to that of the target bridge $X^\star$, has sample paths that are with considerable probability rather different from those of $X^\star$. This then results in small ratios $(\mathrm{d}\mathbb{P}^\star/\mathrm{d}\mathbb{P}^\circ)(\mathrm{X}^\circ)$ with large probability, which in turn leads to small acceptance probabilities and hence to a slow procedure. It is therefore desirable to have proposals that are "close" to the target in an appropriate sense. In this paper we construct such proposals for the multi-dimensional setting.

1.2. Guided proposals

We will consider so-called *guided proposals*, according to the terminology suggested in Papaspiliopoulos and Roberts (2012). This means that our proposals are realizations of a process X° that solves an SDE of the form (1.1) as well, but with a drift term that is adapted in order to force the process X° to hit the point v at time T.

An early paper suggesting guided proposals is Clark (1990) (a paper that seems to have received little attention in the statistics community). Clark (1990) considers the case $d = 1$ and σ constant and advocates using proposals from the SDE $\mathrm{d}X_t^\circ = b(X_t^\circ)\,\mathrm{d}t + \frac{v - X_t^\circ}{T-t}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$. Note that here the guiding drift term that drives the process to v at time T is exactly the drift term of a Brownian bridge. In addition the drift b of the original process appears. The idea is that this ensures that before time T, the proposal behaves similar to the original diffusion X. Delyon and Hu (2006) have generalized the work of Clark (1990) in two important directions. Firstly, they allow non-constant σ using proposals $X^\triangledown$ satisfying the SDE

$$\mathrm{d}X_t^\triangledown = \left(b(t, X_t^\triangledown) + \frac{v - X_t^\triangledown}{T-t} \right) \mathrm{d}t + \sigma(t, X_t^\triangledown)\,\mathrm{d}W_t. \tag{$\triangledown$}$$

This considerably complicates proving that the laws of X° and the target bridge $X^\star$ are absolutely continuous. Further, Delyon and Hu (2006) consider the alternative proposals $X^\vartriangle$ satisfying the SDE

$$\mathrm{d}X_t^\vartriangle = \frac{v - X_t^\vartriangle}{T-t}\,\mathrm{d}t + \sigma(t, X_t^\vartriangle)\,\mathrm{d}W_t. \tag{$\vartriangle$}$$

where the original drift of X is disregarded. This is a popular choice in practice especially with a discretization scheme known as the Modified Brownian Bridge. Both proposals have their individual drawbacks, see Section 1.3.

Another important difference is that they consider the multi-dimensional case. With more degrees of freedom a proposal process that is not appropriately chosen has a much higher chance of not being similar to the target process, leading to very low acceptance probabilities and hence slow simulation procedures. In higher dimensions the careful construction of the proposals is even more important for obtaining practically feasible procedures than in dimension one.

Our approach is inspired by the ideas in Clark (1990) and Delyon and Hu (2006). However, we propose to adjust the drift in a different way, allowing more flexibility in constructing an appropriate guiding term. This is particularly aimed at finding procedures with higher acceptance probabilities in the multi-dimensional case. To explain the approach in more detail we recall that, under weak assumptions the target diffusion bridge $X^\star$ is characterized as the solution to the SDE

$$\mathrm{d}X_t^\star = b^\star(t, X_t^\star)\,\mathrm{d}t + \sigma(t, X_t^\star)\,\mathrm{d}W_t, \qquad X_0^\star = u, \qquad t \in [0, T), \tag{$\star$}$$

where

$$b^\star(t, x) = b(t, x) + a(t, x)\nabla_x \log p(t, x; T, v) \tag{$\star\star$}$$

and $a(t,x) = \sigma(t,x)\sigma'(t,x)$. In the bridge SDE the term $a(t,x)\nabla_x \log p(t,x;T,v)$ is added to the original drift to direct $X^\star$ towards v from the current position $X_t^\star = x$ in just the right manner. Since equation ($\star$) contains the unknown transition densities of the original process X it cannot be employed directly for simulation. We propose to replace this unknown density by one coming from an auxiliary diffusion process with known transition densities. So the proposal process is going to be the solution X° of the SDE

$$\mathrm{d}X_t^\circ = b^\circ(t, X_t^\circ)\,\mathrm{d}t + \sigma(t, X_t^\circ)\,\mathrm{d}W_t, \quad X_0^\circ = u, \tag{$\circ$}$$

where

$$b^\circ(t,x) = b(t,x) + a(t,x)\nabla_x \log \widetilde{p}(t,x;T,v) \tag{$\circ\circ$}$$

and $\widetilde{p}(s,x;t,v)$ is the transition density of a diffusion process $\widetilde{X}$ for which above expression is known in closed form. We note that in general our proposals are different from those defined in Delyon and Hu (2006). First of all the diffusion $a(t,x)$ of the original process appears in the drift of the proposal process X° and secondly we have additional freedom since we can choose the process $\widetilde{X}$.

The paper contains two main theoretical results. In the first we give conditions under which the process X° is indeed a valid proposal process in the sense that its distribution $\mathbb{P}^\circ$ (viewed as Borel measure on $C([0,T],\mathbb{R}^d)$) is equivalent to the law $\mathbb{P}^\star$ of the target process $X^\star$ and we derive an expression for the Radon-Nikodym derivative of the form

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) \propto \exp\left(\int_0^T G(s, X_s^\circ)\,\mathrm{d}s\right),$$

where the functional G does not depend on unknown or inaccessible objects. In the second theorem we show that the assumptions of the general result are fulfilled if in ($\circ\circ$) we choose the transition density $\widetilde{p}$ of a process $\widetilde{X}$ from a large class of *linear processes*. This is a suitable class, since linear processes have tractable transition densities.

1.3. Comparison of proposals

Numerical experiments presented Van der Meulen and Schauer (2015) show that our approach can indeed substantially increase acceptance rates in a Metropolis-Hastings sampler, especially in the multi-dimensional setting. Already in a simple one-dimensional example however we can illustrate the advantage of our method.

Consider the solution X of the SDE,

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \tfrac{1}{2}\,\mathrm{d}W_t, \quad X_0 = u \quad \text{with} \quad b(x) = \beta_1 - \beta_2 \sin(8x).$$

The corresponding bridge $X^\star$ is obtained by conditioning X to hit the point $v \in \mathbb{R}$ at time $T > 0$. We take $u = 0, v = \frac{\pi}{2}$ and consider either the case $\beta_1 = \beta_2 = 2$ or $\beta_1 = 2, \beta_2 = 0$. We want to compare the three mentioned proposals ($\triangledown$),($\vartriangle$) and ($\circ$) in these two settings. The drift b satisfies the assumptions for applying the Exact Algorithm of Beskos and Roberts (2005), but numerical experiments revealed the rejection probability is close to 1 in this particular example. Besides, our main interest lies in comparing proposals that are suited for simulating general diffusion bridges in the multivariate case as well. A simple choice for the guided proposal (○) is obtained by taking $\widetilde{X}$ to be a scaled Brownian motion with constant drift ϑ. This gives $b^\circ(s,x) = b(x) + \frac{v-x}{T-s} - \vartheta$ as the drift of the corresponding guided proposal. Here we can choose ϑ freely. In fact, far more flexibility can be obtained by choosing $\widetilde{X}$ a linear process as in theorem 2. In particular, we could take ϑ to depend on t, resulting in an infinite dimensional class of proposals. For illustration purposes, in this example we show that just taking a scaled Brownian motion with constant drift ϑ for $\widetilde{X}$ is already very powerful.

If $\beta_2 = 0$ the process X is simply a Brownian motion with drift. It is folklore that the corresponding bridge $X^\star$ is then in fact the standard Brownian bridge from u to v, independent of the constant β_1 (see for instance Gasbarra et al. (2007)). So in that case both proposal (△) and proposal (○) with $\vartheta = \beta_1$ coincide with the target bridge. However, the drift $b^\triangledown$ of the proposal (▽) is off by $|b^\star(s,x) - b^\triangledown(s,x)| = |\beta_1|$ leading to bad acceptance rates if $\beta_1 \neq 0$, even for small values of T. This seems to be the prime reason that proposal (▽) is rarely used in practice.

Now if $\beta_2 = 2$, both (▽) and (△) fail to capture the true dynamics of (⋆). Roughly speaking, for (△) the proposals fail to capture the multimodality of the marginal distributions of the true bridge, while proposals with (▽) arrive at values close to v too early due to the mismatch between pulling term and drift. On the other hand the proposals (○) can be quite close to the target bridge for good choices of ϑ, see figure 1. Two effects are in place: incorporating the true drift into the proposal results in the correct local behaviour of the proposal bridge (multimodality in this particular example). Further, an appropriate choice of ϑ reduces the mismatch between the drift part and guiding part of the proposal. The additional freedom in (○) by choice of ϑ will be especially useful, if one can find good values for ϑ in a systematic way. We now explain how this can be accomplished.

Let $\mathbb{P}^\circ_\vartheta$ denote the law of X°. One option to choose ϑ in a systematic way is to take the information projection $\mathbb{P}^\circ_{\vartheta_{\text{opt}}}$ defined by

$$\vartheta_{\text{opt}} = \operatorname{argmin}_\vartheta D_{\text{KL}}(\mathbb{P}^\star \| \mathbb{P}^\circ_\vartheta)$$

Here, the Kullback-Leibler divergence is given by

$$D_{\text{KL}}(\mathbb{P}^\star \| \mathbb{P}^\circ_\vartheta) = \int \log\left(\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ_\vartheta}\right) \mathrm{d}\mathbb{P}^\star.$$

This is a measure how much information is lost, when $\mathbb{P}^\circ_\vartheta$ is used to approximate $\mathbb{P}^\star$. This expression is not of much direct use, as it depends on the unknown measure $\mathbb{P}^\star$. However, given a sample X° from $\mathbb{P}^\circ_{\vartheta_0}$ using a reference parameter ϑ_0, the gradient of $D_{\text{KL}}(\mathbb{P}^\star \| \mathbb{P}^\circ_\vartheta)$ can be approximated by

$$\nabla_\vartheta \log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ_\vartheta}(X^\circ) \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ_{\vartheta_0}}(X^\circ).$$

This in turn can be used in an iterative stochastic gradient descent algorithm (details are given in the appendix). The value $\vartheta = 1.36$ used in Figure 1 was obtained in this way. From the trace plot of the gradient descent algorithm displayed in figure 2 it appears the algorithm settles near the optimal value shown in the right-hand figure.

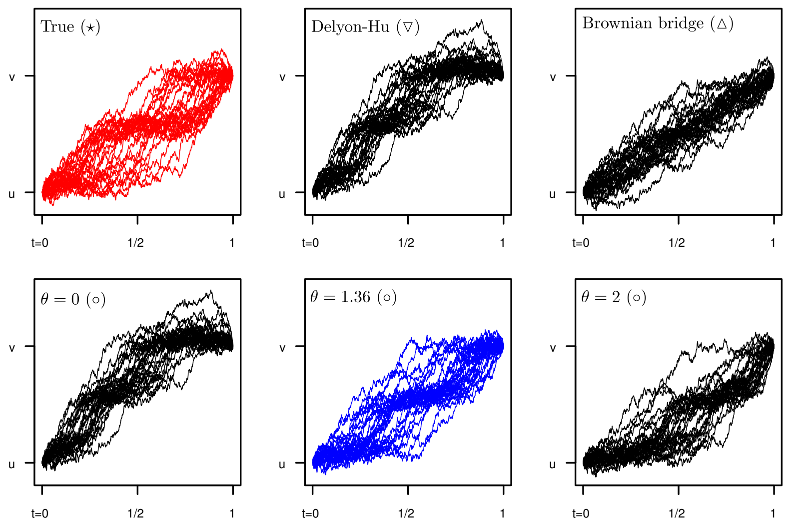

Figure 1. Samples from the true distribution of the bridge compared to different proposals for the example $b(x) = 2 - 2\sin(8x)$. Top row: True bridge, proposals with drift $b^{\triangledown}(t,x) = b(x) + \frac{v-x}{T-t}$ and $b^{\triangle}(t,x) = \frac{v-x}{T-t}$. Bottom row: $b^{\circ}(s,x) = b(x) + \frac{v-x}{T-t} - \vartheta$ for different values of ϑ. The top-middle figure and bottom-left figure coincide.

1.4. Contribution of this paper

In this paper we propose a novel class of proposals for generating diffusion bridges that can be used in Markov Chain Monte Carlo and importance sampling algorithms. We stress that these are not special cases of the proposals from Delyon and Hu (2006) (specified in equations ($\triangledown$) and ($\triangle$)). An advantage of this class is that the drift of the true diffusion process is taken into account while avoiding the drawbacks of proposals of the form ($\triangledown$). This is enabled by the increased flexibility for constructing a pulling term in the drift of the proposal. A particular feature of our choice is that no Itō-integral appears in the likelihood ratio between the true bridge and proposal process. Furthermore, the dispersion coefficient σ does not need to be invertible. In a companion paper (Van der Meulen and Schauer (2015)) we show how guided proposals can be used for Bayesian estimation of discretely observed diffusions.

1.5. Organization

The main results of the paper are presented in Section 2. proofs are given in Sections 3–6.

