Efficient numerical algorithms for the generalized Langevin equation

Leimkuhler, Benedict; Sachs, Matthias

DOI:
10.1137/20M138497X

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Leimkuhler, B & Sachs, M 2022, 'Efficient numerical algorithms for the generalized Langevin equation', *SIAM Journal on Scientific Computing*, vol. 44, no. 1, pp. A364-A388. https://doi.org/10.1137/20M138497X

SIAM J. SCI. COMPUT.
Vol. 44, No. 1, pp. A364–A388

EFFICIENT NUMERICAL ALGORITHMS FOR THE GENERALIZED LANGEVIN EQUATION*

BENEDICT LEIMKUHLER† AND MATTHIAS SACHS‡

Abstract. We study the design and implementation of numerical methods to solve the generalized Langevin equation (GLE) focusing on the sampling properties of the numerical integrators. For this purpose, we cast the GLE in an extended phase space formulation and derive a family of splitting methods that generalize existing Langevin dynamics integration methods. We show exponential convergence in law and the validity of a central limit theorem for the Markov chains obtained via these integration methods, we show that a suggested integration scheme is consistent with asymptotic limits of the exact dynamics and can reproduce (in the short memory limit) a superconvergence property for the analogous splitting of underdamped Langevin dynamics. We then apply our proposed integration method to several model systems, including a Bayesian inference problem. We demonstrate in numerical experiments that our method outperforms other proposed GLE integration schemes in terms of the accuracy of sampling. Moreover, using a parameterization of the memory kernel in the GLE as proposed by Ceriotti, Bussi, and Parrinello *Phys. Rev. Lett.*, 6 (2010), pp. 1170–1180, our experiments indicate that the obtained GLE-based sampling scheme can, in some cases, outperform state-of-the-art sampling schemes based on underdamped Langevin dynamics in terms of robustness and efficiency.

Key words. generalized Langevin dynamics, Markov chain Monte Carlo, symmetric splitting methods

AMS subject classifications. 68Q25, 68R10, 68U05

DOI. 10.1137/20M138497X

1. Introduction.

In this article we study numerical discretization schemes for a generalized Langevin equation (GLE) of the form

$$\text{(GLE)}\qquad \begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{M}^{-1}\boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) - \int_0^t \boldsymbol{K}(t-s)\boldsymbol{M}^{-1}\boldsymbol{p}(s)\mathrm{d}s + \boldsymbol{\eta}(t), \end{aligned}$$

where the dynamical variables $\boldsymbol{q} \in \mathbb{R}^n, \boldsymbol{p} \in \mathbb{R}^n$ denote the positions and momenta of a Hamiltonian system with energy function

$$H(\boldsymbol{q},\boldsymbol{p}) = U(\boldsymbol{q}) + \frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p}. \tag{1.1}$$

The mass matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ is assumed to be symmetric positive definite, and U is a smooth confining potential function (i.e., $U \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ and $U(\boldsymbol{q}) \to \infty$ as $\|\boldsymbol{q}\| \to \infty$)

*Submitted to the journal's Methods and Algorithms for Scientific Computing section December 17, 2020; accepted for publication (in revised form) August 30, 2021; published electronically February 14, 2022.

https://doi.org/10.1137/20M138497X

Funding: This work was supported by the European Research Council grant 320823. The first author was further supported by the Engineering and Physical Sciences Research Council under grant EPSRC EP/P006175/1, "Data-Driven Coarse-Graining using Space-Time Diffusion Maps." The work of the second author was supported by the Statistical and Applied Mathematical Sciences Institute under grant DMS-1638521 and by Duke University.

†School of Mathematics, University of Edinburgh, Edinburgh EH93FD, UK (b.leimkuhler@ed.ac.uk, http://kac.maths.ed.ac.uk/~bl/).

‡Department of Mathematics, University of British Columbia, Vancouver V6T 1Z2, BC, Canada, and School of Mathematics, University of Birmingham, Watson Building, Ring Rd N, Birmingham B15 2TS, United Kingdom (m.sachs@bham.ac.uk, https://sites.google.com/view/matthiassachs).

such that $\int_{\mathbb{R}^n} \mathrm{e}^{-U(\boldsymbol{q})}\mathrm{d}\boldsymbol{q} < \infty$. $\boldsymbol{K} : [0,\infty) \to \mathbb{R}^{n\times n}$ is a matrix-valued (generalized) function, which is referred to as the memory kernel, and $\boldsymbol{\eta}$ is a stationary Gaussian process in $\mathbb{R}^n$ with zero mean, i.e., $\mathbb{E}[\boldsymbol{\eta}(t)] = \mathbf{0}$ for all $t \geq 0$.

The dynamical variable $\boldsymbol{\eta}$ models a random force, which is such that a fluctuation dissipation relation holds. That is, the autocovariance function of the random force and the memory kernel $\boldsymbol{K}$ coincide up to a constant prefactor:

$$\mathbb{E}[\boldsymbol{\eta}(s+t)\boldsymbol{\eta}^\top(s)] = \beta^{-1}\boldsymbol{K}(t) \text{ for all } t, s > 0,$$

where $\beta^{-1} = Tk_B > 0$ with T being the temperature of the modeled system and k_B denoting Boltzmann's constant. In its general form (GLE) the GLE is a non-Markovian dynamical model, meaning that the evolution of the state of the described system depends not only on the state itself but on the state history. The *underdamped Langevin equation* is obtained as a special Markovian variant of the GLE when the memory kernel is chosen as $\boldsymbol{K}(t) = \widehat{\boldsymbol{\Gamma}}\delta(t)$ and $\eta = \sqrt{2\beta^{-1}}\widehat{\boldsymbol{\Gamma}}^{1/2}\dot{\boldsymbol{W}}$, where $\widehat{\boldsymbol{\Gamma}}^{1/2}$ is a matrix root of $\widehat{\boldsymbol{\Gamma}}$, and $\delta(\cdot)$ denotes the Dirac delta function, $\dot{\boldsymbol{W}}$ is a Gaussian white noise in $\mathbb{R}^n$ with independent components, i.e., $\dot{\boldsymbol{W}} = [\dot{W}_i]_{1\leq i\leq n}$, such that $\dot{W}_i \sim \mathcal{N}(0,1)$ and $\mathbb{E}[\dot{W}_i(t)\dot{W}_j(s)] = \delta_{ij}\delta(t-s)$, and $\widehat{\boldsymbol{\Gamma}}$ is a symmetric positive definite matrix which is commonly referred to as the friction matrix. Under this parameterization (GLE) simplifies to the Itô diffusion

$$\dot{\boldsymbol{q}} = \boldsymbol{M}^{-1}\boldsymbol{p}, \quad \dot{\boldsymbol{p}} = -\nabla_{\boldsymbol{q}}U(\boldsymbol{q}) - \widehat{\boldsymbol{\Gamma}}\boldsymbol{M}^{-1}\boldsymbol{p} + \sqrt{2\beta^{-1}}\widehat{\boldsymbol{\Gamma}}^{1/2}\dot{\boldsymbol{W}}. \tag{LD}$$

1.1. The GLE as a dynamical model. Traditionally, the GLE is widely used in thermodynamics to model the dynamics of an open system which exchanges energy with one or more heat baths. The equation can be formally derived via the Mori–Zwanzig projection formalism [39, 55]. As such, it provides a dynamical description of the projection of a physical system onto a finite subset of its degrees of freedom. In the absence of a clear scale separation in the time evolution of explicitly modeled degrees of freedom and the traverse degrees of freedom, Markovian approximations in the form of (LD) fail to reproduce the dynamical properties of the system. Incorporation of memory effects via the stochastic integro-differential equation (GLE) are key to an accurate description of the system dynamics in such a setup. As such the GLE is used as a dynamical model in a wide range of applications, including coarse-grained meso- and macroscale molecular particle dynamics models [18, 33, 34], simulation of solids [23, 42], nonequilibrium dynamics in open systems with temperature gradient [15, 48, 43], and complex fluids and (anomalous) diffusive transport in soft matter [16, 36, 52].

