

## MCMC METHODS FOR DIFFUSION BRIDGES

ALEXANDROS BESKOS

*Department of Statistics, University of Warwick  
Coventry, CV4 7AL, UK  
a.beskos@warwick.ac.uk*

GARETH ROBERTS

*Department of Statics, University of Warwick  
Coventry, CV4 7AL, UK  
gareth.o.roberts@warwick.ac.uk*

ANDREW STUART

*Mathematics Institute, University of Warwick  
Coventry, CV4 7AL, UK  
a.m.stuart@warwick.ac.uk*

JOCHEN VOSS

*Mathematics Institute, University of Warwick  
Coventry, CV4 7AL, UK  
jochen.voss@warwick.ac.uk*

We present and study a Langevin MCMC approach for sampling nonlinear diffusion bridges. The method is based on recent theory concerning stochastic partial differential equations (SPDEs) reversible with respect to the target bridge, derived by applying the Langevin idea on the bridge pathspace. In the process, a Random-Walk Metropolis algorithm and an Independence Sampler are also obtained. The novel algorithmic idea of the paper is that proposed moves for the MCMC algorithm are determined by discretising the SPDEs in the time direction using an *implicit* scheme, parameterised by  $\theta \in [0, 1]$ . We show that the resulting infinite-dimensional MCMC sampler is well defined only if  $\theta = 1/2$ , when the MCMC proposals have the correct quadratic variation. Previous Langevin-based MCMC methods used explicit schemes, corresponding to  $\theta = 0$ . The significance of the choice  $\theta = 1/2$  is inherited by the finite-dimensional approximation of the algorithm used in practice. We present numerical results illustrating the phenomenon and the theory that explains it. Diffusion bridges (with additive noise) are representative of the family of laws defined as a change of measure from Gaussian distributions on arbitrary separable Hilbert spaces; the analysis in this paper can be readily extended to target laws from this family and an example from signal processing illustrates this fact.

*Keywords:* Diffusion Bridge; MCMC; Langevin Sampling; Gaussian Measure; SDE on Hilbert Space; Implicit Euler Scheme; Quadratic Variation

AMS Subject Classification: 65C05, 60H35

## 1. Introduction

We present an MCMC approach for generating paths of nonlinear diffusions subject to constraints on their initial and ending points, i.e. diffusion bridges; we consider the case of additive non-degenerate noise only. Diffusion bridges are prototypical of a variety of conditioned diffusions arising in applications and it is hence of some interest to understand the behaviour of MCMC methods in this context. Furthermore, the corresponding target bridges are representative of the family of distributions defined as a change of measure from Gaussian laws on arbitrary separable Hilbert spaces. The algorithms and the analysis in the paper, though demonstrated in the diffusion context, can be readily extended to target laws in this family and thus encompass a vast range of possible applications. The key transferable idea in the paper is the use of well-chosen implicit proposals which ensure that the resulting MCMC method is well defined on the Hilbert space, and hence has convergence and mixing properties that do not degenerate under the finite-dimensional approximation of the Hilbert space required in practice.

The MCMC sampler is defined on the space of paths satisfying the desired end-point constraints and, in stationarity, delivers correlated samples from the target law. Understanding the MCMC method on pathspace yields insight into the behaviour of practical algorithms which necessarily discretise the paths and approach the infinite-dimensional situation via vanishing discretisation increment. In this sense our work complements existing work on the behaviour of MCMC methods in high dimensions (see [21] for a review). The methodology builds on recent results on the derivation of infinite-dimensional Langevin SDEs with a pre-specified equilibrium law corresponding to a conditioned diffusion process [11, 12, 13, 18, 24]. The Langevin equation is discretised to provide a proposal for a Metropolis-Hastings algorithm on the pathspace.

The dynamics of the diffusion are determined by the SDE:

$$\frac{dX}{du} = AX + f(X) + B \frac{dW}{du}, \quad u \in [0, 1], \quad (1.1)$$

with:

$$f = -BB^\top \nabla V, \quad (1.2)$$

for some  $V: \mathbb{R}^d \mapsto \mathbb{R}$ . Also,  $A \in \mathbb{R}^{d \times d}$ ,  $B \in \mathbb{R}^{d \times d}$  is invertible, and  $W$  is the standard  $d$ -dimensional Wiener process. Generating uncorrelated numerical approximations of (1.1) for a given starting point is relatively straightforward, see e.g. [15]. It is, however, considerably harder to sample from conditional laws of (1.1), since their dynamics are typically intractable. In this paper we consider the diffusion-bridge sampling problem and develop an MCMC method for the numerical solution of (1.1) under the conditions:

$$X(0) = x^-, \quad X(1) = x^+, \quad (1.3)$$

for arbitrary  $x^-, x^+ \in \mathbb{R}^d$ . We denote by  $\Pi$  the distribution of the target bridge (1.1)-(1.3) and by  $\tilde{\Pi}$  the (Gaussian) distribution of the corresponding bridge for

vanishing nonlinear term  $f \equiv 0$ . Under standard regularity conditions,  $\Pi$  and  $\tilde{\Pi}$  are equivalent (we denote this by  $\Pi \approx \tilde{\Pi}$ ) with density provided by Girsanov's theorem ([17]). The gradient form of the drift in (1.2) and Itô's Lemma provide the following expression for  $\Pi$  used in the sequel:

$$\frac{d\Pi}{d\tilde{\Pi}}(x) \propto \exp\{-\langle 1, \Psi(x) \rangle\}, \quad x \in \mathcal{H}, \quad (1.4)$$

where  $\langle \cdot, \cdot \rangle$  is the inner product on the Hilbert space  $\mathcal{H} = L^2([0, 1], \mathbb{R}^d)$  and  $\Psi : \mathbb{R}^d \mapsto \mathbb{R}$  is given by:

$$\Psi(z) = |B^{-1}f(z)|^2/2 + \operatorname{div}f(z)/2 + f(z)^\top (BB^\top)^{-1}Az, \quad z \in \mathbb{R}^d. \quad (1.5)$$

See [12, 18] for analytical calculations in the case of bridge diffusions; see [8, 12, 27] for other connections between SPDEs and different forms of conditioned diffusions. As stated above, the target  $\Pi$  defined via the Girsanov transformation (1.4) is representative of the generalised family of target laws  $\mathbf{\Pi}$  defined via a change of measure from a Gaussian law as follows:

$$\frac{d\mathbf{\Pi}}{d\text{Gaussian}}(x) \propto \exp\{-\Phi(x)\}, \quad x \in \mathbf{H}, \quad (1.6)$$

where  $\mathbf{H}$  is an arbitrary Hilbert space and  $\Phi : \mathbf{H} \mapsto \mathbb{R}$ . The theory that follows can also be applied to target laws from this family. We revisit this important point in Section 5 and in the conclusions; furthermore, in Section 6 we include a numerical example from a signal processing application.

Diffusion-bridge sampling is a major ingredient of parametric inference methods for diffusion processes, see e.g. [2, 3, 20] for applications in finance and econometrics. Numerical methods based on the  $h$ -transform [23], yielding an initial-value problem with the desired bridging property, are inefficient since they require the identification of the transition density of the unconditional process. The approach followed in [3, 20] is to build MCMC samplers (namely, Independence Samplers) by proposing independent samples from a simpler bridge process (such as Brownian bridge). The same idea is used in a discretised context in [5, 6]. The method of this paper is fundamentally different and based on extending the Langevin sampling idea to an infinite-dimensional setting, although one family of MCMC methods contain the Independence Sampler as a special case. We expect that our method will be competitive in situations where the target paths are far from Gaussian, when Independence Samplers are inefficient.

On Euclidean spaces and given a target density  $\pi : \mathbb{R}^n \mapsto \mathbb{R}$  w.r.t. the  $n$ -dimensional Lebesgue measure, the Langevin method is based on the SDE

$$\frac{dX}{dt} = \nabla \log \pi(X) + \sqrt{2} \frac{dW}{dt}. \quad (1.7)$$

This is known to be reversible w.r.t.  $\pi$ .<sup>a</sup> Assuming that (1.7) is ergodic, realisations

<sup>a</sup>Throughout, we use  $t$  to denote algorithmic time and  $u$  to denote the real time that parameterises the bridge path. Real time  $u$  will be the spatial variable in the SPDEs which generalize (1.7) for targets  $\pi$  in infinite dimensions.

4 Beskos et al.

of its solution over long time intervals will provide (in a limiting sense) samples from  $\pi$ . In practice, (1.7) must be discretised; the induced bias can then be removed using a Metropolis-Hastings correction specifying an accept/reject rule for the discretised transitions. The result is a discrete-time Markov chain with invariant density  $\pi$ . The complete MCMC method is referred to as the Metropolis-adjusted Langevin algorithm (MALA) in the literature ([19]).

The methodology of this paper extends the Langevin MCMC sampler to the infinite-dimensional problem of sampling from the bridge law  $\Pi$  given by (1.4). The target process is now treated as a probability density on the pathspace  $L^2([0, 1], \mathbb{R}^d)$  invariant under a path-valued Langevin SDE, namely a stochastic partial differential equation (SPDE), developed in [12, 18]. In fact, we will use two SPDEs, with distinct dynamics, the second one derived from the initial Langevin SPDE after appropriate application of a Green's operator. We discretise the Langevin SPDEs in the algorithmic time direction to obtain path proposals for the MCMC method. The key new algorithmic idea of this paper is to use a form of *implicit* discretisation scheme. The scheme will introduce an implicitness parameter  $\theta$ ;  $\theta = 0$  yields the standard Euler-Maruyama scheme. Critically, the pathspace MCMC sampler (both versions of it, corresponding to the two Langevin SPDEs) is well defined only when  $\theta = 1/2$ . Roughly speaking, the specification  $\theta = 1/2$  provides paths with the correct quadratic variation. The significance of the choice  $\theta = 1/2$  is inherited by the finite-dimensional approximation of the algorithm used in practice (derived by discretising the paths in  $L^2([0, 1], \mathbb{R}^d)$ ). We demonstrate this fact with a numerical study.

The structure of the paper is as follows. In Section 2 we review the Langevin approach as applied to finite-dimensional Euclidean spaces. In Section 3 we define the two Langevin SPDEs reversible with respect to the target diffusion bridge. In Section 4 we construct the two corresponding versions of MALA in the context of the infinite-dimensional diffusion-bridge targets and prove that the algorithm is well-defined if and only if  $\theta = 1/2$ . In Section 5 we show that, in the process, we have additionally defined a Random-Walk Metropolis and an Independence Sampler on the pathspace. We summarise the algorithms. In Section 6 we show numerical examples illustrating the effect of the implicitness parameter  $\theta$  in practical applications and demonstrating superiority of  $\theta = 1/2$ . In particular, we apply MALA to sample a double-well process conditioned by both a bridge condition and then by a continuous-time observation of a functional of the path. We finish with some conclusions in Section 7. For completeness we review, in Appendix A, the definition of the Gaussian distribution on an arbitrary Hilbert space and some of its properties relevant to this paper. Proofs have been collected in Appendix B.

## 2. The Finite-Dimensional Langevin Sampler

Consider the following density w.r.t. the  $n$ -dimensional Lebesgue measure:

$$\pi(x) \propto \exp\{x^\top Lx/2 - \psi(x)\}, \quad x \in \mathbb{R}^n, \quad (2.1)$$

for some function  $\psi: \mathbb{R}^n \mapsto \mathbb{R}$  and a negative, symmetric matrix  $L \in \mathbb{R}^{n \times n}$  such that the expression in (2.1) is integrable. For this target density, the Langevin SDE given in (1.7) is:

$$\frac{dX}{dt} = LX - \nabla\psi(X) + \sqrt{2} \frac{dW}{dt} . \quad (2.2)$$

The density  $\pi$  is invariant for (2.2) since it solves the relevant steady Fokker-Planck equation ([9]). Using again the Fokker-Planck equation one can show that the SDE:

$$\frac{dX}{dt} = C \{LX - \nabla\psi(X)\} + \sqrt{2C} \frac{dW}{dt} , \quad (2.3)$$

for any symmetric, positive-definite matrix  $C$ , has invariant density  $\pi$ . Assuming ergodicity, simulating the Langevin SDE over long time intervals provides a method (in the limit, as  $t \rightarrow \infty$ ) for sampling from  $\pi$ . Critically for most applications,  $\pi$  needs to be known only up to a normalising constant. For results on the convergence properties of the Langevin SDE and conditions on  $\pi$  that guarantee ergodicity see e.g. [16, 22].

In practice, one must consider a discrete-time version of the continuous-time process. The simplest such approximation is the (explicit) Euler scheme ([15]). However, to construct efficient algorithms in the bridge sampling context, we will use an *implicit* scheme. For the case of the Langevin SDE (2.2) and given a current position  $x \in \mathbb{R}^n$ , the implicit scheme obtains the location  $y$  of the path after time  $\Delta t > 0$  via the formula:

$$y - x = \{(1 - \theta)Lx + \theta Ly - \nabla\psi(x)\}\Delta t + \sqrt{2\Delta t}\xi, \quad \xi \sim \mathcal{N}(0, I) , \quad (2.4)$$

for some  $\theta \in [0, 1]$ . When  $\theta = 0$ , the update corresponds to the explicit Euler scheme. We incorporate the implicitness parameter  $\theta$  only in the linear part of the drift to allow for analytical tractability. Since  $I - \theta\Delta tL$  is symmetric positive-definite, uniqueness of the location  $y$  is guaranteed. Convergence of such implicit methods for SDEs is investigated in [15]. The dynamics of the discrete-time approximation will differ from the actual discrete-time dynamics of the SDE: decreasing  $\Delta t$  reduces the bias, but increases the number of steps to reach stationarity and the correlation among the different samples from  $\pi$  once stationarity has been reached.