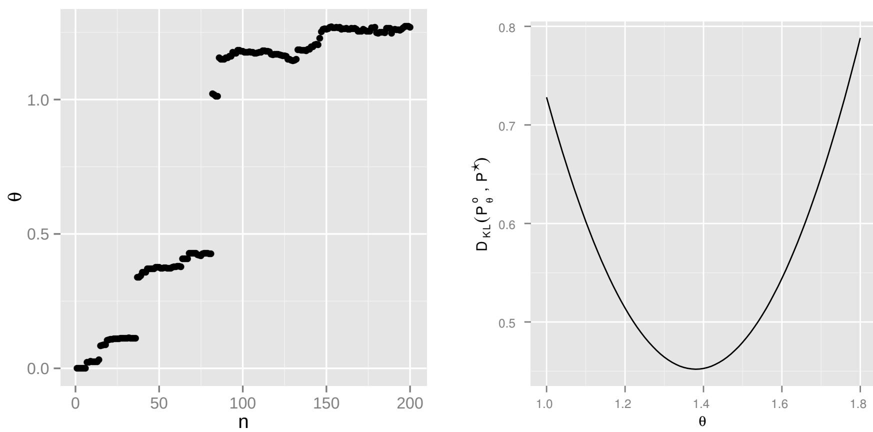

Figure 2. Left: trace plot of ϑ using the stochastic gradient descent algorithm. Right: $\vartheta \mapsto D_{\mathrm{KL}}(\mathbb{P}^\star \| \mathbb{P}^\circ_\vartheta)$, estimated with 100000 simulated bridges.

1.6. General notations and conventions

1.6.1. Vector- and matrix norms

The transpose of a matrix A is denoted by A'. The determinant and trace of a square matrix A are denoted by $|A|$ and $\mathrm{tr}(A)$ respectively. For vectors, we will always use the Euclidean norm, which we denote by $\|x\|$. For a $d \times d'$ matrix A, we denote its Frobenius norm by $\|A\|_F = (\sum_{i=1}^d \sum_{j=1}^{d'} A_{ij}^2)^{1/2}$. The spectral norm, the operator norm induced by the Euclidean norm will de denoted by $\|A\|$, so

$$\|A\| = \sup\{\|Ax\|,\ x \in \mathbb{R}^{d'} \text{ with } \|x\| = 1\}.$$

Both norms are submultiplicative, $\|Ax\| \leq \|A\|_F \|x\|$ and $\|Ax\| \leq \|A\| \|x\|$. The identity matrix will be denoted by Id.

1.6.2. Derivatives

For $f\colon \mathbb{R}^m \to \mathbb{R}^n$ we denote by $\mathrm{D}f$ the $m \times n$-matrix with element (i, j) given by $\mathrm{D}_{ij} f(x) = (\partial f_j / \partial x_i)(x)$. If $n = 1$, then $\mathrm{D}f$ is the column vector containing all partial derivatives of f, that is $\nabla_x f$ from the first section. In this setting we write the i-th element of $\mathrm{D}f$ by $\mathrm{D}_i f(x) = (\partial f / \partial x_i)(x)$ and denote $\mathrm{D}^2 f = \mathrm{D}(\mathrm{D}f)$ so that $\mathrm{D}^2_{ij} f(x) = \partial^2 f(x)/(\partial x_i \partial x_j)$. If $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ does not depend on x, then $\mathrm{D}(Ax) = A'$. Further, for $f\colon \mathbb{R}^n \to \mathbb{R}^n$ we have

$$\mathrm{D}(f(x)' A f(x)) = (\mathrm{D}f(x))'(A + A')f(x).$$

Derivatives with respect to time are always denoted as $\partial/\partial t$.

1.6.3. Inequalities

We write $x \lesssim y$ to denote that there is a universal (deterministic) constant $C > 0$ such that $x \leq Cy$.

2. Main results

2.1. Setup

We continue to use the notation of the introduction, so the process X is the unconditioned process defined as the solution to the SDE (1.1). We assume throughout that the functions b and σ are Lipschitz in both arguments, satisfy a linear growth condition in their second argument and that σ is uniformly bounded. These conditions imply in particular that the SDE has a unique strong solution (e.g. Karatzas and Shreve (1991)). The auxiliary process $\widetilde{X}$ whose transition densities are used in the proposal process is defined as the solution of an SDE like (1.1) as well, but with drift $\widetilde{b}$ instead of b and dispersion $\widetilde{\sigma}$ instead of σ. The functions $\widetilde{b}$ and $\widetilde{\sigma}$ are assumed to satisfy the same Lipschitz, linear growth and boundedness conditions as b and σ. We write $a = \sigma\sigma'$ and $\widetilde{a} = \widetilde{\sigma}\widetilde{\sigma}'$.

The processes X and $\widetilde{X}$ are assumed to have smooth transition densities with respect to Lebesgue measure. More precisely, denoting the law of the process X started in x at time s by $\mathbb{P}^{(s,x)}$, we assume that that for $0 \leq s < t$ and $y \in \mathbb{R}^d$

$$\mathbb{P}^{(s,x)}(X_t \in \,\mathrm{d}y) = p(s, x; t, y)\,\mathrm{d}y$$

and similarly for the process $\widetilde{X}$, whose transition densities are denoted by $\widetilde{p}$ instead of p. The infinitesimal generators of X and $\widetilde{X}$ are denoted by $\mathcal{L}$ and $\widetilde{\mathcal{L}}$, respectively, so that

$$(\mathcal{L}f)(s,x) = \sum_{i=1}^{d} b_i(s,x)\,\mathrm{D}_i f(s,x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(s,x)\,\mathrm{D}^2_{ij} f(s,x), \tag{2.1}$$

for $f \in C^{1,2}(\mathbb{R} \times \mathbb{R}^d, \mathbb{R})$, and similarly for $\widetilde{\mathcal{L}}$ (with $\widetilde{b}$ and $\widetilde{a}$). Under regularity conditions, which we assume to be fulfilled, we have that the transition densities of $\widetilde{X}$ satisfy Kolmogorov's backward equation:

$$\frac{\partial}{\partial s}\widetilde{p}(s,x;t,y) + (\widetilde{\mathcal{L}}\widetilde{p})(s,x;t,y) = 0$$

(here $\widetilde{\mathcal{L}}$ acts on s, x). (See for instance Karatzas and Shreve (1991), p. 368, for sufficient regularity conditions.)

We fix a time horizon $T > 0$ and a point $v \in \mathbb{R}^d$ such that for all $s \leq T$ and $x \in \mathbb{R}^d$ it holds that $p(s, x; T, v) > 0$ and $\widetilde{p}(s, x; T, v) > 0$. The target bridge process $X^\star = (X^\star_t : t \in [0, T])$ is defined by conditioning the original process X to hit the point v at time T. The proposal process $X^\circ = (X^\circ_t : t \in [0, T])$ is defined as the solution of ($\circ$)–($\circ\circ$). In the results ahead we will impose conditions on the transition densities $\widetilde{p}$ of $\widetilde{X}$ that imply that this SDE has a unique solution. All processes are assumed to be defined on the canonical path space and $(\mathcal{F}_t)$ is the corresponding canonical filtration.

For easy reference, the following table briefly describes the various processes around.

X	original, unconditioned diffusion process
$X^\star$	corresponding bridge, conditioned to hit v at time T, defined through ($\star$)
X°	proposal process defined through ($\circ$)
$\widetilde{X}$	auxiliary process whose transition densities $\widetilde{p}$ appear in the definition of X°.

We denote the laws of X, $X^\star$ and X° viewed as measures on the space $C^d([0,t], \mathbb{R}^d)$ of continuous functions from $[0,t]$ to $\mathbb{R}^d$ equipped with Borel-σ-algebra by $\mathbb{P}_t$, $\mathbb{P}_t^\star$ and $\mathbb{P}_t^\circ$ respectively. In case $t = T$ we drop the subscript T.

2.2. Main results

The end-time T and the end-point v of the conditioned diffusion will be fixed throughout. To emphasize the dependence of the transition density on the first two arguments and to shorten notation, we will often write

$$p(s,x) = p(s,x;T,v).$$

Motivated by the guiding term in the drift of $X^\star$ (see ($\star\star$)), we further introduce the notations

$$\boxed{R(s,x) = \log p(s,x), \quad r(s,x) = \mathrm{D}R(s,x), \quad H(s,x) = -\mathrm{D}^2 R(s,x).}$$

Here D acts on x. Similarly the functions $\widetilde{R}$, $\widetilde{r}$ and $\widetilde{H}$ are defined by starting with the transition densities $\widetilde{p}$ in the place of p.

The following proposition deals with the laws of the processes X, X° and $X^\star$ on the interval $[0,t]$ for $t < T$ (strict inequality is essential). Equivalence of these laws is clear from Girsanov's theorem. The proposition gives expressions for the corresponding Radon-Nikodym derivatives, which are derived using Kolmogorov's backward equation. The proof of this result can be found in Section 3.

***Proposition* 1.** *Assume for all $x, y \in \mathbb{R}^d$ and $t \in [0,T)$*

$$\|\widetilde{r}(t,x)\| \lesssim 1 + \frac{\|x - v\|}{T - t}, \qquad \|\widetilde{r}(t,y) - \widetilde{r}(t,x)\| \lesssim \frac{\|y - x\|}{T - t}. \tag{2.2}$$

Define the process ψ by

$$\psi(t) = \exp\left(\int_0^t G(s, X_s^\circ)\,\mathrm{d}s\right), \quad t < T, \tag{2.3}$$

where

$$G(s,x) = (b(s,x) - \widetilde{b}(s,x))'\widetilde{r}(s,x) \\ - \frac{1}{2}\mathrm{tr}\left([a(s,x) - \widetilde{a}(s,x)]\left[\widetilde{H}(s,x) - \widetilde{r}(s,x)\widetilde{r}(s,x)'\right]\right).$$

Then for $t \in [0, T)$ the laws $\mathbb{P}_t$, $\mathbb{P}_t^\circ$ and $\mathbb{P}_t^\star$ are equivalent and we have

$$\frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\mathbb{P}_t^\circ}(X^\circ) = \frac{\widetilde{p}(0,u;T,v)}{\widetilde{p}(t,X_t^\circ;T,v)}\,\psi(t),$$
$$\frac{\mathrm{d}\mathbb{P}_t^\star}{\mathrm{d}\mathbb{P}_t^\circ}(X^\circ) = \frac{\widetilde{p}(0,u;T,v)}{p(0,u;T,v)}\frac{p(t,X_t^\circ;T,v)}{\widetilde{p}(t,X_t^\circ;T,v)}\,\psi(t). \tag{2.4}$$

Proposition 1 is not of much use for simulating diffusion bridges unless its statements can be shown to hold in the limit $t \uparrow T$ as well. One would like to argue that in fact we have equivalence of measures on the whole interval $[0, T]$ and that

$$\frac{\mathrm{d}\mathbb{P}_T^\star}{\mathrm{d}\mathbb{P}_T^\circ}(X^\circ) = \frac{\widetilde{p}(0,u;T,v)}{p(0,u;T,v)}\,\psi(T). \tag{2.5}$$

As $\psi(T)$ does not depend on p, samples from X° can then be used as proposals for $X^\star$ in a Metropolis-Hastings sampler, for instance. Numerical evaluation of $\psi(T)$ is somewhat simplified by the fact that no stochastic integral appears in its expression. To establish (2.5) we need to put appropriate conditions on the processes X and $\widetilde{X}$ that allow us to control the behaviour of the bridge processes X^* and X° near time T.

***Assumption* 1.** *For the auxiliary process $\widetilde{X}$ we assume the following:*

(i) For all bounded, continuous functions $f\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ the transition densities $\widetilde{p}$ of $\widetilde{X}$ satisfy

$$\lim_{t \uparrow T} \int f(t,x)\widetilde{p}(t,x;T,v)\,\mathrm{d}x = f(T,v). \tag{2.6}$$

(ii) For all $x, y \in \mathbb{R}^d$ and $t \in [0,T)$, the functions $\widetilde{r}$ and $\widetilde{H}$ satisfy

$$\|\widetilde{r}(t,x)\| \lesssim 1 + \|x - v\|(T-t)^{-1}$$
$$\|\widetilde{r}(t,x) - \widetilde{r}(t,y)\| \lesssim \|y - x\|(T-t)^{-1}$$
$$\|\widetilde{H}(t,x)\| \lesssim (T-t)^{-1} + \|x - v\|(T-t)^{-1}.$$

(iii) There exist constants $\widetilde{\Lambda}, \widetilde{C} > 0$ such that for $0 < s < T$,

$$\widetilde{p}(s,x;T,v) \le \widetilde{C}(T-s)^{-d/2}\exp\left(-\widetilde{\Lambda}\frac{\|v-x\|^2}{T-s}\right)$$

uniformly in x.

Roughly speaking, Assumption 1 requires that the process $\widetilde{X}$, which we choose ourselves, is sufficiently nicely behaved near time T.

Assumption 2. *For $M > 1$ and $u \geq 0$ define $g_M(u) = \max(1/M, 1 - Mu)$. There exist constants $\Lambda, C > 0$, $M > 1$ 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 $\|\mu_t(s,x) - x\| < M(t-s)\|x\|$ and $\|\mu_t(s,x)\|^2 \geq g_M(t-s)\|x\|^2$, so that for all $s < t \leq T$ and $x, y \in \mathbb{R}^d$,*

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

Assumption 2 refers to the generally unknown transition densities of X. In case the drift of X is bounded, assumption 2 is implied by the stronger Aronson's inequality (cf. Aronson (1967)). However, assumption 2 also holds for example for linear processes which in general have unbounded drift.

