SIAM J. SCI. COMPUT.
Vol. 44, No. 4, pp. A2089–A2120

ON IRREVERSIBLE METROPOLIS SAMPLING RELATED TO LANGEVIN DYNAMICS*

ZEXI SONG† AND ZHIQIANG TAN†

Abstract. There has been considerable interest in designing Markov chain Monte Carlo algorithms by exploiting numerical methods for Langevin dynamics, which includes Hamiltonian dynamics as a deterministic case. A prominent approach is Hamiltonian Monte Carlo (HMC), where a leapfrog discretization of Hamiltonian dynamics is employed. We investigate a recently proposed class of irreversible sampling algorithms, called Hamiltonian assisted Metropolis sampling (HAMS), which uses an augmented target density similarly as in HMC but involves a flexible proposal scheme and a carefully formulated acceptance-rejection scheme to achieve generalized reversibility. We show that as the step size tends to 0, the HAMS proposal satisfies a class of stochastic differential equations including Langevin dynamics as a special case. We provide theoretical results for HAMS, including algebraic properties of the acceptance probability, the stationary variance from the HAMS proposal, and the expected acceptance rate under a product Gaussian target distribution and the convergence rate under standard Gaussian. From these results, we derive default choices of tuning parameters for HAMS such that only the step size needs to be tuned in applications. Various relatively recent algorithms for Langevin dynamics are also shown to fall in the class of HAMS proposals up to negligible differences. Our numerical experiments on sampling high-dimensional latent variables confirm that the HAMS algorithms consistently achieve superior performance compared with several Metropolis-adjusted algorithms based on popular integrators of Langevin dynamics.

Key words. generalized reversibility, Hamiltonian Monte Carlo, Langevin dynamics, Markov chain Monte Carlo, Metropolis–Hastings sampling

MSC codes. 65C05, 60J22

DOI. 10.1137/21M1423701

1. Introduction.

Stochastic simulations are widely used in scientific computing across various fields (e.g., Brooks et al., 2011). Examples include molecular dynamics (MD) simulations in physics and chemistry and posterior simulations in Bayesian statistical analysis. For convenience, we distinguish two basic modes of stochastic simulations, although ideas from the two modes can be combined such as in Metropolis-adjusted MD algorithms (e.g., Bou-Rabee, 2014).

One mode of simulations involves generating Markov chains as numerical discretizations of continuous-time processes defined by stochastic differential equations (SDEs). In particular, consider underdamped Langevin dynamics defined by the SDE:

$$\mathrm{d}x_t = u_t\,\mathrm{d}t, \quad \mathrm{d}u_t = -\eta\, u_t\,\mathrm{d}t - \nabla U(x_t)\,\mathrm{d}t + \sqrt{2\eta}\,\mathrm{d}W_t, \tag{1}$$

where x_t is a position variable, u_t is a momentum variable, $U(x)$ is a potential function, $\eta \geq 0$ is a friction coefficient, and W_t is the standard Brownian motion. See the end of this section for our notation. The stationary distribution of (1) has the augmented density defined as

$$\pi(x,u) \propto \exp\{-H(x,u)\} = \exp\left\{-U(x) - u^{\mathrm{T}}u/2\right\}, \tag{2}$$

where $H(x,u) = U(x) + u^{\mathrm{T}}u/2$ is called the Hamiltonian. The marginal distribution of x is called the Boltzmann distribution with density $\pi(x) \propto \exp\{-U(x)\}$. The

*Submitted to the journal's Methods and Algorithms for Scientific Computing section June 1, 2021; accepted for publication (in revised form) April 4, 2022; published electronically July 26, 2022.
https://doi.org/10.1137/21M1423701

†Department of Statistics, Rutgers University; Piscataway, NJ 08854 USA (zexisong@stat.rutgers.edu, ztan@stat.rutgers.edu).

momentum u can be considered an auxiliary variable, with a standard Gaussian distribution. For simplicity, unit mass and temperature are used, and the Boltzmann constant is set to 1. Various algorithms have been proposed for Langevin dynamics in computational physics and related fields, including early development (van Gunsteren and Berendsen, 1982; Brünger, Brooks, and Karplus, 1984) and more recent contributions (Mannella, 2004; Bussi and Parrinello, 2007; Melchionna, 2007; Goga et al., 2012; Leimkuhler and Matthews, 2012; Grønbech-Jensen and Farago, 2013). For small step sizes, the Markov chains can be shown to provide pathwise accurate approximations to the solutions of (1) and be ergodic with stationary distributions which are close to $\pi(x)$. Hence these algorithms can be used to estimate both dynamic properties under Langevin SDE (1) and equilibrium expectations from Boltzmann distribution $\pi(x)$.

Another mode of simulations is Monte Carlo simulations, also known as Markov chain Monte Carlo (MCMC). The primary goal in MCMC is to sample from a target distribution defined with a density function in the form $\pi(x) \propto \exp\{-U(x)\}$ for some analytically tractable function $U(x)$, which can be interpreted as a potential function. For MCMC, Markov chains are directly constructed in discrete time such that the associated stationary distribution gives exactly the target distribution $\pi(x)$. One of the main workhorses in MCMC is Metropolis–Hastings sampling (Metropolis et al., 1953; Hastings, 1970). Given the current variable x_0, the Metropolis–Hastings algorithm generates x^* from a proposal density $Q(x^*|x_0)$ and then accepts $x_1 = x^*$ as the next variable with probability

$$\min\left\{1, \frac{\pi(x^*)Q(x_0|x^*)}{\pi(x_0)Q(x^*|x_0)}\right\} \tag{3}$$

or rejects x^* and sets $x_1 = x_0$, where $\pi(x^*)/\pi(x_0)$ can be evaluated as $\exp\{-U(x^*) + U(x_0)\}$ without requiring the normalizing constant. The update from x_0 to x_1 defines a Markov transition $K(x_1|x_0)$, depending on both the proposal density and the acceptance-rejection step, such that reversibility is satisfied: $\pi(x_0)K(x_1|x_0) = \pi(x_1)K(x_0|x_1)$. This condition is also called detailed balance in physics. As a result, the Markov chain generated by Metropolis–Hastings sampling is reversible with stationary distribution $\pi(x)$.

Metropolis–Hastings sampling provides a versatile framework, where different choices of the proposal density Q lead to different methods. For example, random walk Metropolis is obtained when the proposal x^* is generated by adding a Gaussian noise to x_0. To exploit gradient information in $\pi(x)$, various algorithms have been developed by using discretizations of physics-based SDEs as proposal schemes. For the Metropolis-adjusted Langevin algorithm (MALA) (Besag, 1994; Roberts and Tweedie, 1996), the proposal x^* is defined as Euler's discretization of the overdamped Langevin process, $\mathrm{d}x_t = -\eta^{-1}\nabla U(x_t)\,\mathrm{d}t + \sqrt{2\eta^{-1}}\,\mathrm{d}W_t$, which can be deduced from (1) in the high-friction limit with the momentum dropped out. Hamiltonian Monte Carlo (HMC) can be formulated as a Metropolis–Hastings algorithm where the proposal x^* is defined from a leapfrog discretization of Hamiltonian dynamics, (1) with $\eta = 0$, initialized by x_0 and a resampled momentum (Duane et al., 1987; Neal, 2011). For these methods, the Markov chain obtained is reversible for $\pi(x)$ in the position space.

To induce irreversible sampling, discretizations of the irreversible Langevin dynamics (1) can be used as proposal schemes, but acceptance-rejection needs to be

carefully determined, in the (x, u) space.[1] It is instructive to examine the guided Monte Carlo (GMC) algorithm (Horowitz, 1991; Bou-Rabee and Vanden-Eijnden, 2012; Ottobre et al., 2016). Each iteration in GMC can be described as follows, given the current variables (x_0, u_0):

- Sample $Z_0 \sim \mathcal{N}(\mathbf{0}, I)$, and compute $u^+ = cu_0 + \sqrt{1-c^2}Z_0$ with $c = \mathrm{e}^{-\eta\epsilon}$ and (x^*, u^-) from (x_0, u^+) by one or multiple leapfrog updates with a step size ϵ.
- Set $(x_1, u_1) = (x^*, u^-)$ with probability $\alpha = \min\{1, \pi(x^*, u^-)/\pi(x_0, u^+)\}$, or set $(x_1, u_1) = (x_0, -u^+)$ with the remaining probability.

The proposal step, $(x_0, u_0) \to (x^*, u^-)$, forms an integrator for (1) based on splitting into an Ornstein–Uhlenbeck process, $\mathrm{d}x_t = 0$ and $\mathrm{d}u_t = -\eta\, u_t\, \mathrm{d}t + \sqrt{2\eta}\, \mathrm{d}W_t$, and a Hamiltonian dynamics, $\mathrm{d}x_t = u_t\, \mathrm{d}t$ and $\mathrm{d}u_t = -\nabla U(x_t)\, \mathrm{d}t$. However, the acceptance-rejection step differs from standard Metropolis–Hastings sampling in that the acceptance probability α depends on the augmented density ratio at (x^*, u^-) versus (x_0, u^+) instead of (x_0, u_0), and the next momentum u_1 is defined as u^- upon acceptance but $-u^+$ in the case of rejection. In fact, each iteration in GMC can be decomposed into two updates, each leaving the target $\pi(x, u)$ invariant. The first is a momentum update, $(x_0, u_0) \to (x_0, u^+)$. The second is a Metropolized leapfrog integrator, $(x_0, u^+) \to (x_1, u_1)$, which can be justified as a composition of two transitions, each satisfying detailed balance for $\pi(x, u)$. Nevertheless, the entire transition kernel in GMC is irreversible for $\pi(x, u)$.[2] See Neal (2011, section 5.5.3) for related discussion.

Recently, Song and Tan (2021) proposed a new class of irreversible sampling algorithms, called Hamiltonian assisted Metropolis sampling (HAMS), using the augmented target density (2) similarly as in GMC. There are two main ingredients in HAMS, in parallel to those in Metropolis–Hastings sampling. Given the current variables (x_0, u_0), HAMS first generates (x^*, u^*) from a proposal density $Q(x^*, u^*|x_0, u_0)$, as defined by (6)–(8) later. Then HAMS performs acceptance-rejection: Set $(x_1, u_1) = (x^*, u^*)$ with acceptance probability α, and set $(x_1, u_1) = (x_0, -u_0)$ with the remaining probability, where

$$\alpha = \min\left\{1, \frac{\pi(x^*, u^*)Q(x_0, -u_0|x^*, -u^*)}{\pi(x_0, u_0)Q(x^*, u^*|x_0, u_0)}\right\}. \tag{4}$$

Compared with the usual formula (3), the momentum is negated in defining the backward proposal in acceptance probability (4). The Markov chain defined by the HAMS update is irreversible and satisfies the following generalized reversibility (or generalized detailed balance):

$$\pi(x_0, u_0)K(x_1, u_1|x_0, u_0) = \pi(x_1, u_1)K(x_0, -u_0|x_1, -u_1), \tag{5}$$

where $K(x_1, u_1|x_0, u_0)$ denotes the transition kernel for the HAMS chain, depending on both the proposal and acceptance-rejection schemes.

[1] Inserting a consistent discretization of (1) in the usual Metropolis–Hastings algorithm may lead to a high rejection rate (Scemama et al., 2006; Ma et al., 2018). Ottobre, Pillai, and Spiliopoulos (2020) studied sampling algorithms where a discretization of an irreversible SDE is combined with the Metroplis–Hastings accept-reject scheme. For a product Gaussian target, such algorithms are shown to perform differently in two asymptotic regimes.

[2] The Metropolized leapfrog can be decomposed as follows: First update $(x_0, u^+) \to (x^*, -u^-)$ or (x_0, u^+) with probability α or $1-\alpha$, and then set (x_1, u_1) by negating the momentum. The transition kernel in GMC is irreversible in (x, u), including the special cases of $c = 0$ or 1. MALA or HMC combines one or multiple leapfrog updates with momentum resampling and produces a reversible chain in the position space.

More broadly, a framework of generalized Metropolis–Hastings sampling is formulated in Song and Tan (2021) to achieve generalized reversibility, where the acceptance probability (4) is used in conjunction with a general proposal density Q, including but not restricted to the HAMS proposal. For example, the second step in GMC, i.e., the Metropolized leapfrog integrator described above, can be obtained as a special case, where the proposal Q is defined by the (deterministic) leapfrog discretization of Hamiltonian dynamics, and the ratio in Q in (4) reduces to 1 due to time reversibility of leapfrog.[3] The proposal Q can also be defined by a discretization of Langevin dynamics (1), which together with (4) leads to various Metropolis-adjusted algorithms. See supplemental section SM1 for details.

In this article, we further investigate HAMS in several directions, together with existing algorithms related to Langevin dynamics. Our main findings can be summarized as follows. First, we show that under an appropriate parametrization with a step size tending to 0, the HAMS proposal satisfies a class of SDEs which include Langevin dynamics as a special case (section 3). In fact, two specific versions of HAMS using one noise per iteration, called HAMS-A and HAMS-B in Song and Tan (2021), are associated with two extreme cases, respectively, the underdamped Langevin SDE (1) and a different SDE in which the position x appears to be dampened instead of the momentum u as in (1).

Second, we study HAMS in general with two noise vectors used per iteration from various perspectives, including the acceptance probability, the stationary variance, the expected acceptance rate under a product Gaussian target or harmonic oscillator (sections 4–6), and the convergence rate under standard Gaussian (section 7). From these studies, we derive default choices of tuning parameters for HAMS such that only the step size needs to be tuned in applications. We also identify a particular version of HAMS which uses two noise vectors per iteration and exhibits a favorable trade-off between the expected acceptance rate and the convergence rate. This algorithm, called HAMS-k, is close to HAMS-A and incorporates a small amount of friction on the position as in HAMS-B in addition to dampening the momentum. Theoretically, our analysis reveals that HAMS-k (including HAMS-A) tends to achieve higher expected acceptance rates than other algorithms with the same step size when the target distribution can be made roughly product Gaussian with many variance components near 1 after preconditioning. This provides an explanation for the superior performance of HAMS reported in Song and Tan (2021) and the simulation study here.

Third, we demonstrate that various relatively recent algorithms for Langevin dynamics can all be put into the class of HAMS proposals up to negligible differences compared with the associated leading terms of the step size (section 8). Examples include Mannella's leapfrog (Mannella, 2004; Burrage, Lenane, and Lythe, 2007), OBABO (Bussi and Parrinello, 2007; Bou-Rabee and Vanden-Eijnden, 2010), stochastic position Verlet (Melchionna, 2007), impulsive Langevin leapfrog (Goga et al., 2012), BAOAB and ABOBA (Leimkuhler and Matthews, 2012), and the Grønbech-Jensen–Farago (GJF) algorithm (Grønbech-Jensen and Farago, 2013). These results not only shed new light on the relationship between the existing algorithms but also attest to the broad scope of HAMS.

Finally, we conduct several numerical experiments (section 9) to compare different versions of HAMS and Metropolis-adjusted OBABO, BAOAB, and ABOBA

[3]The GMC algorithm itself, although leaving $\pi(x, u)$ invariant, does not satisfy generalized reversibility (except when $\eta = 0$). The symmetrized version, Metropolis-adjusted OBABO algorithm (Bussi and Parrinello, 2007), can be put in the HAMS class, satisfying generalized reversibility, as shown in Song and Tan (2021).

algorithms, which are derived by incorporating acceptance-rejection in the framework of generalized Metropolis–Hastings sampling mentioned earlier. The newly identified HAMS-k algorithms lead to the smallest errors in configurational sampling in our double well experiment. Moreover, the HAMS algorithms consistently achieve superior performance in terms of two measures of effective sample sizes in two experiments on sampling high-dimensional latent variables.

Related work. Sampling algorithms related to Langevin dynamics have been extensively studied, including those based on underdamped Langevin dynamics and Hamiltonian dynamics. We discuss directly related work to ours in addition to the earlier discussion.

For SDE-based Markov processes, introducing an appropriately chosen nonreversible component is known to improve sampling performance in the case of multivariate Gaussian equilibrium distributions (Hwang, Hwang-Ma, and Sheu, 1993; Lelièvre, Nier, and Pavliotis, 2013) and in general settings (Duncan, Lelièvre, and Pavliotis, 2016; Guillin and Monmarche, 2016; Rey-Bellet and Spiliopoulos, 2016). For example, a nonreversible perturbation of the overdamped Langevin dynamics can be obtained by adding a skew-symmetric drift term in the position space. In contrast, the underdamped Langevin dynamics are directly defined in the position-momentum space, and the evenness of the momentum distribution is crucial for achieving generalized reversibility (Song and Tan, 2021).