1.2. Application in sampling. Besides its application as a dynamical model, the GLE has been used in molecular sampling to design (approximate) Markov chain Monte Carlo methods with enhanced sampling properties [11, 12, 40, 54, 9], and there are theoretical results showing that GLE-based sampling schemes [41] and annealing schemes [13] can exhibit better convergence properties in comparison to schemes based on an (underdamped) Langevin equation. Indeed, under certain conditions (see subsection 1.4.1) on the memory kernel $\boldsymbol{K}$, the process defined by (GLE) is exponentially ergodic with unique invariant measure given by the Gibbs–Boltzmann distribution

$$\pi(\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}) = \frac{1}{Z}\mathrm{e}^{-\beta H(\boldsymbol{q},\boldsymbol{p})}\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p} \tag{1.2}$$

so that, in particular, the process $\boldsymbol{q}(t), t \geq 0$ can be used to draw samples from the marginal measure

$$\pi_q(\mathrm{d}\boldsymbol{q}) \propto \mathrm{e}^{-\beta U(\boldsymbol{q})}\mathrm{d}\boldsymbol{q}. \tag{1.3}$$

In a nutshell, the idea behind the enhanced sampling methods developed in [11, 12] is to equip the GLE with a memory kernel of the form $\boldsymbol{K}(t) = K(t)\boldsymbol{I}_n$, where K is a scalar-valued generalized function which is constructed such that mixing times associated to (GLE) with $U(\boldsymbol{q}) = \frac{1}{2}\omega\boldsymbol{q}^2$ are minimized over a prescribed frequency range $\omega \in [\omega_{\min}, \omega_{\max}] \subset (0, \infty)$ assuming $\boldsymbol{M} = \boldsymbol{I}_n$. The Fourier transform of the obtained generalized function $K(t)$ resembles the frequency response of a high-pass filter. The corresponding convolution term thus results in weaker damping of low-frequency components and stronger damping of high-frequency components. Consequently, the combined effect of the convolution term and random force on the momentum trajectory is similar to a low-pass filter, where slower modes evolve almost ballistically and faster modes are sufficiently strongly thermostatted to suppress resonance effects in numerical discretizations of the dynamics. Even though this construction assumes the target π_q to be Gaussian, the enhanced sampling properties have been shown in practice to extend to the non-Gaussian case [12, 10]. In particular, in sampling problems where π_q is ill-conditioned (in the sense that the associated covariance matrix has a large condition number), such constructed sampling schemes may result in dramatically improved sampling efficiency and robustness in comparison to schemes obtained from a discretization of an underdamped Langevin equation [12, 4]. Importantly, mixing properties of such constructed GLEs are invariant under orthogonal transformations of the underlying coordinate system (i.e., under change of variables of the form $\boldsymbol{q} \mapsto \boldsymbol{U}\boldsymbol{q}, \boldsymbol{p} \mapsto \boldsymbol{U}\boldsymbol{p}$, where $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ is some orthogonal matrix); see [10].

1.3. Scope and main results of this article. The purpose of this article is to provide a class of numerical integrators with well-understood theoretical properties which are suitable for simulation of the GLE in a wide range of applications. Within this class of numerical integrators, we identify one scheme, gle-BAOAB, which we show analytically and in numerical experiments to outperform previously proposed schemes in terms of numerical discretization error and numerical stability.

The fact that in any computer simulation only finite memory is available means that any computer simulation of (GLE) inevitably results in a quasi-Markovian process. That is, the obtained process is Markovian in some (generally obscure) state space. Here, we focus on a particular class of quasi-Markovian instances of (GLE), whose Markovian form can be related directly to a certain class of Itô diffusion processes in an extended state space (see subsection 1.4). Instead of attempting to directly discretize (GLE), we numerically integrate the corresponding equivalent stochastic differential equation (SDE). This approach allows us, in applications where the solution process of (GLE) is not quasi-Markovian (e.g., if $\boldsymbol{K}(t)$ has the form of a power law), to clearly separate the error due to time discretization of the process from the error induced by a quasi-Markovian approximation.

The focus of this article is on the construction and analysis of integration schemes for the quasi-Markovian approximation. The proposed time discretizations (section 2) of the equivalent SDE are constructed as symmetric stochastic splitting schemes. The decomposition (of the associated generator) which is used as the basis for these stochastic splitting schemes is suggested by results on stochastic splitting schemes for the underdamped Langevin equation [8, 28, 30, 1, 2].

In terms of the analysis, we focus on the properties of the Markov chain obtained by such time discretizations of the quasi-Markovian approximation. That is, we discuss the existence of a stepsize-dependent invariant measure π_h of that Markov chain (here, $h > 0$ denotes the stepsize of the time discretization) and provide conditions for geometric convergence and the validity of a central limit theorem (section 3).

Moreover, we provide a detailed analysis of the marginal measure $\pi_h(\mathrm{d}\boldsymbol{q})$ of the position variable $\boldsymbol{q}$ and its convergence to the exact marginal measure $\pi(\mathrm{d}\boldsymbol{q})$ as $h \to 0$ (subsection 3.2). While the approximation accuracy of the measure $\pi_h(\mathrm{d}\boldsymbol{q})$ is obviously highly relevant for sampling applications, we emphasize that an accurate approximation of the marginal measure $\pi(\mathrm{d}\boldsymbol{q})$ is also relevant for accurately recovering dynamical properties [29, 30].

Another aspect covered in this article is the behavior of the introduced stochastic splitting schemes in certain limits of parametrization (section 4). We show that the obtained numerical schemes behave consistently with well-known scaling limits (summarized in subsection 1.5) of the continuous dynamics and reduce to numerical integrators of the corresponding limiting dynamics with well-known numerically favorable properties. This has important implications both for sampling applications and for applications where the GLE is used as a dynamical model. In the former case, our results ensure that the convergence order of the stepsize-dependent error incurred in the invariant measure π_h is not reduced in the respective limits. In fact, we show that in the overdamped limit of (GLE) we obtain an increase of convergence order (fourth order instead of second order). A property which—in accordance with previous results [28, 30]—we refer to as *superconvergence*. In the latter case, our results ensure that the dynamical properties of the simulated dynamics remain consistent with the underdamped/white noise limit of (GLE).

In the remainder of this section, we set up the basic framework for studying the Markovian reformulation of the GLE. We briefly review the above-mentioned homogenization results for the continuous dynamics as well as some results on the ergodic properties of the continuous dynamics.

1.4. Quasi-Markovian generalized Langevin equations (QGLEs). Consider the SDE defined on the extended space $\Omega_{\boldsymbol{x}} = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$,

$$\text{(QGLE)} \qquad \begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{M}^{-1}\boldsymbol{p}\,\mathrm{d}t, \\ \begin{pmatrix} \dot{\boldsymbol{p}} \\ \dot{\boldsymbol{s}} \end{pmatrix} &= \begin{pmatrix} -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \\ \boldsymbol{0} \end{pmatrix} - \boldsymbol{\Gamma} \begin{pmatrix} \boldsymbol{M}^{-1}\boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix} + \sqrt{\beta^{-1}}\boldsymbol{\Sigma}\dot{\boldsymbol{W}}, \end{aligned}$$

where $\boldsymbol{\Gamma}, \boldsymbol{\Sigma}$ are block matrices of the form

$$\boldsymbol{\Gamma} := \begin{bmatrix} \boldsymbol{\Gamma}_{1,1} & \boldsymbol{\Gamma}_{1,2} \\ \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2} \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}, \quad \boldsymbol{\Sigma} := \begin{bmatrix} \boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)},$$

with $m \geq n$ and where $\dot{\boldsymbol{W}}$ is a Gaussian white noise in $\mathbb{R}^{n+m}$ with independent components, i.e., $\dot{\boldsymbol{W}} = [\dot{W}_i]_{1\leq i\leq n+m}$, such that $\dot{W}_i \sim \mathcal{N}(0,1)$ and $\mathbb{E}[\dot{W}_i(t)\dot{W}_j(s)] = \delta_{ij}\delta(t-s)$.