Assumption 3. *There exist an $\varepsilon \in (0, 1/6)$ and an a.s. finite random variable M such that for all $t \in [0,T]$, it a.s. holds that*

$$\|X_t^\circ - v\| \leq M(T-t)^{1/2-\varepsilon}.$$

This third assumption requires that the proposal process X° does not only converge to v as $t \uparrow T$, as it obviously should, but that it does so at an appropriate speed. A requirement of this kind can not be essentially avoided, since in general two bridges can only be equivalent if they are pulled to the endpoint with the same force. Theorem 2 below asserts that this assumption holds in case $\widetilde{X}$ is a linear process, provided its diffusion coefficient coincides with that of the process X at the final time T.

We can now state the main results of the paper.

Theorem 1. *Suppose that Assumptions 1, 2 and 3 hold and that $\widetilde{a}(T,v) = a(T,v)$. Then the laws of the bridges $X^\star$ and X° are equivalent on $[0,T]$ and (2.5) holds, with ψ as in Proposition 1.*

We complement this general theorem with a result that asserts, as already mentioned, that Assumptions 1 and 3 hold for a class of processes $\widetilde{X}$ given by linear SDEs.

Theorem 2. *Assume $\widetilde{X}$ is a linear process with dynamics governed by the stochastic differential equation*

$$\mathrm{d}\widetilde{X}_t = \widetilde{B}(t)\widetilde{X}_t \,\mathrm{d}t + \widetilde{\beta}(t)\,\mathrm{d}t + \widetilde{\sigma}(t)\,\mathrm{d}W_t, \tag{2.7}$$

for non-random matrix and vector functions $\widetilde{B}$, $\widetilde{\beta}$ and $\widetilde{\sigma}$.

(i) If $\widetilde{B}$ and $\widetilde{\beta}$ are continuously differentiable on $[0,T]$, $\widetilde{\sigma}$ is Lipschitz on $[0,T]$ and there exists an $\eta > 0$ such that for all $s \in [0,T]$ and all $y \in \mathbb{R}^d$,

$$y'\widetilde{a}(s)y \geq \eta\|y\|^2,$$

then $\widetilde{X}$ satisfies Assumption 1.

(ii) Suppose moreover that $\widetilde{a}(T) = a(T, v)$, *that there exists an* $\varepsilon > 0$ *such that for all* $s \in [0,T]$, $x \in \mathbb{R}^d$ *and* $y \in \mathbb{R}^d$

$$y' a(s,x) y \geq \varepsilon \|y\|^2, \tag{2.8}$$

and that b is of the form $b(s,x) = B(s,x)x + \beta(s,x)$, *where* B *is a bounded matrix-valued function and* β *is a bounded vector-valued function. Then there exists an a.s. finite random variable* M *such that, a.s.,*

$$\|X_t^\circ - v\| \leq M \sqrt{(T-t) \log\log\left(\frac{1}{T-t} + e\right)}$$

for all $t \in [0,T]$. *In particular, Assumption 3 holds for any* $\varepsilon > 0$.

The proofs of Theorems 1 and 2 can be found in Sections 4–6.

***Remark* 1.** *Extending absolute continuity of* $X^\star$ *and* X° *on* $[0, T-\varepsilon]$ *(*$\varepsilon > 0$*) to absolute continuity on* $[0,T]$ *is a subtle issue. This can already be seen from a very simple example in the one-dimensional case. Suppose* $d = d' = 1$, $v = 0$, $b \equiv 0$ *and* $\sigma(t,x) \equiv 1$. *That is,* $X^\star$ *is the law of a Brownian bridge from* 0 *at time* 0 *to* 0 *at time* T *satisfying the stochastic differential equation*

$$\mathrm{d}X_t^\star = -\frac{X_t^\star}{T-t}\,\mathrm{d}t + \mathrm{d}W_t.$$

Suppose we take $\widetilde{X}_t = \widetilde{\sigma}\,\mathrm{d}W_t$, *so that* X° *satisfies the stochastic differential equation*

$$\mathrm{d}X_t^\circ = -\frac{1}{\widetilde{\sigma}^2}\frac{X_t^\circ}{T-t}\,\mathrm{d}t + \mathrm{d}W_t.$$

It is a trivial fact that X° *and* $X^\star$ *are absolutely continuous on* $[0,T]$ *if* $\widetilde{\sigma} = 1$ *(this also follows from theorem 2). It is natural to wonder whether this condition is also necessary. The answer to this question is yes, as we now argue. Lemma 6.5 in Hida and Hitsuda (1993) gives a general result on absolute continuity of Gaussian measures. From this result it follows that* X° *and* $X^\star$ *are absolutely continuous on* $[0,T]$ *if and only if for the symmetrized Kullback-Leibler divergences*

$$d_t = \mathbb{E}\left[\log \frac{\mathrm{d}\mathbb{P}_t^\star}{\mathrm{d}\mathbb{P}_t^\circ}(X^\star)\right] + \mathbb{E}\left[\log \frac{\mathrm{d}\mathbb{P}_t^\circ}{\mathrm{d}\mathbb{P}_t^\star}(X^\circ)\right]$$

it holds that $\sup_{t\in[0,T)} d_t < \infty$. *We consider the second term. Denoting* $\alpha = 1/\widetilde{\sigma}^2$, *Girsanov's theorem gives*

$$\log \frac{\mathrm{d}\mathbb{P}_t^\circ}{\mathrm{d}\mathbb{P}_t^\star}(X^\circ) = \int_0^t (1-\alpha)\frac{X_s^\circ}{T-s}\,\mathrm{d}W_s + \frac{1}{2}\int_0^t (\alpha-1)^2 \left(\frac{X_s^\circ}{T-s}\right)^2 \mathrm{d}s$$

By Itō's formula $\frac{X_t^\circ}{T-t} = (1-\alpha)\int_0^t \frac{X_s^\circ}{(T-s)^2}\,\mathrm{d}s + \int_0^t \frac{-1}{T-s}\,\mathrm{d}W_s$.
This is a linear equation with solution

$$\frac{X_t^\circ}{T-t} = -(T-t)^{-1+\alpha} \int_0^t (T-s)^{-\alpha}\,\mathrm{d}W_s,$$

hence

$$\mathbb{E}\left[\left(\frac{X_t^\circ}{T-t}\right)^2\right] = (T-t)^{-2+2\alpha} \int_0^t (T-s)^{-2\alpha}\, \mathrm{d}s$$

For $t < T$, $\int_0^t \mathbb{E}\left[\left(\frac{X_s^\circ}{T-s}\right)^2\right] \mathrm{d}s < \infty$, *so* $\mathbb{E}\left[\int_0^t \frac{X_s^\circ}{T-s}\, \mathrm{d}W_s\right] = 0$. *Therefore*

$$\mathbb{E}\left[\log \frac{\mathrm{d}\mathbb{P}_t^\circ}{\mathrm{d}\mathbb{P}_t^\star}(X^\circ)\right] = \frac{1}{2}(\alpha-1)^2 \int_0^t (T-s)^{-2+2\alpha} \int_0^s (T-\tau)^{-2\alpha}\, \mathrm{d}\tau\, \mathrm{d}s.$$

Unless, $\alpha = 1$, *this diverges for* $t \uparrow T$. *We conclude that the laws of* $X^\star$ *and* X° *are singular if* $\alpha \neq 1$.

Remark 2. *For implementation purposes integrals in likelihood ratios and solutions to stochastic differential equations need to be approximated on a finite grid. This is a subtle numerical issue as the drift of our proposal bridge has a singularity near its endpoint. In a forthcoming work Van der Meulen and Schauer (2015) we show how this problem can be dealt with. The main idea in there is the introduction of a time-change and space-scaling of the proposal process that allows for numerically accurate discretisation and evaluation of the likelihood.*

3. Proof of Proposition 1

We first note that by equation (2.2), $\widetilde{r}$ is Lipschitz in its second argument on $[0,t]$ and satisfies a linear growth condition. Hence, a unique strong solution of the SDE for X° exists.

By Girsanov's theorem (see e.g. Liptser and Shiryaev (2001)) the laws of the processes X and X° on $[0,t]$ are equivalent and the corresponding Radon-Nikodym derivative is given by

$$\frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\mathbb{P}_t^\circ}(X^\circ) = \exp\Big(\int_0^t \beta_s'\, \mathrm{d}W_s - \frac{1}{2}\int_0^t \|\beta_s\|^2\, \mathrm{d}s\Big),$$

where W is a Brownian motion under $\mathbb{P}_t^\circ$ and $\beta_s = \beta(s, X_s^\circ)$ solves $\sigma(s, X_s^\circ)\beta(s, X_s^\circ) = b(s, X_s^\circ) - b^\circ(s, X_s^\circ)$. (Here we lightened notation by writing β_s instead of $\beta(s, X_s^\circ)$. In the remainder of the proof we follow the same convention and apply it to other processes as well.) Observe that by definition of $\widetilde{r}$ and b° we have $\beta_s = -\sigma_s'\widetilde{r}_s$ and $\|\beta_s\|^2 = \widetilde{r}_s' a_s \widetilde{r}_s$, hence

$$\frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\mathbb{P}_t^\circ}(X^\circ) = \exp\Big(-\int_0^t \widetilde{r}_s'\sigma_s\, \mathrm{d}W_s - \frac{1}{2}\int_0^t \widetilde{r}_s' a_s \widetilde{r}_s\, \mathrm{d}s\Big).$$

Denote the infinitesimal operator of X° by $\mathcal{L}^\circ$. By definition of X° and $\widetilde{R}$ we have $\mathcal{L}^\circ\widetilde{R} = \mathcal{L}\widetilde{R} + \widetilde{r}' a \widetilde{r}$. By Itō's formula, it follows that

$$\widetilde{R}_t - \widetilde{R}_0 = \int_0^t \Big(\frac{\partial}{\partial s}\widetilde{R} + \mathcal{L}\widetilde{R}\Big)\, \mathrm{d}s + \int_0^t \widetilde{r}_s' a_s \widetilde{r}_s\, \mathrm{d}s + \int_0^t \widetilde{r}_s'\sigma_s\, \mathrm{d}W_s.$$

Combined with what we found above we get $\frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\mathbb{P}_t^\circ}(X^\circ) = \mathrm{e}^{-(\widetilde{R}_t-\widetilde{R}_0)}\mathrm{e}^{\int_0^t G_s\,\mathrm{d}s}$, where $\left(\frac{\partial}{\partial s}\widetilde{R} + \mathcal{L}\widetilde{R}\right) + \frac{1}{2}\widetilde{r}' a\widetilde{r}$. By Lemma 1 ahead the first term between brackets on the right-hand-side of this display equals $\mathcal{L}\widetilde{R} - \widetilde{\mathcal{L}}\widetilde{R} - \frac{1}{2}\widetilde{r}'\widetilde{a}\widetilde{r}$. Substituting this in the expression for G gives

$$G = (b-\widetilde{b})'\widetilde{r} - \frac{1}{2}\mathrm{tr}\left((a-\widetilde{a})\widetilde{H}\right) + \frac{1}{2}\widetilde{r}'(a-\widetilde{a})\widetilde{r},$$

which is as given in the statement of the theorem. Since $-(\widetilde{R}_t - \widetilde{R}_0) = \log \widetilde{p}(0,u)/\widetilde{p}(t,X_t^\circ)$, we arrive at the first assertion of the proposition.

To prove the second assertion, let $0 = t_0 < t_1 < t_2 < \cdots < t_N < t < T$ and define $x_0 = u$. If g is a bounded function on $\mathbb{R}^{(N+1)}$, then standard calculations show $\mathbb{E}\left[g(X_{t_1}^\star,\ldots,X_{t_N}^\star,X_t^\star)\frac{1}{p(t,X_t^\star)}\right] = \mathbb{E}\left[g(X_{t_1},\ldots,X_{t_N},X_t)\frac{1}{p(0,u)}\right]$, using the abbreviation $p(t,x) = p(t,x;T,v)$. Since the grid and g are arbitrary, this proves that for $t < T$,

$$\frac{\mathrm{d}\mathbb{P}_t^\star}{\mathrm{d}\mathbb{P}_t}(X) = \frac{p(t,X_t;T,v)}{p(0,u;T,v)}. \tag{3.1}$$

Combined with the first statement of the proposition, this yields the second one.