A Metropolis-Hastings correction, employed in such cases, removes the bias and allows for larger jumps in the state space. According to this theory, if  $q(x, y)$  is a transition density on  $\mathbb{R}^n \times \mathbb{R}^n$  proposing a move from the current location  $x$  to  $y$ , then the corrected update:

$$x' = \begin{cases} y, & \text{with probability } \alpha(x, y) \\ x, & \text{otherwise} \end{cases} \quad (2.5)$$

where

$$\alpha(x, y) = 1 \wedge \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} , \quad (2.6)$$

defines the transitions of a discrete-time Markov chain with invariant density  $\pi$ . When  $q(x, y)$  corresponds to the Langevin transition dynamics (2.4) as determined by the explicit Euler scheme ( $\theta = 0$ ) the resulting sampling method is the Metropolis-adjusted Langevin algorithm (MALA), see e.g. [19] and the references therein. MALA will typically converge to stationarity faster than competitive MCMC methods (like for instance, the ‘vanilla’ Random-Walk Metropolis) since it is tailored to the specific target  $\pi$  and employs transitions that respect the invariance of  $\pi$ , up to the discretisation error. For  $\theta \neq 0$  the algorithm is, to the best of our knowledge, new. In fact, choosing  $\theta = 1/2$  will be the key for the construction of efficient algorithms in the infinite-dimensional context of interest in this paper.

For an overview of the properties of MALA see [21]. In that paper, efficiency is examined in the case of  $n$ -dimensional product measures, in the limit  $n \rightarrow \infty$ ; the results give valuable insight into the behaviour of MCMC methods in high dimensions. The work in [21] indicates that, as one might expect, the precise nature of the high-dimensional measure has significant effect on the properties of the MCMC sampler. In particular, in the product set-up of [21] the time step  $\Delta t$  of MALA must be scaled as  $\mathcal{O}(n^{-\frac{1}{3}})$ , hence shrinks to zero in the limit. In contrast, we show that for the bridge targets of this paper, a fixed positive  $\Delta t$  may be used in the infinite-dimensional limit, provided that  $\theta = 1/2$ . Thus, when  $\theta = 1/2$  we anticipate that the time to converge to stationarity, the mixing time in stationarity and other qualitative measures of cost, will be independent of the dimension  $n$  (which in our context refers to an  $n$ -dimensional approximation of the target bridge used in practice). In contrast, for  $\theta = 0$  these times typically grow as  $n \rightarrow \infty$ , see [21].

### 3. The Langevin SPDEs

We proceed to the infinite-dimensional diffusion bridge sampling set-up. We write down the two Langevin SPDEs, defined in [12, 18, 24] and building on ideas appearing in [7], which are invariant with respect to the target bridge law  $\Pi$ . Recall that  $\tilde{\Pi}$  denotes the corresponding Gaussian bridge. Let  $\mathcal{C}$  be the covariance operator of  $\tilde{\Pi}$  (see Appendix A for a review on Gaussian laws on Hilbert spaces). Henceforth, we set  $\mathcal{H} = L^2([0, 1], \mathbb{R}^d)$ .

The first Langevin SPDE, with solution  $x = x(t, u)$ , is as follows:

$$\begin{aligned} \partial_t x &= (\partial_u + A^\top)(BB^\top)^{-1}(\partial_u - A)x - \nabla \Psi(x) + \sqrt{2} \partial_t w, \\ x(t, 0) &= x^-, \quad x(t, 1) = x^+, \quad t \in [0, \infty), \\ x(0, u) &= x_0(u). \end{aligned} \tag{3.1}$$

The first of these equations refers to the evolution dynamics with  $\partial_t w$  denoting space-time white noise. The rest give the necessary boundary and initial value conditions. The second SPDE, termed the *preconditioned*<sup>b</sup> SPDE, is derived from (3.1) by using the covariance operator  $\mathcal{C}$  in the same way that we used the matrix  $C$  in

<sup>b</sup>In statistical terms, it can be thought of as a reparameterization of the initial setting.

the finite-dimensional case (2.3); this correspondence will be made obvious below when we write the SPDEs as  $\mathcal{H}$ -valued SDEs. The preconditioned Langevin SPDE is as follows:

$$\begin{aligned} \partial_t x &= -x + y(t, u) + \sqrt{2} \partial_t \tilde{w} , \\ (\partial_u + A^\top)(BB^\top)^{-1}(\partial_u - A)y &= \nabla \Psi(x) , \\ y(t, 0) &= x^-, \quad y(t, 1) = x^+ , \\ x(0, u) &= x_0(u) . \end{aligned} \tag{3.2}$$

Here  $\tilde{w}_t$ ,  $t \geq 0$ , is a  $\mathcal{C}$ -Wiener process, i.e. a Wiener process on  $\mathcal{H}$  with increments  $\tilde{w}_{t+h} - \tilde{w}_t$  distributed according to  $\mathcal{N}(0, h\mathcal{C})$ , for any  $t, h \geq 0$ . The linear differential operator

$$\mathcal{L} = (\partial_u + A^\top)(BB^\top)^{-1}(\partial_u - A)$$

appearing above is closely related with the covariance operator  $\mathcal{C}$  of the reference Gaussian law  $\tilde{\Pi}$ . When  $\mathcal{L}$  is restricted on the domain:

$$\mathcal{D} = \{x \in H^2([0, 1], \mathbb{R}^d) \mid x(0) = x(1) = 0\}$$

it is true that:

$$\mathcal{L} = (-\mathcal{C})^{-1} , \tag{3.5}$$

see [11]. Here  $H^2([0, 1], \mathbb{R}^d)$  denotes the Sobolev subspace, dense in  $\mathcal{H}$ , consisting of the functions whose derivatives (in the generalised sense) up to second order are in  $\mathcal{H}$ .

**Remark 3.1.** Without loss of generality we will assume that:

$$x^- = x^+ = 0 , \tag{3.6}$$

so, in fact:

$$\tilde{\Pi} \equiv \mathcal{N}(0, \mathcal{C}) .$$

The general case reduces to (3.6) after considering the process  $X - m$ , where  $X$  is the target bridge, and  $m$  the (analytically identifiable) mean of the linear bridge corresponding to  $f \equiv 0$ . If  $x$  solves (3.1) then  $x(t, \cdot) - m$  will sample in the limit from  $X - m$  and its dynamics will be given by (3.1) but now with  $x^- = x^+ = 0$  and  $\nabla \Psi(\cdot + m)$  in the place of  $\nabla \Psi(\cdot)$ . The shift by  $m$  will not affect the linear part of the drift since as shown in [11]:

$$\begin{aligned} (\partial_u + A^\top)(BB^\top)^{-1}(\partial_u - A)m &= 0 , \\ m(0) &= x^-, \quad m(1) = x^+ . \end{aligned}$$

□

The two SPDEs may be thought of as SDEs on a Hilbert space. Equations (3.1), (3.2) correspond to the  $\mathcal{H}$ -valued SDEs:

$$\frac{dx}{dt} = \mathcal{L}x - \nabla\Psi(x) + \sqrt{2} \frac{dw}{dt}, \quad (3.8)$$

$$\frac{dx}{dt} = -x - \mathcal{C}\{\nabla\Psi(x)\} + \sqrt{2\mathcal{C}} \frac{dw}{dt}, \quad (3.9)$$

respectively. The operator  $\nabla\Psi : \mathcal{H} \rightarrow \mathcal{H}$  is the Nemitski operator found from  $\nabla\Psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and extending it to an operator in  $\mathcal{H}$  via  $\nabla\Psi(x)(u) := \nabla\Psi(x(u))$ . Here  $w_t$  is a cylindrical Wiener process on  $\mathcal{H}$ ; see [4] for more technical details. In the sequel, we will use these expressions to develop our methodology. The correspondence between the two alternative expressions (3.2), (3.9) of the same preconditioned SPDE is easily understood through the relation  $\mathcal{L} = (-\mathcal{C})^{-1}$ . Existence and uniqueness of a mild solution for (3.8), a strong solution for (3.9), and ergodicity for both cases, can be deduced under the following conditions on  $\Psi$ .

**Assumption 3.1.** *The mapping  $\Psi : \mathbb{R}^d \mapsto \mathbb{R}$  in (1.5) is continuously differentiable and there exists constant  $C > 0$  such that the Nemitski operator  $\nabla\Psi : \mathcal{H} \rightarrow \mathcal{H}$  satisfies:*

- i)  $|\nabla\Psi(x) - \nabla\Psi(y)| \leq C|x - y|$ ,  $x, y \in \mathcal{H}$ ;*
- ii)  $|\nabla\Psi(x)| \leq C$ ,  $x \in \mathcal{H}$ .*

See [12], where weaker conditions on  $\Psi$  are also provided. For the analysis in the rest of the paper we work under Assumption 3.1.

**Remark 3.2.** We have presented the initial Langevin SPDE (3.1) by referring to the appropriate literature. We have yet to justify the prefix ‘Langevin’. The correspondence between the pairs of equations (2.2), (3.8) and (2.3), (3.9) is already suggestive. The drift of the Langevin SDE (3.8) on the Hilbert space is formed in [13] after copying the finite-dimensional paradigm (1.7): instead of the gradient one has to consider the *variational derivative* of the log-density (1.4). Reversibility of (3.8) and (3.9) w.r.t. the target  $\Pi$  in (1.4) is proved in [12]. See [13] for further discussion.

#### 4. The Pathspace MCMC Samplers

This section is the heart of the paper. We construct path-valued MCMC methods that sample from the target diffusion bridge  $\Pi$ . To do this, we will first discretise the Langevin SPDEs in the algorithmic  $t$ -direction to obtain proposals for the MCMC sampler. We use an implicit Euler scheme parameterised by  $\theta$ . By introducing an additional parameter,  $\alpha \in \{0, 1\}$ , multiplying the nonlinear term in the proposal, we define two alternate proposals. As well as the Langevin proposal (corresponding to  $\alpha = 1$ ), another member of this family ( $\alpha = 0$ ) leads to a natural definition of a

Random Walk proposal on pathspace; we illustrate this in Section 5. We will then provide some motivation concerning the significance of  $\theta$  by showing, in Proposition 4.1, its effect on the quadratic variation of the bridge paths. It will become apparent that  $\theta = 1/2$  is the unique choice that can deliver non-trivial MCMC algorithms on pathspace. We prove this in Theorem 4.1 where we find the analytical expression for the acceptance probability of the proposed paths.

#### 4.1. MCMC Proposals on Pathspace

The  $\mathcal{H}$ -valued Langevin SDEs, non-preconditioned (3.8) and preconditioned (3.9), will be now discretised to provide refined proposals for the Metropolis-Hastings accept/reject rule. Analogously to the finite-dimensional case (2.4), given the implicitness parameter  $\theta \in [0, 1]$  and some time increment  $\Delta t > 0$ , the Langevin proposal  $y$  for the update of a current path  $x$  is formally provided by the scheme:

$$y - x = \{(1 - \theta)\mathcal{L}x + \theta\mathcal{L}y - \alpha\nabla\Psi(x)\}\Delta t + \sqrt{2\Delta t}\xi_0, \quad (4.1)$$

with  $\xi_0 \sim \mathcal{N}(0, \mathbf{I})$ , for the case of (3.8) and by:

$$y - x = \{-(1 - \theta)x - \theta y - \alpha\mathcal{C}(\nabla\Psi(x))\}\Delta t + \sqrt{2\Delta t}\xi, \quad (4.2)$$

with  $\xi \sim \mathcal{N}(0, \mathcal{C})$ , for the case of (3.9). As stated earlier, we have deliberately introduced the tuning parameter  $\alpha \in \{0, 1\}$  into the equations to distinguish Langevin-like and Random Walk-like proposals. In the case  $\alpha = 1$  the proposal is an implicit time-discretization of the SPDE; the convergence of such approximation methods for SPDEs is investigated in [10, 14].

Note that, at present, equation (4.1) should not be conceived in a rigorous manner. The identity mapping  $\mathbf{I}$  is not a nuclear operator on  $\mathcal{H}$ , so the Gaussian law  $\mathcal{N}(0, \mathbf{I})$  is not formally defined (see Appendix A). Also,  $x$  will be typically everywhere non-differentiable, so the differentiation operation  $x \mapsto \mathcal{L}x$  is unjustifiable. Yet, the solution of (4.1) with respect to  $y$  given below is well defined.

We now write, in (4.3) and (4.4), the solutions of (4.1) and (4.2) respectively. Also, for convenience, we introduce some labels.

$$\text{IA : } y = \mathcal{A}_\theta x + \mathcal{B}_\theta(\xi + \mathcal{C}^{1/2}h), \quad (4.3)$$

$$\text{PIA : } y = a_\theta x + b_\theta(\xi + \mathcal{C}h). \quad (4.4)$$

In both cases  $\xi \sim \mathcal{N}(0, \mathcal{C})$ . We have set:

$$\begin{aligned} \mathcal{A}_\theta &= (\mathcal{C} + \theta\Delta t\mathbf{I})^{-1}(\mathcal{C} - (1 - \theta)\Delta t\mathbf{I}), & \mathcal{B}_\theta &= \sqrt{2\Delta t}(\mathcal{C} + \theta\Delta t\mathbf{I})^{-1}\mathcal{C}^{1/2}, \\ a_\theta &= \frac{1 - (1 - \theta)\Delta t}{1 + \theta\Delta t}, & b_\theta &= \frac{\sqrt{2\Delta t}}{1 + \theta\Delta t}, \\ h &= -\alpha\sqrt{\Delta t/2}\nabla\Psi(x). \end{aligned} \quad (4.5)$$

IA stands for ‘Implicit Algorithm’ and PIA for ‘Preconditioned IA’. For the case of (4.3), we first multiplied both sides of (4.1) with  $\mathcal{C}$ , replaced  $\mathcal{C}^{1/2}\xi_0$  with  $\xi \sim \mathcal{N}(0, \mathcal{C})$  and then solved the resulting equation for  $y$ . If  $\theta \in (0, 1]$  the operator  $\mathcal{C} + \theta\Delta t\mathbf{I}$

is onto with a bounded inverse since its eigenvalues are bounded away from zero. Thus, the update  $y$  in (4.3) is well defined for  $\theta \in (0, 1]$ . For (4.4) it is well-defined for  $\theta \in [0, 1]$ .