In what follows we first provide a set of sufficient conditions which ensure that the SDE (QGLE) can be rewritten in the form of the stochastic integro-differential equation (GLE) and possesses an invariant measure which is such that its marginal in $\boldsymbol{q}, \boldsymbol{p}$ coincides with the Gibbs–Boltzmann distribution.

Assumption 1.

(i) There exists a symmetric positive definite matrix $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ such that

$$\boldsymbol{\Gamma} \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} + \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} \boldsymbol{\Gamma}^T = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T. \tag{1.4}$$

(ii) The real parts of all eigenvalues of the matrix

$$\boldsymbol{\Gamma}_{\boldsymbol{M}} := \boldsymbol{\Gamma} \begin{pmatrix} \boldsymbol{M}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_m \end{pmatrix} \tag{1.5}$$

are positive. That is, $-\boldsymbol{\Gamma}_{\boldsymbol{M}}$ is a stable matrix.

(iii) The matrices $\boldsymbol{\Gamma}_{1,1}$ and $\boldsymbol{M}$ commute.

A derivation of the following proposition can be found in [12] (see also [31]).

PROPOSITION 1.1. *Let Assumption* 1 *hold, and let* $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ *be as specified therein. The SDE (QGLE) conserves the probability measure* $\pi(\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{s})$ *with density*

$$\rho_{\boldsymbol{Q},\beta}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) \propto e^{-\beta[U(\boldsymbol{q})+\frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p}+\frac{1}{2}\boldsymbol{s}^T\boldsymbol{Q}^{-1}\boldsymbol{s}]}. \tag{1.6}$$

If, further, $\boldsymbol{s}(0) \sim \mathcal{N}(\mathbf{0},\boldsymbol{Q})$ *(independent of* $\boldsymbol{q}(0),\boldsymbol{p}(0),\dot{\boldsymbol{W}}$*), then the SDE (QGLE) can be rewritten in the form of a stochastic integro-differential equation (GLE) with*

$$\boldsymbol{K}(t) = \boldsymbol{K}_{\boldsymbol{\Gamma}}(t) := \boldsymbol{\Gamma}_{1,1}\delta(t) - \boldsymbol{\Gamma}_{1,2}e^{-t\boldsymbol{\Gamma}_{2,2}}\boldsymbol{\Gamma}_{2,1}. \tag{1.7}$$

We refer to GLEs whose memory kernel is of the form specified in (1.7) as Quasi-Markovian generalized Langevin equations (QGLEs). While more general parametrizations of (QGLE) are possible, we focus in this article on two classes of memory kernels which can be characterized by some additional constraints on the form of $\boldsymbol{\Gamma}$ as summarized in Assumption 2. Memory kernels falling into either of these two classes (or which are positive linear combinations of instances of either class) make up most of the common parameterizations of (QGLE) appearing in the literature (e.g., [36, 27, 19, 26, 39, 3, 35]).

Assumption 2. The matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{Q}$ are such that either

(i) $\boldsymbol{\Gamma}_{1,1} = \mathbf{0}$ and $\boldsymbol{\Gamma}_{1,2}\boldsymbol{Q} = -\boldsymbol{\Gamma}_{2,1}^T$

or

(ii) $\boldsymbol{\Gamma}_{1,1}$ is symmetric positive definite, $\boldsymbol{Q} = \boldsymbol{I}_m$, and $\boldsymbol{\Gamma}_{1,2} = \boldsymbol{\Gamma}_{2,1}^T$.

Parametrization according to Assumption 2 (i) allows for the representation of memory kernels of the form

$$\boldsymbol{K}(t) = \boldsymbol{I}_n \sum_{l=1}^{R} c_l e^{-a_l t}\cos(b_l t), \qquad c_l, a_l > 0,\ b_l \geq 0,\ R \in \mathbb{N}.$$

By choosing the matrices $\boldsymbol{\Gamma}_{1,2}$ and $\boldsymbol{Q}$ appropriately, one can additionally include cross-correlations between components. Typically such parameterizations are used in applications where the GLE is being treated as a dynamical model; see subsection 1.1.

Parametrization according to Assumption 2 (ii) allows for the representation of memory kernels whose Fourier transform is similar to the frequency response of a high-pass filter. Such parameterizations are most commonly used in sampling applications; see subsection 1.2. For example, the memory kernel $\boldsymbol{K}(t) = \boldsymbol{I}_n(\gamma\delta(t) - \lambda\frac{\gamma}{\tau}e^{-t/\tau}), \lambda \in [0,1), \gamma > 0, \tau > 0$ used in [54] can be represented in the form (1.7) using $\boldsymbol{\Gamma}_{1,1} = \gamma\boldsymbol{I}_n$, $\boldsymbol{\Gamma}_{1,2} = \boldsymbol{\Gamma}_{2,1} = \sqrt{\lambda\gamma/\tau}\boldsymbol{I}_n$, and $\boldsymbol{\Gamma}_{2,2} = \tau^{-1}\boldsymbol{I}_n$.

1.4.1. Ergodicity and central limit theorem for QGLEs. Under additional conditions on the coefficients of (QGLE), the solution process is ergodic with unique invariant measure π and satisfies a central limit theorem. This follows from the fact that the associated semigroup of the dynamics converges exponentially in a suitably weighted L^∞ functional space as made precise below.

In order to state this result, we first need to set some notation. For prescribed $\mathcal{K} \in \mathcal{C}^\infty(\Omega_{\boldsymbol{x}}, [1,\infty))$ with $\mathcal{K}(x) \to +\infty$ as $\|x\| \to \infty$, we define the set of functions

$$L^\infty_{\mathcal{K}}(\Omega_{\boldsymbol{x}}) := \left\{ \varphi : \Omega_{\boldsymbol{x}} \to \mathbb{R},\ \text{measurable} \ : \sup_{x \in \Omega_{\boldsymbol{x}}} \left| \frac{\varphi(x)}{\mathcal{K}(x)} \right| < \infty \right\},$$

which, when equipped with the norm $\varphi \mapsto \|\varphi\|_{L^\infty_{\mathcal{K}}} := \sup_{\boldsymbol{x} \in \Omega_{\boldsymbol{x}}} |\varphi(\boldsymbol{x})/\mathcal{K}(\boldsymbol{x})|$, forms a Banach space. We denote by $\mathcal{C}_P(\Omega_{\boldsymbol{x}}, \mathbb{R}) = \bigcap_{l \in \mathbb{N}} L^\infty_{\mathcal{K}_l}(\Omega_{\boldsymbol{x}})$ with $\mathcal{K}_l(\boldsymbol{x}) = 1 + |\boldsymbol{x}|^{2l}$ the set of at most polynomially growing real-valued functions and by $\mathcal{C}^p_P(\Omega_{\boldsymbol{x}}, \mathbb{R}) \subseteq \mathcal{C}_P(\Omega_{\boldsymbol{x}}, \mathbb{R})$ the set of real-valued functions whose partial derivatives up to order $p \in \mathbb{N}$ exist and grow at most polynomially, i.e.,

$$\varphi \in \mathcal{C}^p_P(\Omega_{\boldsymbol{x}}, \mathbb{R}) \iff \partial_{\boldsymbol{x}_{i_1}} \dots \partial_{\boldsymbol{x}_{i_m}} \varphi \in \mathcal{C}_P(\Omega_{\boldsymbol{x}}, \mathbb{R})$$

for any differential operator $\partial_{\boldsymbol{x}_{i_1}} \dots \partial_{\boldsymbol{x}_{i_m}}$ with $m \le p$. Unless stated otherwise, we consider operators introduced in the following to be defined on the core:

$$\mathcal{C}^\infty_P(\Omega_{\boldsymbol{x}}, \mathbb{R}) := \bigcap_{p=1}^{\infty} \mathcal{C}^p_P(\Omega_{\boldsymbol{x}}, \mathbb{R}). \tag{1.8}$$