Lemma 1. *$\widetilde{R}$ satisfies the equation*

$$\frac{\partial}{\partial s}\widetilde{R} + \widetilde{\mathcal{L}}\widetilde{R} = -\frac{1}{2}\widetilde{r}'\widetilde{a}\widetilde{r}.$$

Proof. First note that

$$\mathrm{D}_{ij}^2\widetilde{R}(s,x) = \frac{\mathrm{D}_{ij}^2\widetilde{p}(s,x)}{\widetilde{p}(s,x)} - \left(\mathrm{D}_i\widetilde{R}(s,x)\right)\left(\mathrm{D}_j\widetilde{R}(s,x)\right). \tag{3.2}$$

Next, Kolmogorov's backward equation is given by

$$\frac{\partial}{\partial s}\widetilde{p}(s,x) + \left(\widetilde{\mathcal{L}}\widetilde{p}\right)(s,x) = 0.$$

Dividing both sides by $\widetilde{p}(s,x)$ and using (2.1) we obtain

$$\frac{\partial}{\partial s}\widetilde{R}(s,x) = -\sum_{i=1}^d \widetilde{b}_i(s,x)\,\mathrm{D}_i\widetilde{R}(s,x) - \frac{1}{2}\sum_{i,j=1}^d \widetilde{a}_{ij}(s,x)\frac{\mathrm{D}_{ij}^2\widetilde{p}(s,x)}{\widetilde{p}(s,x)}$$

Now substitute (3.2) for the second term on the right-hand-side and re-order terms to get the result. □

4. Proof of Theorem 1

Auxiliary lemmas used in the proof are gathered in Subsection 4.1 ahead. As before we use the notation $p(s, x) = p(s, x; T, v)$ and similar for $\widetilde{p}$. Moreover, we define $\bar{p} = \frac{\widetilde{p}(0,u)}{p(0,u)}$. The main part of the proof consists in proving that $\bar{p}\psi(T)$ is indeed a Radon-Nikodym derivative, i.e. that it has expectation 1. For $\varepsilon \in (0, 1/6)$ as in Assumption 3, $m \in \mathbb{N}$ and a stochastic process $Z = (Z_t,\, t \in [0, T])$, define

$$\sigma_m(Z) = T \wedge \inf_{t\in[0,T]} \{|Z_t - v| \geq m(T - t)^{1/2-\varepsilon}\}.$$

We suppress the dependence on ε in the notation. We write $\sigma_m = \sigma_m(X), \sigma_m^\star = \sigma_m(X^\star)$, and $\sigma_m^\circ = \sigma_m(X^\circ)$. Note that $\sigma_m^\circ \uparrow T$ holds in probability, by Assumption 3.

By Proposition 1, for any $t < T$ and bounded, $\mathcal{F}_t$-measurable f, we have

$$\mathbb{E}\left[f(X^\star)\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\right] = \mathbb{E}\left[f(X^\circ)\bar{p}\,\psi(t)\right]. \tag{4.1}$$

By Corollary 1 in Subsection 4.1, for each $m \in \mathbb{N}$, $\sup_{0\leq t\leq T} \psi(t)$ is uniformly bounded on the event $\{T = \sigma_m^\circ\}$. Hence, by dominated convergence,

$$\mathbb{E}\left[\bar{p}\,\psi(T)\mathbf{1}_{T=\sigma_m^\circ}\right] = \lim_{t\uparrow T}\mathbb{E}\left[\bar{p}\,\psi(t)\mathbf{1}_{t\leq\sigma_m^\circ}\right] \leq \lim_{t\uparrow T}\mathbb{E}\left[\bar{p}\,\psi(t)\right] = \lim_{t\uparrow T}\mathbb{E}\left[\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\right] = 1.$$

Here the final two equalities follow from equation (4.1) and Lemma 3, respectively. Taking the limit $m \to \infty$ we obtain $\mathbb{E}\left[\bar{p}\,\psi(T)\right] \leq 1$, by monotone convergence. For the reverse inequality note that by similar arguments as just used we obtain

$$\mathbb{E}\left[\bar{p}\,\psi(T)\right] \geq \mathbb{E}\left[\bar{p}\,\psi(T)\mathbf{1}_{T=\sigma_m^\circ}\right] = \lim_{t\uparrow T}\mathbb{E}\left[\bar{p}\,\psi(t)\mathbf{1}_{t\leq\sigma_m^\circ}\right] = \lim_{t\uparrow T}\mathbb{E}\left[\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\mathbf{1}_{t\leq\sigma_m^\star}\right].$$

By Lemma 5, the right-hand-side of the preceding display tends to 1 as $m \to \infty$. We conclude that $\bar{p}\,\mathbb{E}\left[\psi(T)\right] = 1$.

To complete the proof we note that by equation (4.1) and Lemma 3 we have $\bar{p}\,\mathbb{E}\left[\psi(t)\right] \to 1$ as $t \uparrow T$. In view of the preceding and Scheffé's Lemma this implies that $\psi(t) \to \psi(T)$ in L^1-sense as $t \uparrow T$. Hence for $s < T$ and a bounded, $\mathcal{F}_s$-measurable functional g,

$$\mathbb{E}\left[g(X^\circ)\bar{p}\psi(T)\right] = \lim_{t\uparrow T}\mathbb{E}\left[g(X^\circ)\frac{\widetilde{p}(t, X_t^\circ)}{p(t, X_t^\circ)}\left(\bar{p}\frac{p(t, X_t^\circ)}{\widetilde{p}(t, X_t^\circ)}\psi(t)\right)\right].$$

Proposition 1 implies that for $t > s$, the expectation on the right equals

$$\mathbb{E}\left[g(X^\star)\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\right].$$

By Lemma 3 this converges to $\mathbb{E}\,g(X^\star)$ as $t \uparrow T$ and we find that $\mathbb{E}\,g(X^\circ)\bar{p}\psi(T) = \mathbb{E}\,g(X^\star)$. Since $s < t$ and g are arbitrary, this completes the proof.

4.1. Auxiliary results used in the proof of Theorem 1

Lemma 2. *Suppose Assumptions 1(iii) and 2 apply. For*

$$f_t(s,x) = \int p(s,x;t,z)\,\widetilde{p}(t,z;T,v)\,\mathrm{d}z \qquad 0 \le s < t < T,\ x \in \mathbb{R}^d, \tag{4.2}$$

there exist positive constants c and λ such that

$$f_t(s,x) \le c(T-s)^{-d/2} \exp\left(-\lambda \frac{\|v-x\|^2}{T-s}\right).$$

Proof. Let $C, \widetilde{C}, \Lambda$ and $\widetilde{\Lambda}$ be the constant appearing in assumptions 1(iii) and 2. Define $\bar{\Lambda} = \min(\Lambda, \widetilde{\Lambda})/2$. Denote by $\varphi(z;\mu,\Sigma)$ the $N(\mu,\Sigma)$-density, evaluated at z. Then there exists a $\bar{C} > 0$ such that

$$\begin{aligned} f_t(s,x) &\le \bar{C} \int \varphi(z;\mu_t(s,x), \bar{\Lambda}^{-1}(t-s)\mathrm{Id}_d)\varphi(v-z;0,\bar{\Lambda}^{-1}(T-t)\mathrm{Id}_d)\,\mathrm{d}z \\ &= \bar{C}\varphi(v;\mu_t(s,x), \bar{\Lambda}^{-1}(T-s)\mathrm{Id}_d). \end{aligned}$$

Using the second assumed bound on $\mu_t(s,x)$ and the fact that $g_M(t-s) \ge 1/M$ we get

$$\|v-\mu_t(s,x)\|^2 \ge M^{-1}\|v-x\|^2 + (1-g_M(t-s))\|v\|^2 - 2v'(\mu_t(s,x) - g_M(t-s)x).$$

By Cauchy-Schwarz, the triangle inequality and the first assumed inequality we find

$$|v'(\mu_t(s,x) - g_M(t-s)x)| \le \|v\|\|x\|\left(M(t-s) + 1 - g_M(t-s)\right).$$

We conclude that

$$\begin{aligned} \frac{\|v-\mu_t(s,x)\|^2}{T-s} \ge& \frac{1}{M}\frac{\|v-x\|^2}{T-s} + \frac{1-g_M(t-s)}{T-s}\|v\|^2 \\ &- 2\left(\frac{M(t-s)}{T-s} + \frac{1-g_M(t-s)}{T-s}\right)\|v\|\|x\|. \end{aligned}$$

By definition of g_M, the multiplicative terms appearing in front of $\|v\|^2$ and $\|v\|\|x\|$ are both bounded. As there exist constants $D_1 > 0$ and $D_2 \in \mathbb{R}$ such that the third term on the right-hand-side can be lower bounded by $D_1\|v-x\|^2 + D_2$ the result follows. □

The following lemma is similar to Lemma 7 in Delyon and Hu (2006).

Lemma 3. *Suppose Assumptions 1(i), 1(iii) and 2 apply. If $0 < t_1 < t_2 < \cdots < t_N < t < T$ and $g \in C_b(\mathbb{R}^{Nd})$, then*

$$\lim_{t\uparrow T} \mathbb{E}\left[g(X^\star_{t_1},\dots,X^\star_{t_N})\frac{\widetilde{p}(t,X^\star_t)}{p(t,X^\star_t)}\right] = \mathbb{E}\left[g(X^\star_{t_1},\dots,X^\star_{t_N})\right].$$

Lemma 4. *Assume*

1. $b(s,x)$, $\widetilde{b}(s,x)$, $a(s,x)$ *and* $\widetilde{a}(s,x)$ *are locally Lipschitz in* s *and globally Lipschitz in* x*;*
2. $\widetilde{a}(T,v) = a(T,v)$.

Then for all x *and for all* $s \in [0,T)$,

$$\|b(s,x) - \widetilde{b}(s,x)\| \lesssim 1 + \|x - v\| \tag{4.3}$$

and

$$\|a(s,x) - \widetilde{a}(s,x)\|_F \lesssim (T - s) + \|x - v\|. \tag{4.4}$$

If in addition $\widetilde{r}$ *and* $\widetilde{H}$ *satisfy the bounds*

$$\begin{aligned}&\|\widetilde{r}(s,x)\| \lesssim 1 + \|x - v\|(T-s)^{-1}\\ &\|\widetilde{H}(s,x)\|_F \lesssim (T-s)^{-1} + \|x - v\|(T-s)^{-1},\end{aligned}$$

then

$$|G(s,x)| \lesssim 1 + (T-s) + \|x - v\| + \frac{\|x-v\|}{T-s} + \frac{\|x-v\|^2}{T-s} + \frac{\|x-v\|^3}{(T-s)^2}.$$

Proof. Since $|\mathrm{tr}\,(AB)| \le \|A\|_F\|B\|_F$ and $\|AB\|_F \le \|A\|_F\|B\|_F$ for compatible matrices A and B, we have

$$\begin{aligned}|G(s,x)| \le \|b(s,x) - \widetilde{b}(s,x)\|\|\widetilde{r}(s,x)\| + &\\ \|a(s,x) - \widetilde{a}(s,x)\|_F &\left(\|\widetilde{H}(s,x)\|_F + \|\widetilde{r}(s,x)\|^2\right).\end{aligned} \tag{4.5}$$

Bounding $\|b(s,x) - \widetilde{b}(s,x)\|$ proceeds by using the assumed Lipschitz properties for b and $\widetilde{b}$. We have

$$\begin{aligned}\|b(s,x) - \widetilde{b}(s,x)\| &\le \|b(s,x) - b(s,v)\| + \|b(s,v) - \widetilde{b}(s,v)\| + \|\widetilde{b}(s,v) - \widetilde{b}(s,x)\|\\ &\le L_b\|x - v\| + \|b(s,v) - \widetilde{b}(s,v)\| + L_{\widetilde{b}}\|v - x\|,\end{aligned}$$

where L_b and $L_{\widetilde{b}}$ denote Lipschitz constants. Since $b(\cdot, v)$ and $\widetilde{b}(\cdot, v)$ are continuous on $[0,T]$, we have $\|b(s,v) - \widetilde{b}(s,v)\| \lesssim 1$. This inequality together with preceding display gives (4.3).

Bounding $\|a(s,x) - \widetilde{a}(s,x)\|_F$ proceeds by using the assumed Lipschitz properties for a and $\widetilde{a}$ together with $\widetilde{a}(T,v) = a(T,v)$. We have

$$\begin{aligned}\|a(s,x) - \widetilde{a}(s,x)\|_F &\le \|a(s,x) - a(T,x)\|_F + \|a(T,x) - a(T,v)\|_F + \|a(T,v) - \widetilde{a}(T,v)\|_F\\ &\quad + \|\widetilde{a}(T,v) - \widetilde{a}(s,v)\|_F + \|\widetilde{a}(s,v) - \widetilde{a}(s,x)\|_F\\ &\lesssim (T-s) + \|x - v\|.\end{aligned}$$

The final result follows upon plugging in the derived estimates for $\|b(s,x) - \widetilde{b}(s,x)\|$ and $\|a(s,x) - \widetilde{a}(s,x)\|_F$ into equation (4.5) and subsequently using the bounds on $\widetilde{r}$ and $\widetilde{H}$ from the assumptions of the lemma. ☐

Corollary 1. *Under the conditions of Lemma 4, for all $\varepsilon \in (0, 1/6)$ there is a positive constant K (not depending on m) such that for all $t \in [0, T)$*

$$\psi(t)\mathbf{1}_{t\leq\sigma_m^\circ} \leq \exp\left(Km^3\right).$$

Proof. On the event $\{t \leq \sigma_m^\circ\}$ we have

$$\|X_s^\circ - v\| \leq m(T-s)^{1/2-\varepsilon} \qquad \text{for all } s \in [0, t].$$

Together with the result of Lemma 4, this implies that there is a constant $C > 0$ (that does not depend on m) such that for all $s \in [0, t]$

$$\begin{aligned}|G(s, X_s^\circ)| &\leq C\left(1 + m(T-s)^{1/2-\varepsilon} + m(T-s)^{-1/2-\varepsilon} + m^2(T-s)^{-2\varepsilon} + m^3(T-s)^{-1/2-3\varepsilon}\right)\\ &\leq Cm^3\left(1 + (T-s)^{1/2-\varepsilon} + (T-s)^{-1/2-3\varepsilon}\right).\end{aligned}$$