#### 4.2. Proposed Paths and Quadratic Variation

We exhibit the central importance of  $\theta$  for IA and PIA by showing its effect on the quadratic variation of the paths. We begin with the definition of the quadratic variation  $[x]$  of a path  $x \in \mathcal{H}$ :

$$[x] := \lim_{N \rightarrow \infty} \sum_{i=1}^{2^N} |x(i2^{-N}) - x((i-1)2^{-N})|^2 .$$

The above limit refers to a.s. convergence w.r.t. a relevant path distribution. Consider first the case when  $x \sim \tilde{\Pi}$ . By definition, the Gaussian law  $\tilde{\Pi}$  is equivalent to the  $d$ -dimensional Brownian bridge law (considered on the unit length interval  $[0, 1]$ ) with diffusion coefficient matrix  $B = (b_{ij})$ , so  $[x]$  will be a.s. constant:

$$[x] = \sum_{i,j} b_{ij}^2 . \tag{4.7}$$

From the equivalence  $\Pi \approx \tilde{\Pi}$ , the same will be true for paths  $x \sim \Pi$ . The following result identifies the change in the quadratic variation of an input path  $x \sim \Pi$  when we apply the IA and PIA proposal updates.

**Proposition 4.1.** *Let  $x \sim \Pi$ . Consider the proposed path  $y$ , separately for IA and PIA in (4.3), (4.4) respectively. In both cases,  $[y]$  will be a.s. constant with explicit values:*

$$\begin{aligned} \text{IA : } [y] &= \frac{(1-\theta)^2}{\theta^2} [x] , \\ \text{PIA : } [y] &= \left( 1 + \frac{1-2\theta}{(1+\theta\Delta t)^2} \Delta t^2 \right) [x] . \end{aligned}$$

**Proof.** See Appendix B. □

Therefore, when  $\theta \neq 1/2$ , both IA and PIA generate proposed paths of the wrong quadratic variation, which will necessarily be rejected w.p.1. For  $\theta \neq 1/2$ , the quadratic variation provides a statistic for the magnitude of the discrepancy of the proposed paths from the target distribution. The precise formulas in Proposition 4.1 will be important for understanding, in Section 6, the behaviour of the corresponding finite-dimensional algorithms used in practice.

### 4.3. MCMC Acceptance Probability on Pathspace

We will identify the appropriate accept/reject ratio for the proposed path updates that will guarantee reversibility of the Metropolis-adjusted dynamics with respect to the target bridge  $\Pi$ . The complexity of the state space generates singularities not taken into consideration in cases when the standard Metropolis-Hastings (M-H) or MCMC theory is presented. For instance, the proposed transitions can be mutually singular for varying starting point. We thus refer to a generalised definition of the M-H algorithm developed in [25]. We denote by  $Q(x, dy)$  the transition probability kernel corresponding to either of (4.3), (4.4).

We define the bivariate path distribution:

$$\mu(dx, dy) = \Pi(dx)Q(x, dy) , \quad (4.8)$$

and its transpose  $\mu^\top(dx, dy) = \mu(dy, dx)$ . Theorem 2 of [25] states that if  $\mu, \mu^\top$  are mutually singular (denoted by  $\mu \perp \mu^\top$ ) then the M-H acceptance probability will have to be  $\mu$ -a.s. zero. When  $\mu \approx \mu^\top$  with density

$$r(x, y) = \frac{d\mu}{d\mu^\top}(x, y) , \quad (4.9)$$

the acceptance probability is as follows:

$$a(x, y) = 1 \wedge r(y, x) , \quad (4.10)$$

and  $a(\cdot, \cdot)$  will be  $\mu$ -a.s. positive. Clearly, the standard acceptance probability expression in (2.6) is a special case of the generalised expression given here.

We exploit these results in Theorem 4.1 below. We show that, for both IA and PIA,  $\theta = 1/2$  implies  $\mu \approx \mu^\top$ . That  $\theta \neq 1/2$  implies  $\mu \perp \mu^\top$  has already been shown through the quadratic variation result of Proposition 4.1. Then, for  $\theta = 1/2$ , we identify the density (4.9). Instrumental in these constructions is the consideration of the bivariate Gaussian measure:

$$\tilde{\mu}(dx, dy) := \tilde{\Pi}(dx)\tilde{Q}(x, dy) , \quad (4.11)$$

and its transpose  $\tilde{\mu}^\top(dx, dy) = \tilde{\mu}(dy, dx)$ , where  $\tilde{Q}(x, dy)$  is equal to  $Q(x, dy)$  for vanishing nonlinear term  $\Psi \equiv 0$  and  $\tilde{\Pi}$  is the Gaussian reference measure used to define the target measure. The proof of the theorem is based on two results. First, the density  $d\mu/d\tilde{\mu}$  can be found using theory in the literature on perturbations of Gaussian measures. Then,  $\tilde{\mu}$  is symmetric, that is  $\tilde{\mu} \equiv \tilde{\mu}^\top$ . So in fact:

$$\frac{d\mu}{d\mu^\top}(x, y) = \frac{d\mu}{d\tilde{\mu}}(x, y) / \frac{d\mu^\top}{d\tilde{\mu}^\top}(x, y) = \frac{d\mu}{d\tilde{\mu}}(x, y) / \frac{d\mu}{d\tilde{\mu}}(y, x) . \quad (4.12)$$

**Theorem 4.1.** *Let  $\mu = \mu(dx, dy)$  in (4.8) refer to the IA or PIA proposed path update in (4.3) or (4.4) respectively. Then:*

- i) If  $\theta \neq 1/2$  then  $\mu \perp \mu^\top$  and the M-H acceptance probability  $a(x, y)$  will be  $\mu$ -a.s. zero, whereas if  $\theta = 1/2$  then  $\mu \approx \mu^\top$  and  $a(x, y)$  will be  $\mu$ -a.s. positive.*

ii) For  $\theta = 1/2$ , the acceptance probability is:

$$a(x, y) = 1 \wedge \frac{\rho(y, x)}{\rho(x, y)} \quad (4.13)$$

where  $\log \rho(x, y)$ , for each of the two cases, is given by the expression:

$$\begin{aligned} \text{IA: } & c - \langle 1, \Psi(x) \rangle - \frac{\alpha^2 \Delta t}{4} \langle \nabla \Psi(x), \nabla \Psi(x) \rangle \\ & - \frac{\alpha}{2} \langle \nabla \Psi(x), y - x \rangle - \frac{\alpha \Delta t}{4} \langle \nabla \Psi(x), \mathcal{C}^{-1}(y + x) \rangle, \\ \text{PIA: } & c - \langle 1, \Psi(x) \rangle - \frac{\alpha^2 \Delta t}{4} \langle \mathcal{C}^{1/2} \{ \nabla \Psi(x) \}, \mathcal{C}^{1/2} \{ \nabla \Psi(x) \} \rangle \\ & - \frac{\alpha}{2} \langle \nabla \Psi(x), y - x \rangle - \frac{\alpha \Delta t}{4} \langle \nabla \Psi(x), y + x \rangle, \end{aligned}$$

for the same constant  $c \in \mathbb{R}$ .

**Proof.** See Appendix B for the detailed proof. Here we only make some comments on the integral  $\langle \nabla \Psi(x), \mathcal{C}^{-1}(y + x) \rangle$ . By definition,  $\mathcal{C}^{-1}$  is a second order differential operator, so one should be careful when applying it on non-differentiable paths. We note first that for a fixed  $h \in \mathcal{H}$  and  $\xi \sim \mathcal{N}(0, \mathcal{C})$ , the random variable  $\langle h, \mathcal{C}^{-1/2} \xi \rangle$  makes sense as an  $L^2$ -limit w.r.t.  $\mathcal{N}(0, \mathcal{C})$ ; see Theorem A.1 in Appendix A. Thus,  $\langle h, \mathcal{C}^{-1/2} \xi \rangle$  can be understood as a stochastic integral. Using equation (4.5) for  $\theta = 1/2$  and  $\Psi = 0$ , we find that under the reference measure  $\tilde{\mu}$ :

$$z := \mathcal{C}^{-1/2}(y + x) = 2\Delta t \mathcal{C}^{1/2}(x - y) + \sqrt{8\Delta t} \xi.$$

One can then show that the random variable  $\langle \nabla \Psi(x), \mathcal{C}^{-1/2} z \rangle$  is also well defined as an  $L^2$ -limit w.r.t. the law of  $z$ . See the proof of the theorem for more details.  $\square$

The shape of the average acceptance probability as a function of  $\Delta t$  is of particular interest. Practitioners often tune algorithms to a desirable acceptance probability, so a continuous acceptance probability curve with limiting values 1 (resp. 0) at  $\Delta t = 0$  (resp.  $\Delta t = \infty$ ) would imply that any such probability can be attained. We prove such a behaviour in Proposition 4.2 below for PIA. For IA, the technical difficulties appearing when attempting a corresponding proof (due largely to the presence of the stochastic integral in the expression of the IA acceptance probability in Theorem 4.1) are beyond the scope of this paper. In any case, the mathematical expression in Theorem 4.1 does not give any indication of discontinuity. We observe experimentally the desirable shape for the average acceptance probability curve (for both IA, PIA) in numerical examples in the next section.

**Proposition 4.2.** *For  $\theta = 1/2$ , the average Metropolis-Hastings acceptance probability:*

$$A(\Delta t) := \mathbb{E}[a(x, y)]$$

of the PIA proposed update  $y$  in (4.4) when  $x$  is distributed according to the equilibrium law  $\Pi$  has the following properties:

- a) If  $\Psi \equiv 0$  then  $A(\Delta t) = 1$  for any  $\Delta t$ .
- b) If  $\Psi : \mathbb{R}^d \mapsto \mathbb{R}$  and  $\nabla \Psi : \mathbb{R}^d \mapsto \mathbb{R}^d$  are continuous, then  $A(\cdot)$  is continuous and  $\lim_{\Delta t \rightarrow 0} A(\Delta t) = 1$ . If, additionally,  $\alpha \|\nabla \Psi(x) - \nabla \Psi(y_\infty)\| > 0$  w.p.1, where:

$$y_\infty = -x - 2\alpha \mathcal{C}\{\nabla \Psi(x)\}, \quad (4.15)$$

then  $\lim_{\Delta t \rightarrow \infty} A(\Delta t) = 0$ .

**Proof.** See Appendix B. □

Thus for  $\theta = \frac{1}{2}$  and  $\alpha = 1$ , and in the general nonlinear case, one can obtain any desirable acceptance probability with the appropriate  $\Delta t$ .

## 5. Classification of Infinite-Dimensional Samplers

We have so far defined two Metropolis-Hastings Markov chains on the bridge pathspace. We now show that particular choices for the tuning parameters  $\alpha$  and  $\Delta t$  provide algorithms that can be classified in terms of familiar MCMC samplers used in general applications. Throughout this section  $\theta = 1/2$ .

By construction,  $\alpha = 1$  corresponds to the (implicit version of the) Metropolis-adjusted Langevin Algorithm (MALA). It then turns out that the choice  $\alpha = 0$  provides natural analogues of the Random-Walk Metropolis (RWM) for our pathspace set-up. Indeed, for Euclidean state-spaces when the reference measure is the Lebesgue measure, RWM is characterised by the symmetricity property  $q(x, y) = q(y, x)$ , where  $q(x, y)$  is the transition density of the proposed update. In our context, the infinite-dimensional pathspace dynamics are determined w.r.t. the reference Gaussian law  $\tilde{\Pi}$ . So, the analogue of the above symmetricity property will be:

$$\tilde{\Pi}(dx)Q(x, dy) = \tilde{\Pi}(dy)Q(y, dx),$$

We saw in Section 4.3 that this measure relation is true if  $\alpha = 0$  (the bivariate linear law  $\tilde{\mu}(dx, dy)$  defined there, which coincides with  $\tilde{\Pi}(dx)Q(x, dy)$  when  $\alpha = 0$ , is symmetric). In agreement with standard RWM, when  $\alpha = 0$  the Metropolis-Hastings acceptance probability in Theorem 4.1, for both IA and PIA, simplifies to

$$1 \wedge \frac{\Pi(y)}{\Pi(x)}, \quad (5.2)$$

for  $\Pi(x) = (d\Pi/d\tilde{\Pi})(x)$  given in (1.4). Another familiar sampler, the Independence Sampler (InS), can be deduced from PIA for  $\alpha = 0$  and  $\Delta t = 2$ . For these choices the proposal is  $y = \xi \sim \mathcal{N}(0, \mathcal{C})$ , independently of the current path  $x$ . Independence

samplers on pathspace have already been used in the literature, see [20, 6, 5]. They propose global moves in the pathspace as opposed to *local* moves suggested by MALA and RWM developed in this paper (for the first time, to the best of our knowledge, in the context of the target laws of the paper).