In particular, the infinitesimal generator, $\mathcal{L}_{\mathrm{GLE}}$, of (QGLE) when constrained to this set of test functions takes the form

$$\mathcal{L}_{\mathrm{GLE}} = -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \cdot \nabla_{\boldsymbol{p}} + \boldsymbol{M}^{-1}\boldsymbol{p} \cdot \nabla_{\boldsymbol{q}} - \boldsymbol{\Gamma}_{\boldsymbol{M}} \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix} \cdot \nabla_{\boldsymbol{z}} + \frac{\beta^{-1}}{2} \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T : \nabla^2_{\boldsymbol{z}}, \tag{1.9}$$

where $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T : \nabla^2_{\boldsymbol{z}} = \sum_{i=1}^M \sum_{j=1}^M [\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T]_{i,j} \partial_{\boldsymbol{z}_i} \partial_{\boldsymbol{z}_j}$, $M = n + m$. Here, as well as in the remainder of this article, $\boldsymbol{z} = (\boldsymbol{p}, \boldsymbol{s})$ is used as shorthand for the combined vector of momenta and auxiliary variables. For $t \ge 0$, we denote the evolution operator associated with the SDE (QGLE) as $e^{t\mathcal{L}_{\mathrm{GLE}}}$, i.e., $(e^{t\mathcal{L}_{\mathrm{GLE}}}\varphi)(x) = \mathbb{E}[\varphi(\boldsymbol{x}(t)) \,|\, \boldsymbol{x}(0) = x]$, where the expectation is with respect to the driving Wiener process, $\boldsymbol{W}$ of (QGLE), and $\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s})$ is used as a shorthand for the combined vector of positions, momenta, and auxiliary variables.

Assumption 3.

(i) The matrices $\boldsymbol{\Gamma}_{\boldsymbol{M}}$ and $\boldsymbol{\Sigma}$ are such that the operator $\partial_t - \mathcal{L}_{\mathrm{GLE}}$ is hypoelliptic. (See [31, Proposition 7] for sufficient algebraic conditions on $\boldsymbol{\Gamma}_{\boldsymbol{M}}, \boldsymbol{\Sigma}$ for hypoellipticity of the operator.) In particular,

$$\mathbb{R}^M = \bigcup_{k=0}^{M} \bigcup_{i=1}^{M} \boldsymbol{\Gamma}^k_{\boldsymbol{M}} \Sigma_i, \quad \boldsymbol{\Sigma} = (\Sigma_1, \dots, \Sigma_M),$$

where $M = n + m$ and $\Sigma_i \in \mathbb{R}^M$ for $1 \le i \le M$.

(ii) The potential function U is of the form $U(\boldsymbol{q}) = U_1(\boldsymbol{q}) + U_2(\boldsymbol{q})$, where $U_1(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T \boldsymbol{\Omega} \boldsymbol{q}$, with $\boldsymbol{\Omega} \in \mathbb{R}^{n \times n}$ being a symmetric positive definite matrix, and $U_2 \in \mathbb{C}^\infty(\mathbb{R}^n, \mathbb{R})$ is such that its derivatives are uniformly bounded in $\mathbb{R}^n$, i.e., $\sup_{\boldsymbol{q} \in \mathbb{R}^n} \|\partial_{\boldsymbol{q}_{i_1}} \partial_{\boldsymbol{q}_{i_2}} \dots \partial_{\boldsymbol{q}_{i_k}} U_2(\boldsymbol{q})\| < \infty$ for any $k \in \mathbb{N}$, and $i_1, \dots, i_k \in \{1, \dots, n\}$.

Under the above stated assumptions exponential convergence of the associated semigroup and a central limit theorem for trajectory averages can be established.

PROPOSITION 1.2 ([31]). *Let Assumptions* 1 *to* 3 *be satisfied.*

(i) *For any* $l \in \mathbb{N}$*, there exist constants* $\kappa_l > 0, C_l > 0$ *such that*

$$\text{for all } t \geq 0 \text{ for all } \varphi \in L^\infty_{\mathcal{K}_l}, \ \left\| \mathrm{e}^{t\mathcal{L}_{\mathrm{GLE}}}\varphi - \int \varphi \,\mathrm{d}\pi \right\|_{L^\infty_{\mathcal{K}_l}} \leq C_l \mathrm{e}^{-t\kappa_l} \left\| \varphi - \int \varphi \,\mathrm{d}\pi \right\|_{L^\infty_{\mathcal{K}_l}}.$$

(ii) *Let* $\overline{\varphi}_t := t^{-1}\int_0^t \varphi(\boldsymbol{x}(t))\mathrm{d}t$*. If* $\varphi \in L^\infty_{\mathcal{K}_l}(\Omega_{\boldsymbol{x}})$ *for some* $l \in \mathbb{N}$*, then there is a finite* $\sigma^2_\varphi > 0$ *so that*

$$\sqrt{t}\left(\overline{\varphi}_t - \int \varphi \,\mathrm{d}\pi\right) \xrightarrow[t\to+\infty]{\text{law}} \mathcal{N}(0, \sigma^2_\varphi). \tag{1.10}$$

If not explicitly stated otherwise, we assume throughout the remainder of this article that the parameterization of (QGLE) is such that Assumptions 1 to 3 are all satisfied.

1.5. Limiting dynamics. The underdamped and overdamped Langevin equations can be obtained as limits of the Markovian reformulation of the GLE. In what follows we briefly review two key results from [44] and [35] (see also [46] for related long-time equilibration estimates), which we will later show to hold in slightly modified form for the discretized dynamics. For this purpose we consider the following rescaled process, which is obtained from (QGLE) by a change of variable corresponding to a time rescaling as $(\boldsymbol{q}^\lambda(t), \boldsymbol{p}^\lambda(t), \boldsymbol{s}^\lambda(t)) = (\boldsymbol{q}(\lambda t), \boldsymbol{p}(\lambda t), \boldsymbol{s}(\lambda t))$, with $\lambda > 0$:

$$\begin{aligned} \dot{\boldsymbol{q}}^\lambda &= \lambda \boldsymbol{p}^\lambda, \\ \begin{pmatrix} \dot{\boldsymbol{p}}^\lambda \\ \dot{\boldsymbol{s}}^\lambda \end{pmatrix} &= \left[\lambda \begin{pmatrix} -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}^\lambda) \\ \boldsymbol{0} \end{pmatrix} - \lambda \boldsymbol{\Gamma}^\mu \begin{pmatrix} \boldsymbol{p}^\lambda \\ \boldsymbol{s}^\lambda \end{pmatrix}\right] + \sqrt{\lambda\beta^{-1}}\boldsymbol{\Sigma}^\mu \dot{\boldsymbol{W}}, \end{aligned} \tag{QGLE-scaled}$$

with $\boldsymbol{s}^\lambda(0) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ and

$$\boldsymbol{\Gamma}^\mu = \begin{pmatrix} \boldsymbol{0} & -\mu_1 \boldsymbol{D}_a \\ \mu_1 \boldsymbol{D}_a & \mu_2 \boldsymbol{D}_b \end{pmatrix} \in \mathbb{R}^{2n\times 2n}, \qquad \boldsymbol{\Sigma}^\mu = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \sqrt{2\mu_2}\boldsymbol{D}_b \end{pmatrix} \in \mathbb{R}^{2n\times 2n}, \tag{1.11}$$

where $\boldsymbol{D}_a := \mathrm{diag}(a_1, a_2, \ldots, a_n)$ and $\boldsymbol{D}_b := \mathrm{diag}(b_1, b_2, \ldots, b_n)$ are positive diagonal matrices. In the view of the stochastic integro-differential equation (GLE) the form of the matrices $\boldsymbol{\Gamma}^\mu, \boldsymbol{\Sigma}^\mu$ corresponds to a rescaling of the memory kernel $\boldsymbol{K}(t) = \mathrm{diag}(a_1^2\mathrm{e}^{-tb_1}, \ldots, a_n^2\mathrm{e}^{-tb_n})$ as $\boldsymbol{K}^\mu(t) = \mu_1^2\boldsymbol{K}(\mu_2 t)$. Without loss of generality we assume $\boldsymbol{M} = \boldsymbol{I}_n$ here.[1]

For the parameter choice $\lambda = 1, \mu_1 = \epsilon^{-1}$, and $\mu_2 = \epsilon^{-2}$, the rescaled process (1.11) converges weakly to the solution of an underdamped Langevin equation as $\epsilon \to 0$ as detailed in the following proposition.