Hence,

$$\psi(t)\mathbf{1}_{t\leq\sigma_m^\circ} \leq \exp\left(Cm^3\int_0^T\left(1 + (T-s)^{1/2-\varepsilon} + (T-s)^{-1/2-3\varepsilon}\right)\mathrm{d}s\right) \leq \exp\left(Km^3\right),$$

for some constant K. □

Lemma 5. *Suppose Assumptions 1(i), 1(iii) and 2 apply. Then*

$$\lim_{m\to\infty}\lim_{t\uparrow T}\mathbb{E}\left[\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\mathbf{1}_{t\leq\sigma_m^\star}\right] = 1.$$

Proof. First,

$$\mathbb{E}\left[\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\mathbf{1}_{t\leq\sigma_m^\star}\right] = \mathbb{E}\left[\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\right] - \mathbb{E}\left[\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\mathbf{1}_{t>\sigma_m^\star}\right].$$

Hence, by Lemma 3, it suffices to prove that the second term tends to 0. For $t < T$

$$\begin{aligned}p(0, u)\,\mathbb{E}\left[\frac{\widetilde{p}(t, X_t^\star)}{p(t, X_t^\star)}\mathbf{1}_{t>\sigma_m^\star}\right] &= \mathbb{E}\left[\widetilde{p}(t, X_t)\mathbf{1}_{t>\sigma_m}\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\widetilde{p}(t, X_t)\mathbf{1}_{t>\sigma_m} \mid \mathcal{F}_{\sigma_m}\right]\right] = \mathbb{E}\left[\mathbf{1}_{t>\sigma_m}\mathbb{E}\left[\widetilde{p}(t, X_t) \mid \mathcal{F}_{\sigma_m}\right]\right]\\ &= \mathbb{E}\left[\mathbf{1}_{t>\sigma_m}\int p(\sigma_m, X_{\sigma_m}; t, z)\widetilde{p}(t, z)\,\mathrm{d}z\right] = \mathbb{E}\left[\mathbf{1}_{t>\sigma_m}f_t(\sigma_m, X_{\sigma_m})\right],\end{aligned}$$

where f_t is defined in equation (4.2). Here we used (3.1) and the strong Markov property. By Lemma 2,

$$\mathbb{E}\left[f_t(\sigma_m, X_{\sigma_m})\right] \lesssim \mathbb{E}\left[(T-\sigma_m)^{-d/2}\exp\left(-\lambda\frac{\|v - X_{\sigma_m}\|^2}{T-\sigma_m}\right)\right].$$

Since $\|v - X_{\sigma_m}\| = m(T-\sigma_m)^{1/2-\varepsilon}$, the right-hand-side can be bounded by a constant times $\mathbb{E}\left[(T-\sigma_m)^{-d/2}\exp\left(-\lambda m^2 (T-\sigma_m)^{-2\varepsilon}\right)\right]$. Note that this expression does not depend on t. The proof is concluded by taking the limit $m \to \infty$. Trivially, $T - \sigma_m \in [0,T]$, so that the preceding display can be bounded by

$$C \sup_{\tau \in [0,\infty)} \tau^{-d/2} \exp\left(-\lambda m^2 \tau^{-2\varepsilon}\right) \leq C \left(\frac{d}{4\lambda m^2 \mathrm{e}\varepsilon}\right)^{\frac{d}{4\varepsilon}}.$$

This tends to 0 as $m \to \infty$. □

5. Proof of Theorem 2(i)

It is well known (see for instance Liptser and Shiryaev (2001)) that the linear process $\widetilde{X}$ is a Gaussian process that can be described in terms of the fundamental $d \times d$ matrix $\Phi(t)$, which satisfies

$$\Phi(t) = \mathrm{Id} + \int_0^t \widetilde{B}(\tau)\Phi(\tau)\,\mathrm{d}\tau.$$

We define $\Phi(t,s) = \Phi(t)\Phi(s)^{-1}$,

$$\mu_t(s,x) = \Phi(t,s)x + \int_s^t \Phi(t,\tau)\widetilde{\beta}(\tau)\,\mathrm{d}\tau \tag{5.1}$$

and

$$K_t(s) = \int_s^t \Phi(t,\tau)a(\tau)\Phi(t,\tau)'\,\mathrm{d}\tau. \tag{5.2}$$

To simplify notation, we use the convention that whenever the subscript t is missing, it has the value of the end time T. So we write $\mu(s,x) = \mu_T(s,x)$ and $K(s) = K_T(s)$. The Gaussian transition densities of the process $\widetilde{X}$ can be explicitly expressed in terms of the objects just defined. In particular we have

$$\widetilde{R}(s,x) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|K(s)| - \frac{1}{2}(v-\mu(s,x))'K(s)^{-1}(v-\mu(s,x)). \tag{5.3}$$

This will allow us to derive explicit expressions for all the functions involved in Assumption 1.

For future purposes, we state a number of properties of $\Phi(t,s)$, which are well known in literature on linear differential equations (proofs can be found for example in Sections 2.1.1 up till 2.1.3 in Chicone (1999)).

- $\Phi(t,s)\Phi(s,\tau) = \Phi(t,\tau)$, $\Phi(t,s)^{-1} = \Phi(s,t)$ and $\frac{\partial \Phi}{\partial s}(t,s) = -\Phi(t,s)B(s)$.
- There is a constant $C \geq 0$ such that for all $s,t \in [0,T]$, $\|\Phi(t,s)\| \leq C$ (this is a consequence of Gronwall's lemma).
- $|\Phi(t,s)| = \exp\left(\int_s^t \mathrm{tr}(\widetilde{B}(u))\,\mathrm{d}u\right)$ (Liouville's formula).
- If $\widetilde{B}(t) \equiv \widetilde{B}$ does not depend on t, $\Phi(t,s) = \exp(\widetilde{B}(t-s)) = \sum_{k=0}^{\infty} \frac{1}{k!}\widetilde{B}^k (t-s)^k$.

By Theorem 1.3 in Chicone (1999), we have that the mappings $(t,s,x) \mapsto \mu_t(s,x)$ and $(t,s) \mapsto \Phi_t(s)$ are continuously differentiable.

The following lemma provides the explicit expressions for the functions $\widetilde{r}$ and $\widetilde{H}$.

Lemma 6. *For $s \in [0,T)$ and $x \in \mathbb{R}^d$*

$$\widetilde{r}(s,x) = \mathrm{D}\widetilde{R}(s,x) = \Phi(T,s)' K(s)^{-1}(v - \mu(s,x))$$

and

$$\begin{aligned}\widetilde{H}(s,x) = \widetilde{H}(s) &= -\mathrm{D}\widetilde{r}(s,x) = \Phi(T,s)'K(s)^{-1}\Phi(T,s) \\ &= \left(\int_s^T \Phi(s,\tau)\widetilde{a}(\tau)\Phi(s,\tau)'\,\mathrm{d}\tau\right)^{-1}.\end{aligned} \tag{5.4}$$

Moreover, we have the relation $\widetilde{r}(s,x) = \widetilde{H}(s)(v(s) - x)$ where

$$v(s) = \Phi(s,T)v - \int_s^T \Phi(s,\tau)\widetilde{\beta}(\tau)\,\mathrm{d}\tau. \tag{5.5}$$

Proof. We use the conventions and rules on differentiations outlined in Section 1.6. Since $K(s)$ is symmetric

$$\begin{aligned}\widetilde{r}(s,x) &= -\mathrm{D}(v - \mu(s,x))K(s)^{-1}(v-\mu(s,x)) \\ &= \Phi(T,s)'K(s)^{-1}(v - \mu(s,x)),\end{aligned}$$

where we used $\mathrm{D}\mu(s,x) = \Phi(s)'$.

By equation (5.1),

$$v - \mu(s,x) = v - \Phi(T,s)x - \int_s^T \Phi(T,\tau)\widetilde{\beta}(\tau)\,\mathrm{d}\tau. \tag{5.6}$$

The expression for $\widetilde{H}$ now follows from

$$\begin{aligned}\widetilde{H}(s) &= -\mathrm{D}(\Phi(T,s)'K(s)^{-1}(v-\mu(s,x))) \\ &= \mathrm{D}(\Phi(T,s)'K(s)^{-1}\Phi(T,s)x) = \Phi(T,s)'K(s)^{-1}\Phi(T,s),\end{aligned}$$

where the second equality follows from equation (5.6).

The final statement follows upon noting that

$$\begin{aligned}\widetilde{r}(s,x) &= \Phi(T,s)'K(s)^{-1}\Phi(T,s)\Phi(s,T)(v-\mu(s,x)) \\ &= \widetilde{H}(s)\Phi(s,T)(v - \mu(s,x)) = \widetilde{H}(s)(v(s) - x).\end{aligned}$$

The last equality follows by multiplying equation (5.6) from the left with $\Phi(s,T)$. □

In the following three subsections we use the explicit computations of the preceding lemma to verify Assumption 1, in order to complete the proof statement (i) of Theorem 2.

5.1. Assumption 1(i)

Lemma 7. *If $f\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ is bounded and continuous then*

$$\lim_{t\to T}\int f(t,z)\widetilde{p}(t,z;T,v)\,\mathrm{d}z = f(T,v).$$

Proof. The log of the transition density of a linear process is given in equation (5.3). Using v as defined in (5.5) and the expression for μ as given in (5.1), we get

$$\mu(t,x) = \Phi(T,t)\left(x+\Phi(t,T)v - v(t)\right) = \Phi(T,t)(x-v(t)) + v.$$

This gives

$$A(t,x) := (v-\mu(t,x))' K(t)^{-1}(v-\mu(t,x)) = (\Phi(T,t)(x-v(t))'K(t)^{-1}\Phi(T,t)(x-v(t))$$

It follows that we can write

$$\int f(t,x)\widetilde{p}(t,x;T,v)\,\mathrm{d}z = \int \frac{f(t,x)}{\sqrt{|K(t)|}}(2\pi)^{-d/2}\exp\left(-\frac{1}{2}A(t,x)\right)\,\mathrm{d}x.$$

Upon substituting $z=\Phi(T,t)(x-v(t))$ this equals

$$\int f(t,\Phi(t,T)z+v(t))(2\pi)^{-d/2}\frac{1}{\sqrt{|K(t)|}}\exp\left(-\frac{1}{2}z'K(t)^{-1}z\right)|\Phi(t,T)|\,\mathrm{d}z.$$

We can rewrite this expression as $\mathbb{E}\left[W_t\right]$ where

$$W_t = |\Phi(t,T)| f(t,\Phi(t,T)Z_t + v(t)).$$

and Z_t denotes a random vector with $N(0,K(t))$-distribution. As $t\uparrow T$, Z_t converges weakly to a Dirac mass at zero. As $\Phi(t,T)$ converges to the identity matrix and $v(t)\to v$, we get that $\Phi(t,T)Z_t + v(t)$ converges weakly to v. By the continuous mapping theorem and continuity of f, W_t converges weakly to $f(T,v)$. Since the limit is degenerate, this statement holds for convergence in probability as well. By boundedness of f, we get $\mathbb{E}\left[W_t\right]\to f(T,v)$. □

5.2. Assumption 1(ii)

Lemma 8. *There exists a positive constant C such that for all $s\in[0,T)$ and $x,y\in\mathbb{R}^d$*

$$(T-s)\|\widetilde{H}(s)\| \le C, \tag{5.7}$$

$$\|\widetilde{r}(s,x)\| \le C\left(1+\frac{\|v-x\|}{T-s}\right), \tag{5.8}$$

$$\|\widetilde{r}(s,y)-\widetilde{r}(s,x)\| \le C\frac{\|y-x\|}{T-s} \tag{5.9}$$

$$\frac{\|v-x\|}{T-s} \le C\left(1+\|\widetilde{r}(s,x)\|\right). \tag{5.10}$$

Proof. In the proof, we use the relations proved in Lemma 6. From this lemma it follows that

$$\widetilde{H}(s)^{-1} = \int_s^T \Phi(s,\tau)a(\tau)\Phi(s,\tau)^T \,\mathrm{d}\tau.$$

Since $\Phi(s,\tau)$ is uniformly bounded and $\tau \mapsto \widetilde{a}(\tau)$ is continuous, it easily follows that $y'\widetilde{H}(s)^{-1}y \leq \widetilde{c}(T-s)\|y\|^2$ for all $y \in \mathbb{R}^d$. By uniform ellipticity of $\widetilde{a}$, there exists a constant $c_1 > 0$ such that for all $y \in \mathbb{R}^d$

$$y'\Phi(s,\tau)\widetilde{a}(\tau)\Phi(s,\tau)'y \geq c_1 y'\Phi(s,\tau)\Phi(s,\tau)'y.$$

Secondly, there exists a constant $c_2 > 0$ such that $y'\Phi(s,\tau)\Phi(s,\tau)'y \geq c_2\|y\|^2$ uniformly in $s, \tau \in [0,T]$. To see this, suppose this second claim is false. Then for each $n \in \mathbb{N}$ there are $s_n, \tau_n \in [0,T]$, $y_n \in \mathbb{R}^d \setminus \{0\}$ such that $\|\Phi(s_n,\tau_n)'y_n\|^2 \leq \frac{1}{n}\|y_n\|^2$, or letting $z_n = y_n/\|y_n\|$,

$$\|\Phi(s_n,\tau_n)'z_n\|^2 \leq \frac{1}{n}.$$

By compactness of the set $[0,T]^2 \times \{z \in \mathbb{R}^d, \|z\| = 1\}$ and by continuity of Φ, there exists a convergent subsequence $s_{n_i}, \tau_{n_i}, z_{n_i} \to s^*, \tau^*, z^*$, such that, $\|\Phi(s^*,\tau^*)'z^*\|^2 = 0$ with $z^* \neq 0$. This contradicts Liouville's formula.