For Gaussian diffusions, the optimal nonreversible drift term in maximizing the spectral gap has been fully characterized by Lelièvre, Nier, and Pavliotis, (2013). See also Bou-Rabee and Eberle (2020, section 4.2). Those results are concerned with convergence of the associated SDEs under an arbitrary multivariate Gaussian target, hence mainly relevant to discretization-based algorithms as the step size tends to zero. For comparison, our study of tuning parameters to optimize convergence rates in section 7 is restricted to a multivariate standard Gaussian target but applicable to HAMS and related algorithms with a fixed step size. As expected, our results agree with those in Lelièvre, Nier, and Pavliotis, (2013) in the limit of small step sizes.

Scemama et al. (2006) and Bussi and Parrinello (2007) discussed Metropolized integrators based on Langevin dynamics to achieve generalized reversibility, which can be explicitly defined as (5) for the Markov transition kernel. The discretization schemes such as OBABO are specifically chosen with two noise vectors used per iteration to induce a proper transition density in the (x, u) space. However, our formulation of generalized Metropolis–Hastings sampling allows discretization schemes using only one noise vector, for example, to derive HAMS-A/B and Metropolis-adjusted BAOAB and ABOBA algorithms.

Bou-Rabee and Vanden-Eijnden (2010) formally studied the pathwise convergence of the Metropolis-adjusted OBABO algorithm to the solutions of SDE (1) in addition to the ergodicity with respect to the stationary distribution $\pi(x, u)$. On the other hand, Bou-Rabee and Vanden-Eijnden (2012) and Bou-Rabee (2014) mainly discussed Metropolis-adjusted MD algorithms, where the Metropolized leapfrog integrator is used as in GMC to discretize Hamiltonian dynamics in the splitting of (1) including in the OBABO splitting.

Recently, there has been considerable research on nonasymptotic analysis of Langevin-related algorithms. Examples include Dalalyan (2017) and Durmus and Moulines (2019) on the unadjusted overdamped Langevin algorithm, Mangoubi and Smith (2019), Chen and Vempala (2019), and Bou-Rabee and Eberle (2021) on unadjusted HMC, and Cheng et al. (2018), Eberle, Guillin, and Zimmer (2019), Dalalyan

and Riou-Durand (2020), and Cao, Lu, and Wang (2020) on unadjusted underdamped Langevin algorithms. In addition, Metropolis-adjusted algorithms have also been studied, including Chen et al. (2020) on HMC and Dwivedi et al. (2019) on MALA.

Organization. We provide a review of HAMS in section 2. Then we study SDE limits in section 3, algebraic properties of the acceptance probability, the stationary variance implied by the HAMS proposal, and the expected acceptance rate with acceptance-rejection under a product Gaussian target in sections 4–6, and the convergence rate quantified by the spectral radius under a standard Gaussian target in section 7. We discuss matching with existing algorithms in section 8, present numerical studies in section 9, and give a conclusion in section 10. Additional information is collected in the appendix and online supplement.

Notation. Assume that a target density $\pi(x)$ is defined on $\mathbb{R}^d$. The potential energy function $U(x)$ is defined such that $\pi(x) \propto \exp\{-U(x)\}$. Denote the gradient of $U(x)$ as $\nabla U(x)$ and Hessian $\nabla^2 U(x)$. The (multivariate) normal distribution is denoted as $\mathcal{N}(\mu, \Sigma)$ with mean μ and variance-covariance matrix (or, for short, variance matrix) Σ, and the density function as $\mathcal{N}(\cdot|\mu, \Sigma)$. A mean-zero product Gaussian distribution is defined in the form $\mathcal{N}(\mathbf{0}, \mathrm{diag}(\gamma_1^{-1}, \ldots, \gamma_d^{-1}))$, where $(\gamma_1^{-1}, \ldots, \gamma_d^{-1})$ are marginal variances. Write $\mathbf{0}$ for a vector or matrix with all 0 entries and I for an identity matrix of appropriate dimensions.

2. Review of HAMS.

We give a description of HAMS, a class of irreversible sampling algorithms in Song and Tan (2021). Throughout, we write the current variables as (x_0, u_0), a proposal as (x^*, u^*), and the next variables as (x_1, u_1) after the acceptance-rejection step.

Given the current variables (x_0, u_0), HAMS generates a proposal (x^*, u^*) as follows.

- Sample

$$Z_0 = \begin{pmatrix} Z_0^{(1)} \\ Z_0^{(2)} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, 2A - A^2), \quad \text{with } A = \begin{pmatrix} a_1 I & a_2 I \\ a_2 I & a_3 I \end{pmatrix}, \tag{6}$$

  where each I is a $k \times k$ identity matrix, with k the dimension of x, and a_1, a_2, a_3 are scalar coefficients such that $\mathbf{0} \le A \le 2I$ (in the positive semidefinite sense), hence ensuring that $2A - A^2$ is a variance-covariance matrix.
- Compute

$$\begin{pmatrix} x^* \\ u^\dagger \end{pmatrix} = \begin{pmatrix} x_0 \\ -u_0 \end{pmatrix} - A \begin{pmatrix} \nabla U(x_0) \\ -u_0 \end{pmatrix} + \begin{pmatrix} Z_0^{(1)} \\ Z_0^{(2)} \end{pmatrix}, \tag{7}$$

$$u^* = u^\dagger + \phi(x^* - x_0 - \nabla U(x^*) + \nabla U(x_0)), \tag{8}$$

  where ϕ is a scalar coefficient.

There are four tuning parameters, a_1, a_2, a_3, and ϕ. The proposal scheme (6)–(8) is derived in several steps (Song and Tan, 2021): first applying an auxiliary variable argument and an overrelaxation technique to obtain a reversible proposal, introducing negation of the momentum to obtain (7), and incorporating the new gradient $\nabla U(x^*)$ in the momentum update (8).

To describe the acceptance-rejection scheme, it is helpful to rewrite the update formulas (7) and (8) as follows:

$$\tilde{Z}^{(1)} = Z_0^{(1)} - a_1 \nabla U(x_0) + a_2 u_0, \qquad \tilde{Z}^{(2)} = Z_0^{(2)} - a_2 \nabla U(x_0) + a_3 u_0, \tag{9}$$

$$x^* = x_0 + \tilde{Z}^{(1)}, \tag{10}$$

$$u^* = -u_0 + \tilde{Z}^{(2)} + \phi(\tilde{Z}^{(1)} + \nabla U(x_0) - \nabla U(x^*)). \tag{11}$$

Equations (9)–(11) determine a forward transition from (x_0, u_0) to (x^*, u^*), depending on the noise vector Z_0. For a backward transition, we compute the new noise vector $Z^* = (\,(Z^{(1)*})^{\mathrm{T}}, (Z^{(2)*})^{\mathrm{T}}\,)^{\mathrm{T}}$, with

$$Z^{(1)*} = \tilde{Z}^{(1)} - a_1 \nabla U(x^*) - a_2 u^*, \qquad Z^{(2)*} = \tilde{Z}^{(2)} - a_2 \nabla U(x^*) - a_3 u^*. \tag{12}$$

Then as shown in Song and Tan (2021), the mapping from (x_0, u_0) to (x^*, u^*) in (9)–(11) using the noise vector Z_0 is also satisfied from $(x^*, -u^*)$ to $(x_0, -u_0)$ using the noise vector $-Z^*$ in (12). The forward and backward transitions can be illustrated as

$$\begin{pmatrix} x_0 \\ u_0 \end{pmatrix} \xrightarrow{Z_0} \begin{pmatrix} x^* \\ u^* \end{pmatrix}, \qquad \begin{pmatrix} x^* \\ -u^* \end{pmatrix} \xrightarrow{-Z^*} \begin{pmatrix} x_0 \\ -u_0 \end{pmatrix}, \tag{13}$$

where the two arrows denote the *same* mapping, depending on Z_0 or $-Z^*$.

Once (x^*, u^*) are obtained, HAMS sets the next variables $(x_1, u_1) = (x^*, u^*)$ with probability α and $(x_1, u_1) = (x_0, -u_0)$ with the remaining probability $1 - \alpha$, where

$$\begin{aligned} \alpha &= \min\left[1, \frac{\exp\{-H(x^*, u^*)\}\mathcal{N}(Z^*|\mathbf{0}, 2A - A^2)}{\exp\{-H(x_0, u_0)\}\mathcal{N}(Z_0|\mathbf{0}, 2A - A^2)}\right] \\ &= \min[1, \exp\{G(x_0, u_0, Z_0) - G(x^*, u^*, Z^*)\}] = \min[1, \exp(-\Delta G)], \end{aligned} \tag{14}$$

with $\Delta G = G(x^*, u^*, Z^*) - G(x_0, u_0, Z_0)$ and $G(x, u, Z) = H(x, u) + \frac{1}{2} Z^{\mathrm{T}}(2A - A^2)^{-1} Z$. We can view G as a generalized Hamiltonian, being analogous to the Hamiltonian H but also incorporating the noise vector Z. Notice that in the case of rejection, while the configuration x remains the current x_0, the momentum is reset by negating the current u_0.

There are two desirable properties simultaneously achieved by HAMS. First, the HAMS algorithm produces irreversible Markov chains with the augmented density $\pi(x, u)$ as a stationary distribution. In fact, HAMS can be understood as an example of generalized Metropolis–Hastings sampling as discussed in section 1. The proposal density in the forward transition is $Q(x^*, u^*|x_0, u_0) = \mathcal{N}(Z_0|\mathbf{0}, 2A - A^2)$ by the change of variables where, given (x_0, u_0), Z_0 is the noise vector transformed to (x^*, u^*) by (9)–(11) and the Jacobian determinant of the transformation from Z_0 to (x^*, u^*) is 1.[4] The proposal density in the backward transition is $Q(x_0, -u_0|x^*, -u^*) = \mathcal{N}(-Z^*|\mathbf{0}, 2A - A^2)$, because the transition from $(x^*, -u^*)$ to $(x_0, -u_0)$ is determined by the same mapping as (x_0, u_0) to (x^*, u^*), only with the noise vector Z_0 replaced by $-Z^*$, as illustrated in (13). Hence, by Proposition 3 in Song and Tan (2021), HAMS satisfies the generalized detailed balance (5) and admits $\pi(x, u)$ as a stationary distribution, where $K(x_1, u_1|x_0, u_0)$ denotes the transition kernel from (x_0, u_0) to (x_1, u_1), defined by both the proposal and acceptance-rejection schemes.

Second, the HAMS algorithm becomes rejection-free; i.e., the proposal (x^*, u^*) is always accepted when the target density $\pi(x)$ is standard Gaussian. In this case, the proposal scheme (7)–(8), with $\nabla U(x) = x$, reduces to

[4] The transformation $Z_0 \to (x^*, u^*)$ can be decomposed as $(Z_0^{(1)}, Z_0^{(2)}) \to (\tilde{Z}^{(1)}, \tilde{Z}^{(2)})$ by (9) and $(\tilde{Z}^{(1)}, \tilde{Z}^{(2)}) \to (x^*, u^*)$ by (10)–(11). The first transformation is a translation, whereas the second transformation has a Jacobian matrix which is triangular with both diagonal components 1.

$$\begin{pmatrix} x^* \\ u^* \end{pmatrix} = (I - A) \begin{pmatrix} x_0 \\ -u_0 \end{pmatrix} + \begin{pmatrix} Z_0^{(1)} \\ Z_0^{(2)} \end{pmatrix}, \tag{15}$$

which, by definition of Z_0 in (6), yields an irreversible vector autoregressive (VAR) process in (x, u) with $\mathcal{N}(\mathbf{0}, I)$ as a stationary distribution. Moreover, the acceptance probability in (14) can be directly calculated to be $\alpha \equiv 1$. A subtle point is that the rejection-free property of HAMS under a standard Gaussian target $\pi(x)$ depends on using both the proposal scheme (7)–(8) and the generalized Metropolis–Hastings probability (14) derived from (4). If the standard Metropolis–Hastings probability similar to (3) were used, HAMS would not achieve the rejection-free property under standard Gaussian; otherwise, the resulting Markov chain would be reversible, which contradicts the irreversibility of (15).

Two special cases of HAMS are further investigated in Song and Tan (2021), where the noise variance matrix $2A - A^2$ is singular (hence only a single noise vector is needed) and a specific choice of ϕ is derived. The first is called HAMS-A, where A is singular with

$$a_1 = a,\ a_3 = b,\ a_2 = \sqrt{ab},\ \phi = \sqrt{ab}/(2 - a), \tag{16}$$

subject to $a, b > 0$ and $a + b \leq 2$. The second is called HAMS-B, where $2I - A$ is singular with

$$a_1 = 2 - \tilde{a},\ a_3 = 2 - \tilde{b},\ a_2 = \sqrt{\tilde{a}\tilde{b}},\ \phi = \sqrt{\tilde{b}/\tilde{a}}, \tag{17}$$

subject to $\tilde{a}, \tilde{b} > 0$ and $\tilde{a} + \tilde{b} \leq 2$. In addition, a concrete choice of b given a in HAMS-A and that of $\tilde{b}$ given $\tilde{a}$ in HAMS-B are identified by studying the lag-1 stationary auto-covariance matrix of HAMS under a standard Gaussian target $\pi(x)$. Extensions of these choices of ϕ and (a, b) or $(\tilde{a}, \tilde{b})$ are studied in sections 4 and 7.

3. Appropriate SDE limits. The HAMS algorithm is derived from the perspective of MCMC as in Metropolis–Hastings sampling, where a discrete-time Markov chain is simulated by generating and then accepting or rejecting a proposal. In this section, we show that under an appropriate parametrization depending on a step size, HAMS leads to continuous-time limits characterized by SDEs related to Langevin dynamics.

We introduce the following parametrization of (a_1, a_2, a_3) for the matrix A in (6):

$$a_1 = 2 - c_1 \left(1 + \sqrt{1 - \epsilon^2}\right), \quad a_3 = c_2 \left(1 + \sqrt{1 - \epsilon^2}\right), \quad a_2 = \epsilon\sqrt{c_1 c_2}, \tag{18}$$

where $\epsilon \in [0, 1]$ is a step size and $c_1 \geq 0$ and $c_2 \geq 0$ are carryover coefficients. For $c_1 > 0$ and $c_2 > 0$, the constraint $\mathbf{0} \leq A \leq 2I$ is satisfied if and only if $0 < c_1 \leq 1$ and $0 < c_2 \leq 1$. To derive an SDE limit, we further impose the relationship

$$c_1 = \mathrm{e}^{-\eta_1 \epsilon/2}, \quad c_2 = \mathrm{e}^{-\eta_2 \epsilon/2}, \quad \phi = \mathcal{O}(\epsilon), \tag{19}$$

where $\eta_1 \geq 0$ and $\eta_2 \geq 0$ are friction coefficients similarly as in Langevin dynamics, and the form of ϕ can be flexible, subject to being of order $\mathcal{O}(\epsilon)$. For any fixed $\epsilon > 0$, the preceding relationship about (c_1, c_2) and (η_1, η_2) is one-to-one and hence unrestricted. However, as $\epsilon \to 0$, the relationship (19) implies that $c_1 \to 1$ and $c_2 \to 1$ for any fixed (η_1, η_2).

PROPOSITION 1. *For a target density $\pi(x)$ on $\mathbb{R}^d$, suppose that $\nabla^2 U(x)$ exists, and the spectral norm $\|\nabla^2 U(x)\|$ is bounded in x by a constant. Then the HAMS proposal (x^*, u^*) in (7)–(8) with the parametrization (18)–(19) and fixed (η_1, η_2), up to higher-order terms as $\epsilon \to 0$, is equivalent to Euler's discretization for the following SDE:*

$$\begin{pmatrix} \mathrm{d}x_t \\ \mathrm{d}u_t \end{pmatrix} = -\begin{pmatrix} \eta_1 I & -I \\ I & \eta_2 I \end{pmatrix} \begin{pmatrix} \nabla U(x_t) \\ u_t \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sqrt{2\eta_1}\, \mathrm{d}W_t^{(1)} \\ \sqrt{2\eta_2}\, \mathrm{d}W_t^{(2)} \end{pmatrix}, \tag{20}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are independent, standard Brownian motions.

Informally, we say that the HAMS proposal satisfies the SDE (20) as $\epsilon \to 0$. The preceding result can be generalized under weaker conditions on the spectral norm $\|\nabla^2 U(x)\|$. Nevertheless, the present form already gives several informative implications.