PROPOSITION 1.3 (white noise limit, [44]). *Let* $\lambda = 1, \mu_1 = \epsilon^{-1}$*, and* $\mu_2 = \epsilon^{-2}$*. For finite* $T > 0$ *we have*

$$(\boldsymbol{q}^\lambda(t), \boldsymbol{p}^\lambda(t)) \xrightarrow[\varepsilon\to 0]{\text{law}} (\boldsymbol{q}(t), \boldsymbol{p}(t)),$$

uniformly in $t \in [0, T]$*, where* $(\boldsymbol{q}(t), \boldsymbol{p}(t))_{t\geq 0}$ *denotes the solution process of the underdamped Langevin equation* (LD) *with* $\widehat{\boldsymbol{\Gamma}} = \boldsymbol{D}_a^2\boldsymbol{D}_b^{-1}$*,* $\boldsymbol{M} = \boldsymbol{I}_n$*, and initial values* $(\boldsymbol{q}(0), \boldsymbol{p}(0)) = (\boldsymbol{q}^\lambda(0), \boldsymbol{p}^\lambda(0))$.

[1] Note that (QGLE) can always be transformed to a system with an isotropic mass matrix by applying a change of variable of the form $\boldsymbol{q} \mapsto \boldsymbol{M}^{1/2}\boldsymbol{q}$, $\boldsymbol{p} \mapsto \boldsymbol{M}^{-1/2}$ to the process resulting in modifications as $-\nabla U \mapsto -\boldsymbol{M}^{-1/2}\nabla U$, $\boldsymbol{\Gamma} \mapsto \mathrm{diag}(\boldsymbol{M}^{-1/2}, \boldsymbol{I}_m)\,\boldsymbol{\Gamma}\,\mathrm{diag}(\boldsymbol{M}^{-1/2}, \boldsymbol{I}_m)$, and $\boldsymbol{\Sigma} \mapsto \mathrm{diag}(\boldsymbol{M}^{-1/2}, \boldsymbol{I}_m)\boldsymbol{\Sigma}$ of the force, friction matrix, and diffusion matrix, respectively.

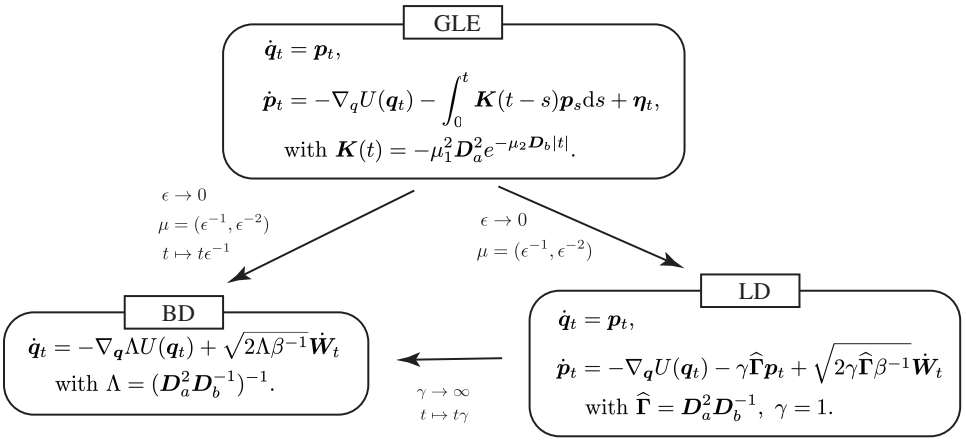

FIG. 1. *Diagram of Langevin limits. The overdamped limit of the GLE (GLE → BD) is detailed in Proposition* 1.4. *The white noise limit (GLE → LD) is detailed in Proposition* 1.3. *The overdamped limit of the underdamped Langevin equation (LD → BD) is a well-known result in the literature (see, e.g.,* [45]*).*

Similarly, when considering the scaling $\lambda = \mu_1 = \mu_2 = \epsilon^{-1}$ one can show that the solution of (QGLE-scaled) converges weakly to the solution of an overdamped Langevin equation of the form

$$\dot{\boldsymbol{q}} = -\Lambda \nabla U(\boldsymbol{q}) + \sqrt{2\beta^{-1}}\Lambda^{1/2}\dot{\boldsymbol{W}}. \tag{BD}$$

PROPOSITION 1.4 (overdamped limit). *Let* $\lambda = \mu_1 = \mu_2 = \epsilon^{-1}$. $\boldsymbol{s}^\lambda(0) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$, *and* $(\boldsymbol{q}^\lambda(0), \boldsymbol{p}^\lambda(0)) \in \mathbb{R}^{2n}$. *For finite* $T > 0$ *we have*

$$\boldsymbol{q}^\lambda(t) \xrightarrow[\varepsilon \to 0]{\text{law}} \boldsymbol{q}(t),$$

uniformly in $t \in [0, T]$, *where* $(\boldsymbol{q}(t))_{t \geq 0}$ *denotes the solution process of the overdamped Langevin equation* (BD) *with* $\Lambda = (\boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1})^{-1}$.

Proof. This result is a direct consequence of Theorem IV.1 in [35]. The same limit has also been studied in a less general setup in [50]. Refer to Figure 1 for a diagram illustrating these different limits. ∎

2. Symmetric stochastic splitting methods for the QGLE. In this section we present the basic construction and implementation of the class of proposed stochastic splitting schemes for the Markovian reformulation (QGLE) of the GLE as well as an elementary analysis of the incurred weak error.

2.1. Construction of numerical methods based on splitting. As mentioned in the introduction, we construct splitting schemes using a similar procedure as that employed for the underdamped Langevin equation in [28, 30]. Such schemes are based on a decomposition of the generator of the underdamped Langevin equation as $\mathcal{L}_{\mathrm{LD}} = \mathcal{L}_{\mathrm{A}} + \mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\widehat{\mathrm{O}}}$, where $\mathcal{L}_{\mathrm{A}} = -\nabla U(\boldsymbol{q}) \cdot \nabla_{\boldsymbol{p}}$, $\mathcal{L}_{\mathrm{B}} = -\boldsymbol{p} \cdot \boldsymbol{M}^{-1} \nabla_{\boldsymbol{q}}$ and $\mathcal{L}_{\widehat{\mathrm{O}}} = -\widehat{\boldsymbol{\Gamma}}\boldsymbol{M}^{-1} \cdot \nabla \boldsymbol{p} + \beta^{-1}\widehat{\boldsymbol{\Gamma}} : \nabla_{\boldsymbol{p}}^2$. By applying a Strang splitting with stepsize $h > 0$ twice (typically first to treat the Liouville operator $\mathcal{L}_{\mathrm{A}} + \mathcal{L}_{\mathrm{B}}$ associated with the Hamiltonian vector field and then subsequently to compute the combination of that operator with the term $\mathcal{L}_{\mathrm{O}}$), a symmetric stochastic splitting scheme with the following associated evolution operator is obtained:

$$\hat{\mathcal{P}}_h^{\text{ld-OBABO}} = \exp\left((h/2)\mathcal{L}_{\widehat{\mathrm{O}}}\right)\exp\left((h/2)\mathcal{L}_{\mathrm{B}}\right)\exp\left(h\mathcal{L}_{\mathrm{A}}\right)\exp\left((h/2)\mathcal{L}_{\mathrm{B}}\right)\exp\left((h/2)\mathcal{L}_{\widehat{\mathrm{O}}}\right).$$

Similarly, by either changing the ordering within the Strang splitting or by changing the pair of operators selected for the first application of the Strang splitting, other splitting schemes can be obtained which are uniquely identified by palindromes of the form XYZYX, where $\mathrm{X},\mathrm{Y},\mathrm{Z} \in \{\mathrm{A},\mathrm{B},\widehat{\mathrm{O}}\}$ are distinct placeholders. (The symmetry of this decomposition is not essential but typically improves the accuracy and efficiency of the resulting scheme with little added computational cost.)