Summarising, particular instances of IA and PIA yield the following identifiable MCMC proposals:

$$\begin{aligned}
 \text{MALA:} & \quad (\alpha, \theta) = (1, 1/2), & y &= \mathcal{A}_{.5} x + \mathcal{B}_{.5} (\xi + \mathcal{C}^{1/2} h) \\
 \text{P-MALA:} & \quad (\alpha, \theta) = (1, 1/2), & y &= a_{.5} x + b_{.5} (\xi + \mathcal{C} h) \\
 \text{RWM:} & \quad (\alpha, \theta) = (0, 1/2), & y &= \mathcal{A}_{.5} x + \mathcal{B}_{.5} \xi \\
 \text{P-RWM:} & \quad (\alpha, \theta) = (0, 1/2), & y &= a_{.5} x + b_{.5} \xi \\
 \text{InS:} & \quad (\alpha, \theta, \Delta t) = (0, 1/2, 2), & y &= \xi
 \end{aligned}$$

for  $h = -\sqrt{\frac{\Delta t}{2}} \nabla \Psi(x)$ . The prefix (P-) distinguishes preconditioned dynamics and  $\mathcal{A}_{.5}, \mathcal{B}_{.5}, a_{.5}, b_{.5}$  follow the definition of  $\mathcal{A}_\theta, \mathcal{B}_\theta, a_\theta, b_\theta$  in (4.5) when  $\theta = 1/2$  and, therefore, depend only on the time increment  $\Delta t$ . Note that:

$$\mathcal{A}_{.5}^2 + \mathcal{B}_{.5}^2 = \text{I}, \quad a_{.5}^2 + b_{.5}^2 = 1,$$

for any  $\Delta t > 0$ , a property which has been instrumental for the definition of the MCMC samplers in Theorem 4.1 and is an immediate consequence of selecting  $\theta = 1/2$ . The importance of this identity can be easily realised in the context of P-RWM, when it provides the appropriate weighting of two independent Brownian(-like) paths that delivers a path of the correct quadratic variation.

Note here that the above defined algorithms can also be applied to targets from the more general family of distributions in (1.6) determined as change of measure from Gaussian laws  $\mathcal{N}(0, \mathcal{C})$  on general Hilbert spaces  $\mathbf{H}$ . For RWM, P-RWM the proposals will be exactly the same as above, considering now the corresponding covariance operator of the reference Gaussian law; the acceptance probability will be as in (5.2) using the relevant target law. The only difference is that the Nemitski operator  $\nabla \Psi(x)$  is replaced by the derivative  $D\Phi : \mathbf{H} \rightarrow \mathbf{H}$  in all proposals with  $\alpha = 1$  and in the resulting acceptance probabilities.

## 6. Numerical Application

We will focus on Langevin dynamics and apply MALA and P-MALA to conditioned nonlinear diffusion processes. Our objective is to illustrate the central role of the parameter  $\theta$  for the algorithms. We consider the scalar diffusion bridge process:

$$\begin{aligned}
 \frac{dX}{du} &= f(X) + \sigma \frac{dW}{du}, \quad u \in [0, U], \\
 X(0) &= x^-, \quad X(U) = x^+,
 \end{aligned} \tag{6.1}$$

where the drift  $f$  is the (negative) gradient of a double-well potential with two stable equilibrium points at  $-1$  and  $+1$ :

$$f(x) = -\left(\frac{(x-1)^2(x+1)^2}{1+x^2}\right)' = x\left(\frac{8}{(1+x^2)^2} - 2\right). \tag{6.2}$$

At the end of the section we also consider MALA algorithms for the same diffusion, but conditioned by observation of a functional of the solution, illustrating that the ideas in this paper have applicability beyond the case of bridge diffusions.

### 6.1. The Langevin Algorithm in Practice

In the previous section we defined infinite-dimensional MALA and P-MALA algorithms that sample from target bridges by proposing moves in pathspace with step-size  $\Delta t$ . The samplers are well-defined only for the unique choice  $\theta = 1/2$  of the implicitness parameter. In practice, bridge paths will have to be discretised along the  $u$ -direction, so an increment  $\Delta u > 0$  will be necessarily introduced. In this finite-dimensional context, corresponding MALA and P-MALA algorithms are well-defined for any  $\theta \in [0, 1]$ . We will present these practical algorithms and examine the effect of the choice  $\theta = 1/2$  in this realistic set-up. Of course, for  $\theta = 1/2$  the algorithm will correspond to the finite-dimensional approximation of the pathspace-valued MALA and P-MALA defined earlier.

We construct the finite-dimensional MALA for the case of the target (6.1). Following Remark 3.1 we assume that  $x^- = x^+ = 0$ . We proceed as follows.

- a) The Langevin SPDE (3.1) is discretised in  $u$ -direction with grid spacing  $\Delta u = U/N$  for some  $N \in \mathbb{N}$ . The result is the following SDE on  $\mathbb{R}^{N-1}$ :

$$\frac{dX^N}{dt} = \frac{1}{\sigma^2} A X^N - \Psi'(X^N) + \sqrt{\frac{2}{\Delta u}} \frac{dW}{dt}, \quad (6.3)$$

where we have used the discretised version of the Laplace operator  $\partial_u^2$ :

$$A = \frac{1}{\Delta u^2} \begin{pmatrix} -2 & 1 & & & \\ & 1 & -2 & 1 & \\ & & & \ddots & \\ & & & & 1 & -2 & 1 \\ & & & & & & 1 & -2 \end{pmatrix}. \quad (6.4)$$

Alternatively, one can obtain this SDE by applying the familiar finite-dimensional Langevin dynamics (1.7) to a discretised version of the target law  $\Pi$ . Indeed, following the Girsanov expression (1.4) of  $\Pi$ , we derive its approximation:

$$\pi(x) = C_1 \exp \left\{ - \sum_{i=1}^N \Psi(x_{i-1}) \Delta u + \frac{\Delta u}{2\sigma^2} \langle x, Ax \rangle \right\}, \quad x \in \mathbb{R}^{N-1}, \quad (6.5)$$

for  $x_0 = x^- = 0$ ,  $x_N = x^+ = 0$  and a normalising constant  $C_1 \in \mathbb{R}$ . The quadratic term comes from the discrete time dynamics of the Brownian bridge  $\tilde{\Pi}$  appearing at the Girsanov formula. We retrieve the SDE (6.3) by applying the Langevin rule (2.3) with  $C = 1/\Delta u$  to this target  $\pi$ .

- b) The SDE (6.3) is discretised in  $t$ -direction with step size  $\Delta t$ , using the implicit  $\theta$ -method as in (2.4). Given a current vector location  $x \in \mathbb{R}^{N-1}$ , this leads to the proposed update  $y$  such that:

$$L_\theta y = R_\theta x - \Psi'(x) \Delta t + \sqrt{\frac{2\Delta t}{\Delta u}} \xi, \quad \xi \sim \mathcal{N}(0, \mathbf{I}), \quad (6.6)$$

where

$$L_\theta = I - \frac{\theta \Delta t}{\sigma^2} A, \quad R_\theta = I + \frac{(1-\theta) \Delta t}{\sigma^2} A.$$

- c) The Metropolis-Hastings correction (2.6) is then applied to the resulting discretisation. The target law in the current context is  $\pi$  given above, and the transition density is:

$$q(x, y) = C_2 \exp\left(-\frac{\Delta u}{4\Delta t} |L_\theta y - R_\theta x + \Psi'(x) \Delta t|^2\right), \quad x, y \in \mathbb{R}^{N-1},$$

corresponding to the recursion (6.6).

We follow the same procedure for P-MALA. Instead of (6.3), we now get:

$$\frac{dX^N}{dt} = -X^N - C \Psi'(X^N) + \sqrt{\frac{2C}{\Delta u}} \frac{dW}{dt}, \quad (6.8)$$

where  $C = -\sigma^2 A^{-1}$ . The rest of the algorithm proceeds as for MALA; we do not give further details.

## 6.2. Numerical Results

We apply MALA and P-MALA to the target diffusion bridge (6.1)-(6.2). The theory in Section 4 suggests that unless  $\theta = 1/2$ , when there is a limiting infinite-dimensional algorithm for  $\Delta u \rightarrow 0$ , both algorithms will eventually degenerate as  $\Delta u \rightarrow 0$ . In Fig.1 and Fig.2 we show the empirical average acceptance probability in equilibrium, produced from a long run of MALA and P-MALA, for various grid sizes  $\Delta t$  and  $\Delta u$  and various choices of the implicitness parameter  $\theta$ .

Fig.1 shows that in the case of MALA minor deviations from the threshold value  $\theta = 1/2$  can have dramatic effect on the behaviour of the algorithm even for moderate  $\Delta u$ . The middle and bottom panel show that one could empirically conjecture about the existence of a well-defined limiting algorithm as  $\Delta u \rightarrow 0$  only when  $\theta = 1/2$ . Fig.2 shows a different behaviour for P-MALA. The algorithm does not degenerate for  $\theta \neq 1/2$  and the values of  $\Delta u$  used in the graphs. In fact, for these  $\Delta u$ , there is no apparent difference among the acceptance probability curves over the various choices of  $\theta$ . A first line of explanation for this contrasting behaviour of MALA and P-MALA can be provided by the quadratic variation expressions in Proposition 4.1. In the limiting pathspace context, the distortion, when  $\theta \neq 1/2$ , at the quadratic variation of the paths is of order  $\mathcal{O}(\Delta t^2)$  for P-MALA as opposed to  $\mathcal{O}(1)$  for MALA. Thus, in the case of MALA, the difference at the quadratic

variation is sufficiently large for any  $\Delta t$  to lead to almost certain rejection even for moderate  $\Delta u$ . For P-MALA, one must use  $\Delta u$  of higher order to capture the effect of  $\theta \neq 1/2$ . The following proposition gives the precise statements about the effect of the increments  $\Delta u$ ,  $\Delta t$  on the properties of the algorithms, in the linear case.

**Proposition 6.1.** *Consider the Langevin algorithms MALA and P-MALA as applied to the Gaussian target corresponding to the density  $\pi$  in (6.5) for vanishing non-linear term  $\Psi \equiv 0$ . Let  $A(\theta, \Delta u, \Delta t)$  denote the average acceptance probability of the proposed vector  $y$  when the current vector  $x$  is distributed according to the equilibrium law  $\pi$ . It is then true that:*

- a) *If  $\theta = 1/2$  then for both algorithms  $A(\theta, \Delta u, \Delta t) = 1$  for any  $\Delta u$ ,  $\Delta t$ .*
- b) *If  $\theta \neq 1/2$ , let  $\Delta t = \beta \Delta u^\alpha$  for some  $\alpha, \beta > 0$ . The asymptotic behaviour of  $A(\theta, \Delta u, \Delta t)$  is as below:*
  - 1) *for  $\alpha < \alpha_0$ :  $\lim_{\Delta u \downarrow 0} A(\theta, \Delta u, \Delta t) = 0$ ,*
  - 2) *for  $\alpha = \alpha_0$ :  $\lim_{\Delta u \downarrow 0} A(\theta, \Delta u, \Delta t) \in (0, 1)$ ,*
  - 3) *for  $\alpha > \alpha_0$ :  $\lim_{\Delta u \downarrow 0} A(\theta, \Delta u, \Delta t) = 1$ ,*

*where, for each of the two algorithms, the threshold rate  $\alpha_0$  is specified as follows:*

$$\begin{aligned} \text{MALA: } \alpha_0 &= 7/3, \\ \text{P-MALA: } \alpha_0 &= 1/3. \end{aligned}$$

**Proof.** See Appendix B. □

So, P-MALA will degenerate ( $\theta \neq 1/2$ ) if  $\Delta u$  is of the order  $\Delta t^\alpha$  for  $\alpha > 3$ . For MALA, the corresponding values of  $\Delta u$  are much larger:  $\Delta t^\alpha$ ,  $\alpha > 3/7$ . To capture the degeneration of P-MALA, keeping the same choices for the increment  $\Delta u$  as in Fig.1 and Fig.2 we should focus on larger values of  $\Delta t$ . In Fig.3 we show numerical results from an application of P-MALA, where the only difference with Fig.2 is that the length of the target bridge is now  $U = 1$  (compared to  $U = 10$  of the previous graphs). This choice provides non-negligible acceptance probabilities around the values of  $\Delta t$  which distinguish the algorithms for  $\theta = 1/2$  and  $\theta \neq 1/2$ .

### 6.3. A Signal Processing Example

In this section we illustrate the effect of the implicitness parameter  $\theta$  when solving a filtering/smoothing problem. We use the MALA algorithm to sample from the conditional distribution of an (unknown) signal  $\{x_t\}_{t \in [0, U]}$ , given an observation of a path  $\{y_t\}_{t \in [0, U]}$ . We assume that the joint evolution of signal and observation is described by the SDE

$$\begin{aligned} \frac{dX}{du} &= f(X) + \frac{dW_1}{du}, \quad u \in [0, U], \\ \frac{dY}{du} &= X + \frac{dW_2}{du}, \quad u \in [0, U], \end{aligned} \tag{6.9}$$

where  $f$  is the double-well drift from (6.2). The distribution of  $x$  given  $y$  is known as the smoothing distribution and can be written as a measure change from the Gaussian distribution (Kalman-Bucy smoother) found when  $f \equiv 0$ . Fig. 4 shows that the dichotomy between the cases  $\theta = 1/2$  and  $\theta \neq 1/2$ , illustrated in the previous figures for bridge diffusions, also occurs for this smoothing problem. The figure illustrates the MALA method using the appropriate underlying SPDE derived in [12, 13]. For  $\theta = 1/2$  the acceptance rates do not significantly change when  $\Delta u$  is decreased whereas for  $\theta \neq 1/2$  decreasing  $\Delta u$  requires a simultaneous decrease in  $\Delta t$ .