First, the SDE (20) falls into a general class of SDEs which admit the augmented density $\pi(x, u) \propto \exp\{-H(x, u)\}$ as a stationary distribution (Duncan, Lelièvre, and Pavliotis, 2016; Ma et al., 2018), where $H(x, u) = U(x) + u^{\mathrm{T}} u/2$. In fact, (20) can be put into the form

$$\mathrm{d}z_t = -(D + Q)\nabla H(z_t)\mathrm{d}t + (2D)^{1/2}\mathrm{d}W_t,$$

where $z_t = (x_t^{\mathrm{T}}, u_t^{\mathrm{T}})^{\mathrm{T}}$, $W_t = (W_t^{(1)\mathrm{T}}, W_t^{(2)\mathrm{T}})^{\mathrm{T}}$, D is a positive semidefinite matrix, and Q is a skew-symmetric matrix, defined as follows:

$$D = \begin{pmatrix} \eta_1 I & \mathbf{0} \\ \mathbf{0} & \eta_2 I \end{pmatrix}, \quad Q = \begin{pmatrix} \mathbf{0} & -I \\ I & \mathbf{0} \end{pmatrix}.$$

Moreover, by Theorem 2 in Ma et al. (2018), the Markov process (20) satisfies generalized reversibility with respect to $\pi(x, u)$, where the backward process is defined by the SDE:

$$\begin{pmatrix} \mathrm{d}x_t \\ \mathrm{d}u_t \end{pmatrix} = -\begin{pmatrix} \eta_1 I & I \\ -I & \eta_2 I \end{pmatrix} \begin{pmatrix} \nabla U(x_t) \\ u_t \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sqrt{2\eta_1}\, \mathrm{d}W_t^{(1)} \\ \sqrt{2\eta_2}\, \mathrm{d}W_t^{(2)} \end{pmatrix}. \tag{21}$$

Interestingly, the Markov process defined by (21) is equivalent to that defined by (20) but with u_t and $W_t^{(2)}$ replaced by $-u_t$ and $-W_t^{(2)}$; that is, (21) is equivalent to

$$\begin{pmatrix} \mathrm{d}x_t \\ -\mathrm{d}u_t \end{pmatrix} = -\begin{pmatrix} \eta_1 I & -I \\ I & \eta_2 I \end{pmatrix} \begin{pmatrix} \nabla U(x_t) \\ -u_t \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sqrt{2\eta_1}\, \mathrm{d}W_t^{(1)} \\ -\sqrt{2\eta_2}\, \mathrm{d}W_t^{(2)} \end{pmatrix}.$$

Hence the backward process (21) is stochastically the same as the forward process (20) except for the sign reversal of the momentum. This relationship between the forward and backward processes represents a continuous-time limit of that between the forward and backward transitions of the proposals in HAMS, as illustrated in (13). The generalized reversibility of (20) as a limit of the HAMS proposal implies that HAMS with the parametrization (18)–(19) leads to an acceptance rate which tends to 1 in the limit of $\epsilon \to 0$. Similarly as emphasized in Ma et al. (2018), the limiting acceptance rate of 1 would not be obtained if the HAMS proposal were plugged into standard Metropolis–Hastings sampling. This point also echoes the discussion after (15) about the rejection-free property of HAMS under standard Gaussian.

Second, it is interesting to examine two special cases of the SDE (20). On one hand, setting $\eta_1 = 0$ in (20) yields the underdamped Langevin dynamics, i.e., (1) with $\eta = \eta_2$:

$$\begin{pmatrix} \mathrm{d}x_t \\ \mathrm{d}u_t \end{pmatrix} = -\begin{pmatrix} \mathbf{0} & -I \\ I & \eta_2 I \end{pmatrix} \begin{pmatrix} \nabla U(x_t) \\ u_t \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \mathbf{0} \\ \sqrt{2\eta_2}\mathrm{d}W_t^{(2)} \end{pmatrix}. \tag{22}$$

On the other hand, setting $\eta_2 = 0$ in (20) leads to the SDE,

$$\begin{pmatrix} \mathrm{d}x_t \\ \mathrm{d}u_t \end{pmatrix} = -\begin{pmatrix} \eta_1 I & -I \\ I & \mathbf{0} \end{pmatrix} \begin{pmatrix} \nabla U(x_t) \\ u_t \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sqrt{2\eta_1}\mathrm{d}W_t^{(1)} \\ \mathbf{0} \end{pmatrix}. \tag{23}$$

The case $\eta_1 = \eta_2 = 0$ corresponds to the deterministic Hamiltonian dynamics. There are interesting differences between (22) and (23), in relation to the Hamiltonian dynamics.[5] The underdamped Langevin dynamics (22) is widely used to describe the motions of particles in the presence of frictions, where the momentum u is dampened by a friction force and a Gaussian white noise. By comparison, the SDE (23) is mathematically well defined, indicating that the configuration x is affected by the force field as in overdamped Langevin, in addition to the momentum. To our knowledge, the physical meaning of (23) remains to be studied.

Third, the two special cases of $\eta_1 = 0$ or $\eta_2 = 0$ are directly related to HAMS-A or HAMS-B, respectively. In fact, HAMS-A can be obtained by taking $\eta_1 = 0$ or equivalently $c_1 = 1$ in (18) and the choice of ϕ as described in (16):

$$a_1 = 1 - \sqrt{1-\epsilon^2}, \quad a_2 = \epsilon\sqrt{c_2}, \quad a_3 = c_2\left(1 + \sqrt{1-\epsilon^2}\right), \quad \phi = \frac{\epsilon\sqrt{c_2}}{1+\sqrt{1-\epsilon^2}}. \tag{24}$$

Moreover, HAMS-B can be obtained by taking $\eta_2 = 0$ or equivalently $c_2 = 1$ and the choice of ϕ as described in in (17):

$$a_1 = 2 - c_1\left(1 + \sqrt{1-\epsilon^2}\right),\ a_2 = \epsilon\sqrt{c_1},\ a_3 = 1 + \sqrt{1-\epsilon^2},\ \phi = \frac{\epsilon}{\sqrt{c_1}\left(1+\sqrt{1-\epsilon^2}\right)}. \tag{25}$$

See the appendix for simplification of the proposal scheme (6)–(8) for HAMS-A/B. We record the following consequence of Proposition 1 for HAMS-A and HAMS-B.

COROLLARY 1. *Under the setting of Proposition* 1*, in the limit of* $\epsilon \to 0$*, the HAMS-A proposal* (x^*, u^*) *defined in* (83)–(85) *with* $c_2 = \mathrm{e}^{-\eta_2\epsilon/2}$ *and fixed* η_2 *satisfies the underdamped Langevin SDE* (22) *with* $\eta_1 = 0$*, and the HAMS-B proposal* (x^*, u^*) *defined in* (86)–(88) *with* $c_1 = \mathrm{e}^{-\eta_1\epsilon/2}$ *and fixed* η_1 *satisfies the SDE* (23) *with* $\eta_2 = 0$.

The preceding result sheds new light on differences between HAMS-A and HAMS-B. The parametrization (24) for HAMS-A is the same as used in Song and Tan (2021). But the parametrization (25) for HAMS-B differs slightly from that used in Song and Tan (2021) and has the conceptual advantage of inducing the SDE (23) with $c_1 = \mathrm{e}^{-\eta_1\epsilon/2}$ and any fixed η_1 as $\epsilon \to 0$. See the supplement (section SM3.2) for further discussion.

[5] The two dynamics (22) and (23) can be distinguished as follows (Bou-Rabee and Eberle, 2020, Lemma 4.11): Consider an isotropic Gaussian target $\pi(x)$ defined as $\mathcal{N}(\mathbf{0}, \gamma^{-1}I)$. For fixed $\eta > 0$, the optimal convergence characterized by maximizing the spectral gap over (η_1, η_2) for the SDE (20) subject to $\eta_1, \eta_2 \geq 0$ and $\eta_1 + \eta_2 \leq \eta$ is achieved by $(\eta_1, \eta_2) = (0, \eta)$ if $\gamma \leq 1$ and $\eta \leq 2\gamma^{1/2}$ or $(\eta_1, \eta_2) = (\eta, 0)$ if $\gamma \geq 1$ and $\eta \leq 2\gamma^{-1/2}$. In the case $\eta > \min(2\gamma^{1/2}, 2\gamma^{-1/2})$, the optimal choices of η_1 and η_2 can be shown to be both nonzero.

4. Default choice of ϕ. The SDE (20) is informative about the behavior of HAMS with the parametrization (18)–(19) in the limit of $\epsilon \to 0$, where $\phi = \mathcal{O}(\epsilon)$ can be of a flexible form. To derive a specific choice of ϕ with fixed ϵ, we study the generalized Hamiltonian difference, ΔG, in the acceptance probability (14) under a univariate Gaussian target $\mathcal{N}(0, \gamma^{-1})$, which is called a harmonic oscillator in physics. As discussed in section 2, when the target density $\pi(x)$ is standard Gaussian, HAMS is rejection-free: The acceptance probability (14) is always 1, or equivalently ΔG is always 0. But when the target density is $\mathcal{N}(0, \gamma^{-1})$ with $\gamma \neq 1$, HAMS is no longer rejection-free. We seek a choice of ϕ such that ΔG is minimally affected as γ deviates from 1.

PROPOSITION 2. *Suppose that the target density $\pi(x)$ is $\mathcal{N}(0, \gamma^{-1})$. Then ΔG defined in* (14) *can be expressed as a quadratic form:*

$$\Delta G = \left(x_0, u_0, Z_0^{(1)}, Z_0^{(2)}\right) D(\gamma) \left(x_0, u_0, Z_0^{(1)}, Z_0^{(2)}\right)^{\mathrm{T}},$$

where $D(\gamma)$ is a 4×4 matrix. For $i, j = 1, 2, 3, 4$, the (i, j)th entry of $D(\gamma)$ is $d_{ij}(\gamma)$, a polynomial of γ, with coefficients depending on (a_1, a_2, a_3, ϕ). The coefficient of the leading term of $d_{44}(\gamma)$ is always 0. Furthermore, for any a_1, a_2, a_3 such that $\mathbf{0} \leq A \leq 2I$, the coefficients of the leading terms of $d_{11}(\gamma), d_{22}(\gamma), d_{33}(\gamma)$ are simultaneously minimized in absolute values by the choice $\phi = a_2/(2 - a_1)$.

Proposition 2 can also be extended to a product Gaussian target $\pi(x)$ with variance matrix $\mathrm{diag}(\gamma_1^{-1}, \ldots, \gamma_d^{-1})$, where ΔG is a sum of quadratic forms in the ith coordinates of $x_0, u_0, Z_0^{(1)}$ and $Z_0^{(2)}$ for $i = 1, \ldots, d$. The result gives a default choice of ϕ for HAMS in a unified manner. In the special cases of HAMS-A and HAMS-B, the choice $\phi = a_2/(2 - a_1)$ is easily seen to agree with those derived in Song and Tan (2021) as stated in (24) and (25). The derivation of ϕ in Song and Tan (2021) is similar as above for HAMS-A but involves a seemingly different angle for HAMS-B, where the choice of ϕ in (25) is unique in ensuring that the two backward noise vectors $Z_1^{(1)*}$ and $Z_2^{(2)*}$ in (12) are proportional to each other.

As another interesting consequence of the default choice of ϕ, the quantity ΔG can be substantially simplified even for a general target density $\pi(x)$.

COROLLARY 2. *Suppose that $\phi = a_2/(2 - a_1)$ is used. Then for a general target density $\pi(x)$ on $\mathbb{R}^d$, ΔG in the acceptance probability* (14) *reduces to*

(26)
$$\Delta G = U(x^*) - U(x_0) + \frac{\{\nabla U(x_0) + \nabla U(x^*)\}^{\mathrm{T}} \left[a_1\{\nabla U(x_0) + \nabla U(x^*)\} - 2\left(a_2 u_0 + Z_0^{(1)}\right)\right]}{2(2 - a_1)}.$$

Particularly, for a product Gaussian target distribution $\mathcal{N}(\mathbf{0}, \mathrm{diag}(\gamma_1^{-1}, \ldots, \gamma_d^{-1}))$, ΔG reduces to $\Delta G = \sum_{i=1}^d \Delta G_i$, where

$$\Delta G_i = \frac{a_1\gamma_i(\gamma_i - 1)}{2(2 - a_1)} \left(a_2 u_{0i} + Z_{0i}^{(1)} - a_1\gamma_i x_{0i}\right) \left(a_2 u_{0i} + Z_{0i}^{(1)} + (2 - a_1\gamma_i) x_{0i}\right), \quad (27)$$

with x_{0i}, u_{0i} and $Z_{0i}^{(1)}$ representing the ith coordinate of the vectors x_0, u_0 and $Z_0^{(1)}$.

We remark that the expressions (26) and (27) depend on only (a_1, a_2) and $(x_0, u_0, Z_0^{(1)})$ but not a_3 or $Z_0^{(2)}$, even though the proposal u^* depends on a_3 and $Z_0^{(2)}$. The expression (26) allows simple calculation of the acceptance probability (14) in implementation of HAMS. The expression (27) is instrumental for our analysis of acceptance rates in section 6.

5. Stationary variance under product Gaussian. In section 4, we derive a default choice of ϕ by exploiting an algebraic property of ΔG under a product Gaussian target. In this and the following section, we also consider a product Gaussian target distribution but turn to study stochastic properties of HAMS and understand impacts of different choices for the tuning parameters (a_1, a_2, a_3). To focus on main ideas with succinct notation, we present the discussion here for a univariate Gaussian target. The results can be easily extended to a product Gaussian target distribution.

We study the stationary variance of the HAMS proposal (6)–(8), applied iteratively without performing acceptance-rejection, under a univariate Gaussian target $\mathcal{N}(0, \gamma^{-1})$. A similar approach can be found in Burrage, Lenane, and Lythe (2007) in comparing various methods for solving the underdamped Langevin SDE (22). In this setting, the updates (6)–(8) can be equivalently written as an order-1 VAR process,

$$\begin{pmatrix} x^* \\ u^* \end{pmatrix} = \Phi \begin{pmatrix} x_0 \\ u_0 \end{pmatrix} + \zeta, \tag{28}$$

where Φ is the coefficient matrix and ζ represents noise terms independent of (x_0, u_0). Detailed expressions are given in the supplement (section SM3.5).

PROPOSITION 3. *Suppose that the target density $\pi(x)$ is $\mathcal{N}(0, \gamma^{-1})$, and the choice $\phi = a_2/(2-a_1)$ is used in HAMS. Then the stationary variance of the HAMS proposal* (6)–(8) *or equivalently the VAR process* (28) *is*

$$\mathrm{Var}\begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} \frac{a_1-2}{\gamma(a_1\gamma-2)} & 0 \\ 0 & 1 \end{pmatrix}. \tag{29}$$

It is interesting to compare the stationary variance matrix (29) with the variance matrix from the augmented density $\pi(x, u)$, that is,

$$\Gamma = \begin{pmatrix} \gamma^{-1} & 0 \\ 0 & 1 \end{pmatrix}. \tag{30}$$

At stationarity, the HAMS proposal leads to $\mathrm{Var}(u)$ and $\mathrm{Cov}(x, u)$ which are the same as the target values in (30). But $\mathrm{Var}(x)$ differs from γ^{-1} unless $\gamma = 1$ or $a_1 = 0$. The first case $\gamma = 1$ confirms the rejection-free property of HAMS under standard Gaussian, as discussed in section 2. The latter case $a_1 = 0$ is degenerate, where a_2 must also be 0 by the positive semidefiniteness of A, and hence the update of x becomes nonergodic, $x^* = x_0$.

To study the order of error in $\mathrm{Var}(x)$, we use the parameterization of a_1 in (18)–(19) and take a Taylor expansion with respect to the step size ϵ:

$$\mathrm{Var}(x) = \frac{a_1 - 2}{\gamma(a_1\gamma - 2)} = \frac{1}{\gamma} + \frac{\gamma - 1}{\gamma} \cdot \frac{\eta_1}{2}\epsilon + \frac{\gamma - 1}{\gamma} \cdot \left(\frac{1 + (\gamma - 1/2)\eta_1^2}{4}\right)\epsilon^2 + \mathcal{O}(\epsilon^3). \tag{31}$$

The leading error term in $\mathrm{Var}(x)$ is $\{(\gamma - 1)/\gamma\}\eta_1\epsilon/2$ for $\gamma \neq 1$. For HAMS-A with $\eta_1 = 0$ as in (24), this term vanishes, and the overall error is $\mathcal{O}(\epsilon^2)$. More generally, if we set $\eta_1 = k\epsilon$ for some $k \geq 0$, then the expansion (31) becomes

$$\mathrm{Var}(x) = \frac{a_1 - 2}{\gamma(a_1\gamma - 2)} = \frac{1}{\gamma} + \frac{\gamma - 1}{\gamma}\left(\frac{1}{4} + \frac{k}{2}\right)\epsilon^2 + \mathcal{O}(\epsilon^3), \tag{32}$$

and hence the overall error is also $\mathcal{O}(\epsilon^2)$. Nevertheless, for any fixed γ, the coefficient of the leading error term in (32), $\frac{\gamma-1}{\gamma}(\frac{1}{4} + \frac{k}{2})$, is minimized in absolute values by

taking $k = 0$, corresponding to HAMS-A. Therefore HAMS-A is the best within the HAMS class when considering the stationary variance in the univariate (or product) Gaussian setting.