This construction can be easily generalized to the Markovian reformulation of the GLE (QGLE) by using the fact that (QGLE) structurally resembles the underdamped Langevin equation (LD). That is, we consider a decomposition of $\mathcal{L}_{\mathrm{GLE}}$ as $\mathcal{L}_{\mathrm{GLE}} = \mathcal{L}_{\mathrm{A}} + \mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\mathrm{O}}$, where $\mathcal{L}_{\mathrm{A}}, \mathcal{L}_{\mathrm{B}}$ are defined as above and

$$\mathcal{L}_{\mathrm{O}} = -\boldsymbol{\Gamma}\begin{pmatrix}\boldsymbol{M}^{-1}\boldsymbol{p}\\ \boldsymbol{s}\end{pmatrix}\cdot\nabla_{\boldsymbol{z}} + \frac{\beta^{-1}}{2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T : \nabla_{\boldsymbol{z}}^2.$$

The only difference between this decomposition and the decomposition of the operator $\mathcal{L}_{\mathrm{LD}}$ is that $\mathcal{L}_{\mathrm{O}}$ corresponds to the generator of a linear SDE in $\boldsymbol{p}$ and $\boldsymbol{s}$, whereas the operator $\mathcal{L}_{\widehat{\mathrm{O}}}$ in the otherwise identical decomposition of $\mathcal{L}_{\mathrm{LD}}$ is the generator of a linear SDE in $\boldsymbol{p}$ only. Thus, symmetric splitting schemes for (QGLE) can be constructed in the same way as for the underdamped Langevin equation resulting in numerical integrators with associated evolution operators of the form

$$\hat{\mathcal{P}}_h^{\text{gle-XYZYX}} = \exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{X}}\right)\exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{Y}}\right)\exp\left(h\mathcal{L}_{\mathrm{Z}}\right)\exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{Y}}\right)\exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{X}}\right), \tag{2.1}$$

where $\mathrm{X},\mathrm{Y},\mathrm{Z} \in \{\mathrm{A},\mathrm{B},\mathrm{O}\}$ are again distinct placeholders.

2.2. Implementation. By construction the numerical integrator for the associated evolution operator $\hat{\mathcal{P}}_h^{\text{gle-XYZYX}}$ is of the form

$$\hat{\Phi}_h^{\text{gle-XYZYX}} = \Phi_{h/2}^{\mathrm{X}} \circ \Phi_{h/2}^{\mathrm{Y}} \circ \Phi_h^{\mathrm{Z}} \circ \Phi_{h/2}^{\mathrm{Y}} \circ \Phi_{h/2}^{\mathrm{X}}, \tag{2.2}$$

where $\Phi_h^{\mathrm{X}}, \mathrm{X} \in \{\mathrm{A},\mathrm{B},\mathrm{O}\}$ are the solution maps of the differential equations associated with the operators $\mathcal{L}_{\mathrm{X}}, \mathrm{X} \in \{\mathrm{A},\mathrm{B},\mathrm{O}\}$, respectively. A practical implementation of the above-described splitting schemes therefore requires that each differential equation associated with the operators $\mathcal{L}_{\mathrm{A}}, \mathcal{L}_{\mathrm{B}}$, and $\mathcal{L}_{\mathrm{O}}$ can be solved exactly. Indeed, in the case of the operators $\mathcal{L}_{\mathrm{A}}$ and $\mathcal{L}_{\mathrm{B}}$ the solution of the associated differential equations $\dot{\boldsymbol{q}} = \boldsymbol{M}^{-1}\boldsymbol{p}$, and $\dot{\boldsymbol{p}} = -\nabla U(\boldsymbol{q})$ correspond to Euler updates of the form

$$\Phi_h^{\mathrm{A}} : (\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) \mapsto (\boldsymbol{q} + h\boldsymbol{M}^{-1}\boldsymbol{p}, \boldsymbol{p}, \boldsymbol{s}), \quad \Phi_h^{\mathrm{B}} : (\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) \mapsto (\boldsymbol{q}, \boldsymbol{p} - h\nabla_{\boldsymbol{q}} U(\boldsymbol{q}), \boldsymbol{s}), \tag{2.3}$$

respectively. The solution of the SDE associated with the operator $\mathcal{L}_{\mathrm{O}}$,

$$\dot{\boldsymbol{z}} = -\boldsymbol{\Gamma}_{\boldsymbol{M}}\boldsymbol{z} + \sqrt{\beta^{-1}}\boldsymbol{\Sigma}\dot{\boldsymbol{W}}, \tag{2.4}$$

coincides in law with $\boldsymbol{z}(h) = \boldsymbol{F}_h\boldsymbol{z}(0) + \boldsymbol{S}_h\mathscr{R}$ (see, e.g., [17, 45]), where $\mathscr{R} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n+m})$ denotes a vector of independent and standard normal distributed random variables in $\mathbb{R}^{n+m}$, $\boldsymbol{F}_h = \exp(-h\boldsymbol{\Gamma}_{\boldsymbol{M}})$ is the matrix exponential of the matrix $-h\boldsymbol{\Gamma}_{\boldsymbol{M}}$, and $\boldsymbol{S}_h$ solves the equation

$$\boldsymbol{S}_h\boldsymbol{S}_h{}^T = \beta^{-1}\left[\begin{pmatrix}\boldsymbol{M} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}\end{pmatrix} - \boldsymbol{F}_h\begin{pmatrix}\boldsymbol{M} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}\end{pmatrix}\boldsymbol{F}_h{}^T\right].$$

Algorithm 1. gle-BAOAB

input: $(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s})$
$\boldsymbol{p} \leftarrow \boldsymbol{p} - \frac{h}{2}\nabla U(\boldsymbol{q})$;
$\boldsymbol{q} \leftarrow \boldsymbol{q} + \frac{h}{2}\boldsymbol{M}^{-1}\boldsymbol{p}$;
$\begin{bmatrix}\boldsymbol{p}\\ \boldsymbol{s}\end{bmatrix} \leftarrow \boldsymbol{F}_h \begin{bmatrix}\boldsymbol{p}\\ \boldsymbol{s}\end{bmatrix} + \boldsymbol{S}_h\mathscr{R}$;
$\boldsymbol{q} \leftarrow \boldsymbol{q} + \frac{h}{2}\boldsymbol{M}^{-1}\boldsymbol{p}$;
$\boldsymbol{p} \leftarrow \boldsymbol{p} - \frac{h}{2}\nabla U(\boldsymbol{q})$;
output: $(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s})$

The corresponding stochastic flow map which updates the combined state vector $\boldsymbol{x}$ accordingly is of the form

$$\Phi_h^{\mathrm{O}} : (\boldsymbol{q}, \boldsymbol{z}) \mapsto (\boldsymbol{q}, \boldsymbol{F}_h\, \boldsymbol{z} + \boldsymbol{S}_h\mathscr{R}),$$

where $\mathscr{R}$ is independently resampled at every application of Φ_h^{O}. With the definition of the updates $\Phi_h^{\mathrm{A}}, \Phi_h^{\mathrm{B}}, \Phi_h^{\mathrm{O}}$ at hand, one can find explicit algorithmic forms for the integration map $\hat{\Phi}_h^{\text{gle-XYZYX}}$. We provide an algorithmic implementation of $\hat{\Phi}_h^{\text{gle-BAOAB}}$ in Algorithm 1.