## 7. Conclusions

We have demonstrated the potential of Langevin-based MCMC methods in an infinite-dimensional context. The methodology introduced relies on a special implicit discretization of the Langevin equation and applies for target measures which are defined via measure change from a reference Gaussian law. This results in MCMC methods which are robust under finite-dimensional approximation of the infinite-dimensional space. We have exposed the details in the case where the target measure is a diffusion bridge with non-degenerate additive noise, for pedagogical reasons. However the general setting is highlighted in Section 5.

In the context of bridge sampling the algorithm allows for tuned local moves in the pathspace and is expected to behave better than previous methods based on Independence Samplers, especially in various singular limits such as small diffusivity, or large length of the path. We have shown that standard numerical methods for finite-dimensional SDEs (implicit scheme, preconditioning) are of particular importance when employed in the SPDE setting arising for pathspace sampling. The implicitness parameter choice  $\theta = 1/2$  is necessary for the infinite-dimensional algorithm and optimal for its practical implementation in a (high but) finite-dimensional approximation. There are two versions of the algorithm, one preconditioned. Different choices of path quantities lead to situations in which either method can be superior in terms of rate of convergence of the empirical measure, but both methods have the desirable property that the rate of convergence is not mesh-dependent, for  $\theta = \frac{1}{2}$ . We have also introduced Langevin and Random Walk versions of the algorithm; again either can be more efficient, depending on the precise problem being studied, but both give rise to convergence rates which are mesh-independent.

The method, as developed in this paper, considers target bridges where the drift is a gradient plus linear term. Furthermore, the noise is additive. Future research will be directed towards extending its applicability beyond this setting. See [12] for discussion about the form of the Langevin SPDE in more general settings. In a wider perspective, the methodology employed here can potentially be applied to a large family of sampling problems for conditioned diffusions, including the problem of signal processing illustrated in Section 6.

Another perspective on our work concerns its relation to existing analysis of

MCMC methods in high dimensional spaces. Our work provides further insight into this important general problem and provides a number of open avenues for investigation. In particular, it is of interest to extend the work in this paper to cover general target distributions defined as change of measure from Gaussian laws  $\mathcal{N}(0, \mathcal{C})$ . It is not necessarily the case that implicit methods will always be favourable. This is because the additional cost of inverting linear combinations of the identity operator and  $-\mathcal{L} = \mathcal{C}^{-1}$  must be traded against the improvement in the rate of convergence that stems from choosing  $\theta = \frac{1}{2}$ . It is also of interest to include the study of proposals other than the Langevin one. As demonstrated here, a natural analogue of Random Walk Metropolis in this pathspace setting is to make proposals from the time-discrete SPDE with the nonlinear term  $\Psi$  set to zero.

Finally, we believe that working with the infinite-dimensional setting provides a better insight into the properties of the algorithms used in practice and a clear understanding of the effect of the involved parameters (as for instance  $\theta$  in our case). We anticipate that this idea could also be relevant in other cases involving infinite-dimensional systems and their approximation.

### Acknowledgements

The computing facilities for producing Fig.1-4 were provided by the Centre for Scientific Computing of the University of Warwick. The authors are grateful to EPSRC and ONR for financial support.

## APPENDIX

### A. Gaussian Law on Hilbert Space: An Excerpt.

Let  $\mathbf{H}$  denote an arbitrary separable Hilbert space with inner product  $\langle \cdot, \cdot \rangle$ .

**Definition A.1.** An  $\mathbf{H}$ -valued random variable (r.v.)  $\xi$  is called Gaussian (and its distribution a Gaussian distribution), if for any  $h \in \mathbf{H}$  the r.v.  $\langle \xi, h \rangle$  follows a Gaussian distribution on  $\mathbb{R}$ .

We will consider only non-degenerate Gaussian distributions: Gaussian laws excluding Dirac measures. We denote by  $\mathcal{L}(\mathbf{H})$  the set of continuous, linear operators from  $\mathbf{H}$  to  $\mathbf{H}$ .

**Proposition A.1.** *If  $\xi$  is a Gaussian r.v. on  $\mathbf{H}$  then there exist an  $m \in \mathbf{H}$  and a positive, self-adjoint, nuclear operator  $\Sigma \in \mathcal{L}(\mathbf{H})$  such that:*

$$\begin{aligned} \mathbb{E}[\langle \xi, h \rangle] &= \langle m, h \rangle, \quad h \in \mathbf{H}, \\ \mathbb{E}[\langle \xi - m, h \rangle \langle \xi - m, g \rangle] &= \langle h, \Sigma g \rangle, \quad h, g \in \mathbf{H}. \end{aligned}$$

*$m, \Sigma$  are uniquely determined. Conversely, for  $m, \Sigma$  as above, there exists a unique (up to its probability law) Gaussian r.v. satisfying the above moment conditions.*

We thus denote a Gaussian distribution on  $\mathbf{H}$  as  $\mathcal{N}(m, \Sigma)$  for appropriate  $m \in \mathbf{H}$ ,  $\Sigma \in \mathfrak{L}(\mathbf{H})$ . From standard operator theory, a positive, self-adjoint, nuclear operator  $\Sigma \in \mathfrak{L}(\mathbf{H})$  is associated with an eigensequence  $\{(\lambda_i, e_i)\}$  so that  $\Sigma e_i = \lambda_i e_i$  for all  $i \geq 1$ , with the property that  $\{\lambda_i\}$  is a bounded sequence of positive reals with  $\sum \lambda_i < \infty$ , and  $\{e_i\}$  is an orthonormal basis for  $\mathbf{H}$ , i.e.  $\langle e_i, e_j \rangle = \delta_{ij}$ .

We proceed with the Karhunen-Loève expansion of Gaussian laws. Given an orthonormal base  $\{e_i\}$  of  $\mathbf{H}$ , the Fourier expansion for an arbitrary  $x \in \mathbf{H}$  is as follows:

$$x = \sum_{i=0}^{\infty} \langle x, e_i \rangle e_i .$$

We set  $x_i = \langle x, e_i \rangle$ . From the standard isometry of  $\mathbf{H}$  and  $\ell^2$ ,  $\sum_i x_i^2 < \infty$ . Assume now that  $x \sim \mathcal{N}(m, \Sigma)$  and  $\{(\lambda_i, e_i)\}$  is the eigensequence corresponding to  $\Sigma$ . Proposition A.1 implies that  $\langle x, e_i \rangle \sim \mathcal{N}(m_i, \lambda_i)$  with  $m_i = \langle m, e_i \rangle$ , independently of  $\langle x, e_j \rangle$  for any  $j \neq i$ . We can now write the Karhunen-Loève expansion: if  $\{\xi_i\}$  is a sequence of iid standard Gaussian variables then:

$$\mathcal{N}(m, \Sigma) = \sum_{i=1}^{\infty} \{ \sqrt{\lambda_i} \xi_i + m_i \} e_i ,$$

in the sense that the right-hand random series converges in  $L^2$  to a random element of law  $\mathcal{N}(m, \Sigma)$ . Therefore, a Gaussian law on an arbitrary Hilbert space can be equivalently represented by a product of scalar Gaussian laws:

$$\mathcal{N}(m, \Sigma) \equiv \prod_{i=1}^{\infty} \mathcal{N}(m_i, \lambda_i) .$$

We will now state a result on the absolute continuity properties of Gaussian laws. In the light of the Karhunen-Loève expansion above such properties can be directly implied by known results for product laws on  $\mathbb{R}^\infty$ .

**Lemma A.1.** *The product Gaussian laws  $\prod_{i=1}^{\infty} \mathcal{N}(\mu_i, \sigma_i)$  and  $\prod_{i=1}^{\infty} \mathcal{N}(0, \tau_i)$  are either equivalent or mutually singular; they are equivalent if and only if:*

$$\sum_{i=1}^{\infty} \mu_i^2 / \sigma_i < \infty, \quad \sum_{i=1}^{\infty} \left( \frac{\sigma_i}{\tau_i} - 1 \right)^2 < \infty .$$

The result follows from the Hellinger distance of product laws, see Proposition 2.20 of [4]. It can also be derived by Kakutani's dichotomy, see e.g. [26]. In the theorem that follows,  $\text{Im}\Sigma^{1/2}$  denotes the image set of the operator  $\Sigma^{1/2}$ .

**Theorem A.1.** *The Gaussian laws  $\mu = \mathcal{N}(m, \Sigma)$  and  $\nu = \mathcal{N}(0, \Sigma)$  are equivalent if and only if  $m \in \text{Im}\Sigma^{1/2}$ . In the case of equivalence:*

$$\frac{d\mu}{d\nu}(x) = \exp \left\{ \langle h, \Sigma^{-1/2} x \rangle - \frac{1}{2} \langle h, h \rangle \right\}, \quad x \in \mathbf{H} ,$$

where  $h = \Sigma^{-1/2}m$ ; the random element  $x \mapsto \langle h, \Sigma^{-1/2}x \rangle$  is defined as the  $L^2(\nu)$ -limit:

$$\sum_{i=1}^{\infty} \frac{\langle h, e_i \rangle \langle x, e_i \rangle}{\sqrt{\lambda_i}}, \quad x \in \mathbf{H},$$

where  $\{(\lambda_i, e_i)\}$  is the eigensequence associated with  $\Sigma$ .

This is Theorem 2.21 of [4]. The proof of the absolute continuity follows directly from Lemma A.1 since  $m \in \text{Im}\Sigma^{1/2}$  iff  $\sum_i m_i^2/\lambda_i < \infty$ . The operator  $\Sigma^{1/2}$  is not onto, so  $\langle h, \Sigma^{-1/2}x \rangle$  cannot make sense as a regular inner product and corresponds to the extension of the Itô stochastic integral to general Hilbert spaces.

## B. Proofs

*Proof of Proposition 4.1:*

We will prove the quadratic variation statements assuming that  $x \sim \tilde{\Pi}$  and  $h \equiv 0$ , where  $h$  is the nonlinear term appearing in the definition of the proposed paths in (4.3), (4.4). The result will then also hold for the required setting since a.s. properties are preserved among equivalent probability measures. In our case  $\Pi \approx \tilde{\Pi}$  and, from Theorem A.1 in Appendix A, the laws of  $\xi$ ,  $\xi + \mathcal{C}^{1/2}h$  and  $\xi + \mathcal{C}h$ , where  $\xi \sim \mathcal{N}(0, \mathcal{C})$ , are all equivalent.

For the case of PIA the result follows directly from the independence of  $x$  and  $\xi$ . The quadratic variation of the proposed path  $a_\theta x + b_\theta \xi$ , for scalars  $a_\theta, b_\theta$ , will be simply:

$$[y] = a_\theta^2 [x] + b_\theta^2 [\xi],$$

for the quadratic variation  $[x] = [\xi]$  given in (4.7). Substituting  $a_\theta, b_\theta$  with their expressions in (4.5) gives the required result.

For IA, we will have to work with the Karhunen-Loève expansion of Gaussian laws reviewed in Appendix A. Let  $\{\lambda_i\}$  be the eigenvalues of the covariance operator  $\mathcal{C}$  and  $\{e_i\}$  the orthonormal base of  $\mathcal{H}$  formed by the corresponding eigenvectors of  $\mathcal{C}$ , thus  $\mathcal{C}e_i = \lambda_i e_i$ . From (4.3) we find the  $i$ -th Fourier coefficient of the proposed path  $y$ :

$$\langle y, e_i \rangle = \frac{\lambda_i - \Delta t(1 - \theta)}{\lambda_i + \Delta t\theta} \langle x, e_i \rangle + \frac{\sqrt{2\Delta t\lambda_i}}{\lambda_i + \Delta t\theta} \langle \xi, e_i \rangle. \quad (\text{B.1})$$

We have assumed that  $x \sim \tilde{\Pi} \equiv \mathcal{N}(0, \mathcal{C})$ . From the Karhunen-Loève expansion of this Gaussian law,  $\langle x, e_i \rangle \sim \mathcal{N}(0, \lambda_i)$  independently of the rest of the Fourier coefficients. Similarly  $\langle \xi, e_i \rangle \sim \mathcal{N}(0, \lambda_i)$  independently of each other and of  $x$ . From (B.1) we now trivially get that  $\langle y, e_i \rangle \sim \mathcal{N}(0, a_i \lambda_i)$  where:

$$a_i = \frac{(\lambda_i - \Delta t(1 - \theta))^2 + 2\Delta t\lambda_i}{(\lambda_i + \Delta t\theta)^2}$$

22 Beskos et al.

so the (Gaussian) distribution of  $y$  corresponds to that of the product of scalar Gaussian laws  $\prod_{i=1}^{\infty} \mathcal{N}(0, a_i \lambda_i)$ . From the properties of a covariance operator,  $\lambda_i > 0$  for any  $i \geq 1$  with  $\sum_i \lambda_i < \infty$ . Using this, we find that:

$$\lim_{i \rightarrow \infty} a_i = (1 - \theta)^2 / \theta^2 =: a; \quad \sum_{i=1}^{\infty} (a_i / a - 1)^2 < \infty .$$

Lemma A.1 in Appendix A now implies the equivalence:

$$\prod_{i=1}^{\infty} \mathcal{N}(0, a_i \lambda_i) \approx \prod_{i=1}^{\infty} \mathcal{N}(0, a \lambda_i) .$$

Recall that the left-hand side law is the distribution of  $y$ ; the right-hand side law is trivially identified as the distribution of  $\sqrt{a} x$  which has a.s. constant quadratic variation. Therefore,  $[y]$  is also a.s. constant and given as:

$$[y] = a [x] = \frac{(1 - \theta)^2}{\theta^2} [x] .$$

The proof is now complete. □

*Proof of Theorem 4.1:*

We prove the statements i) and ii), first for IA.