We remark that under univariate (or product) Gaussian, several existing algorithms for simulating Langevin dynamics are known to yield the correct variance γ^{-1} for x, including Mannella's leapfrog (Mannella, 2004; Burrage, Lenane, and Lythe, 2007), the GJF algorithm (Grønbech-Jensen and Farago, 2013), and BAOAB and ABOBA (Leimkuhler and Matthews, 2012; Leimkuhler and Matthews, 2013). But the variance for u obtained from these algorithms is shown to be $(1-\gamma^{-1}\epsilon^2/4)^{-1}$, hence with error of order $\mathcal{O}(\epsilon^2)$ from the exact variance 1 for u. For harmonic oscillators, the variances of x and u are related to the configurational and kinetic temperatures (Farago, 2019); see the supplement, section SM4.1. Nevertheless, these results are primarily of interest in the MD settings where no acceptance-rejection is performed. When using HAMS or Metropolis-adjusted versions of these algorithms for sampling from the augmented density $\pi(x, u)$, the acceptance-rejection step is defined, depending on both x and u. We provide further discussion from the sampling perspective in section 6.

6. Expected acceptance rate under product Gaussian. Section 5 investigates the stationary variance under a product Gaussian target when using the HAMS proposal without acceptance-rejection. In this section, we study the expected acceptance rate of HAMS under product Gaussian while incorporating the acceptance-rejection step. Our analysis under univariate Gaussian $\mathcal{N}(0, \gamma^{-1})$ reveals that HAMS-A achieves an expected rejection rate which is the smallest by the leading term of order $O(\epsilon^3)$ among the HAMS class, in agreement with the best accuracy of the stationary variance obtained by HAMS-A without acceptance-rejection. The leading rejection rate is of lower order $O(\epsilon^{5/2})$ for Metropolized BAOAB and ABOBA or of order $O(\epsilon^3)$ for Metropolized OBABO but with a larger prefactor for γ near 1 when compared with HAMS-A. Such differences are also shown to translate into meaningful advantages of HAMS-A under product Gaussian.

First, we provide a useful result about the expected acceptance rate of HAMS under a general target density $\pi(x)$. A similar result is discussed in Neal (2011) and Calvo, Sanz-Alonso, and Sanz-Serna (2019) for HMC, which uses, as a proposal, a deterministic integrator such as the leapfrog integrator for the Hamiltonian dynamics. By comparison, the HAMS proposal (6)–(8) is a stochastic mapping, depending on a noise vector Z.

LEMMA 1. *Assume that the HAMS chain is stationary, with* $(x_0, u_0) \sim \pi(x, u)$. *Then the expected acceptance rate is*

$$\mathrm{E}[\alpha] = 2\mathrm{P}[\Delta G < 0] + \mathrm{P}[\Delta G = 0],$$

where α *is the acceptance probability defined in* (14).

The preceding result includes the term $\mathrm{P}[\Delta G = 0]$, which is important for HAMS because $\mathrm{P}[\Delta G = 0] = 1$ for HAMS under a standard Gaussian target. Such a term is absent in the related result for HMC in Calvo, Sanz-Alonso, and Sanz-Serna (2019), where the probability that the change in the Hamiltonian, ΔH, equals zero is assumed to be zero.

Next, we consider a univariate Gaussian target. The expected acceptance rate of HAMS can be monotonically linked to the expectation of the generalized Hamiltonian difference ΔG. Interestingly, Calvo, Sanz-Alonso, and Sanz-Serna (2019) shows that

under univariate Gaussian, the expected acceptance rate of HMC satisfies a similar formula as (33) below, with ΔG replaced by ΔH. However, the rescaling argument used in Calvo, Sanz-Alonso, and Sanz-Serna (2019) to directly transfer the formula from standard to nonstandard Gaussian for HMC is not applicable to HAMS, partly because (33) holds trivially for HAMS with $\alpha \equiv 1$ and $\Delta G \equiv 0$ under standard Gaussian.

PROPOSITION 4. *Suppose that the target density $\pi(x)$ is $\mathcal{N}(0, \gamma^{-1})$, and the choice $\phi = a_2/(2 - a_1)$ is used in HAMS. Then the expected acceptance rate of HAMS in stationarity is*

$$\mathrm{E}[\alpha] = 1 - \frac{2}{\pi} \arctan\left(\sqrt{\frac{\mathrm{E}[\Delta G]}{2}}\right). \tag{33}$$

To assess the order of acceptance rates, we use the parameterization of a_1 in (18)–(19) and apply the following formula of $\mathrm{E}[\Delta G]$ from the proof of Proposition 4:

$$\mathrm{E}[\Delta G] = \frac{a_1^3(\gamma - 1)^2\gamma}{2(2 - a_1)}. \tag{34}$$

For fixed $\eta_1 > 0$, a series expansion of (33) with respect to ϵ gives

$$\mathrm{E}[\alpha] = 1 - \frac{\sqrt{\gamma(\gamma-1)^2}}{\sqrt{2}\pi}\eta_1^{3/2}\epsilon^{3/2} + \frac{\sqrt{\gamma(\gamma-1)^2}}{8\sqrt{2}\pi}(\eta_1^2 - 6)\eta_1^{1/2}\epsilon^{5/2} + \mathcal{O}(\epsilon^{7/2}). \tag{35}$$

The leading error term in (35) is $\mathcal{O}(\epsilon^{3/2})$ for $\gamma \neq 1$. For HAMS-A with $\eta_1 = 0$, both the first two error terms vanish in (35), and a separate expansion of (33) shows that $\mathrm{E}[\alpha] = 1 - \mathcal{O}(\epsilon^3)$. If we set $\eta_1 = k\epsilon$ for $k \geq 0$, the expected acceptance rate is also $1 - \mathcal{O}(\epsilon^3)$.

COROLLARY 3. *With $\eta_1 = k\epsilon$ for $k \geq 0$ in the parameterization (18)–(19), the expected acceptance rate of HAMS under univariate Gaussian $\mathcal{N}(0, \gamma^{-1})$ satisfies*

$$\mathrm{E}[\alpha] = 1 - \frac{(1 + 2k)^{3/2}}{4\pi}\sqrt{\gamma(\gamma - 1)^2} \cdot \epsilon^3 + \mathcal{O}(\epsilon^5). \tag{36}$$

For any fixed γ, a smaller k leads to a smaller prefactor in the leading error term in (36), and hence a higher expected acceptance rate as $\epsilon \to 0$ under univariate Gaussian, with $k = 0$ (i.e., HAMS-A) being optimal in the HAMS class.

As mentioned in section 5, several existing algorithms lead to the correct variance for x but incur errors in the variance for u under univariate Gaussian. In the supplement, section SM1, we derive Metropolized versions of BAOAB and ABOBA (Leimkuhler and Matthews, 2012) and OBABO (Bussi and Parrinello, 2007) and then study the corresponding expected acceptance rates under univariate Gaussian. Interestingly, for each method with appropriately defined ΔG, identity (33) remains valid in relating $\mathrm{E}[\alpha]$ to $\mathrm{E}[\Delta G]$.

COROLLARY 4. *Under univariate Gaussian $\mathcal{N}(0, \gamma^{-1})$, the expected acceptance rate of Metropolis-adjusted BAOAB or ABOBA satisfies*

$$\mathrm{E}[\alpha] = 1 - \frac{\sqrt{2}}{4\pi}\gamma\sqrt{\eta} \cdot \epsilon^{5/2} + \mathcal{O}(\epsilon^{7/2}), \tag{37}$$

whereas the expected acceptance rate of Metropolis-adjusted OBABO satisfies

$$\mathrm{E}[\alpha] = 1 - \frac{\gamma^{3/2}}{4\pi} \cdot \epsilon^3 + \mathcal{O}(\epsilon^9). \tag{38}$$

The expected acceptance rate (37) for Metropolis-adjusted BAOAB or ABOBA deviates from 1 by a lower order $\mathcal{O}(\epsilon^{5/2})$, compared with $\mathcal{O}(\epsilon^3)$ for HAMS-A or HAMS with $\eta_1 = k\epsilon$. The expected acceptance rate (38) for Metropolis-adjusted OBABO deviates from 1 by a leading term of the same order $\mathcal{O}(\epsilon^3)$ as in (36) for HAMS with $\eta_1 = k\epsilon$, but the prefactor $\gamma^{3/2}/(4\pi)$ is larger than in (36) for γ near 1. The leading term in (36) reduces to 0 due to the rejection-free property of HAMS under standard Gaussian ($\gamma = 1$).

Finally, we investigate the expected acceptance rate under a product Gaussian target distribution $\mathcal{N}(\mathbf{0}, \mathrm{diag}(\gamma_1^{-1}, \ldots, \gamma_d^{-1}))$. In contrast with section 5, analysis of acceptance rates under product Gaussian is more involved than under univariate Gaussian, because acceptance-rejection is determined jointly over all coordinates. The acceptance probability of a generalized Metropolis–Hastings sampler including HAMS is $\alpha = \min(1, \exp(-\Delta G))$, where $\Delta G = \sum_{i=1}^d \Delta G_i$ and each ΔG_i depends on γ_i, respectively. For HAMS, ΔG_i is given by (27) under product Gaussian. We consider the asymptotic regime where $d \to \infty$ and $\epsilon \to 0$ as a function of d and use the following result on the limiting behavior of expected acceptance rates, which can be proved similarly as Theorem 2 in Calvo, Sanz-Alonso, and Sanz-Serna (2019). For simplicity, the dependency of quantities on the dimension d is suppressed in the notation.

LEMMA 2. *In the general scenario described above, assume that*

$$\mathrm{E}[\Delta G_i^2] = 2\mu_i + \mathcal{O}(\mu_i^2), \quad \mathrm{E}[\Delta G_i^3] = \mathcal{O}(\mu_i^2), \quad \mathrm{E}[\Delta G_i^4] = \mathcal{O}(\mu_i^2), \tag{39}$$

where $\mu_i = \mathrm{E}[\Delta G_i]$. *In addition, assume that*

$$\lim_{d\to\infty} \max_{1\le i\le d} \mathrm{E}[\Delta G_i] = 0, \tag{40}$$

and for some $\mu \in [0, \infty)$,

$$\lim_{d\to\infty} \mathrm{E}[\Delta G] = \lim_{d\to\infty} \sum_{i=1}^d \mathrm{E}[\Delta G_i] = \mu. \tag{41}$$

Then as $d \to \infty$, *the following results hold.*

- *The random variables* ΔG *in stationarity converge in distribution to* $\mathcal{N}(\mu, 2\mu)$.
- *The expected acceptance rates also converge:* $\mathrm{E}[\alpha] \to 2\Phi(-\sqrt{\mu/2})$, *where* Φ *is the cumulative distribution function of standard Gaussian.*

For HAMS, combining Lemma 2 and formula (34) for $\mathrm{E}[\Delta G_i]$ leads to Corollary 5. Condition (39) can be directly verified, whereas (40)–(41) impose conditions on $(\gamma_1, \ldots, \gamma_d)$.

COROLLARY 5. *For HAMS under product Gaussian, suppose that for some* $m, \tau \in [0, \infty)$,

$$\lim_{d\to\infty} \frac{a_1^3}{4} d = m, \quad \lim_{d\to\infty} \frac{1}{d} \sum_{i=1}^d \gamma_i(\gamma_i - 1)^2 = \tau, \quad \lim_{d\to\infty} \max_{1\le i\le d} \frac{\gamma_i(\gamma_i-1)^2}{d} = 0.$$

With $\eta_1 = k\epsilon$ *for* $k \ge 0$ *in* (18)–(19), *the first condition above is* $\lim_{d\to\infty}(k + \frac{1}{2})^3\epsilon^6 d/4 = m$. *Take* $\mu = m\tau$. *Then in stationarity as* $d \to \infty$,

$$\Delta G \to_{\mathcal{D}} \mathcal{N}(\mu, 2\mu) \quad \textit{and} \quad \mathrm{E}[\alpha] \to 2\Phi\left(-\sqrt{\mu/2}\right). \tag{42}$$

Under similar conditions, Corollary 5 can also be extended to Metropolis-adjusted BAOAB, ABOBA, and OBABO, with the corresponding definitions of μ.

COROLLARY 6. *For Metropolis-adjusted BAOAB and ABOBA under product Gaussian, suppose that for some* $m, \tau \in [0, \infty)$,

$$\lim_{d\to\infty} \frac{\eta\epsilon^5}{16} d = m, \quad \lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \gamma_i^2 = \tau, \quad \lim_{d\to\infty} \frac{\epsilon}{d}\sum_{i=1}^{d} \gamma_i^3 = 0, \quad \lim_{d\to\infty} \max_{1\le i\le d} \frac{\gamma_i^2}{d} = 0.$$

Then (42) *holds with* $\mu = m\tau$. *For Metropolis-adjusted OBABO under product Gaussian, suppose that for some* $m, \tau \in [0, \infty)$,

$$\lim_{d\to\infty} \frac{\epsilon^6}{32} d = m, \quad \lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \gamma_i^3 = \tau, \quad \lim_{d\to\infty} \max_{1\le i\le d} \frac{\gamma_i^3}{d} = 0.$$

Then (42) *holds with* $\mu = m\tau$.

The preceding results reveal an interesting comparison between HAMS and related algorithms. Using the parameterization (18)–(19) with $\eta_1 = k\epsilon$ for $k \ge 0$, the requirement $\lim_{d\to\infty} a_1^3 d/4 = m$ in Corollary 5 is satisfied by taking $\epsilon = \mathcal{O}(d^{-1/6})$. For Metropolis-adjusted OBABO, the condition $\lim_{d\to\infty} \epsilon^6 d/32 = m$ is also satisfied by taking $\epsilon = \mathcal{O}(d^{-1/6})$. However, for Metropolis-adjusted BAOAB and ABOBA, in order to achieve a nonzero limit of $\mathrm{E}[\alpha]$ in Corollary 6, we require $\epsilon = \mathcal{O}(d^{-1/5})$ which is worse than HAMS and Metropolis-adjusted OBABO. Furthermore, while the limits of the expected acceptance rate $\mathrm{E}[\alpha]$ take the same form of (42), the exact expressions of $\mu = m\tau$ differ between the aforementioned methods. The quantity τ is derived from the prefactors of the leading error terms in (36)–(38). For HAMS with $\eta_1 = k\epsilon$ including HAMS-A, τ is approximately the average of $\gamma_i(\gamma_i - 1)^2$ over the coordinates instead of direct powers of γ_i as in the other methods. Therefore, if a high proportion of γ_i's are close to 1, HAMS with $\eta_1 = k\epsilon$ tends to achieve higher expected acceptance rates than the others. This provides an explanation for superior performance of HAMS when the target distribution can be made to resemble product Gaussian with many variance components near 1 after preconditioning, as reported in Song and Tan (2021) and the simulation study later.

7. Convergence rate under standard Gaussian. In sections 5–6, we mainly study stationary properties of HAMS under product Gaussian, where the HAMS chain is assumed to be stationary without or with acceptance-rejection. In this section, we examine how the convergence rate can be optimized within the HAMS class under standard Gaussian, taken to be univariate without loss of generality. The analysis is facilitated by the fact that the HAMS chain under standard Gaussian is rejection-free and reduces to a VAR process. Our investigation reveals an interesting trade-off between different versions of HAMS and leads to a specific choice of c_2 given (ϵ, c_1) or c_1 given (ϵ, c_2), which can be used as the default choices when the target distribution can be transformed into roughly standard Gaussian after preconditioning.

Consider the standard Gaussian setting, where HAMS is rejection-free. In this case, the HAMS proposal reduces to the order-1 VAR process (15) or equivalently

$$\begin{pmatrix} x^* \\ u^* \end{pmatrix} = \Phi \begin{pmatrix} x_0 \\ u_0 \end{pmatrix} + Z_0, \tag{43}$$

where $Z_0 \sim \mathcal{N}(\mathbf{0}, 2A - A^2)$ and $\Phi = \left(\begin{smallmatrix} 1-a_1 & a_2 \\ -a_2 & a_3-1 \end{smallmatrix}\right)$, depending on $A = \left(\begin{smallmatrix} a_1 & a_2 \\ a_2 & a_3 \end{smallmatrix}\right)$ such that $\mathbf{0} \le A \le 2I$ as stated in (6). The convergence rate of the VAR process (43) is known

to be captured by the spectral radius of the coefficient matrix Φ, which is defined as the maximum modulus of its eigenvalues (Roberts and Sahu, 1997, Theorem 1). A smaller spectral radius of Φ leads to faster convergence for the VAR process (43).