2.3. Weak convergence order. A numerical scheme with associated evolution operator $\mathcal{P}_h$ is said to have global weak order p when applied to (QGLE) with $\boldsymbol{x}(0) \sim \pi_0$ if, for all $\varphi \in \mathcal{C}_P^\infty(\Omega_{\boldsymbol{x}}, \mathbb{R})$ and for all $T > 0$, there exists a constant $C(T, \varphi) > 0$ such that

$$\left|\left(\mathcal{P}_h^n\varphi\right)(x) - \left(\mathrm{e}^{nh\mathcal{L}_{\mathrm{GLE}}}\varphi\right)(x)\right| \le h^p C(T, \varphi)$$

for all $t_n = nh \in [0, T]$, for π_0-almost all x, and for all sufficiently small $h > 0$. Here, as well as in what follows, $\mathcal{P}^n = \mathcal{P}^{n-1}\mathcal{P}$ with $\mathcal{P}^0 = \mathrm{Id}$ denotes the nth power of the evolution operator $\mathcal{P}$. The above discretization schemes all have weak order $p = 2$.

PROPOSITION 2.1. *Let Assumption* 3 *be satisfied. Then, any symmetric stochastic splitting scheme with evolution operator* $\hat{\mathcal{P}}_h^{\text{gle-XYZYX}}$ *has global weak order* 2.

Proof. This result is a direct consequence of Theorem 2 in [38] which provides a set of sufficient conditions for the local weak error to coincide with that of the global weak order. We therefore only provide a brief outline of the proof. By Taylor expanding (at $h = 0$) both $(\mathrm{e}^{nh\mathcal{L}_{\mathrm{GLE}}}\varphi)(x)$ and $(\mathcal{P}_h^n\varphi)(x)$ with $\mathcal{P}_h = \hat{\mathcal{P}}_h^{\text{gle-XYZYX}}$ and comparing powers in h, we obtain an expansion of the local weak error as

$$\left|\Big(\mathcal{P}_h\varphi\Big)(x) - \left(\mathrm{e}^{h\mathcal{L}_{\mathrm{GLE}}}\varphi\right)(x)\right| = h^3\mathcal{A}_3\varphi(x) + h^5 r_{\varphi,\delta,5}(x),$$

with

$$\begin{aligned}\mathcal{A}_3 = \frac{1}{12}\Big(&[\mathcal{L}_{\mathrm{Z}}, [\mathcal{L}_{\mathrm{Z}}, \mathcal{L}_{\mathrm{Y}}]] + [\mathcal{L}_{\mathrm{Y}} + \mathcal{L}_{\mathrm{Z}}, [\mathcal{L}_{\mathrm{Y}} + \mathcal{L}_{\mathrm{Z}}, \mathcal{L}_{\mathrm{X}}]] \\ &- \frac{1}{2}[\mathcal{L}_{\mathrm{Y}}, [\mathcal{L}_{\mathrm{Y}}, \mathcal{L}_{\mathrm{Z}}]] - \frac{1}{2}[\mathcal{L}_{\mathrm{X}}, [\mathcal{L}_{\mathrm{X}}, \mathcal{L}_{\mathrm{Y}} + \mathcal{L}_{\mathrm{Z}}]]\Big),\end{aligned} \tag{2.5}$$

where $[A, B] = AB - BA$ denotes the commutator of the two linear operators A, B. In other words the convergence order of the local weak error is 2. Assumption 3 ensures that for $\varphi \in \mathcal{C}_P^\infty(\Omega_{\boldsymbol{x}}, \mathbb{R})$ the remainder term $r_{\varphi,h,5}$ as well as $\mathcal{A}_3\varphi$ are both

contained in $\mathcal{C}_P^\infty(\Omega_{\boldsymbol{x}}, \mathbb{R})$. For sufficiently small stepsize $h > 0$ the existence of a suitable Lyapunov function (see proof of Theorem 3.1) ensures that moments, $\mathbb{E}\big[\|\widehat{\boldsymbol{x}}_k\|^{2m}\big]$, of any order $m \in \mathbb{N}$ are uniformly bounded in the iteration index $k \in \mathbb{N}$. All together, the conditions of [38, Theorem 2] are met, which implies that the global weak convergence order coincides with the local weak convergence order. ∎

3. Error analysis of ergodic averages. As in the case of the underdamped Langevin equation and the overdamped Langevin equation, Markov chains of the discretized dynamics

$$\widehat{\boldsymbol{x}}_{k+1} = \hat{\Phi}_h^{\text{gle-XYZYX}}(\widehat{\boldsymbol{x}}_k),\ \widehat{\boldsymbol{x}}_0 = \boldsymbol{x}(0),\ k \in \mathbb{N}, \tag{3.1}$$

can be used as approximate Markov chain Monte Carlo methods for the computation of expectations with respect to the extended Gibbs–Boltzmann distribution π. That is, expectations of observables $\varphi \in L^2(\pi)$, where $L^2(\pi) := \{\varphi : \Omega_{\boldsymbol{x}} \to \mathbb{R} \text{ measurable} \mid \int \varphi^2 \mathrm{d}\pi < \infty\}$, are approximately computed as trajectory averages of the form $\widehat{\varphi}_N := \frac{1}{N}\sum_{k=1}^{N-1} \varphi(\widehat{\boldsymbol{x}}_k)$ from a finite trajectory $(\boldsymbol{x}_k)_{k=1,\dots,N}$. Such approximate computations are performed under the premise that $(\widehat{\boldsymbol{x}}_k)_{k\in\mathbb{N}}$ is ergodic with respect to the invariant measure $\pi_h \approx \pi$ so that

$$\lim_{N\to\infty} \widehat{\varphi}_N = \int \varphi(\boldsymbol{x})\pi_h(\mathrm{d}\boldsymbol{x})$$

for almost all realizations of the Markov chain $(\widehat{\boldsymbol{x}}_k)_{k\in\mathbb{N}}$. In this section, we provide theoretical justification for such a computation by showing that the above-mentioned assumptions are indeed satisfied. We first show in subsection 3.1 that the proposed numerical schemes result in ergodic Markov chains. Moreover, we show the validity of a central limit theorem for the Monte Carlo error in the following decomposition of the approximation error:

$$\widehat{\varphi}_N - \mathbb{E}_{x\sim\pi}\left[\varphi(x)\right] = \underbrace{\left(\widehat{\varphi}_N - \mathbb{E}_{x\sim\pi_h}\left[\varphi(x)\right]\right)}_{\text{Monte Carlo error}} + \underbrace{\left(\mathbb{E}_{x\sim\pi_h}\left[\varphi(x)\right] - \mathbb{E}_{x\sim\pi}\left[\varphi(x)\right]\right)}_{\text{Systematic bias}}. \tag{3.2}$$

In the next subsection (subsection 3.2) we provide an analysis of the stepsize-dependent systematic bias.

3.1. Ergodic properties and central limit theorem. In addition to showing the existence and uniqueness of the invariant measure π_h, we show geometric ergodicity of the Markov chain $(\widehat{\boldsymbol{x}}_k)_{k\in\mathbb{N}}$. Geometric ergodicity is equivalent to exponential convergence of the corresponding evolution operator $\hat{\mathcal{P}}_h^{\text{gle-XYZYX}}$ in some suitably weighted L_∞ space. By [7], the latter property implies the validity of a central limit theorem.