*Proofs for IA:*

i) We will use the bivariate Gaussian law  $\tilde{\mu}(dx, dy) = \tilde{\Pi}(dx) \tilde{Q}(x, dy)$ . It is useful to write side-by-side the definition of the involved kernels:

$$\begin{aligned} Q(x, dy) &: y = \mathcal{A}_\theta x + \mathcal{B}_\theta(\xi + \mathcal{C}^{1/2}h) \\ \tilde{Q}(x, dy) &: y = \mathcal{A}_\theta x + \mathcal{B}_\theta \xi , \end{aligned} \tag{B.6}$$

for  $\xi \sim \mathcal{N}(0, \mathcal{C})$ . From Theorem A.1 in Appendix A, the laws of  $\xi + \mathcal{C}^{1/2}h$  and  $\xi$  are equivalent. Thus,  $Q(x, dy) \approx \tilde{Q}(x, dy)$  for any  $x \in \mathcal{H}$ . Since also  $\Pi \approx \tilde{\Pi}$ , it follows directly from Fubini's theorem ([1]) that  $\mu \approx \tilde{\mu}$  and, equivalently,  $\mu^\top \approx \tilde{\mu}^\top$ . Therefore, investigation of equivalence or singularity for the nonlinear laws  $\mu, \mu^\top$  has now been reduced to exploring the same properties for the centred Gaussian bivariate laws  $\tilde{\mu}$  and  $\tilde{\mu}^\top$ . From the definition of the Gaussian distribution on a Hilbert space, see Appendix A, we find that the covariance operators of  $\tilde{\mu}$  and  $\tilde{\mu}^\top$  are equal to:

$$\begin{pmatrix} \mathcal{C} & \mathcal{C} \mathcal{A}_\theta^\top \\ \mathcal{A}_\theta \mathcal{C}^\top & \mathcal{C}_1 \end{pmatrix}, \quad \begin{pmatrix} \mathcal{C}_1 & \mathcal{A}_\theta \mathcal{C}^\top \\ \mathcal{C} \mathcal{A}_\theta^\top & \mathcal{C} \end{pmatrix},$$

respectively, where:

$$\mathcal{C}_1 = \mathcal{B}_\theta \mathcal{C} \mathcal{B}_\theta^\top + \mathcal{A}_\theta \mathcal{C} \mathcal{A}_\theta^\top .$$

One can easily verify that  $\mathcal{C} = \mathcal{C}_1$  iff  $\theta = 1/2$ . Also, for any  $\theta$ ,  $\mathcal{A}_\theta \mathcal{C}^\top = \mathcal{C} \mathcal{A}_\theta^\top$ . Thus, if  $\theta = 1/2$  the centred Gaussian laws  $\tilde{\mu}$ ,  $\tilde{\mu}^\top$  coincide which, from the relation  $\mu \approx \tilde{\mu}$ , implies that  $\mu \approx \mu^\top$ . Singularity for  $\theta \neq 1/2$  follows directly from Proposition 4.1.

ii) From the symmetricity of  $\tilde{\mu}$ ,  $a(x, y)$  is as in (4.13) for:

$$\rho(x, y) = \frac{d\mu}{d\tilde{\mu}}(x, y) .$$

See (4.12) for an explanatory calculation. Recall the conditional structure of the two laws,  $\mu(dx, dy) = \Pi(dx)Q(x, dy)$  and  $\tilde{\mu}(dx, dy) = \tilde{\Pi}(dx)\tilde{Q}(x, dy)$ . The density of  $\Pi$  w.r.t.  $\tilde{\Pi}$  is available from Girsanov's theorem, and is given in (1.4). We now fix some  $x \in \mathcal{H}$  and consider the definition of the kernels  $Q(x, dy)$  and  $\tilde{Q}(x, dy)$  in (B.6). From Theorem A.1 in the Appendix, the density of the law of  $\xi + \mathcal{C}^{1/2}h$  w.r.t. the law of  $\xi$ , with  $\xi \sim \mathcal{N}(0, \mathcal{C})$ , is:

$$\rho_x(z) = \exp \left\{ \langle h, \mathcal{C}^{-1/2}z \rangle - \frac{1}{2} \langle h, h \rangle \right\} ,$$

for the random element  $\langle h, \mathcal{C}^{-1/2} \cdot \rangle$  defined as an  $L^2(\tilde{\Pi})$ -limit in Theorem A.1. The laws  $Q(x, dy)$  and  $\tilde{Q}(x, dy)$  are derived via the location-scale transformation:

$$g_x(z) = \mathcal{A}_{.5}x + \mathcal{B}_{.5}z, \quad z \in \mathcal{H} ,$$

of the Gaussian laws  $\xi + \mathcal{C}^{1/2}h$  and  $\xi$  respectively, thus their density is:

$$\frac{dQ(x, \cdot)}{d\tilde{Q}(x, \cdot)}(y) = \rho_x(g_x^{-1}(y)) . \quad (\text{B.11})$$

From Fubini's theorem the bivariate density  $\rho(x, y)$  is the product of the densities (1.4) and (B.11). So:

$$\log \rho(x, y) = c - \langle 1, \Psi(x) \rangle - \frac{1}{2} \langle h, h \rangle + \langle h, \mathcal{C}^{-1/2}g_x^{-1}(y) \rangle , \quad (\text{B.12})$$

where  $\exp(c)$  is the normalising constant of the expression of  $d\Pi/d\tilde{\Pi}$  in (1.4). From the analytical expressions for  $\mathcal{A}_{.5}$ ,  $\mathcal{B}_{.5}$  we can rewrite:

$$g_x^{-1}(y) = \mathcal{C}^{1/2}(y - x)/\sqrt{2\Delta t} + \mathcal{C}^{-1/2}(y + x)\sqrt{\Delta t/8} .$$

Let  $\{(\lambda_i, e_i)\}$  be the eigensequence corresponding to the covariance operator  $\mathcal{C}$ . From the definition of the stochastic integral  $\langle h, \mathcal{C}^{-1/2} \cdot \rangle$  in Theorem A.1:

$$\begin{aligned} \langle h, \mathcal{C}^{-1/2}g_x^{-1}(y) \rangle &= \sum_{i=1}^{\infty} \frac{\langle h, e_i \rangle \langle g_x^{-1}(y), e_i \rangle}{\sqrt{\lambda_i}} = \\ &= \frac{1}{\sqrt{2\Delta t}} \langle h, y - x \rangle + \sqrt{\frac{\Delta t}{8}} \sum_{i=1}^{\infty} \frac{\langle h, e_i \rangle \langle \mathcal{C}^{-1/2}(y + x), e_i \rangle}{\sqrt{\lambda_i}} , \end{aligned}$$

where both limits are in the  $L^2$ -norm w.r.t. the law  $y \sim \tilde{Q}(x, dy)$ . In agreement with the definition of the stochastic integral  $\langle h, \mathcal{C}^{-1/2} \cdot \rangle$ , as the first of the two series above, the second series provides the following definition:

$$\langle h, \mathcal{C}^{-1}(y+x) \rangle := \sum_{i=1}^{\infty} \frac{\langle h, e_i \rangle \langle \mathcal{C}^{-1/2}(y+x), e_i \rangle}{\sqrt{\lambda_i}} .$$

The discussion in the main body of the proof shows that we can make sense of this as an  $L^2$ -limit. From these last two equations and the analytical expression for  $h$  from (4.5), the initial formula for  $\log \rho(x, y)$  in (B.12) can be rewritten as the one given at the statement of the Theorem.

*Proofs for PIA:*

The proof of i) is exactly the same as the proof for IA above, with the exception that the operators  $\mathcal{A}_\theta$  and  $\mathcal{B}_\theta$  should be now replaced by the scalars  $a_\theta$  and  $b_\theta$ . The same holds for ii) as well:  $a_{.5}$  and  $b_{.5}$  should now be used in the place  $\mathcal{A}_{.5}$  and  $\mathcal{B}_{.5}$ .

□

*Proof of Proposition 4.2:*

Part a) of the proposition follows trivially from the expression for the acceptance probability in Theorem 4.1. We proceed to part b). The expected acceptance probability in equilibrium is formally calculated as follows:

$$A(\Delta t) = \int_{\mathcal{H} \times \mathcal{H}} \left( 1 \wedge \frac{\rho(y, x)}{\rho(x, y)} \right) \Pi(dx) Q(x, dy) .$$

The dependence of the integrating measure on the time increment can be removed after rewriting:

$$A(\Delta t) = \int_{\mathcal{H} \times \mathcal{H}} \left( 1 \wedge \frac{\rho(y(\Delta t), x, \Delta t)}{\rho(x, y(\Delta t), \Delta t)} \right) \Pi(dx) \tilde{\Pi}(d\xi) ,$$

since, from the definition of the proposed path  $y$  in (4.4) with  $\theta = 1/2$ ,  $y = y(x, \xi, \Delta t)$  for  $\xi \sim \tilde{\Pi}$ . We simplified  $y(x, \xi, \Delta t)$  to  $y(\Delta t)$  for notational convenience and also wrote  $\rho(\cdot, \cdot, \Delta t)$  instead of  $\rho(\cdot, \cdot)$  to emphasise the dependence of the density  $\rho$  on  $\Delta t$ . The integrand in the above expression is bounded, thus, from the bounded convergence theorem, to prove continuity of  $\Delta t \mapsto A(\Delta t)$  it suffices to prove a.s. continuity of the mappings  $\Delta t \mapsto \rho(y(\Delta t), x, \Delta t)$  and  $\Delta t \mapsto \rho(x, y(\Delta t), \Delta t)$ . A simple inspection of the analytical formulae from Theorem 4.1 and the expression for  $y$  in (4.4), indicates that the logarithm of both of these mappings is a linear combination of inner products between the paths  $x$ ,  $y(\Delta t)$ ,  $\Psi(x)$ ,  $\Psi(y(\Delta t))$ ,  $\nabla \Psi(x)$  and  $\nabla \Psi(y(\Delta t))$  or  $\mathcal{C}^{1/2}$ -transformations of them. Fix some continuous paths  $x, \xi \in \mathcal{H}$ . The continuity of  $\Psi : \mathbb{R}^d \mapsto \mathbb{R}$  and  $\nabla \Psi : \mathbb{R}^d \mapsto \mathbb{R}^d$  and the formula for  $y(\Delta t)$  imply that  $\Delta t \mapsto y(\Delta t)$ ,  $\Delta t \mapsto \nabla \Psi(y(\Delta t))$  are continuous as mappings from  $[0, \infty)$  to  $L^2([0, 1], \mathbb{R}^d)$  while  $\Delta t \mapsto \Psi(y(\Delta t))$  is continuous as a mapping from  $[0, \infty)$  to  $L^2([0, 1], \mathbb{R})$ . Thus, from the continuity of the inner product and  $\mathcal{C}^{1/2}$ , any of the

above inner products are continuous functions of  $\Delta t$ . Continuity of  $\Delta t \mapsto A(\Delta t)$  has now been demonstrated. Clearly,  $A(\Delta t) \rightarrow A(0) = 1$  when  $\Delta t \rightarrow 0$ . Let  $\Delta t \rightarrow \infty$ . Using the analytical formulae, we find that the coefficient of  $(-\Delta t/4)$  at the expression  $\log \rho(y(\Delta t), x, \Delta t) - \log \rho(x, y(\Delta t), \Delta t)$  converges a.s. to:

$$\alpha^2 \| \mathcal{C}^{1/2} \{ \nabla \Psi(y_\infty) \} \|^2 + \alpha \langle \nabla \Psi(y_\infty), x + y_\infty \rangle + \alpha^2 \| \mathcal{C}^{1/2} \{ \nabla \Psi(x) \} \|^2 ,$$

for  $y_\infty = \lim_{\Delta t \rightarrow \infty} y(\Delta t)$ , in  $L^2([0, 1], \mathbb{R}^d)$  or pointwise, given explicitly in (4.15). Using the formula in (4.15), we find that the above expression is actually equal to:

$$\alpha^2 \| \mathcal{C}^{1/2} \{ \nabla \Psi(y_\infty) - \nabla \Psi(x) \} \|^2 .$$

Therefore, if  $\alpha \| \nabla \Psi(x) - \nabla \Psi(y_\infty) \| > 0$  w.p.1, then:

$$\lim_{\Delta t \rightarrow \infty} \frac{\rho(y(\Delta t), x, \Delta t)}{\rho(x, y(\Delta t), \Delta t)} = 0, \text{ w.p.1 ,}$$

which implies that  $\lim_{\Delta t \rightarrow \infty} A(\Delta t) = 0$ .

□

*Proof of Proposition 6.1:*

From (6.5), the target law is now the  $(N - 1)$ -dimensional Gaussian distribution  $\pi \equiv \mathcal{N}(0, -\frac{\sigma^2}{\Delta u} A^{-1})$ . One can verify that the eigenvalues of  $A$  are  $-\lambda_1, \dots, -\lambda_{N-1}$  where:

$$\lambda_i = \frac{2(1 - \cos(i\pi/N))}{(\Delta u)^2} .$$

Let  $\Lambda$  be the diagonal matrix with  $i$ -th diagonal element  $\lambda_i$  and  $Z$  the orthonormal matrix such that  $Z^\top (-A) Z = \Lambda$ .