If the coefficients a_1, a_2, a_3 are all free to choose, subject to $\mathbf{0} \le A \le 2I$, then the spectral radius of Φ can be made equal to 0 by setting $a_3 = a_1$ and $a_2 = \pm(1 - a_1)$ for $a_1 \in [.5, 1.5]$. The corresponding VAR process (43) can be shown to converge to $\mathcal{N}(\mathbf{0}, I)$ after 2 iterations for any initial value. However, such choices of A are incompatible with the parametrization (18)–(19), where $a_1 \to 0$ and $a_3 \to 2$ as $\epsilon \to 0$ for appropriate SDE limits.

To obtain a meaningful solution, we seek to minimize the spectral radius of Φ over possible choices of (a_2, a_3) while fixing a_1 and $\nu \equiv a_2^2/a_3$. For the parametrization (18), this program corresponds to optimizing the choice of c_2 while fixing (ϵ, c_1), which indicates that both $a_1 = 2 - c_1(1 + \sqrt{1+\epsilon^2})$ and $\nu = a_2^2/a_3 = c_1(1 - \sqrt{1-\epsilon^2})$ are fixed.

PROPOSITION 5. *Suppose that the target density $\pi(x)$ is $\mathcal{N}(0,1)$. For any fixed $0 < a_1 < 2$ and $\nu \equiv a_2^2/a_3 \ge 0$ such that $\nu \le a_1 \le 1 + \nu$, the convergence rate of the HAMS process* (43) *is optimized, or the spectral radius of Φ is minimized over (a_2, a_3) by the choices*

$$a_3^* = \left(\sqrt{\nu + 2 - a_1} - \sqrt{\nu}\right)^2, \quad a_2^* = \pm\sqrt{\nu a_3^*}, \tag{44}$$

with the minimum spectral radius given by $\frac{|a_3^* - a_1|}{2}$.

There are two constraints on the fixed values of (a_1, ν) in the preceding result. The first constraint $\nu \le a_1$ is needed to ensure $a_1 a_3 \ge a_2^2 = \nu a_3$, and hence $A \ge \mathbf{0}$. The second constraint $a_1 \le 1 + \nu$ is equivalent to requiring $c_1 \ge 1/2$ in the parametrization (18) by the expression of (a_1, ν) mentioned above. If this constraint were relaxed, then the optimal choice of (a_2, a_3) would be of a different form than (44) by extending the proof of Proposition 5. Nevertheless, the constraint $a_1 \le 1 + \nu$ or equivalently $c_1 \ge 1/2$ is automatically satisfied in HAMS-A, with $\nu = a_1$ and $c_1 = 1$. For HAMS-A, the optimal choice (44) given a_1 reduces to $a_3^* = (\sqrt{2} - \sqrt{a_1})^2$ and $a_2^* = \pm\sqrt{a_1 a_3^*}$ in agreement with Song and Tan (2021), Lemma 3. Moreover, for $c_1 = \mathrm{e}^{-\eta_1 \epsilon/2}$ in the parametrization (19) with η_1 bounded, the constraint $c_1 \ge 1/2$ is also satisfied in the practical situation where the step size ϵ is relatively small.

As motivated by the discussions in sections 5–6, we study HAMS with $\eta_1 = k\epsilon$ in (19) for some constant $k \ge 0$. Given $0 < \epsilon < 1$ and $1/2 \le c_1 \le 1$ in the parametrization (18), (a_1, ν) are fixed as mentioned above. The optimal choice of a_3 in (44) translates into

$$a_3^* = \left\{3 - \sqrt{1-\epsilon^2} - 2\sqrt{2}\epsilon\left(1 + \sqrt{1-\epsilon^2}\right)^{-1/2}\right\} c_1. \tag{45}$$

For $\eta_1 = k\epsilon$ and $c_1 = \exp(-k\epsilon^2/2)$, by the expressions of a_1 in (18) and a_3^* in (45), the minimum spectral radius of Φ can be shown as $\epsilon \to 0$ to be

$$\frac{|a_3^* - a_1|}{2} = 1 - \epsilon - k\epsilon^2 + \mathcal{O}(\epsilon^3). \tag{46}$$

From (46), a larger k corresponds to a smaller spectral radius for Φ, hence faster convergence for HAMS under standard Gaussian. By comparison, as seen from (32) and (36), a smaller k corresponds to a more accurate stationary variance without

acceptance-rejection and a higher expected acceptance rate with acceptance-rejection for HAMS when the target density is nonstandard Gaussian. Hence there exists a trade-off in the behavior of HAMS when using $\eta_1 = k\epsilon$ and the associated choice a_3^* for different values of k.

The preceding scheme of optimizing over (a_2, a_3) given (a_1, ν) extends the corresponding scheme for HAMS-A, where ν is identical to a_1 by definition. A similar extension can be obtained for HAMS-B by minimizing the spectral radius of Φ over possible choices of (a_1, a_2) while fixing a_3 and $\tilde{\nu} \equiv a_2^2/(2 - a_1)$. For the parametrization (18), this scheme corresponds to optimizing the choice of c_1 while fixing (ϵ, c_2), which leads to fixed $a_3 = c_2(1 + \sqrt{1 + \epsilon^2})$ and $\tilde{\nu} = a_2^2/(2 - a_1) = c_2(1 - \sqrt{1 - \epsilon^2})$.

PROPOSITION 6. *Suppose that the target density $\pi(x)$ is $\mathcal{N}(0, 1)$. For any fixed $0 < a_3 < 2$ and $\tilde{\nu} \equiv a_2^2/(2 - a_1) \geq 0$ such that $\tilde{\nu} \leq 2 - a_3 \leq 1 + \tilde{\nu}$, the convergence rate of the HAMS process* (43) *is optimized, or the spectral radius of Φ is minimized over (a_1, a_2) by the choices*

$$2 - a_1^* = \left(\sqrt{\tilde{\nu} + a_3} - \sqrt{\tilde{\nu}}\right)^2, \quad a_2^* = \pm\sqrt{\tilde{\nu}(2 - a_1^*)}, \tag{47}$$

with the minimum spectral radius given by $\frac{|a_3 - a_1^*|}{2}$.

For HAMS-B, defined with $c_2 = 1$ as in (25), simple calculation shows that $\tilde{\nu}$ is identical to $2 - a_3$, and hence the optimal choice (47) given a_3 reduces to $2 - a_1^* = (\sqrt{2} - \sqrt{2 - a_3})^2$ and $a_2^* = \pm\sqrt{(2 - a_3)(2 - a_1^*)}$. This result is symmetric to Song and Tan (2021), Lemma 3, where the optimal choices of (a_2, a_3) given a_1 are determined for HAMS-B by the relationship $2 - a_3 = (\sqrt{2} - \sqrt{2 - a_1})^2$. The change that a_1 is tuned given a_3, instead of a_3 given a_1, is due to the parametrization (25) used here for HAMS-B, which is slightly different from that in Song and Tan (2021) as mentioned at the end of section 3.

In connection with the SDEs in section 3, we record the implied choice of η_2 by (44) for HAMS-A with $\eta_1 = 0$ and that of η_1 by (47) for HAMS-B with $\eta_2 = 0$. The optimal choice identified as $\epsilon \to 0$ can also be obtained by specializing the general analysis of optimal convergence for linear SDEs in Lelièvre, Nier, and Pavliotis (2013) to the standard Gaussian setting (Bou-Rabee and Eberle, 2020, Theorem 4.13).[6] Our general results above are derived with any fixed step size under a standard Gaussian target and may still be useful in the case where the target is made roughly standard Gaussian after preconditioning.

COROLLARY 7. *For HAMS-A, the choice of η_2 based on* (44) *is of the order $2 + \mathcal{O}(\epsilon^2)$, and the associated HAMS-A proposal satisfies the underdamped Langevin SDE* (22) *with $\eta_2 = 2$ as $\epsilon \to 0$. For HAMS-B, the choice of η_1 based on* (47) *is of the order $2 + \mathcal{O}(\epsilon^2)$, and the associated HAMS-B proposal satisfies the SDE* (22) *with $\eta_1 = 2$ as $\epsilon \to 0$.*

For any fixed, possibly nonzero η_1, the optimal parameter identified as $\epsilon \to 0$ in Proposition 5 also agrees with that leading to the optimal convergence of SDE (20) under standard Gaussian, which can be deduced from Bou-Rabee and Eberle (2020), Lemma 4.12.

[6] The optimal convergence in Lelièvre, Nier, and Pavliotis (2013) is characterized by maximizing over η_2 the spectral gap for the Langevin SDE (22), defined as the minimum real part of the eigenvalues of the coefficient matrix for $(x_t, u_t)\mathrm{d}t$. The optimal convergence in our analysis corresponds to minimization of the spectral radius of Φ in VAR process (43), defined as the maximum modulus of the eigenvalues of Φ. The optimal choice from our analysis as $\epsilon \to 0$ agrees with that from the SDE-based analysis in the standard Gaussian setting.

COROLLARY 8. *For HAMS with fixed ϵ and η_1 in the parametrization* (18)–(19)*, the choice of η_2 based on* (44) *is of order $2+\eta_1+\mathcal{O}(\epsilon^2)$. The corresponding SDE* (20) *achieves the optimal convergence rate with respect to η_2 for the fixed η_1 under standard Gaussian.*

The SDE-based analysis in Lelièvre, Nier, and Pavliotis (2013) can also be applied with various constraints on the SDE coefficients to a multivariate Gaussian target with a known variance matrix. The optimal choice would agree with that from our analysis (as $\epsilon \to 0$) after preconditioning with the known variance matrix, provided that the same constraints can be properly incorporated. Further comparison of the two approaches can be studied in future work.

To facilitate tuning, the formulas in Propositions 5 and 6 can be used as the default choices given (ϵ, c_1) as in HAMS-A or given (ϵ, c_2) as in HAMS-B. Whenever possible, it is helpful to exploit preconditioning, that is, applying a linear transformation of x based on an approximate variance matrix such that the target density can be roughly aligned with standard Gaussian. Further discussion about preconditioning is provided in the appendix.

8. Matching with existing algorithms. The HAMS class is related to a class of SDEs (20) including underdamped Langevin in section 3 in the limit of a small step size. In this section, we show that various popular algorithms for simulating Langevin dynamics (22) can be put in the HAMS class up to negligible differences which are of higher orders of the step size than the associated leading terms. For the purpose of matching, all physical quantities such as the Boltzmann constant, the temperature, and the mass are set to 1 in the existing algorithms. In addition, to simplify the notation, the target density $\pi(x)$ or the potential function $U(x)$ is assumed to be univariate.

As the existing algorithms are conventionally used for MD simulations without acceptance-rejection, we only discuss how the proposal (x^*, u^*) is defined given the current variables (x_0, u_0). Nevertheless, as described in the supplement (section SM1), an acceptance-rejection step can be incorporated into these algorithms in the framework of generalized Metropolis–Hastings sampling (Song and Tan, 2021). The resulting sampling algorithms are used in our numerical experiments (section 9).

GJF, BAOAB, and impulsive Langevin leapfrog (IL). First, we study three relatively recent algorithms, where a single noise is used in each update. The GJF algorithm (Grønbech-Jensen and Farago, 2013) is defined as follows:

$$x^* = x_0 - \frac{\epsilon^2}{2+\eta\epsilon}\nabla U(x_0) + \frac{2\epsilon}{2+\eta\epsilon}u_0 + \frac{\epsilon}{2+\eta\epsilon}W, \tag{48}$$

$$u^* = \frac{2-\eta\epsilon}{2+\eta\epsilon}u_0 + \frac{\eta\epsilon^2-2\epsilon}{2(2+\eta\epsilon)}\nabla U(x_0) - \frac{\epsilon}{2}\nabla U(x^*) + \frac{2}{2+\eta\epsilon}W, \tag{49}$$

where $W \sim \mathcal{N}(0, 2\eta\epsilon)$. Throughout, $\eta \geq 0$ is the friction coefficient playing the role of η_2 in (22). The BAOAB algorithm (Leimkuhler and Matthews, 2012) is given by

$$\tilde{u} = u_0 - \frac{\epsilon}{2}\nabla U(x_0), \quad \tilde{x} = x_0 + \frac{\epsilon}{2}\tilde{u}, \tag{50}$$

$$\tilde{\tilde{u}} = e^{-\eta\epsilon}\tilde{u} + \sqrt{1-e^{-2\eta\epsilon}}W, \tag{51}$$

$$x^* = \tilde{x} + \frac{\epsilon}{2}\tilde{\tilde{u}}, \quad u^* = \tilde{\tilde{u}} - \frac{\epsilon}{2}\nabla U(x^*), \tag{52}$$

where $W \sim \mathcal{N}(0,1)$. The IL algorithm (Goga et al., 2012) is expressed in terms of half-step momentums, which are denoted as $u_{-\frac{1}{2}}$ and $u_{\frac{1}{2}}$. The update is given as follows:

$$\tilde{u} = u_{-\frac{1}{2}} - \epsilon \nabla U(x_0), \quad \tilde{\tilde{u}} = -\tilde{c}\tilde{u} + \sqrt{\tilde{c}(2-\tilde{c})}W, \tag{53}$$

$$x^* = x_0 + \epsilon\left(\tilde{u} + \frac{1}{2}\tilde{\tilde{u}}\right), \quad u_{\frac{1}{2}} = \tilde{u} + \tilde{\tilde{u}}, \tag{54}$$

where W is a $\mathcal{N}(0,1)$ noise and $0 \le \tilde{c} \le 1$ represents the fraction of momentum lost due to friction with $\tilde{c} = 1 - e^{-\eta\epsilon}$.

PROPOSITION 7. *Suppose that we rescale the momentum in GJF and BAOAB by*

$$u^* \leftarrow \frac{\sqrt{4-\epsilon^2}}{2}u^*, \quad u_0 \leftarrow \frac{\sqrt{4-\epsilon^2}}{2}u_0,$$

and define the full-step momentum in IL by

$$u_0 = \frac{2}{\sqrt{4-\epsilon^2}}\left(u_{-\frac{1}{2}} - \frac{\epsilon}{2}\nabla U(x_0)\right), \quad u^* = \frac{2}{\sqrt{4-\epsilon^2}}\left(u_{\frac{1}{2}} - \frac{\epsilon}{2}\nabla U(x^*)\right).$$

See the appendix for explicit expressions. Then the following results hold.

- *BAOAB and IL are equivalent to each other.*
- *GJF, BAOAB, and IL can be put exactly into the HAMS form* (6)–(8)*, with* (a_1, a_2, a_3, ϕ) *satisfying* (16) *in HAMS-A except for a difference of the order* $\mathcal{O}(\epsilon^2)$ *in* ϕ*.*

From the proof of Proposition 7, the choices of (a_1, a_2, a_3, ϕ) used to match rescaled GJF, BAO, and IL with HAMS-A are of the order

$$a_1 = \frac{\epsilon^2}{2}\left(1 - \frac{\eta\epsilon}{2}\right) + \mathcal{O}(\epsilon^4), \quad a_2 = \epsilon\left(1 - \frac{\eta\epsilon}{2}\right) + \mathcal{O}(\epsilon^3), \quad a_3 = 2\left(1 - \frac{\eta\epsilon}{2}\right) + \mathcal{O}(\epsilon^2), \tag{55}$$

$$\phi = \frac{\epsilon}{2} + \mathcal{O}(\epsilon^2). \tag{56}$$

By comparison, the SDE parametrization (24) for HAMS-A with $\eta_2 = \eta$ satisfies

$$a_1 = \frac{\epsilon^2}{2} + \mathcal{O}(\epsilon^4), \quad a_2 = \epsilon\left(1 - \frac{\eta\epsilon}{4}\right) + \mathcal{O}(\epsilon^3), \quad a_3 = 2\left(1 - \frac{\eta\epsilon}{2}\right) + \mathcal{O}(\epsilon^2), \tag{57}$$

$$\phi = \frac{\epsilon}{2} + \mathcal{O}(\epsilon^2). \tag{58}$$

Interestingly, the choices of (a_1, a_2, a_3) in (55) and (57) agree in the (first) leading terms but not in the second leading terms. This difference does not affect the convergence of all these algorithms to underdamped Langevin SDE (22) as $\epsilon \to 0$.