Integrating over $\tau \in [s,T]$ gives

$$y'\widetilde{H}(s)^{-1}y \geq c(T-s)\|y\|^2, \tag{5.11}$$

where $c = c_1c_2$. Hence, we have proved that

$$c\|y\|^2 \leq y'((T-s)\widetilde{H}(s))^{-1}y \leq \widetilde{c}\|y\|^2.$$

Since $\widetilde{H}$ is symmetric, this says that the eigenvalues of the matrix $((T-s)\widetilde{H}(s))^{-1}$ are contained in the interval $[c, \widetilde{c}]$. This implies that the eigenvalues of $(T-s)\widetilde{H}(s)$ are in $[1/\widetilde{c}, 1/c]$. Since the operator norm of a positive definite matrix is bounded by its largest eigenvalue, it follows that $(T-s)\|\widetilde{H}(s)\| \leq 1/c$.

To prove the second inequality, note that

$$\begin{aligned}
\widetilde{r}(s,x) &= \widetilde{H}(s)(v(s)-x) = \widetilde{H}(s)\left[v(s)-v(T)+v-x\right] \\
&= (T-s)\widetilde{H}(s)\left[-\frac{v(T)-v(s)}{T-s} + \frac{v-x}{T-s}.\right]
\end{aligned}$$

Now

$$v(T)-v(s) = \left(\Phi(T,T)-\Phi(s,T)\right)v + \int_s^T \Phi(s,\tau)\widetilde{\beta}(\tau)\,\mathrm{d}\tau.$$

As $s \mapsto \Phi(s,T)$ is continuously differentiable, we have

$$\|v(T)-v(s)\| \leq C_1(T-s) + \int_s^T \|\Phi(s,\tau)\|\|\widetilde{\beta}(\tau)\|\,\mathrm{d}\tau \leq C_2(T-s).$$

Hence,

$$\|\widetilde{r}(s,x)\| \leq (T-s)\|\widetilde{H}(s)\| \left(C_2 + \frac{\|v-x\|}{T-s}\right)$$

which yields (5.8). Also,

$$\|\widetilde{r}(s,x) - \widetilde{r}(s,y)\| = \|\widetilde{H}(s)(y-x)\| \lesssim \frac{\|y-x\|}{T-s}.$$

For obtaining the fourth inequality of the lemma,

$$\widetilde{H}(s)(v-x) = \widetilde{r}(s,x) + \widetilde{H}(s)(v(T)-v(s)).$$

Upon multiplying both sides by $((T-s)\widetilde{H}(s))^{-1}$ this gives

$$\frac{\|v-x\|}{T-s} \leq \|((T-s)\widetilde{H}(s))^{-1}\|\|r(s,x)\| + \frac{\|v(T)-v(s)\|}{T-s}.$$

Substitution of the derived bounds on $\widetilde{H}(s)^{-1}$ and $v(T)-v(s)$ completes the proof. □

5.3. Assumption 1(iii)

Lemma 9. *There exist positive constants C and Λ such that for all $s \in [0,T)$*

$$\widetilde{p}(s,x;T,v) \leq C(T-s)^{-d/2} \exp\left(-\Lambda \frac{\|v-x\|^2}{T-s}\right). \tag{5.12}$$

Proof. Using the relations from Lemma 6 together with equation (5.3), some straightforward calculations yield

$$\widetilde{R}(s,x) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|K(s)| - \frac{1}{2}\widetilde{r}(s,x)'\widetilde{H}(s)^{-1}\widetilde{r}(s,x).$$

By (5.11), there exists a positive constant $c_1 > 0$ such that

$$\widetilde{r}(s,x)'\widetilde{H}(s)^{-1}\widetilde{r}(s,x) \geq c_1(T-s)\|\widetilde{r}(s,x)\|^2.$$

By equation (5.10) the right-hand-side is lower bounded by

$$c_1 \left\{ \max\left(\frac{\|x-v\|}{\sqrt{T-s}} - c_2\sqrt{T-s}, 0\right)\right\}^2$$

for some positive constant c_2. Now if $a \geq 0$ and $b \in [0,c_2]$, then there exist $c_3, c_4 > 0$ such that $(\max(a-b,0))^2 \geq c_3 a^2 - c_4$ (this is best seen by drawing a picture). Applying this with $a = \|v-x\|/\sqrt{T-s}$ and $b = c_2\sqrt{T-s}$ gives

$$\widetilde{r}(s,x)'\widetilde{H}(s)^{-1}\widetilde{r}(s,x) \geq c_1\left(c_3 \frac{\|v-x\|^2}{T-s} - c_4\right).$$

This yields the exponential bound in (5.12).

Since $\widetilde{H}(s)^{-1} = \Phi(s,T)K(s)\Phi(s,T)^T$ we have $|K(s)| = \frac{|\Phi(T,s)|^2}{|\widetilde{H}(s)|}$. Multiplying both sides by $(T-s)^{-d}$ gives

$$(T-s)^{-d}|K(s)| = \frac{|\Phi(T,s)|^2}{|(T-s)\widetilde{H}(s)|}.$$

Since the eigenvalues of $(T-s)\widetilde{H}(s)$ are bounded by $1/c$ uniformly over $s \in [0,T]$ (see Lemma 8) and the determinant of a symmetric matrix equals the product of its eigenvalues, we get

$$(T-s)^{-d}|K(s)| \geq |\Phi(T,s)|^2 c^d = c^d \exp\left(2\int_s^T \mathrm{tr}(\widetilde{B}(u))\,\mathrm{d}u\right).$$

by Liouville's formula. Now it follows that the right-hand-side of the preceding display is bounded away from zero uniformly over $s \in [0,T]$. $\square$

6. Proof of Theorem 2(ii)

Auxiliary results used in the proof are gathered in Subsection 6.1 ahead.

By (5.10) in Lemma 8 we have $\|x-v\| \lesssim (T-t)(1+\|\widetilde{r}(t,x)\|)$. Therefore we focus on bounding $\|\widetilde{r}(t,x)\|$. Define w to be the positive definite square root of $a(T,v)$. Then it follows from our assumptions that $\|w\| < \infty$ and $\|w^{-1}\| < \infty$, hence we can equivalently derive a bound for $\widetilde{Z}(s,x) = w\,\widetilde{r}(s,x)$. We do this in two steps. First we obtain a preliminary bound by writing an SDE for $\widetilde{Z}$ and bounding the terms in the equation. Next we strengthen the bound using a Gronwall-type inequality.

By Lemma 11, $\widetilde{Z}$ satisfies the stochastic differential equation

$$\mathrm{d}\widetilde{Z}(s,X_s^\circ) = -w\widetilde{H}(s)\sigma(s,X_s^\circ)\,\mathrm{d}W_s + \Upsilon(s,X_s^\circ)\,\mathrm{d}s + \Delta(s,X_s^\circ)\widetilde{Z}(s,X_s^\circ)\,\mathrm{d}s, \tag{6.1}$$

where

$$\Delta(s,X_s^\circ) = w\left(\widetilde{H}(s)\left(\widetilde{a}(s) - a(s,X_s^\circ)\right) - \widetilde{B}(s)\right)w^{-1} \tag{6.2}$$

$$\Upsilon(s,X_s^\circ) = w\widetilde{H}(s)\left(\widetilde{b}(s,X_s^\circ) - b(s,X_s^\circ)\right). \tag{6.3}$$

Define $\widetilde{J}(s) = w\widetilde{H}(s)w$. For Δ we have the decomposition $\Delta = \Delta_1 + \Delta_2 + \Delta_3$, with

$$\begin{aligned}\Delta_1(s,X_s^\circ) &= \frac{1}{T-s}\left(\mathrm{Id} - w^{-1}a(s,X_s^\circ)w^{-1}\right) \\ \Delta_2(s,X_s^\circ) &= \left(\widetilde{J}(s) - \frac{1}{T-s}\right)\left(\mathrm{Id} - w^{-1}a(s,X_s^\circ)w^{-1}\right) \\ \Delta_3(s) &= w\left[\widetilde{H}(s)\left(\widetilde{a}(s) - \widetilde{a}(T)\right) - \widetilde{B}(s)\right]w^{-1}.\end{aligned} \tag{6.4}$$

To see this, we calculate $\Delta_1(s, X_s^\circ) + \Delta_2(s, X_s^\circ) = \widetilde{J}(s)\left(\mathrm{Id} - w^{-1}a(s, X_s^\circ)w^{-1}\right)$ and

$$\Delta(s, X_s^\circ) - \Delta_1(s, X_s^\circ) - \Delta_2(s, X_s^\circ) = w\left[\widetilde{H}(s)\widetilde{a}(s) - \widetilde{B}(s)\right]w^{-1} - \widetilde{J}(s).$$

Upon substituting $\widetilde{J}(s) = w\widetilde{H}(s)a(T, v)w^{-1} = w\widetilde{H}(s)\widetilde{a}(T)w^{-1}$ into this display we end up with exactly $\Delta_3(s)$.

For Υ we have a decomposition $\Upsilon = \Upsilon_1\widetilde{Z} + \Upsilon_2$ with

$$\begin{aligned}\Upsilon_1(s, X_s^\circ) &= w\widetilde{H}(s)(B(s, X_s^\circ) - \widetilde{B}(s))\widetilde{H}^{-1}(s)w^{-1}\\ \Upsilon_2(s, X_s^\circ) &= w\widetilde{H}(s)[\widetilde{\beta}(s) - \beta(s) - (B(s, X_s^\circ) - \widetilde{B}(s))v(s)].\end{aligned}$$

Here, $v(s)$ is as defined in (5.5). To prove the decomposition, first note that Υ, Υ_1 and Υ_2 share the factor $w\widetilde{H}(s)$. Therefore, it suffices to prove that

$$\begin{aligned}&\widetilde{b}(s, x) - b(s, x) - (B(s, x) - \widetilde{B}(s))\widetilde{H}^{-1}(s)w^{-1}\widetilde{Z}(s, x)\\ &\qquad\qquad = \widetilde{\beta}(s) - \beta(s, x) - (B(s, x) - \widetilde{B}(s))v(s).\end{aligned}\tag{6.5}$$

By Lemma 6, $\widetilde{Z}(s, x) = w\widetilde{r}(s, x) = w\widetilde{H}(s)\left(\widetilde{v}(s) - x\right)$. Upon substituting this into the left-hand-side of the preceding display we obtain

$$\left(\widetilde{B}(s) - B(s, x)\right)x + \widetilde{\beta}(s) - \beta(s, x) - \left(B(s, x) - \widetilde{B}(s)\right)\left(\widetilde{v}(s) - x\right),$$

which is easily seen to be equal to the right-hand-side of (6.5). Thus, (6.1) can be written as

$$\begin{aligned}&\mathrm{d}\widetilde{Z}(s, X_s^\circ) = -w\widetilde{H}(s)\sigma(s, X_s^\circ)\,\mathrm{d}W_s\\ &\quad + \left[\Delta_1(s, X_s^\circ) + \Delta_2(s, X_s^\circ) + \Delta_3(s) + \Upsilon_1(s, X_s^\circ)\right]\widetilde{Z}(s, X_s^\circ)\,\mathrm{d}s + \Upsilon_2(s, X_s^\circ)\,\mathrm{d}s.\end{aligned}\tag{6.6}$$

Next, we derive bounds on Δ_1, Δ_2, Δ_3, Υ_1 and Υ_2.

- By Lemma 12 it follows that there is a $\varepsilon_0 \in (0, 1/2)$ such that

$$y'\Delta_1(s, X_s^\circ)y \le \frac{1 - \varepsilon_0}{T - s}\|y\|^2 \quad \text{for all} \quad s \in [0, T) \quad \text{and} \quad y \in \mathbb{R}^d.$$

- By Lemma 13, $\|\widetilde{J}(s) - \mathrm{Id}/(T - s)\|$ is bounded for $s \in [0, T]$. As σ is bounded, this implies Δ_2 can be bounded by deterministic constant $C_1 > 0$.
- For Δ_3, we employ the Lipschitz property of $\widetilde{a}$ to deduce that there is a deterministic constant $C_2 > 0$ such that

$$\|\Delta_3(s)\| \le (T - s)\|\widetilde{H}(s)\|\left\|\frac{\widetilde{a}(s) - \widetilde{a}(T)}{T - s}\right\| + \|\widetilde{B}(s)\| \le C_2.$$

- Since $(s, x) \mapsto B(s, x)$ is assumed to be bounded, there exists a deterministic constant $C_3 > 0$ such that

$$\|\Upsilon_1(s, X_s^\circ)\| \le \|B(s, X_s^\circ) - \widetilde{B}(s)\| \le C_3.$$

- Similarly, using that $s \mapsto \widetilde{v}(s)$ is bounded on $[0, T]$, we have the existence of a deterministic constant C_4 such that

$$(T-s)\|\Upsilon_2(s)\| =$$
$$\|w\|(T-s)\left\|\widetilde{H}(s)\right\|\left[\|\widetilde{\beta}(s)\| + \|\beta(s, X_s^\circ)\| + \|B(s, X_s^\circ) - \widetilde{B}(s)\|\|v(s)\|\right] \leq C_4.$$

Now we set $A(s, x) = \Delta_1(s, x) + \Delta_2(s, x) + \Delta_3(s) + \Upsilon_1(s, x)$ and let $\Psi(s)$ be the principal fundamental matrix at 0 for the corresponding random homogeneous linear system

$$\mathrm{d}\Psi(s) = A(s, X_s^\circ)\Psi(s)\,\mathrm{d}s, \qquad \Psi(0) = \mathrm{Id}. \tag{6.7}$$

Since $s \mapsto A(s, X_s^\circ)$ is continuous for each realization X°, $\Psi(s)$ exists uniquely (Chicone (1999), Theorem 2.4). Using the just derived bounds, for all $y \in \mathbb{R}^d$

$$y' A(s, X_s^\circ) y \leq \frac{1-\varepsilon_0}{T-s}\|y\|^2 + C_1 + C_2 + C_3.$$

By Lemma 14, this implies existence of a positive constant C such that

$$\|\Psi(t)\Psi(s)^{-1}\| \leq C\left(\frac{T-s}{T-t}\right)^{1-\varepsilon_0}, \qquad 0 \leq s \leq t < T.$$

By Lemma 15, for $s < T$ we can represent $\widetilde{Z}$ as

$$\widetilde{Z}(s, X_s^\circ) = \Psi(s)\widetilde{Z}(0, u) + \Psi(s)\int_0^s \Psi(h)^{-1}\Upsilon_2(h)\,\mathrm{d}h - M_s, \tag{6.8}$$

where
$$M_s = \Psi(s)\int_0^s \Psi(h)^{-1} w\widetilde{H}(h)\sigma(h, X_h^\circ)\,\mathrm{d}W_h. \tag{6.9}$$

Bounding $\|\widetilde{Z}(s, X^\circ)\|$ can be done by bounding the norm of each term on the right-hand-side of equation (6.8).