*Proofs for MALA:*

The 1-1 transformation

$$x \mapsto \sqrt{\Delta u} Z^\top x$$

of the original MALA Markov chain corresponds to an MCMC algorithm with the following target and proposal:

$$\begin{aligned} \text{Target: } & \pi \equiv \mathcal{N}(0, \sigma^2 \Lambda^{-1}) \\ \text{Proposal: } & \left( \mathbf{I} + \frac{\theta \Delta t}{\sigma^2} \Lambda \right) y = \left( \mathbf{I} - \frac{(1 - \theta) \Delta t}{\sigma^2} \Lambda \right) x + \sqrt{2 \Delta t} \xi , \end{aligned} \tag{B.21}$$

where  $\xi \sim \mathcal{N}(0, \mathbf{I})$ . Since the acceptance probability of this algorithm is identical to the acceptance probability of the original algorithm, we henceforth assume that we work with the transformed MALA in (B.21). After some calculations one can find that the acceptance probability of (the transformed) MALA is:

$$a(x, y) = 1 \wedge \exp \left\{ \frac{1}{4} (1 - 2\theta) \sigma^{-4} \Delta t \sum_{i=1}^{N-1} \lambda_i^2 (x_i^2 - y_i^2) \right\} . \tag{B.22}$$

26 *Beskos et al.*

Proof of a): For  $\theta = 1/2$ ,  $a(x, y) = 1$  so also  $A(\theta, \Delta u, \Delta t) = 1$ .

Proof of b): We now assume that  $\theta \neq 1/2$  and set  $\Delta t = \beta \Delta u^\alpha$ . Consider the following set:

$$\Gamma = \{(x, y) : \pi(x)q(x, y) < \pi(y)q(y, x)\} .$$

From the definition of the acceptance probability  $a(x, y)$  in (2.6) we get:

$$\mathbb{E}[\alpha(x, y)] = \mathbb{E}[\alpha(x, y) \mathbb{I}_\Gamma(x, y)] + \mathbb{E}[\alpha(x, y) \mathbb{I}_{\Gamma^c}(x, y)] = 2 \mathbb{P}[\Gamma]$$

since one can trivially check that each of the two summands above is equal to the probability of  $\Gamma$ . From (B.22), we can equivalently write:

$$A(\theta, \Delta u, \Delta t) = 2 \mathbb{P}[(1 - 2\theta) \sum_{i=1}^{N-1} \lambda_i^2 (x_i^2 - y_i^2) > 0] . \quad (\text{B.25})$$

We will calculate the expectation and variance of the random variable:

$$Z_N = (1 - 2\theta) \sum_{i=1}^{N-1} \lambda_i^2 (x_i^2 - y_i^2) . \quad (\text{B.26})$$

The dynamics of  $x, y$  are determined by (B.21). One can easily find that:

$$\mathbb{E}[x_i^2] = \frac{\sigma^2}{\lambda_i}, \quad \mathbb{E}[y_i^2] = \frac{\sigma^2}{\lambda_i} + \frac{(1 - 2\theta)\sigma^2 \Delta t^2 \lambda_i}{(\sigma^2 + \theta \Delta t \lambda_i)^2} .$$

Combining the two results we find the expectation:

$$\mathbb{E}[Z_N] = -(1 - 2\theta)^2 \sigma^2 \sum_{i=1}^{N-1} \lambda_i \frac{(\Delta t \lambda_i)^2}{(\sigma^2 + \theta \Delta t \lambda_i)^2} . \quad (\text{B.28})$$

An elementary, but tedious, calculation gives the variance:

$$\begin{aligned} \text{Var}[Z_N] &= 2\sigma^4(1 - 2\theta)^2 \times \\ &\sum_{i=1}^{N-1} \Delta t \lambda_i^3 \frac{(1 - 2\theta)^2 (\Delta t \lambda_i)^3 + 4\theta^2 \sigma^2 (\Delta t \lambda_i)^2 + 8\theta \sigma^4 \Delta t \lambda_i + 4\sigma^6}{(\sigma^2 + \theta \Delta t \lambda_i)^4} . \end{aligned} \quad (\text{B.29})$$

We consider separately the following cases.

- $a < 2$ .

In this case  $\Delta t \lambda_i \rightarrow \infty$  as  $N \rightarrow \infty$ . Hence:

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbb{E}[\Delta u^3 Z_N] &= -(1 - 2\theta)^2 \sigma^2 \lim_{N \rightarrow \infty} \sum_{i=1}^{N-1} \Delta u^2 \lambda_i \frac{(\Delta t \lambda_i)^2}{(\sigma^2 + \theta \Delta t \lambda_i)^2} \Delta u = \\ &= -(1 - 2\theta)^2 \sigma^2 \int_0^U 2(1 - \cos(\pi u/U)) \frac{1}{\theta^2} du < 0 . \end{aligned}$$

For the variance we get:

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{\Delta u} \text{Var} [\Delta u^3 Z_N] &= \\ &= 2\sigma^4(1-2\theta)^2 \lim_{N \rightarrow \infty} \sum_{i=1}^{N-1} \Delta u^4 \lambda_i^2 \frac{(1-2\theta)^2 (\Delta t \lambda_i)^4 + O((\Delta t \lambda_i)^3)}{(\sigma^2 + \theta \Delta t \lambda_i)^4} \Delta u = \\ &= 2\sigma^4(1-2\theta)^2 \int_0^U 4(1 - \cos(\pi u/U))^2 \frac{(1-2\theta)^2}{\theta^4} du . \end{aligned}$$

So, in the limit the expectation of  $\Delta u^3 Z_N$  is negative and its variance is zero. Therefore:

$$\lim_{N \rightarrow \infty} A(\theta, \Delta u, \Delta t) = 2 \lim_{N \rightarrow \infty} \text{P} [\Delta u^3 Z_N > 0] = 0 . \quad (\text{B.30})$$

- $\alpha = 2$ .

Working as before we find that:

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{E} [\Delta u^3 Z_N] &= -(1-2\theta)^2 \sigma^2 \lim_{N \rightarrow \infty} \sum_{i=1}^{N-1} \Delta u^2 \lambda_i \frac{\beta^2 \Delta u^4 \lambda_i^2}{(\sigma^2 + \theta \beta \Delta u^2 \lambda_i)^2} \Delta u = \\ &= -(1-2\theta)^2 \sigma^2 \int_0^U \beta^2 \frac{8(1 - \cos(\pi u/U))^3}{(\sigma^2 + \theta \beta \cdot 2(1 - \cos(\pi u/U)))^2} du < 0 , \end{aligned}$$

and that  $\text{Var} [\Delta u^3 Z_N] / \Delta u$  converges to a finite real. So, (B.30) will apply in this case as well.

- $\alpha > 2$ .

Now,  $\Delta t \lambda_i \rightarrow 0$  as  $N \rightarrow \infty$  and:

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{E} [\Delta u^{7-2\alpha} Z_N] &= -(1-2\theta)^2 \sigma^2 \lim_{N \rightarrow \infty} \sum_{i=1}^{N-1} \beta^2 \Delta u^6 \lambda_i^3 \frac{1}{(\sigma^2 + \theta \Delta t \lambda_i)^2} \Delta u = \\ &= -(1-2\theta)^2 \sigma^2 \int_0^U \beta^2 \left(2(1 - \cos(\pi u/U))\right)^3 \frac{1}{\sigma^4} du < 0 . \end{aligned}$$

For the variance we find:

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{\Delta u^{7-3\alpha}} \text{Var} [\Delta u^{7-2\alpha} Z_N] &= \\ &= 2\sigma^4(1-2\theta)^2 \lim_{N \rightarrow \infty} \sum_{j=1}^{N-1} \beta \Delta u^6 \lambda_j^3 \frac{O(\Delta t \lambda_j) + 4\sigma^6}{(\sigma^2 + \theta \Delta t \lambda_j)^4} \Delta u = \\ &= 2\sigma^4(1-2\theta)^2 \int_0^U \beta \left(2(1 - \cos(\pi u/U))\right)^3 \frac{4}{\sigma^2} du . \end{aligned}$$

So, for  $7 - 3\alpha > 0$  the acceptance probability still converges to zero, and part b)-1) of the Proposition has been proven. For  $7 - 3\alpha = 0$  the distribution of  $\Delta u^{7-2\alpha} Z_N$

converges to a non-degenerate Gaussian distribution with mean and variance given by the above two limits. Since the limiting mean is negative, part b)-2) of the Proposition follows directly. For  $7 - 3\alpha < 0$  the variance of  $\Delta u^{7-2\alpha} Z_N$  increases without bound while the mean converges as above. Thus:

$$\lim_{N \rightarrow \infty} A(\theta, \Delta u, \Delta t) = 2 \lim_{N \rightarrow \infty} \mathbb{P} [\Delta u^{7-2\alpha} Z_N > 0] = 2 \cdot \frac{1}{2} = 1 ,$$

which gives part b)-3) of the Proposition.

*Proofs for P-MALA:*

The original P-MALA algorithm has target  $\pi \equiv \mathcal{N}(0, -\frac{\sigma^2}{\Delta u} A^{-1})$  and proposed transitions determined by the  $\theta$ -solution of (6.8), that is:

$$y - x = -(1 - \theta)\Delta t x - \theta \Delta t y + \sqrt{\frac{2\Delta t C}{\Delta u}} \xi, \quad \xi \sim \mathcal{N}(0, \mathbf{I}) .$$

Recall that  $C = -\sigma^2 A^{-1}$ . The 1-1 transformation:

$$x \mapsto \Lambda^{1/2} \sigma^{-1} \sqrt{\Delta u} Z^\top x$$

of this Markov chain corresponds to the following MCMC algorithm:

$$\begin{aligned} \text{Target:} \quad & \pi \equiv \mathcal{N}(0, \mathbf{I}) \\ \text{Proposal:} \quad & (1 + \theta\Delta t) y = (1 - (1 - \theta)\Delta t) x + \sqrt{2\Delta t} \xi , \end{aligned} \tag{B.34}$$

with  $\xi \sim \mathcal{N}(0, \mathbf{I})$ . Comparing (B.34) with (B.21) we see that the two algorithms coincide if in (B.21) we replace:

$$\Lambda \leftrightarrow \sigma^2 \mathbf{I} .$$

Thus, the acceptance probability of the P-MALA algorithm (B.34) can be analysed in terms of the equations (B.22)-(B.29) just by replacing  $\lambda_i \leftrightarrow \sigma^2$ . The expressions for the expectation and the variance of  $Z_N$  are now greatly simplified. One can easily find that  $\mathbb{E}[Z_N]$  is negative and of order  $\Delta u^{2\alpha-1}$ , and  $\text{Var}[Z_N]$  is of order  $\Delta u^{\alpha-1}$ . Then,  $\Delta u^{1-2\alpha} Z_N$  has a limiting negative mean and variance of order  $\Delta u^{1-3\alpha}$ . The required result for the various values of  $\alpha$  follows directly from the relation:

$$A(\theta, \Delta u, \Delta t) = 2 \mathbb{P} [\Delta u^{1-2\alpha} Z_N > 0] .$$

□

## References

1. Robert B. Ash. *Real analysis and probability*. Academic Press, New York, 1972. Probability and Mathematical Statistics, No. 11.
2. Alexandros Beskos, Omiros Papaspiliopoulos, Gareth O. Roberts, and Paul Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68:1–29, 2006. With discussion and a reply by the authors.
3. Siddhartha Chib, Michael Pitt, and Neil Shephard. Likelihood based inference for diffusion driven models. In preparation, 2004.
4. Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.
5. Garland B. Durham and A. Ronald Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.*, 20(3):297–338, 2002. With comments and a reply by the authors.
6. Ola Elerian, Siddhartha Chib, and Neil Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993, 2001.
7. Tadahisa Funaki. Random motion of strings and related stochastic evolution equations. *Nagoya Math. J.*, 89:129–193, 1983.
8. Tadahisa Funaki and Stefano Olla. Fluctuations for  $\nabla\phi$  interface model on a wall. *Stochastic Process. Appl.*, 94(1):1–27, 2001.
9. C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*, volume 13 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, third edition, 2004.
10. István Gyöngy. Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise. II. *Potential Anal.*, 11(1):1–37, 1999.
11. M. Hairer, A. M. Stuart, J. Voss, and P. Wiberg. Analysis of SPDEs arising in path sampling. I. The Gaussian case. *Commun. Math. Sci.*, 3(4):587–603, 2005.
12. Martin Hairer, Andrew M. Stuart, and Jochen Voss. Analysis of SPDEs arising in path sampling. Part 2: The nonlinear case. *Ann. Appl. Probab.*, 17:1657–1706, 2007.
13. Martin Hairer, Andrew M. Stuart, and Jochen Voss. Sampling the posterior: an approach to non-Gaussian data assimilation. *Phys. D*, 230:50–64, 2007.
14. P. E. Kloeden and S. Shott. Linear-implicit strong schemes for Itô-Galerkin approximations of stochastic PDEs. *J. Appl. Math. Stochastic Anal.*, 14(1):47–53, 2001. Special issue: Advances in applied stochastics.
15. Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
16. J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic*

## 30 REFERENCES

- Process. Appl.*, 101(2):185–232, 2002.
17. B. K. Øksendal. *Stochastic Differential Equations: An Introduction With Applications*. Springer-Verlag, 1998.
  18. Maria G. Reznikoff and Eric Vanden-Eijnden. Invariant measures of stochastic partial differential equations and conditioned diffusions. *C. R. Math. Acad. Sci. Paris*, 340(4):305–308, 2005.
  19. Christian P. Robert and George Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.
  20. G. O. Roberts and O. Stramer. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, 88(3):603–621, 2001.
  21. Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.*, 16(4):351–367, 2001.
  22. Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
  23. L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Ito calculus, Reprint of the second (1994) edition.
  24. Andrew M. Stuart, Jochen Voss, and Petter Wiberg. Conditional path sampling of SDEs and the Langevin MCMC method. *Commun. Math. Sci.*, 2(4):685–697, 2004.
  25. Luke Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998.
  26. David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.
  27. Lorenzo Zambotti. Integration by parts on  $\delta$ -Bessel bridges,  $\delta > 3$  and related SPDEs. *Ann. Probab.*, 31(1):323–348, 2003.