OBABO and Vanden-Eijnden–Ciccotti (VEC). Next we study two algorithms where two noise variables are used in each update. The OBABO algorithm (Bussi and Parrinello, 2007) is given by

$$u^+ = \sqrt{c}u_0 + \sqrt{1-c}W_1, \tag{59}$$

$$\tilde{u} = u^+ - \frac{\epsilon}{2}\nabla U(x_0), \quad x^* = x_0 + \epsilon\tilde{u}, \quad u^- = \tilde{u} - \frac{\epsilon}{2}\nabla U(x^*), \tag{60}$$

$$u^* = \sqrt{c}u^- + \sqrt{1-c}W_2, \tag{61}$$

where $0 \le c \le 1$ is the amount of momentum carryover defined as $c = e^{-\eta\epsilon}$, and $W_1, W_2 \sim \mathcal{N}(0,1)$ independently. The VEC integrator described by Equation 21 in Vanden-Eijnden and Ciccotti (2006) is

(62)
$$x^* = x_0 - \frac{\epsilon^2}{2}\nabla U(x_0) + \frac{2\epsilon - \eta\epsilon^2}{2}u_0 + \frac{\sqrt{2\eta}\epsilon^{3/2}}{2}W_1 + \frac{\sqrt{6\eta}\epsilon^{3/2}}{6}W_2,$$

(63)
$$u^* = \left(1 - \eta\epsilon + \frac{\eta^2\epsilon^2}{2}\right)u_0 + \frac{\eta\epsilon^2 - \epsilon}{2}\nabla U(x_0) - \frac{\epsilon}{2}\nabla U(x^*) + \frac{\sqrt{2\eta\epsilon}}{2}(2 - \eta\epsilon)W_1 - \frac{\sqrt{6}}{6}(\eta\epsilon)^{3/2}W_2,$$

where $W_1, W_2 \sim \mathcal{N}(0,1)$ independently.

PROPOSITION 8. *If the coefficient $\epsilon/2$ is replaced by $\epsilon/(1+\sqrt{1-\epsilon^2})$ for $\nabla U(x_0)$ and $\nabla U(x^*)$ in* (60)*, then OBABO can be put exactly into the HAMS form* (6)–(8)*, with the default choice ϕ in Proposition* 2*. If the coefficient $\frac{\eta\epsilon^2-\epsilon}{2}$ is replaced by $\frac{\eta\epsilon^2-\epsilon}{2} - \frac{\epsilon^3}{4}$ for $\nabla U(x_0)$ in* (63)*, then VEC can be matched with HAMS with default ϕ, except for differences of order $\mathcal{O}(\epsilon^2)$ in the ϕ choice and $\mathcal{O}(\epsilon^3)$ in the variances and covariance of (x^*, u^*) given (x_0, u_0). See the appendix for explicit expressions of the modified algorithms.*

From the proof of Proposition 8, the choices of (a_1, a_2, a_3, ϕ) used to match modified OBABO and VEC with HAMS satisfy the same expansions as in (55)–(56). Even though two noise variables are used per iteration, the leading terms of (a_1, a_2, a_3) satisfy $a_1 a_3 = a_2^2$, characteristic of HAMS-A, which explicitly uses only one noise variable per iteration.

Shifted HAMS. For all the methods discussed so far, including HAMS, x^* is determined using the gradient $\nabla U(x_0)$ evaluated at the current step. There exist other methods, however, which first update x_0 to some intermediate value $\tilde{x}$ and then use $\nabla U(\tilde{x})$ in the expression of x^*. In an attempt to match such methods, we introduce a variation of HAMS, called shifted HAMS, with the following update:

$$\begin{pmatrix} x^* \\ u^* \end{pmatrix} = \begin{pmatrix} x_0 \\ -u_0 \end{pmatrix} - \tilde{A}\begin{pmatrix} \nabla U(\tilde{x}) \\ -u_0 \end{pmatrix} + \begin{pmatrix} Z_0^{(1)} \\ Z_0^{(2)} \end{pmatrix}, \tag{64}$$

where $\tilde{x} = x_0 + bu_0$, $(Z_0^{(1)}, Z_0^{(2)})^{\mathrm{T}} \sim \mathcal{N}(\mathbf{0}, 2A - A^2)$, and

$$A = \begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix}, \quad \tilde{A} = A\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_1 & ba_1 + a_2 \\ a_2 & ba_2 + a_3 \end{pmatrix}.$$

Here b is a scalar tuning parameter involved to define the shifted value $\tilde{x}$ such that the gradient $\nabla U(\tilde{x})$ is used in the update instead of $\nabla U(x_0)$. Taking $b = 0$ in (64) leads back to the original HAMS update (7) before the u^* update. The coefficient matrix $\tilde{A}$ in (64) is derived to achieve the property that when the target density $\pi(x)$ is standard Gaussian with $\nabla U(x) = x$, the update (64) reduces to the same VAR process (15) as the original HAMS.

ABOBA, SPV, and Mannella's leapfrog. The ABOBA algorithm (Leimkuhler and Matthews, 2012) is defined by the following update:

$$\tilde{x} = x_0 + \frac{\epsilon}{2} u_0, \tag{65}$$

$$\tilde{u} = u_0 - \frac{\epsilon}{2}\nabla U(\tilde{x}), \quad \tilde{\tilde{u}} = e^{-\eta\epsilon}\tilde{u} + \sqrt{1 - e^{-2\eta\epsilon}}W, \quad u^* = \tilde{\tilde{u}} - \frac{\epsilon}{2}\nabla U(\tilde{x}), \tag{66}$$

$$x^* = \tilde{x} + \frac{\epsilon}{2} u^*, \tag{67}$$

where $W \sim \mathcal{N}(0, 1)$. The update in the stochastic position Verlet (SPV) algorithm (Melchionna, 2007) is given by

$$\tilde{x} = x_0 + \frac{\epsilon}{2} u_0, \tag{68}$$

$$u^* = e^{-\eta\epsilon} u_0 - \frac{1 - e^{-\eta\epsilon}}{\eta}\nabla U(\tilde{x}) + \sqrt{1 - e^{-2\eta\epsilon}}W, \tag{69}$$

$$x^* = \tilde{x} + \frac{\epsilon}{2} u^*, \tag{70}$$

where $W \sim \mathcal{N}(0, 1)$. Mannella's leapfrog (Mannella, 2004) is given by

$$\tilde{x} = x_0 + \frac{\epsilon}{2} u_0, \tag{71}$$

$$u^* = c_2 \left(c_1 u_0 - \epsilon\nabla U(\tilde{x}) + \sqrt{2\eta}W \right), \tag{72}$$

$$x^* = \tilde{x} + \frac{\epsilon}{2} u^*, \tag{73}$$

where $W \sim \mathcal{N}(0, \epsilon)$, $c_1 = \frac{2-\eta\epsilon}{2}$, and $c_2 = \frac{2}{2+\eta\epsilon}$.

PROPOSITION 9. *Suppose that the coefficient $\epsilon/2$ for u_0 and u^* is replaced by $\epsilon/(1 + \sqrt{1 - \epsilon^2})$ in* (65) *and* (67) *for ABOBA and in* (71) *and* (73) *for Mannella's leapfrog and replaced by*

$$b = \frac{2(1 - e^{-\eta\epsilon})/\eta}{1 + e^{-\eta\epsilon} + \sqrt{(1 + e^{-\eta\epsilon})^2 - 4(1 - e^{-\eta\epsilon})^2/\eta^2}}$$

in (68) *and* (70) *for SPV. Then ABOBA, SPV, and Mannella's leapfrog can be put into the form of* (64) *in shifted HAMS with suitable choices (a_1, a_2, a_3, b), except for differences of order $\mathcal{O}(\epsilon^3)$ in the variances and covariance of (x^*, u^*) given (x_0, u_0).*

From the proof of Proposition 9, the choices of (a_1, a_2, a_3) used to match the three algorithms with shifted HAMS are the same as in (55) for matching other algorithms with HAMS. Moreover, the two choices of b stated in Proposition 9 are both of the order

$$b = \frac{\epsilon}{2} + \mathcal{O}(\epsilon^3), \tag{74}$$

which shares the same leading term, $\frac{\epsilon}{2}$, as the ϕ choice in (56), although the ϵ^2 term vanishes in (74). These observations shed interesting light on the relationship among the existing algorithms in addition to their connections with HAMS.

9. Numerical experiments. We conduct numerical experiments to compare HAMS-A, HAMS-B, and HAMS with $\eta_1 = k\epsilon$, henceforth labeled as HAMS-k, for $k = 1, 2, 3$, and Metropolized versions of BAOAB, ABOBA, and OBABO, which are derived in the framework of generalized Metropolis–Hastings sampling as described in the supplement, section SM1. The target densities include a one-dimensional double well potential and two higher-dimensional latent variable distributions.

9.1. Sampling from a double well. Consider the one-dimensional double well as in Leimkuhler and Matthews (2013):

$$\pi(x) \propto \exp(-T^{-1}U(x)), \quad U(x) = (x^2-1)^2 + x,$$

where both the temperature T and the Boltzmann's constant are set to 1. See the accompanying supplemental file M142370SupMat.pdf [local/web 3.26MB]for a plot of $\pi(x)$. For a unit mass, the augmented density with the momentum u is

$$\pi(x,u) \propto \exp\left\{-T^{-1}\left(U(x) + \frac{1}{2}u^2\right)\right\}. \tag{75}$$

We compare different algorithms for sampling from (75), where acceptance-rejection is included at each iteration. The estimation error is nonmonotonic in the step size. In contrast, the experiment in Leimkuhler and Matthews (2013) is conducted in the MD setting, where every proposal is accepted, and the estimation error increases with the step size.

We follow Leimkuhler and Matthews (2013) and set the friction in underdamped Langevin to one. Thus for HAMS-A and HAMS-1/2/3, we fix $\eta_2 = 1$. For BAOAB, ABOBA, and OBABO, we fix $\eta = 1$. For HAMS-B, which is associated with SDEs with fixed $\eta_2 = 0$ (section 3), we set $\eta_1 = 1$ to reciprocate. We use 8 different step sizes starting from $\epsilon = 0.04$ and increase by 0.04 until $\epsilon = 0.32$. For each ϵ, we collect 10000 draws and repeat this process 3000 times. The starting values of x and u are randomly drawn from Uniform$[-1, 1]$.

We assess the performance using the accuracy in temperature estimates. In fact, the temperature T can be equivalently expressed as the configurational temperatures

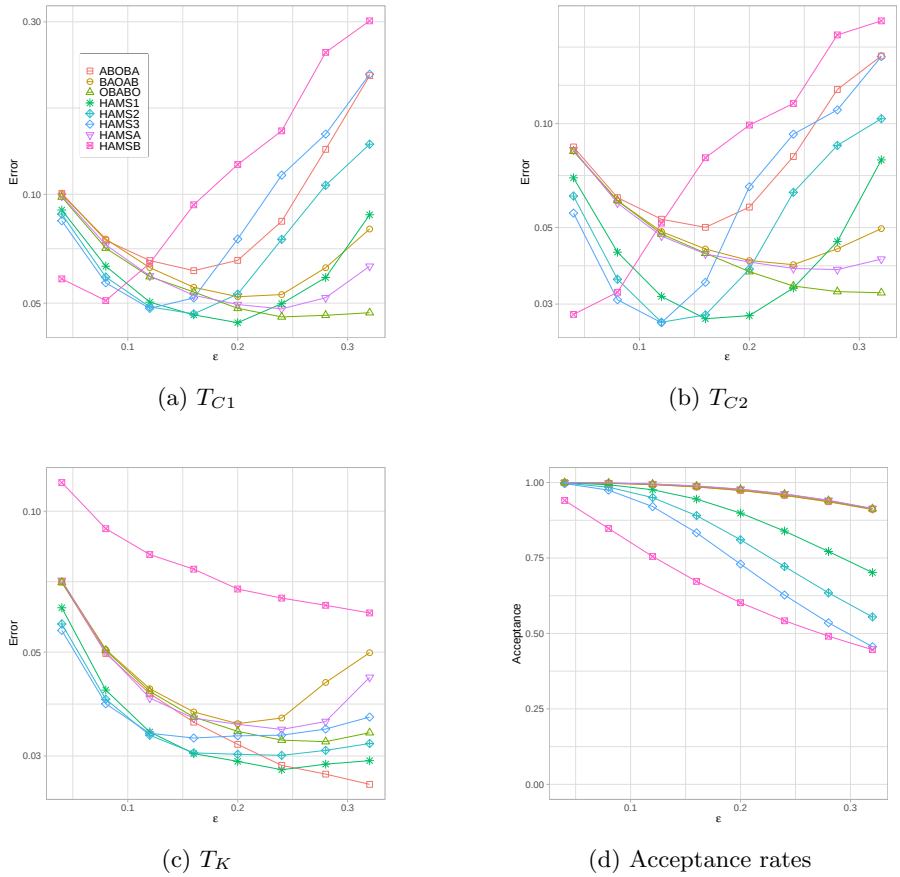
$$T_{C1} = \mathrm{E}[x \cdot \nabla U(x)], \qquad T_{C2} = \frac{\mathrm{E}[(\nabla U(x))^2]}{\mathrm{E}[\nabla^2 U(x)]}, \tag{76}$$

or as the kinetic temperature

$$T_K = \mathrm{E}[u^2]. \tag{77}$$

The expression T_{C1} is used in Leimkuhler and Matthews (2013), whereas T_{C2} is used in Farago (2019). It can be directly shown that the theoretical values of these expressions are the same as T, $T_{C1} = T_{C2} = T_K = T$ (see supplemental section SM4.1). However, due to sampling errors, the empirical estimates of these temperatures can be different. We use root mean squared errors of T_{C1}, T_{C2}, and T_K from repeated experiments as metrics. In the supplement, we also report density estimation and details of error calculation.

In Figure 1, the errors in T_{C1}, T_{C2}, and T_K are plotted on the log scale. First, we examine estimates of the configurational temperatures. There appear to be three groups among the algorithms studied. The first group is HAMS-1/2/3, which leads to smallest errors in both T_{C1} and T_{C2} among all algorithms, when $\epsilon \leq 0.16$. The performance of HAMS-k improves as k increases from 1 to 3, when ϵ is small. In the second group, the error curves of HAMS-A, BAOAB, and OBABO are comparable and those of ABOBA consistently higher. For both T_{C1} and T_{C2}, HAMS-B, in its own group, is the best for the smallest ϵ but as ϵ increases its performance quickly deteriorates. Over the whole range of ϵ, HAMS-1 has the smallest T_{C1} error, whereas both HAMS-2 and HAMS-3 reach the smallest error in T_{C2}.



(a) T_{C1} (b) T_{C2}

(c) T_K (d) Acceptance rates

FIG. 1. *Root mean squared errors in configurational temperatures and kinetic temperature, average acceptance rates for the double well. Results are based on* 3000 *repeated experiments.*

For the kinetic temperature T_K, the same three groups of algorithms emerge as above. When $\epsilon \leq 0.16$, we see two groups each with comparable performance: the first group is HAMS-1/2/3, which outperforms the second group including HAMS-A, BAOAB, ABOBA, and OBABO. As ϵ increases beyond 0.2 the two groups diverge with ABOBA achieving the smallest error overall while BAOAB producing larger errors. In its own group, HAMS-B leads to large errors in T_K, which are decreasing as ϵ increases in the range studied.

We also present the average acceptance rates in Figure 1. The algorithms can also be divided into three groups as above. For the group HAMS-A, BAOAB, ABOBA, and OBABO, which are directly related to underdamped Langevin, the acceptance rates are relatively high across the range of ϵ. For a fixed ϵ, when k increases (with $k = 0$ corresponding to HAMS-A), the acceptance rates of HAMS-k drops, which is consistent with the discussion in section 6. When achieving the smallest T_{C1} and T_{C2} errors as remarked above, HAMS-1/2/3 have much lower acceptance rates compared with the HAMS-A group at the same step sizes. The step sizes leading to the best accuracy in T_{C1} and T_{C2} are higher within the HAMS-A group than within HAMS-1/2/3. This illustrates the interesting behavior of HAMS-k, distinct from existing algorithms purely based on underdamped Langevin.

9.2. Latent variable sampling. We compare the methods by sampling latent variables in a stochastic volatility model and a log-Gaussian Cox model. In order to improve sampling efficiency, we perform preconditioning on the target densities, as described in the appendix for HAMS and in the supplement for other algorithms studied. This allows us to employ the default tuning suggested by Propositions 5 and 6. Consequently each method only depends on a single step size parameter ϵ. We tune ϵ during a burn-in period to achieve roughly 70% acceptance rates for all methods. All simulation details are provided in supplement section SM4.