The norm of the first term can be bounded by $\|\widetilde{Z}(0, u)\|\|\Psi(s)\| \lesssim (T-s)^{\varepsilon_0 - 1}$. The norm of the second one can be bounded by

$$\int_0^s \left(\frac{T-h}{T-s}\right)^{1-\varepsilon_0} \frac{1}{T-h}\|\Upsilon_2(h)(T-h)\|\,\mathrm{d}h \lesssim (T-s)^{\varepsilon_0 - 1}.$$

For the third term, it follows from Lemma 16, applied with $U(s, h) = w\widetilde{H}(h)\sigma(h, X_h^\circ)$, that there is an a.s. finite random variable $\overline{M}$ such that for all $s < T$ $\|M_s\| \leq \overline{M}(T-s)^{\varepsilon_0 - 1}$. Therefore, there exists a random variable $\overline{M}'$ such that

$$\|\widetilde{Z}(s, X_s^\circ)\| \leq \overline{M}'(T-s)^{\varepsilon_0 - 1}. \tag{6.10}$$

We finish the proof by showing that the bound obtained can be improved upon. We go back to equation (6.1) and consider the various terms. By inequality (4.3) and the inequalities of Lemma 8 we can bound

$$\|\Upsilon(s,x)\| \lesssim \|\widetilde{H}(s)\|\,(1+\|x-v\|) \lesssim (T-s)^{-1} + \frac{\|v-x\|}{T-s} \lesssim 1 + (T-s)^{-1} + \|\widetilde{Z}(s,x)\|.$$

Similarly, using inequality (4.4)

$$\|\Delta(s,x)\| \lesssim 1 + \frac{\|v-x\|}{T-s} \lesssim 1 + \|\widetilde{Z}(s,x)\|.$$

The quadratic variation $\langle L\rangle$ of the martingale part $L_t = \int_0^t w\widetilde{H}(s)\sigma(s,X_s^\circ)\,\mathrm{d}W_s$ is given by $\langle L\rangle_t = \int_0^t w\widetilde{H}(s)a(s,X_s^\circ)\widetilde{H}(s)w\,\mathrm{d}s$. Hence, by the boundedness of $\|\widetilde{H}(s)(T-s)\|$ we have

$$\|\langle L\rangle_t\| \lesssim \int_0^t \frac{1}{(T-s)^2}\,\mathrm{d}s = \frac{1}{T-t} - \frac{1}{T} \le \frac{1}{T-t}.$$

By the Dambis-Dubins-Schwarz time-change theorem and the law of the iterated logarithm of Brownian motion, it follows that there exists an a.s. finite random variable N such that $\|L_t\| \le Nf(t)$ for all $t<T$, where

$$f(t) = \sqrt{\frac{1}{T-t}\log\log\left(\frac{1}{T-t}+e\right)}.$$

Taking the norm on the left- and right-hand-side of equation (6.1), applying the derived bounds and using that $\int_0^t (T-s)^{-1}\,\mathrm{d}s \lesssim \sqrt{1/(T-t)}$ we get with $\rho(s) = \|\widetilde{Z}(s,X_s^\circ)\|$ that $\rho(t) \le Nf(t) + C\int_0^t \left(\rho(s)+\rho^2(s)\right)\,\mathrm{d}s$, $t<T$ for some positive constant C. The bound (6.10) derived above implies that ρ is integrable on $[0,T]$. The proof of assertion (ii) of Theorem 2 is now completed by applying Lemma 17.

6.1. Auxiliary results used in the proof of Theorem 2(ii)

Lemma 10. *Define $V(s) = w^{-1}\widetilde{H}(s)^{-1}w^{-1}$ and $V'(s) = \frac{\partial}{\partial s}V(s)$. It holds that $s \mapsto V'(s)$ is Lipschitz on $[0,T]$ and $V'(s) \to -\mathrm{Id}$ as $s \uparrow T$.*

Proof. By equation (5.4)

$$\Phi(T,s)\widetilde{H}(s)^{-1}\Phi(T,s)' = \int_s^T \Phi(T,\tau)\widetilde{a}(\tau)\Phi(T,\tau)'\,\mathrm{d}\tau.$$

Taking the derivative with respect to s on both sides and reordering terms gives

$$\frac{\partial}{\partial s}\widetilde{H}(s)^{-1} = -\widetilde{a}(s) + \widetilde{B}(s)\widetilde{H}(s)^{-1} + \widetilde{H}(s)^{-1}\widetilde{B}(s)',$$

and hence $V'(s) = w^{-1}\left(-\widetilde{a}(s) + \widetilde{B}(s)\widetilde{H}(s)^{-1} + \widetilde{H}(s)^{-1}\widetilde{B}(s)'\right)w^{-1}$. Since $\|\Phi(s,\tau)\| \leq C$ for all $s, \tau \in [0,T]$, it follows that $s \mapsto V'(s)$ is Lipschitz on $[0,T]$. Furthermore, $V'(s) \to -w^{-1}a(T)w^{-1} = -\mathrm{Id}$, as $s \uparrow T$. □

Lemma 11. *We have*

$$\begin{aligned} \mathrm{d}\widetilde{r}(s, X_s^\circ) = &-\widetilde{H}(s)\sigma(s, X_s^\circ)\,\mathrm{d}W_s \\ &+ \widetilde{H}(s)\left(\widetilde{b}(s, X_s^\circ) - b(s, X_s^\circ)\right)\mathrm{d}s \\ &+ \left(\widetilde{H}(s)\left(\widetilde{a}(s) - a(s, X_s^\circ)\right) - \widetilde{B}\right)\widetilde{r}(s, X_s^\circ)\,\mathrm{d}s, \end{aligned}$$

where $\widetilde{B} = D\widetilde{b}$.

Proof. In the proof, we will omit dependence on s and X_s° in the notation. By Itō's formula

$$\mathrm{d}\widetilde{r} = \frac{\partial}{\partial s}\widetilde{r}\,\mathrm{d}s - \widetilde{H}dX^\circ. \tag{6.11}$$

For handling the second term we plug-in the expression for X° from its defining stochastic differential equation. This gives

$$\widetilde{H}\,\mathrm{d}X^\circ = \widetilde{H}b\,\mathrm{d}s + \widetilde{H}a\widetilde{r}\,\mathrm{d}s + \widetilde{H}\sigma\,\mathrm{d}W. \tag{6.12}$$

For the first term, we compute the derivative of $\widetilde{r}(s,x)$ with respect to s. For this, we note that by Lemma 1 $\frac{\partial}{\partial s}\widetilde{R} = -\widetilde{\mathcal{L}}\widetilde{R} - \frac{1}{2}\widetilde{r}'\widetilde{a}\widetilde{r}$, with $\widetilde{\mathcal{L}}\widetilde{R} = \widetilde{b}'\widetilde{r} - \frac{1}{2}\mathrm{tr}\left(\widetilde{a}\widetilde{H}\right)$. Next, we take D on both sides of this equation. Since we assume $\widetilde{R}(s,x)$ is differentiable in (s,x) we have $D\left((\partial/\partial s)\widetilde{R}\right) = (\partial/\partial s)\widetilde{r}$. Further, $D\left(\widetilde{\mathcal{L}}\widetilde{R}\right) = \widetilde{B}\widetilde{r} - \widetilde{H}\widetilde{b}$ and $D\left(\frac{1}{2}\widetilde{r}'\widetilde{a}\widetilde{r}\right) = -\widetilde{H}\widetilde{a}\widetilde{r}$. Therefore, $\frac{\partial}{\partial s}\widetilde{r} = -\widetilde{B}\widetilde{r} + \widetilde{H}\widetilde{b} + \widetilde{H}\widetilde{a}\widetilde{r}$. Plugging this expression together with (6.12) into equation (6.11) gives the result. □

Lemma 12. *There exists an* $\varepsilon_0 \in (0, 1/2)$ *such that for* $0 \leq s < T$, $x, y \in \mathbb{R}^d$

$$y'\Delta_1(s,x)y \leq \left(\frac{1-\varepsilon_0}{T-s}\right)\|y\|^2,$$

with Δ_1 *as defined in* (6.4).

Proof. Let $y \in \mathbb{R}^d$. By (2.8) there is $\varepsilon > 0$ such that

$$y'\Delta_1(s,x)y = y'\left(\frac{1}{T-s}\right)(\mathrm{Id} - w^{-1}a(s,x)w^{-1})y \leq \left(\frac{1}{T-s}\right)\left(y'y - \varepsilon y'\widetilde{a}(T)^{-1}y\right).$$

Since $\widetilde{a}(T) = a(T,v)$ is positive definite, its inverse is positive definite as well. Hence, there exists a $\varepsilon' > 0$ such that $y'\widetilde{a}(T)^{-1}y \geq \varepsilon'\|y\|^2$. This gives $y'\Delta_1(s,x)y \leq \frac{1-\varepsilon\varepsilon'}{T-s}\|y\|^2$. Let $\varepsilon_0 = \varepsilon\varepsilon'$. We can take ε sufficiently small such that $\varepsilon_0 \in (0, 1/2)$. □

Lemma 13. *Let $\widetilde{J}(s) = w\widetilde{H}(s)w$. There exists a $C > 0$ such that*

$$\left\| \widetilde{J}(s) - \frac{1}{T-s}\mathrm{Id} \right\| < C \quad \textit{for all } s < T.$$

Proof. We have

$$\left\| \widetilde{J}(s) - \frac{1}{T-s}\mathrm{Id} \right\| \leq \frac{1}{T-s} \left\| \widetilde{J}(s) \right\| \left\| (T-s)\mathrm{Id} - \widetilde{J}^{-1}(s) \right\|. \tag{6.13}$$

Let $\widetilde{V}(s) = \widetilde{J}(s)^{-1}$ and $\widetilde{V}'(s) = \frac{\partial}{\partial s}\widetilde{V}(s)$. Since $\widetilde{V}(T) = 0$ and $\widetilde{V}'(T) = -\mathrm{Id}$ (see Lemma 10) we can write

$$(T-s)\mathrm{Id} - \widetilde{V}(s) = -\int_s^T \widetilde{V}'(T) + \int_s^T \widetilde{V}'(h)\,\mathrm{d}h.$$

By Lemma 10, $s \mapsto \widetilde{V}'(s)$ is Lipschitz on $[0,T]$ and therefore

$$\left\| (T-s)\mathrm{Id} - \widetilde{V}(s) \right\| \lesssim \int_s^T (T-h)\,\mathrm{d}h = (T-s)^2/2.$$

Substituting the derived bound into (6.13) gives

$$\left\| \widetilde{J}(s) - \frac{1}{T-s}\mathrm{Id} \right\| \lesssim (T-s) \left\| \widetilde{J}(s) \right\| \lesssim (T-s) \left\| \widetilde{H}(s) \right\| \lesssim 1.$$

The last inequality follows from Lemma 8. □

Lemma 14. *Let $\Psi(t)$ be the principal fundamental matrix at 0 for the random homogeneous linear system*

$$\mathrm{d}\Psi(s) = A(s)\Psi(s)\,\mathrm{d}s, \qquad \Psi(0) = I. \tag{6.14}$$

Suppose that the matrix function $A(s)$ is of the form $A(s) = A_1(s) + A_2(s)$, where both A_1 and A_2 are continuous on $[0,T)$. Assume A_2 is bounded and A_1 is such that there are $\varepsilon_0 \in (0, 1/2)$ and $C_1 > 0$ that for all $s \in [0,T)$ and vectors y

$$y' A_1(s) y \leq \left(\frac{1-\varepsilon_0}{T-s} + C_1 \right) \|y\|^2.$$

Then there is a $C > 0$ such that for all $0 \leq s \leq t < T$

$$\|\Psi(t)\Psi(s)^{-1}\| \leq C \left(\frac{T-s}{T-t} \right)^{1-\varepsilon_0}.$$