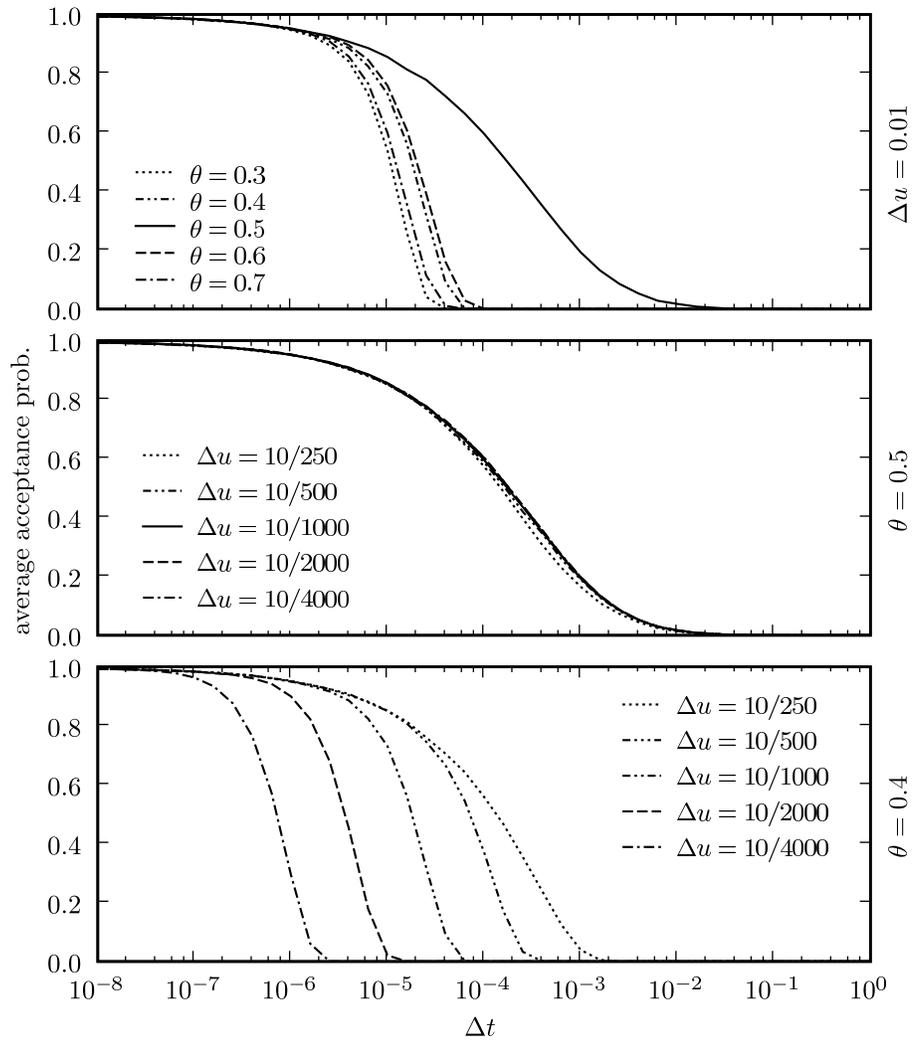


Fig. 1. MALA: Average acceptance probability in equilibrium for sampling paths of the double-well process (6.1) on  $[0, 10]$  with  $x^- = 0$ ,  $x^+ = 0$  and  $\sigma = 1$ . The diagram shows the average acceptance probability as a function of the algorithmic step size  $\Delta t$  for different values of  $\theta$  and  $\Delta u$ .

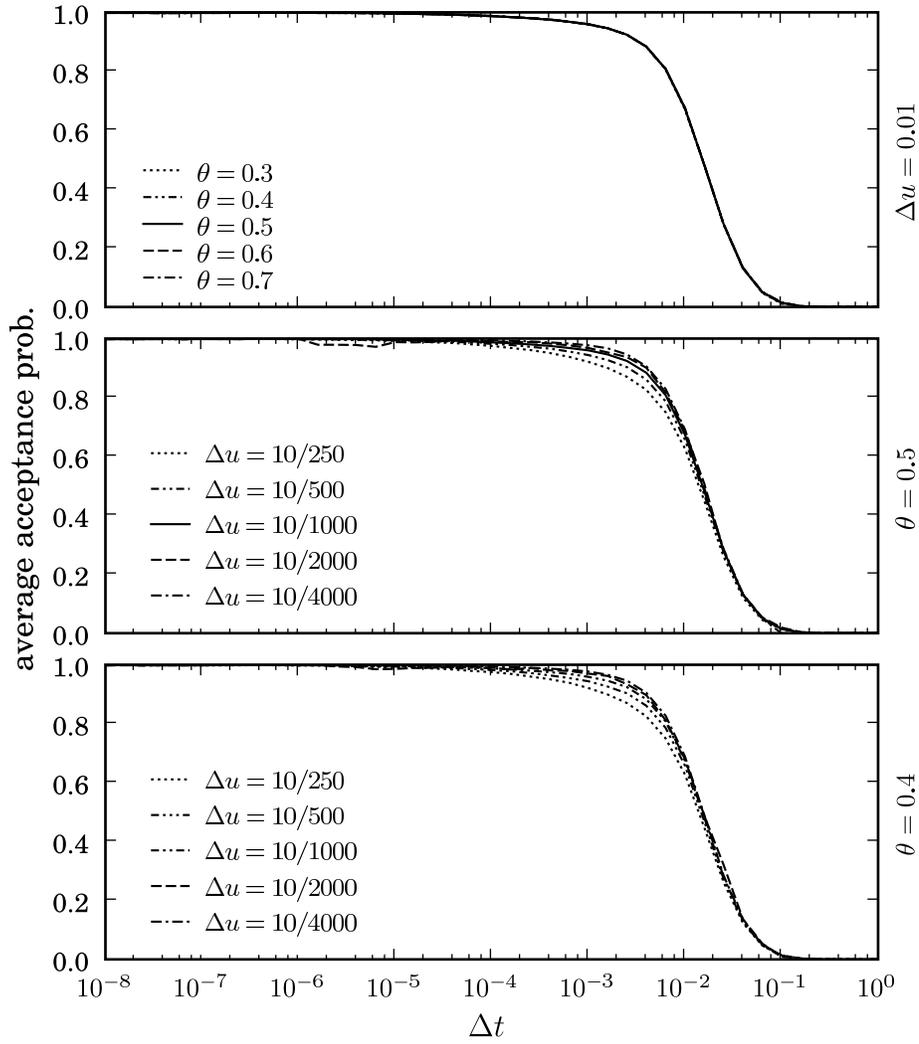


Fig. 2. P-MALA: Graphs similar to those of Fig.1, now derived using the preconditioned algorithm.

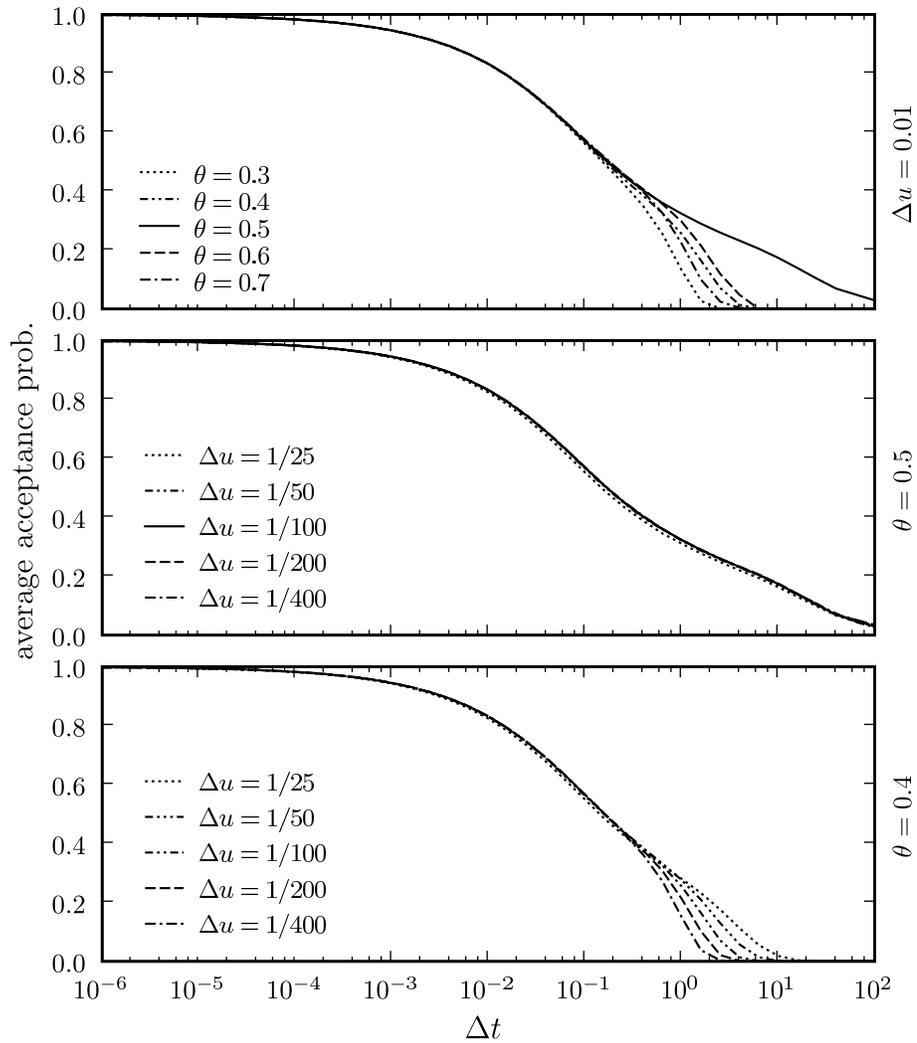


Fig. 3. P-MALA: Graphs from an application of P-MALA, similar to those of Fig.2 with the only difference that the target bridge is now considered on the unit length interval  $[0, 1]$ . Notice the difference at the scale of the  $\Delta t$ -axis compared to the previous graphs.

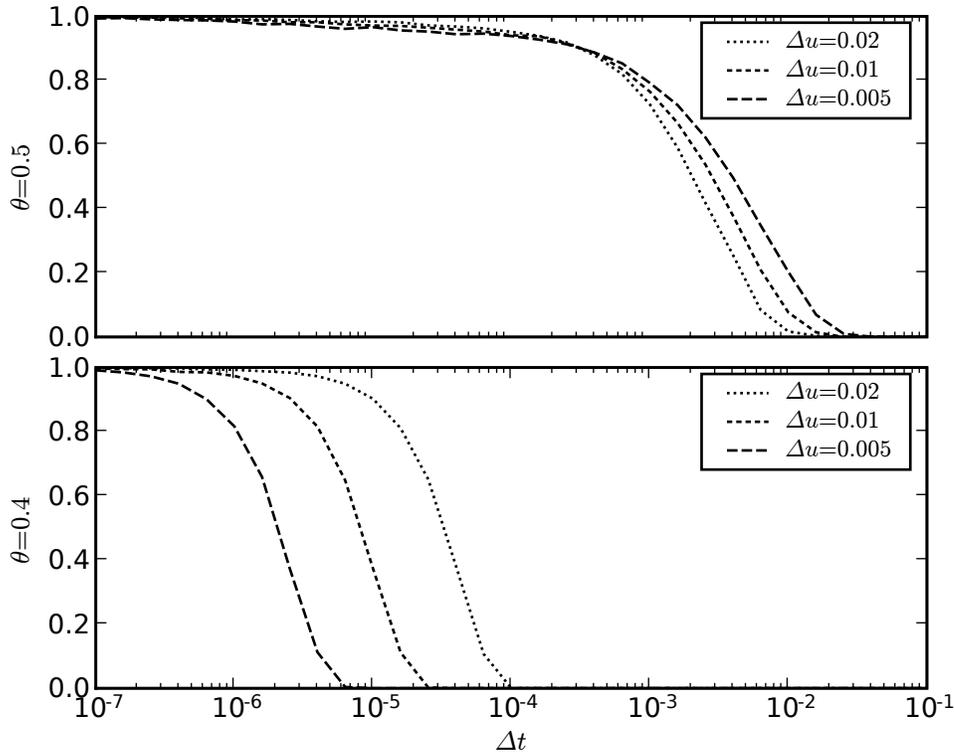


Fig. 4. This figure illustrates how the acceptance rates of the MALA algorithm for a discretised version of the smoothing problem from Section 6.3 depend on the time discretisation step size  $\Delta t$ . The different curves correspond to different space discretisations  $\Delta u$ . The upper panel gives the average acceptance probabilities in equilibrium for  $\theta = 1/2$ . In this case the Metropolis-Hastings algorithm can also be applied to the infinite dimensional problem. The lower panel illustrates the case  $\theta = 0.4$  which only makes sense for the discretised equation. One can see that the method degenerates as  $\Delta u \rightarrow 0$ .