To evaluate MCMC samples, a useful metric is the effective sample size (ESS), $\text{ESS} = n/\{1 + 2\sum_{l=1}^{\infty} \rho(l)\}$, where n is the total number of draws and $\rho(l)$ is the lag-l autocorrelation. We report two estimators of ESS which are both suitable for irreversible Markov chains. The first one is the Bartlett window estimator (labeled as ESS_1):

$$\text{ESS}_1 = \frac{n}{1 + 2\sum_{l=1}^{L} \left(1 - \frac{l}{L}\right) \rho(l)}, \tag{78}$$

where the cutoff value L is a large number (taken to be 3000 in our results). The second one (labeled as ESS_2) is based on the within and between variances from multiple chains in repeated simulations. Suppose that we have m Markov chains each with n draws, denoted as $\{x_{ij} : i = 1, \ldots, n, j = 1, \ldots, m\}$. Then ESS can be estimated by

$$\text{ESS}_2 = n\,\frac{W}{B}, \quad W = \frac{1}{m(n-1)} \sum_{i,j} (x_{ij} - \bar{x}_{\cdot j})^2, \quad B = \frac{n}{m-1} \sum_{j} (\bar{x}_{\cdot j} - \bar{x})^2, \tag{79}$$

where $\bar{x}_{\cdot j} = n^{-1} \sum_{i=1}^{n} x_{ij}$ and $\bar{x} = m^{-1} \sum_{j=1}^{m} \bar{x}_{\cdot j}$. In fact, B/n is an estimator of the variance of the average of n draws, whereas W is an estimator of the marginal variance of x. For relatively large m (50 in our experiments), the estimator ESS_2 can be more reliable than ESS_1 based on within-chain auto-correlations in directly measuring consistency between repeated simulations. Both ESS estimators are computed from each coordinate for a multidimensional distribution. Following Girolami and Calderhead (2011), we report the minimum ESS over all coordinates, adjusted by runtime, as a measure of computational efficiency.

Stochastic volatility model. First, consider a stochastic volatility model studied in Kim, Shephard, and Chib (1998), where latent volatilities are generated as

$$x_t = \varphi x_{t-1} + \theta_t, \quad \theta_t \sim \mathcal{N}(0, \sigma^2), \quad t = 2, 3, \ldots, T, \tag{80}$$

with $x_1 \sim \mathcal{N}(0, \sigma^2/(1 - \varphi^2))$, and the observations are generated as

$$y_t = z_t \beta \exp(x_t/2), \quad z_t \sim \mathcal{N}(0, 1), \qquad t = 1, \ldots, T. \tag{81}$$

Let $\mathbf{x} = (x_1, \ldots, x_T)^{\mathrm{T}}$ and $\mathbf{y} = (y_1, \ldots, y_T)^{\mathrm{T}}$. We generate $T = 1000$ observations from (80)–(81) using parameter values $\beta = 0.65, \sigma = 0.15$, and $\varphi = 0.98$. We fix $\mathbf{y}$ and the parameters and then sample latent variables from $p(\mathbf{x}|\mathbf{y}, \beta, \sigma, \varphi)$. See the accompanying supplemental file M142370SupMat.pdf [local/web 3.26MB]for expressions of gradients and preconditioning matrices used. All algorithms are run for 5000 burn-in iterations, and then 5000 draws are collected. Initial values of latent variables are drawn from standard normal distribution. The simulation process is repeated for 50 times.

TABLE 1

Runtime and ESS comparison for sampling latent variables in the stochastic volatility model. Results are averaged over 50 *repetitions.*

Method	Time (s)	ESS_1 (min, median, max)	$\frac{\text{minESS}_1}{\text{Time}}$	ESS_2 (min, median, max)	$\frac{\text{minESS}_2}{\text{Time}}$
HAMS-A	33.0	(2000, 3728, 7034)	60.56	(563, 1093, 2619)	17.05
HAMS-1	32.1	(2117, 3461, 6349)	65.99	(505, 1032, 2101)	15.73
HAMS-2	32.2	(1936, 3276, 5754)	60.05	(496, 1029, 2247)	15.37
HAMS-3	32.3	(2199, 3221, 6014)	68.11	(461, 988, 2301)	14.27
HAMS-B	33.4	(2301, 3487, 6890)	68.84	(501, 1058, 2997)	14.99
BAOAB	33.8	(466, 801, 1188)	13.79	(128, 235, 481)	3.81
ABOBA	34.1	(443, 756, 1143)	13.00	(132, 224, 538)	3.88
OBABO	32.8	(667, 1050, 1624)	20.31	(141, 318, 709)	4.29

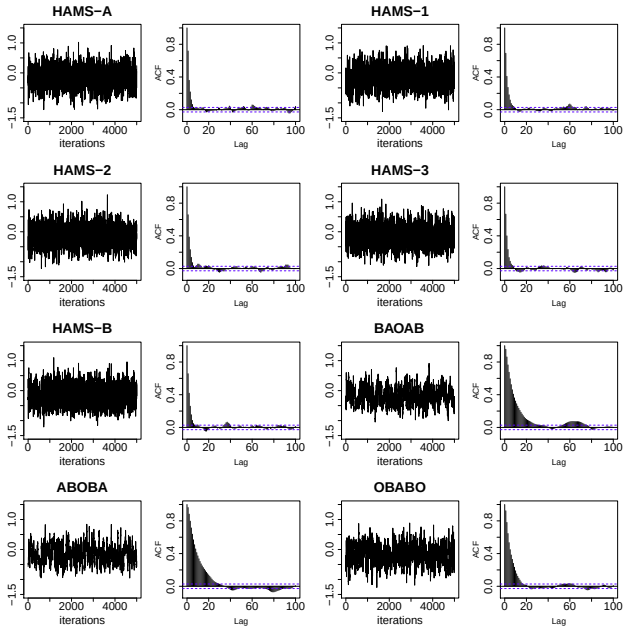


FIG. 2. *Trace and auto-correlation function plots of a latent variable from an individual run for sampling latent variables in the stochastic volatility model.*

Table 1 shows the runtime and ESS comparison. In terms of ESS_1, HAMS-B is the best, followed closely by HAMS-3 and then HAMS-1, which are slightly better than HAMS-A and HAMS-2. On the other hand, in terms of ESS_2, HAMS-A clearly leads all other methods. We also observe that the performance of HAMS-k improves as k decreases, whereas HAMS-B is in between HAMS-2 and HAMS-3. Using either ESS metric, we see that all HAMS variants are superior to BAOAB, ABOBA, and OBABO.

Trace plots in Figure 2 show that HAMS methods have much better mixing than the rest. The average sample means of latent variables are similar across all methods, as shown by Figure SM2 in the supplement. Hence it is more informative to compare the variation among repeated experiments. Figure 3 shows the sample means of latent variables after centering. A thinner spread indicates better consistency. We see that HAMS-A, HAMS-1, and HAMS-B have comparable spread, while BAOAB, ABOBA, and OBABO show much larger variation. The performance of HAMS-2/3 (omitted in Figure 3) is similar to that of HAMS-1.

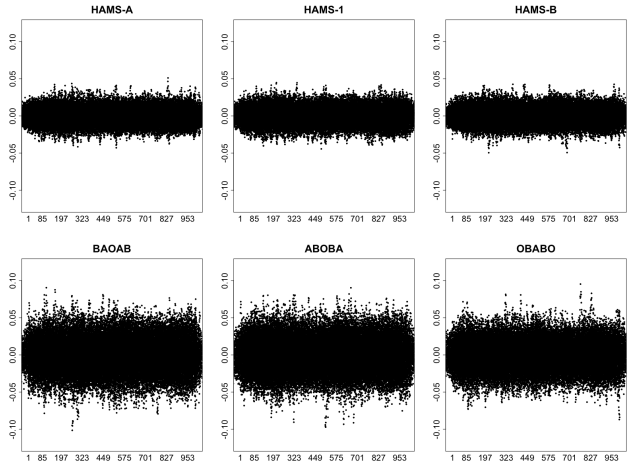


FIG. 3. *Centered plots of sample means of all* 1000 *latent variables over* 50 *repetitions for sampling latent variables in the stochastic volatility model.*

TABLE 2
Runtime and ESS comparison for sampling latent variables in the log-Gaussian Cox model. Results are averaged over 50 *repetitions.*

Method	Time (s)	ESS_1 (min, median, max)	$\frac{\text{minESS}_1}{\text{Time}}$	ESS_2 (min, median, max)	$\frac{\text{minESS}_2}{\text{Time}}$
HAMS-A	528.4	(968, 1467, 4607)	1.83	(218, 444, 1406)	0.41
HAMS-1	530.1	(665, 1142, 3118)	1.25	(175, 344, 937)	0.33
HAMS-2	530.3	(700, 1080, 2740)	1.32	(174, 323, 1011)	0.33
HAMS-3	530.3	(656, 1019, 2546)	1.24	(159, 308, 800)	0.30
HAMS-B	529.8	(606, 938, 2680)	1.14	(142, 279, 804)	0.27
BAOAB	530.5	(316, 494, 972)	0.60	(68, 144, 401)	0.13
ABOBA	536.2	(324, 478, 1080)	0.60	(76, 143, 348)	0.14
OBABO	529.4	(348, 555, 1215)	0.66	(75, 166, 420)	0.14

Log-Gaussian Cox model. Next consider a log-Gaussian Cox model, where the latent variables $\mathbf{x} = (x_{ij})_{i,j=1,\dots,m}$ are associated with an $m \times m$ grid (Christensen, Roberts, and Rosenthal, 2005). Assume that x_{ij}'s are normal with means 0 and a covariance function $C[(i,j),(i',j')] = \sigma^2 \exp(-\sqrt{(i-i')^2 + (j-j')^2}/(m\beta))$. By abuse of notation, we denote $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, C)$ of dimension $n = m^2$. The observations $(y_{ij})_{i,j=1,\dots,m}$ are independently Poisson, where the mean of y_{ij} is $\lambda_{ij} = n^{-1}\exp(x_{ij} + \mu)$, with μ treated as known. The density of latent variables given response $\mathbf{y}$ is

$$p(\mathbf{x}|\mathbf{y}, \beta, \sigma^2, \mu) \propto \exp\left\{-\frac{1}{2}x^{\mathrm{T}}C^{-1}x\right\} \exp\left\{\sum_{i,j}(y_{ij}(x_{ij}+\mu) - \lambda_{ij})\right\}. \tag{82}$$

We take $m = 64$ and generate $n = 64^2 = 4096$ observations using the parameter values $\sigma^2 = 1.91$, $\beta = 1/33$, and $\mu = \log(126) - 0.955$. We fix the simulated $\mathbf{y}$ values and the parameters, and then sample latent variables $\mathbf{x}$ from the density (82). All algorithms are run for 5000 burn-in iterations, and then 5000 draws are collected. We initialize the latent variables from a standard normal distribution. The simulation process is repeated for 50 times.

From Table 2, we see that for the Cox model, HAMS-A is the best in both ESS_1 and ESS_2. In terms of ESS_2, we see that similarly as in Table 1, HAMS-k becomes worse as k increases. However, this does not hold true for ESS_1, where HAMS-2 is slightly better than HAMS-1. Among the HAMS methods, HAMS-B has the lowest ESS in this case. Similarly to the stochastic volatility results, all three non-HAMS methods show inferior performance.

According to trace plots in Figure 4, HAMS methods mix better than the other methods. Furthermore, the auto-correlation function (ACF) of HAMS-A has the fastest decay. The average sample means of latent variables are also aligned across different methods (see Figure SM3 in the supplement). From the centered sample

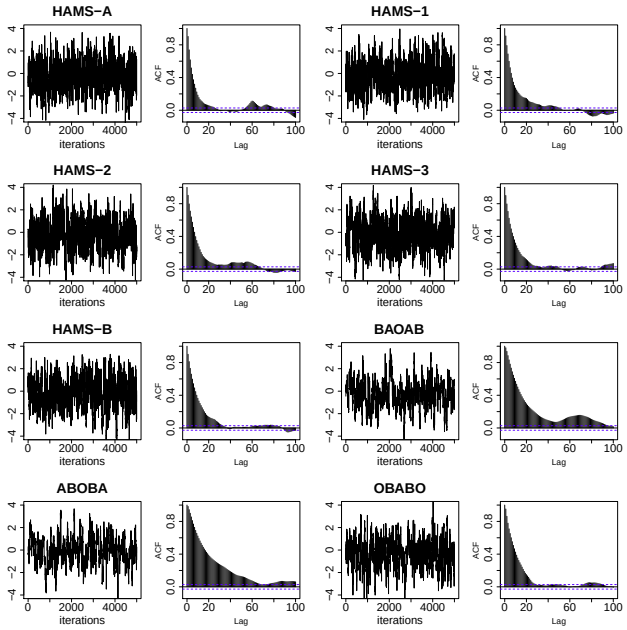


FIG. 4. *Trace and ACF plots of a latent variable from an individual run for sampling latent variables in the log-Gaussian Cox model.*

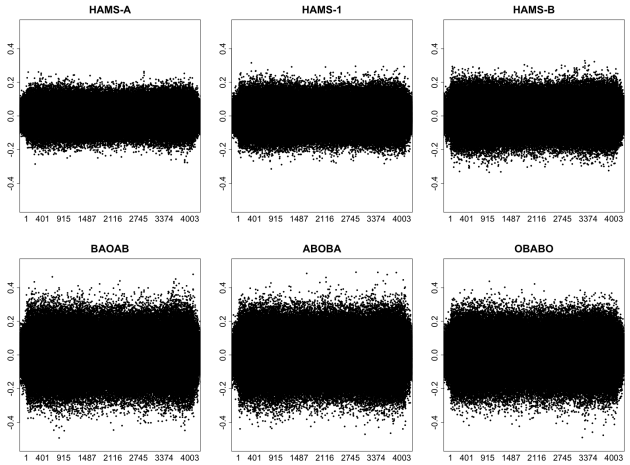


FIG. 5. *Centered plots of sample means of all* 4096 *latent variables over* 50 *repetitions for sampling latent variables in the stochastic volatility model.*

means in Figure 5, we see that HAMS-A has a slight advantage over HAMS-1 and HAMS-B. The three remaining methods are clearly less consistent than HAMS. The performance of HAMS-2/3 (omitted in Figure 5) is similar to that of HAMS-1.

10. Conclusion. We investigate HAMS in several directions, including deriving SDE limits, studying theoretical properties under the product or standard Gaussian setting, and establishing connections to existing algorithms for Langevin dynamics. Further work is needed to study possible extensions of our results to non-Gaussian target distributions, for example, properties from the default choice of ϕ. Recently, convergence properties of underdamped Langevin dynamics and Euler's discretizations have been obtained under general settings (Cheng et al., 2018; Dalalyan and Riou-Durand, 2020; Cao, Lu, and Wang, 2020). It is interesting to study the impact of using improved discretizations including HAMS and those in section 8 and that of incorporating acceptance-rejection. Moreover, investigation of HAMS and related methods is desired in simulation settings more diverse than our current experiments.

11. Appendix.

11.1. Proposal schemes of HAMS-A/B. For the choice (24) for HAMS-A, the proposal scheme (6)–(8) can be simplified to

$$\tilde{u} = \sqrt{c_2}u_0 - \frac{\epsilon}{1+\sqrt{1-\epsilon^2}}\nabla U(x_0) + Z, \tag{83}$$

$$x^* = x_0 + \epsilon\tilde{u}, \tag{84}$$

$$u^* = -u_0 + 2\sqrt{c_2}\tilde{u} + \frac{\epsilon\sqrt{c_2}}{1+\sqrt{1-\epsilon^2}}\left(\nabla U(x_0) - \nabla U(x^*)\right), \tag{85}$$

with $Z \sim \mathcal{N}(\mathbf{0}, (1-c_2)I)$. For the choice (25) for HAMS-B, the proposal scheme (6)–(8) can be simplified to

$$\tilde{u} = \sqrt{c_1}u_0 - \frac{2 - c_1\left(1+\sqrt{1-\epsilon^2}\right)}{\epsilon}\nabla U(x_0) + \frac{\sqrt{c_1}\left(1+\sqrt{1-\epsilon^2}\right)}{\epsilon}Z, \tag{86}$$

$$x^* = x_0 + \epsilon\tilde{u}, \tag{87}$$

$$u^* = u_0 - \frac{\epsilon}{\sqrt{c_1}\left(1+\sqrt{1-\epsilon^2}\right)}\left(\nabla U(x_0) + \nabla U(x^*)\right), \tag{88}$$

with $Z \sim \mathcal{N}(\mathbf{0}, (1-c_1)I)$. Taking $c_2 = 1$ in (83)–(85) or $c_1 = 1$ in (86)–(88) yields

$$\tilde{u} = u_0 - \frac{\epsilon}{1+\sqrt{1-\epsilon^2}}\nabla U(x_0), \quad x^* = x_0 + \epsilon\tilde{u}, \quad u^* = \tilde{u} - \frac{\epsilon}{1+\sqrt{1-\epsilon^2}}\nabla U(x^*),$$

which is the same as the leapfrog discretization of the Hamiltonian dynamics, except with step size $\epsilon/(1+\sqrt{1-\epsilon^2})$ instead of $\epsilon/2$ for momentum updates.