Proof. For $z \in \mathbb{R}^d$, let $Z(t) = \Psi(t)z$, so $\mathrm{d}Z(t) = (A_1(t) + A_2(t))Z(t)\,\mathrm{d}t$. Let $\|A_2(t)\| \leq C_2$ (say). Integrating $\mathrm{d}[Z(u)'Z(u)] = \mathrm{d}[Z(u)']Z(u) + Z(u)'[\mathrm{d}Z(u)] = Z(u)'(A_1 + A_2 + A_1' +$

$A_2')Z(u)\,\mathrm{d}u$ over $[s,t]$ yields

$$\begin{aligned} Z(t)'Z(t) &= Z(s)'Z(s) + \int_s^t Z(h)'(A_1(h) + A_1(h)')Z(h)\,\mathrm{d}h \\ &\quad + \int_s^t Z(h)'(A_2(h) + A_2(h)')Z(h)\,\mathrm{d}h \\ &\leq Z(s)'Z(s) + \int_s^t 2\left(\frac{1-\varepsilon_0}{T-h} + C_1 + C_2\right) Z(h)'Z(h)\,\mathrm{d}h. \end{aligned}$$

From Gronwall's lemma,

$$\|Z_t\|^2 \leq \|Z_s\|^2 \exp\left(2\int_s^t \frac{1-\varepsilon_0}{T-u}\,\mathrm{d}u + 2(t-s)(C_1+C_2)\right).$$

Let $z = \Psi(s)^{-1}x$. For any x with $\|x\| \leq 1$ this implies

$$\|\Psi(t)\Psi(s)^{-1}x\| \leq \|\Psi(s)\Psi(s)^{-1}x\| \left(\frac{T-s}{T-t}\right)^{1-\varepsilon_0} \mathrm{e}^{(t-s)(C_1+C_2)}$$

or $\|\Psi(t)\Psi(s)^{-1}\| \leq \mathrm{e}^{T(C_1+C_2)}\left(\frac{T-s}{T-t}\right)^{1-\varepsilon_0}$. □

Lemma 15. *Suppose Y is a strong solution of the stochastic differential equation $\mathrm{d}Y_t = \alpha_t\,\mathrm{d}W_t + (\beta_t + \gamma_t Y_t)\,\mathrm{d}t$, where $\alpha_t = \alpha(t, Y_t)$, $\beta_t = \beta(t, Y_t)$ and $\gamma_t = \gamma(t, Y_t)$. Let Ψ be the matrix solution to $\mathrm{d}\Psi(t) = \gamma_t\Psi(t)\,\mathrm{d}t$, $\Psi(0) = \mathrm{Id}$ and define the process Y' by*

$$Y_t' = \Psi(t)\left[Y_0 + \int_0^t \Psi(h)^{-1}\beta_h\,\mathrm{d}h + \int_0^t \Psi^{-1}\alpha_h dW_h\right].$$

If $\sup_{s\leq\tau}\|\gamma_s\| < \infty$, then Y and Y' are indistinguishable on $[0,\tau]$.

Proof. By computing $\int_0^t \gamma_s Y'(s)\,\mathrm{d}s$ and using the (stochastic) Fubini theorem it is easy to verify that Y' satisfies the stochastic differential equation

$$\mathrm{d}Y_t' = \alpha_t\,\mathrm{d}W_t + (\beta_t + \gamma_t Y_t')\,\mathrm{d}t.$$

This implies $Y_s' - Y_s = \int_0^t \gamma_s(Y_s' - Y_s)\,\mathrm{d}s$ and thus

$$\sup_{s\leq t}\|Y_s' - Y_s\| \leq \max_{s\leq t}\|\gamma_s\| \int_0^t \sup_{h\leq s}\|Y_h' - Y_h\|\,\mathrm{d}s.$$

By Gronwall's lemma $\sup_{s\leq t}\|Y_s' - Y_s\| \leq 0$, which concludes the proof. □

Lemma 16. *Define $M_t = \Psi(t)\int_0^t \Psi(s)^{-1}U(s)\,\mathrm{d}W_s$, where Ψ satisfies $\mathrm{d}\Psi(s) = A(s)\Psi(s)\,\mathrm{d}s$ and $\Psi(0) = \mathrm{Id}$. Assume $(T-s)\|U(s)\| \lesssim 1$ for $s \in [0,T)$. Assume that the assumptions of*

Lemma 14 hold with $\varepsilon_0 \in (0, 1/2)$ and additionally that there are constants $C_1, C_2 > 0$ such that for all $0 \le s < T$

$$\|A(s)\| \le C_1 \frac{1}{T-s} + C_2. \tag{6.15}$$

Then there exists an a.s. finite random variable N such that for all $0 \le s < T$ $\|M_s\| \le (T-s)^{\varepsilon_0 - 1} N$.

Proof. Let $\gamma \in (\varepsilon_0, 1/2)$ and define

$$M_t^{(\gamma)} = \int_0^t (T-s)^{1-\gamma} U(s)\, \mathrm{d}W_s, \tag{6.16}$$

so that $M_t = \int_0^t (T-s)^{\gamma-1} \Psi(t)\Psi(s)^{-1}\, \mathrm{d}M_s^{(\gamma)}$.

By partial integration,

$$M_t = (T-t)^{\gamma-1} M_t^{(\gamma)} - \Psi(t) \int_0^t M_s^{(\gamma)}\, \mathrm{d}\left((T-s)^{\gamma-1}\Psi(s)^{-1}\right).$$

By straightforward algebra the integral appearing on the right-hand-side can be simplified and we get

$$M_t = (T-t)^{\gamma-1} M_t^{(\gamma)} - \Psi(t) \int_0^t M_s^{(\gamma)} (T-s)^{\gamma-2} \Psi(s)^{-1} \left[(1-\gamma)\mathrm{Id} - (T-s)A(s)\right] \mathrm{d}s.$$

By equation (6.15), $\|(1-\gamma)\mathrm{Id} - (T-s)A(s)\| \le 1 + C_1 + C_2(T-s)$. Therefore,

$$\|M_t\| \le (T-t)^{\gamma-1} \|M_t^{(\gamma)}\| + \sup_{0 \le s \le t} \|M_s^{(\gamma)}\| \int_0^t (T-s)^{\gamma-2} \|\Psi(t)\Psi(s)^{-1}\| \left(1 + C_1 + C_2(T-s)\right) \mathrm{d}s.$$

Using Lemma 14, the integral on the right-hand-side of the preceding display can be bounded by a positive constant times

$$\int_0^t (T-s)^{\gamma-2} \left(\frac{T-s}{T-t}\right)^{1-\varepsilon_0} \mathrm{d}s = (T-t)^{-1+\varepsilon_0} \int_0^t (T-s)^{-1+\gamma-\varepsilon_0}\, \mathrm{d}s.$$

From the choice $\gamma > \varepsilon_0$, this last integral is bounded. So we obtain $\|M_t\| \le (T-t)^{\varepsilon_0 - 1} N$, with $N = C \sup_{0 \le t \le T} \|M_t^{(\gamma)}\|$ for some $C > 0$. It remains to show that N is a.s. finite. By the assumption on U, the quadratic variation of $M^{(\gamma)}$ satisfies, since $\gamma < 1/2$,

$$\left\| \left\langle M^{(\gamma)} \right\rangle_T \right\| \le \int_0^T \frac{1}{(T-s)^{2\gamma}}\, \mathrm{d}s < \infty.$$

Hence, the result follows from the Dambis-Dubins-Schwarz theorem. □

Lemma 17. *Let $f : [0,T) \to [0,\infty)$ be nondecreasing and bounded on any subinterval $[0,\tau]$, $\tau < T$. Suppose ρ is integrable, continuous and nonnegative on $[0,T)$. If*

$$\rho(t) \le f(t) + C \int_0^t \left(\rho(s) + \rho^2(s)\right) \,\mathrm{d}s, \quad t \in [0,T)$$

for some positive constant C, then $\rho \lesssim f$ on $[0,T)$.

For the proof we need the following Gronwall–Bellman type lemma. A proof can be found in Mitrinović et al. (1991) (Chapter XII.3, Theorem 4).

Lemma 18. *Let $\rho(t)$ be continuous and nonnegative on $[0,\tau]$ and satisfy*

$$\rho(t) \le f(t) + \int_0^t h(s)\rho(s)\,\mathrm{d}s, \quad t \in [0,\tau],$$

where h is a nonnegative integrable function on $[0,T)$ and with f nonnegative, nondecreasing and bounded on $[0,\tau]$. Then

$$\rho(\tau) \le f(\tau) \exp\left(\int_0^T h(s)\,\mathrm{d}s\right).$$

Proof of lemma 17. Applying the Gronwall–Bellman lemma with $h(s) = C(1+\rho(s))$ gives that for any $\tau \in [0,T)$,

$$\rho(\tau) \le f(\tau) \exp\left(\int_0^\tau h(s)\,\mathrm{d}s\right) \le f(\tau) \exp\left(\int_0^T C(1+\rho(s))\,\mathrm{d}s\right).$$

The integral on the right-hand-side is finite. □

Appendix A: Information projection and entropy method

The following procedure to find the information projection is similar to the cross entropy method in rare event simulation. The algorithm proceeds by stochastic gradient descent to improve ϑ using samples from proposals with a varying reference value for ϑ (named ϑ_n below), which is updated every K steps.

***Algorithm* 1.**
Initialisation: *Choose a starting value for ϑ, let $n = 1$ and choose decay weights $\alpha(n,k)$.*
Repeat *for $n = 1, 2, \ldots$*

1. **Update** ϑ_n. *Let $\vartheta_n = \vartheta$.*
2. **Sample proposals.** *Sample $m = 1, \ldots, M$ bridge proposals $X^{\circ(m)}$ with parameter ϑ_n.*

3. **Stochastic gradient descent.** *For* $k = 1, \ldots, K$

$$\vartheta \leftarrow \vartheta - \alpha(n,k)\frac{1}{M}\sum_{m=1}^{M}\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ_{\vartheta_n}}(X^{\circ(m)})\nabla_\vartheta \log\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ_{\vartheta}}(X^{\circ(m)}).$$

If $M = 1$ and $K = 1$ this an algorithm of stochastic gradient descent type and $\alpha_n = \alpha_0\frac{\gamma}{\gamma+n}$ would be a standard choice. But depending on the form of $\widetilde{b}_\vartheta$, the update in step 3 might be computationally cheap in comparison with step 2 and one would prefer to sample $M > 1$ bridges in batches and do step 3 for $K > 1$.

In figure 2 we took starting the values $\vartheta = 0$, $\alpha_n = (10 + 2n)^{-1}$ and $M = K = 1$.

References

Aronson, D. G. (1967). Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.* **73**, 890–896.

Bayer, C. and Schoenmakers, J. (2013). Simulation of forward-reverse stochastic representations for conditional diffusions. *Ann. Appl. Probab.* To appear.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12**(6), 1077–1098.

Beskos, A., Roberts, G., Stuart, A. and Voss, J. (2008). MCMC methods for diffusion bridges. *Stoch. Dyn.* **8**(3), 319–350.

Beskos, A. and Roberts, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.* **15**(4), 2422–2444.

Bladt, M. and Sørensen, M. (2012). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Preprint.

Chicone, C. (1999). *Ordinary differential equations with applications*, volume 34 of *Texts in Applied Mathematics*. Springer-Verlag, New York.

Clark, J. (1990). The simulation of pinned diffusions. In *Decision and Control, 1990., Proceedings of the 29th IEEE Conference on*, pp. 1418–1420. IEEE.

Delyon, B. and Hu, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications* **116**(11), 1660 – 1675.

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20**(3), 297–338. With comments and a reply by the authors.

Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69**(4), 959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.* **19**(2), 177–191.

Fearnhead, P. (2008). Computational methods for complex stochastic systems: a review of some alternatives to MCMC. *Stat. Comput.* **18**(2), 151–171.

Gasbarra, D., Sottinen, T. and Valkeila, E. (2007). Gaussian bridges. In *Stochastic analysis and applications*, volume 2 of *Abel Symp.*, pp. 361–382. Springer, Berlin.

Hida, T. and Hitsuda, M. (1993). *Gaussian processes*, volume 120 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the 1976 Japanese original by the authors.

Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.

Lin, M., Chen, R. and Mykland, P. (2010). On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.* **105**(490), 820–838.

Liptser, R. S. and Shiryaev, A. N. (2001). *Statistics of random processes. I*, volume 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, expanded edition. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.

Mitrinović, D. S., Pečarić, J. E. and Fink, A. M. (1991). *Inequalities involving functions and their integrals and derivatives*, volume 53 of *Mathematics and its Applications (East European Series)*. Kluwer Academic Publishers Group, Dordrecht.

Papaspiliopoulos, O. and Roberts, G. (2012). Importance sampling techniques for estimation of diffusion models. In *Statistical Methods for Stochastic Differential Equations*, Monographs on Statistics and Applied Probability, p. 311–337. Chapman and Hall.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88**(3), 603–621.

Stuart, A. M., Voss, J. and Wiberg, P. (2004). Fast communication conditional path sampling of SDEs and the Langevin MCMC method. *Commun. Math. Sci.* **2**(4), 685–697.

Van der Meulen, F. H. and Schauer, M. (2015). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *ArXiv e-prints* .