11.2. Preconditioning. We present a preconditioned HAMS algorithm in Algorithm 1. Let $\hat{\Sigma}$ be an approximation of $\mathrm{Var}(x)$, and take the Cholesky decomposition $\hat{\Sigma}^{-1} = LL^{\mathrm{T}}$ where L is lower triangular. By preconditioning, we apply HAMS to the target density of the transformed variable $\hat{x} = L^{\mathrm{T}}x$, while keeping the momentum variable $u \sim \mathcal{N}(\mathbf{0}, I)$. Algorithm 1 is formulated similarly as the preconditioned HAMS-A/B in Song and Tan (2021) to minimize the number of matrix-by-vector manipulations per iteration for efficient implementation.

11.3. Modified algorithms for matching. We first state the modified algorithms studied in Propositions 7–8 for matching with HAMS. The rescaled GJF update is

Algorithm 1. Preconditioned HAMS

Initialize $x_0, u_0, \hat{x}_0 = L^{\mathrm{T}} x_0$ and $\nabla U(\hat{x}_0) = L^{-1}\nabla U(x_0)$.
for $t = 0, 1, 2, \ldots, N_{iter}$ **do**
  Sample $w \sim \text{Uniform}[0,1]$ and $\binom{Z_0^{(1)}}{Z_0^{(2)}} \sim \mathcal{N}(\mathbf{0}, 2A - A^2)$
  $\xi = a_2 u_t + Z_0^{(1)}, \quad \hat{x}^* = \hat{x}_t - a_1 \nabla U(\hat{x}_t) + \xi$
  Propose $x^* = (L^{\mathrm{T}})^{-1}\hat{x}^*$
  $\nabla U(\hat{x}^*) = L^{-1}\nabla U(x^*), \quad \tilde{\xi} = \nabla U(\hat{x}^*) + \nabla U(\hat{x}_t)$
  $\rho = \exp\{U(x_t) - U(x^*) + \frac{1}{2-a_1}(\tilde{\xi})^{\mathrm{T}}(\xi - \frac{a_1}{2}\tilde{\xi})\}$
  if $w < \min(1, \rho)$ **then**
    $x_{t+1} = x^*, \quad \hat{x}_{t+1} = \hat{x}^*, \quad \nabla U(\hat{x}_{t+1}) = \nabla U(\hat{x}^*)$ # Accept
    $u_{t+1} = \frac{a_1 + a_2^2 + 2a_3 - a_1 a_3 - 2}{2 - a_1} u_t - \frac{a_2}{2-a_1}\tilde{\xi} + \frac{a_2}{2-a_1} Z_0^{(1)} + Z_0^{(2)}$
  else
    $x_{t+1} = x_t, u_{t+1} = -u_t, \hat{x}_{t+1} = \hat{x}_t, \nabla U(\hat{x}_{t+1}) = \nabla U(\hat{x}_t)$ # Reject

$$x^* = x_0 - \frac{\epsilon^2}{2+\eta\epsilon}\nabla U(x_0) + \frac{\epsilon\sqrt{4-\epsilon^2}}{2+\eta\epsilon}u_0 + \frac{\epsilon}{2+\eta\epsilon}W,$$
$$u^* = \frac{2-\eta\epsilon}{2+\eta\epsilon}u_0 + \frac{\eta\epsilon^2 - 2\epsilon}{\sqrt{4-\epsilon^2}(2+\eta\epsilon)}\nabla U(x_0) - \frac{\epsilon}{\sqrt{4-\epsilon^2}}\nabla U(x^*) + \frac{4}{\sqrt{4-\epsilon^2}(2+\eta\epsilon)}W,$$

where $W \sim \mathcal{N}(0, 2\eta\epsilon)$. The rescaled BAOAB update is

$$\tilde{u} = u_0 - \frac{\epsilon}{\sqrt{4-\epsilon^2}}\nabla U(x_0), \quad \tilde{x} = x_0 + \frac{\epsilon\sqrt{4-\epsilon^2}}{4}\tilde{u}, \quad \tilde{\tilde{u}} = e^{-\eta\epsilon}\tilde{u} + \sqrt{\frac{1-e^{-2\eta\epsilon}}{4-\epsilon^2}}W,$$
$$x^* = \tilde{x} + \frac{\epsilon\sqrt{4-\epsilon^2}}{4}\tilde{\tilde{u}}, \quad u^* = \tilde{\tilde{u}} - \frac{\epsilon}{\sqrt{4-\epsilon^2}}\nabla U(x^*),$$

which can be written more succinctly as

$$x^* = x_0 - \frac{\epsilon^2}{4}(1+e^{-\eta\epsilon})\nabla U(x_0) + \frac{\epsilon\sqrt{4-\epsilon^2}}{4}(1+e^{-\eta\epsilon})u_0 + \frac{\epsilon\sqrt{1-e^{-2\eta\epsilon}}}{2}W,$$
$$u^* = e^{-\eta\epsilon}u_0 - \frac{\epsilon e^{-\eta\epsilon}}{\sqrt{4-\epsilon^2}}\nabla U(x_0) - \frac{\epsilon}{\sqrt{4-\epsilon^2}}\nabla U(x^*) + 2\sqrt{\frac{1-e^{-2\eta\epsilon}}{4-\epsilon^2}}W,$$

where $W \sim \mathcal{N}(0,1)$. With the full-step momentum in Proposition 7, the IL update can be shown to be equivalent to rescaled BAOAB (see the supplement, section SM3.13). The rescaled OBABO update is

$$u^+ = \sqrt{c}u_0 + \sqrt{1-c}W_1, \quad \tilde{u} = u^+ - \frac{\epsilon}{1+\sqrt{1-\epsilon^2}}, \nabla U(x_0), \quad x^* = x_0 + \epsilon\tilde{u},$$
$$u^- = \tilde{u} - \frac{\epsilon}{1+\sqrt{1-\epsilon^2}}\nabla U(x^*), \quad u^* = \sqrt{c}u^- + \sqrt{1-c}W_2,$$

where $c = e^{-\eta\epsilon}$ and $W_1, W_2 \sim \mathcal{N}(0,1)$ independently. The modified VEC update is

$$x^* = x_0 - \frac{\epsilon^2}{2}\nabla U(x_0) + \frac{2\epsilon - \eta\epsilon^2}{2}u_0 + \frac{\sqrt{2\eta}\epsilon^{3/2}}{2}W_1 + \frac{\sqrt{6\eta}\epsilon^{3/2}}{6}W_2,$$
$$u^* = \frac{2 - 2\eta\epsilon + \eta^2\epsilon^2}{2}u_0 + \frac{\epsilon^3 + 2\eta\epsilon^2 - 2\epsilon}{4}\nabla U(x_0) - \frac{\epsilon}{2}\nabla U(x^*)$$
$$+ \frac{\sqrt{2\eta\epsilon}}{2}(2-\eta\epsilon)W_1 - \frac{\sqrt{6}}{6}(\eta\epsilon)^{3/2}W_2,$$

where $W_1, W_2 \sim \mathcal{N}(0,1)$ independently.

Next we state the modified algorithms studied in Proposition 9 for matching with shifted HAMS. The modified ABOBA update is

$$\tilde{x} = x_0 + bu_0,\ \tilde{u} = u_0 - \frac{\epsilon}{2}\nabla U(\tilde{x}),\ \tilde{\tilde{u}} = e^{-\eta\epsilon}\tilde{u} + \sqrt{1-e^{-2\eta\epsilon}}W,\ u^*$$
$$= \tilde{\tilde{u}} - \frac{\epsilon}{2}\nabla U(\tilde{x}),\ x^* = \tilde{x} + bu^*,$$

where $W \sim \mathcal{N}(0,1)$ and $b = \frac{\epsilon}{1+\sqrt{1-\epsilon^2}}$. The modified SPV update is

$$\tilde{x} = x_0 + bu_0,\ \ u^* = e^{-\eta\epsilon}u_0 - \frac{1-e^{-\eta\epsilon}}{\eta}\nabla U(\tilde{x}) + \sqrt{1-e^{-2\eta\epsilon}}W,\ \ x^* = \tilde{x} + bu^*,$$

where $W \sim \mathcal{N}(0,1)$ and b is defined in Proposition 9. The modified Mannella's leapfrog is

$$\tilde{x} = x_0 + bu_0,\ \ u^* = c_2\left(c_1u_0 - \epsilon\nabla U(\tilde{x}) + \sqrt{2\eta}W\right),\ \ x^* = \tilde{x} + bu^*,$$

where $W \sim \mathcal{N}(0,\epsilon)$, $c_1 = \frac{2-\eta\epsilon}{2}$, $c_2 = \frac{2}{2+\eta\epsilon}$, and $b = \frac{\epsilon}{1+\sqrt{1-\epsilon^2}}$.

REFERENCES

J. E. Besag, *Comments on "Representations of knowledge in complex systems" by U. Grenander and M.I. Miller*, J. R. Stat. Soc. Ser. B, 56 (1994), pp. 591–592.

N. Bou-Rabee, *Time integrators for molecular dynamics*, Entropy, 16 (2014), pp. 138–162.

N. Bou-Rabee and A. Eberle, *Markov chain Monte Carlo methods*, Lecture Notes in Hausdorff School on MCMC: Recent Developments and New Connections, 2020.

N. Bou-Rabee and A. Eberle, *Mixing Time Guarantees for Unadjusted Hamiltonian Monte Carlo*, preprint, arXiv:2105.00887, 2021.

N. Bou-Rabee and Vanden-Eijnden, *Pathwise accuracy and ergodicity of Metropolized integrators for SDEs*, Comm. Pure Appl. Math., 63 (2010), pp. 655–696.

N. Bou-Rabee and E. Vanden-Eijnden, *A patch that imparts unconditional stability to explicit integrators for Langevin-like equations*, J. Comput. Phys., 231 (2012), pp. 2565–2580.

S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, CRC Press, Boca Raton, FL, 2011.

A. Brünger, C. L. Brooks, and M. Karplus, *Stochastic boundary conditions for molecular dynamics simulations of ST2 water*, Chem. Phys. Lett., 105 (1984), pp. 495–500.

K. Burrage, I. Lenane, and G. Lythe, *Numerical methods for second-order stochastic differential equations*, SIAM J. Sci. Comput., 29 (2007), pp. 245–264.

G. Bussi and M. Parrinello, *Accurate sampling using Langevin dynamics*, Phys. Rev. E, 75 (2007), 056707.

M. P. Calvo, D. Sanz-Alonso, and J. M. Sanz-Serna, *HMC: Avoiding Rejections by Not Using Leapfrog and Some Results on the Acceptance Rate*, preprint, arXiv:1912.03253, 2019.

Y. Cao, J. Lu, and L. Wang, *On Explicit L^2-Convergence Rate Estimate for Underdamped Langevin Dynamics*, preprint, arXiv:1908.04746, 2020.

Y. Chen, R. Dwivedi, M. J. Wainwright, and B. Yu, *Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients*, J. Mach. Learn. Res., 21 (2020), pp. 1–63.

Z. Chen, and S. S. Vempala, *Optimal Convergence Rate of Hamiltonian Monte Carlo for Strongly Logconcave Distributions*, preprint, arXiv:1905.02313, 2019.

X. Cheng, N. S. Chatterji, P. L. Bartlett, and M. I. Jordan, *Underdamped Langevin MCMC: A non-asymptotic analysis*, in Proceedings of the 31st Conference On Learning Theory, Stockholm, Sweeden, 2018, volume 75, pp. 300–323.

O. F. Christensen, G. O. Roberts, and J. S. Rosenthal, *Scaling limits for the transient phase of local Metropolis–Hastings algorithms*, J. R. Stat. Soc. Ser. B, 67 (2005), pp. 253–268.

A. S. Dalalyan, *Theoretical guarantees for approximate sampling from smooth and log-concave densities*, J. R. Stat. Soc. Ser. B, 79 (2017), pp. 651–676.

A. S. Dalalyan and L. Riou-Durand, *On sampling from a log-concave density using kinetic Langevin diffusions*, Bernoulli, 26 (2020), pp. 1956–1988.

S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B, 195 (1987), pp. 216–222.

A. B. Duncan, T. Lelièvre, and G. A. Pavliotis, *Variance reduction using nonreversible Langevin samplers*, J. Stat. Phys., 163 (2016), pp. 457–491.

A. Durmus and E. Moulines, *High-dimensional Bayesian inference via the unadjusted Langevin algorithm*, Bernoulli, 25 (2019), pp. 2854–2882.
R. Dwivedi, Y. Chen, M. J. Wainwright, and B. Yu, *Log-concave sampling: Metropolis-Hastings algorithms are fast*, J. Mach. Learn. Res., 20 (2019), pp. 1–42.
A. Eberle, A. Guillin, and R. Zimmer, *Couplings and quantitative contraction rates for Langevin dynamics*, Ann. Probab., 47 (2019), pp. 1982–2010.
O. Farago, *Langevin thermostat for robust configurational and kinetic sampling*, Phys. A, 534 (2019), 122210.
M. Girolami and B. Calderhead, *Riemann manifold Langevin and Hamiltonian Monte Carlo methods*, J. R. Stat. Soc. Ser. B, 73 (2011), pp. 123–214.
N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink, and H. J. C. Berendsen, *Efficient algorithms for Langevin and DPD dynamics*, J. Chem. Theory Comput., 8 (2012), pp. 3637–3649.
N. Grønbech-Jensen and O. Farago, *A simple and effective Verlet-type algorithm for simulating Langevin dynamics*, Mol. Phys., 111 (2013), pp. 983–991.
A. Guillin and P. Monmarche, *Optimal linear drift for the speed of convergence of an hypoelliptic diffusion*, Electron. Commun. Probab., 21 (2016), pp. 1–14.
W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika, 57 (1970), pp. 97–109.
A. M. Horowitz, *A generalized guided Monte Carlo algorithm*, Phys. Lett. B, 268 (1991), pp. 247–252.
C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu, *Accelerating Gaussian Diffusions*, Ann. Appl. Probab., 3 (1993), pp. 897–913.
S. Kim, N. Shephard, and S. Chib, *Stochastic volatility: Likelihood inference and comparison with ARCH models*, Rev. Econ. Stud., 65 (1998), pp. 361–393.
B. Leimkuhler and C. Matthews, *Rational construction of stochastic numerical methods for molecular sampling*, Appl. Math. Res. Express, 2013 (2012), pp. 34–56.
B. Leimkuhler and C. Matthews, *Robust and efficient configurational molecular sampling via Langevin dynamics*, J. Chem. Phys., 138 (2013), 174102.
T. Lelièvre, F. Nier, and G. A. Pavliotis, *Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion*, J. Stat. Phys., 152 (2013), pp. 237–274.
Y.-A. Ma, E. Fox, T. Chen, and L. Wu, *Irreversible samplers from jump and continuous Markov processes*, Stat. Comput., 29 (2018), pp. 177–202.
O. Mangoubi and A. Smith, *Mixing of Hamiltonian Monte Carlo on strongly log-concave distributions 2: Numerical integrators*, in Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS), Naha, Japan, 2019, pp. 586–595.
R. Mannella, *Quasisymplectic integrators for stochastic differential equations*, Phys. Rev. E, 69 (2004), 041107.
S. Melchionna, *Design of quasisymplectic propagators for Langevin dynamics*, J. Chem. Phys., 127 (2007), 044108.
N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys., 21 (1953), pp. 1087–1092.
R. M. Neal, *MCMC using Hamiltonian dynamics*, in Handbook of Markov Chain Monte Carlo, S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, eds., CRC Press, Boca Raton, FL, 2011.
M. Ottobre, N. S. Pillai, F. J. Pinski, and A. M. Stuart, *A function space HMC algorithm with second order Langevin diffusion limit*, Bernoulli, 22 (2016), pp. 60–106.
M. Ottobre, N. S. Pillai, and K. Spiliopoulos, *Optimal scaling of the MALA algorithm with irreversible proposals for Gaussian targets*, Stoch. Partial Differ. Equ. Anal. Comput., 8 (2020), pp. 311–361.
L. Rey-Bellet and K. Spiliopoulos, *Improving the convergence of reversible samplers*, J. Stat. Phys., 164 (2016), pp. 472–494.
G. O. Roberts and S. K. Sahu, *Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler*, J. R. Stat. Soc. Ser. B, 59 (1997), pp. 291–317.
G. O. Roberts and R. L. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernoulli, 2 (1996), pp. 341–363.
A. Scemama, T. Lelièvre, G. Stoltz, E. Cancès, and M. Caffarel, *An efficient sampling algorithm for variational Monte Carlo*, J. Chem. Phys., 125 (2006), 114105.
Z. Song and Z. Tan, *Hamiltonian assisted Metropolis sampling*, J. Amer. Statist. Assoc., (2021), https://doi.org/10.1080/01621459.2021.1982723.
W. van Gunsteren and H. Berendsen, *Algorithms for Brownian dynamics*, Mol. Phys., 45 (1982), pp. 637–647.
E. Vanden-Eijnden and G. Ciccotti, *Second-order integrators for Langevin equations with holonomic constraints*, Chem. Phys. Lett., 429 (2006), pp. 310–316.