THEOREM 3.1. *Let Assumption* 1 *and Assumption* 3 *be satisfied, and let* $\mathcal{P}_h = \hat{\mathcal{P}}_h^{\text{gle-XYZYX}}$. *Fix* $l \in \mathbb{N}, l > 0$, *and consider*

$$\mathcal{K}_l(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = \left(\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x}\right)^l + 1,\ l \in \mathbb{N}, \tag{3.3}$$

where $\boldsymbol{C} \in \mathbb{R}^{n+m}$ *is a suitably chosen symmetric positive definite matrix (see subsection* SM1.1 *for details). Then, there exists* $h^* > 0$ *such that for any* $h \in (0, h^*)$

1. *the Markov chain associated with* $\mathcal{P}_h$ *has a unique invariant probability measure* π_h, *which admits a density with respect to the Lebesgue measure on* $\Omega_{\boldsymbol{x}}$ *and has finite moments, i.e.,*

$$\int_{\Omega_{\boldsymbol{x}}} \mathcal{K}_l \mathrm{d}\pi_h < \infty \quad \textit{for all} \quad l \in \mathbb{N}; \tag{3.4}$$

2. *there are constants* $C_l > 0$, $r_l \in (0,1)$ *such that*

$$\text{(3.5)} \quad \textit{for all } \varphi \in L^\infty_{\mathcal{K}_l} \quad \textit{for all } k \in \mathbb{N}, \quad \left\| (\mathcal{P}_h^k \varphi) - \mathbb{E}_{\pi_h} \varphi \right\|_{L^\infty_{\mathcal{K}_l}} \leq C_l r_l^k \|\varphi\|_{L^\infty_{\mathcal{K}_l}}.$$

A complete proof of this result can be found in section SM1 of the supplementary material. Here, we provide a brief outline: the proof of the theorem relies on an application of Theorem 1.2 of [20] (see also [37, 6] for similar results) and as such includes the standard steps commonly followed for proving geometric ergodicity of a Markov chain. We first show that under the conditions of Theorem 3.1 a minorization condition is satisfied.

Assumption 4 (minorization condition). Fix any $x_{\max} > 0$. There exists $h^* > 0$ such that for any $h \in (0, h^*)$, there is $\alpha > 0$ so that

$$\text{(3.6)} \qquad \text{for all } \varphi \in \mathcal{C}_0(\Omega_{\boldsymbol{x}}, \mathbb{R}), \qquad \inf_{|\boldsymbol{x}| \leq x_{\max}} \left(\mathcal{P}_h^k \varphi\right)(\boldsymbol{x}) \geq \alpha \int_{\Omega_{\boldsymbol{x}}} \varphi(x) \nu(dx),$$

where $k = 2$ and ν denotes the Lebesgue measure on $\Omega_{\boldsymbol{x}}$.

The validity of this minorization condition ensures that within any compact ball that is centered at the origin, the Markov chain $(\boldsymbol{x}_k)_{k \in \mathbb{N}}$ is mixing within a finite number of steps. As such it already ensures irreducibility of the Markov chain and thus guarantees the uniqueness of the invariant measure π_h provided the latter exists. In order to ensure the existence of an invariant measure and exponential converge to that measure, the existence of a suitable Lyapunov function is shown in the second step of the proof.

Assumption 5 (uniform Lyapunov condition). For any $l \in \mathbb{N}$, $l > 0$, there exist $h^* > 0$ and $a_l, b_l > 0$ such that for any h, $0 < h \leq h^*$,

$$\mathcal{P}_h \mathcal{K}_l \leq e^{-a_l h} \mathcal{K}_l + b_l h.$$

Given the validity of both Assumption 4 and Assumption 5, the remaining statements of Theorem 3.1 then follow as a consequence of [20, Theorem 1.2].

By [32, Corollary 2.26] exponential convergence in the sense of Theorem 3.1 implies that the operator $\mathrm{Id} - \mathcal{P}_h$, when constrained to the subspace

$$L^\infty_{\mathcal{K}_l,0} = \left\{ \varphi \in L^\infty_{\mathcal{K}_l} \ : \int \varphi \pi_h = 0 \right\},$$

is invertible and the corresponding inverse operator is bounded in terms of the operator norm induced by $\|\cdot\|_{L^\infty_{\mathcal{K}_l}}$. By the results in [7], this is sufficient for a functional central limit theorem to hold (3.7).

Corollary 3.2 (central limit theorem). *Let* $l \in \mathbb{N}$ *and* $\varphi \in L^\infty_{\mathcal{K}_l}$. *For sufficiently small* $h > 0$, *there is finite* $\widehat{\sigma}^2_\varphi > 0$ *so that*

$$\text{(3.7)} \qquad \sqrt{N} \left(\widehat{\varphi}_N - \int \varphi \, \mathrm{d}\pi_h \right) \xrightarrow[N \to +\infty]{\text{law}} \mathcal{N}(0, \widehat{\sigma}^2_\varphi).$$

3.2. Analysis of the systematic bias. In this section we provide results regarding the convergence order in h of the discretization bias in ergodic averages of symmetric splitting schemes. As discussed in subsection 2.3, the weak convergence order of these schemes is two, which together with the above shown ergodicity result

implies that also the convergence order of the systematic discretization bias in ergodic averages is at least two, i.e.,

$$\lim_{N\to\infty} \widehat{\varphi}_N = \int \varphi(\boldsymbol{x})\pi_h(\mathrm{d}\boldsymbol{x}) = \int \varphi(\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x}) + O(h^2),$$

as $h \to 0$. In what follows we discuss two special cases where the second order convergence can be improved upon. We first derive the explicit form of the measure π_h of the gle-BAOAB scheme in the situation where the target measure π is Gaussian. Secondly, we analyze the behavior of the discretization error of the gle-BAOAB scheme in the overdamped limit.

Remark 3.3. We point out that the gle-OABAO scheme is similar to the gle-BAOAB scheme in the sense that

$$\left(\hat{\mathcal{P}}_h^{\text{gle-OABAO}}\right)^n = e^{\frac{h}{2}\mathcal{L}_{\mathrm{O}}} e^{\frac{h}{2}\mathcal{L}_{\mathrm{A}}} \left(\hat{\mathcal{P}}_h^{\text{gle-BAOAB}}\right)^{n-1} e^{\frac{h}{2}\mathcal{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathcal{L}_{\mathrm{O}}}$$

for any $n \geq 1$. Under certain technical conditions which are detailed in [30, Lemma 24] this observation allows us to write the invariant measure of gle-OABAO as a transformation (by the adjoint of $e^{\frac{h}{2}\mathcal{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathcal{L}_{\mathrm{O}}}$) of the invariant measure of the gle-BAOAB scheme. Similarly, this observation allows us to obtain Monte Carlo estimates from gle-OABAO trajectories with the same statistical properties as Monte Carlo estimates obtained from a gle-BAOAB trajectory by employing a postprocessing step as detailed in [53].

3.2.1. Systematic bias for quadratic potentials. An important property of the gle-BAOAB scheme is that its invariant measure π_h, when applied to a system with quadratic potential, is such that the marginal in $\boldsymbol{q}$, $\pi_h(\mathrm{d}\boldsymbol{q})$, coincides with the marginal in $\boldsymbol{q}$ of the exact invariant measure $\pi(\mathrm{d}\boldsymbol{q}) \propto e^{-\beta \boldsymbol{q}\boldsymbol{\Omega}^{-1}\boldsymbol{q}}\,\mathrm{d}\boldsymbol{q}$ so that

$$\lim_{N\to\infty} \widehat{\varphi}_N = \int \varphi(\boldsymbol{q})\pi_h(\mathrm{d}\boldsymbol{q})$$

for $\varphi \in L^2(\pi(\mathrm{d}\boldsymbol{q}))$. More specifically, we have the following theorem.

THEOREM 3.4. *Let* $U(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T\boldsymbol{\Omega}\,\boldsymbol{q}$ *with* $\boldsymbol{\Omega} \in \mathbb{R}^{n\times n}$ *symmetric positive definite. The Gaussian measure* $\pi_h(\mathrm{d}\boldsymbol{x}) \propto \exp(-(\boldsymbol{x}-\boldsymbol{\mu}_h)^T\boldsymbol{V}_h^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_h))\mathrm{d}\boldsymbol{x}$*, with*

$$\boldsymbol{\mu}_h = \boldsymbol{0}, \quad \boldsymbol{V}_h = \beta^{-1}\,\mathrm{diag}\left(\boldsymbol{\Omega}^{-1},\ \left(1-h^2/4\right)\boldsymbol{M},\ \boldsymbol{Q}\right), \tag{3.8}$$

is invariant under gle-BAOAB *with* $\boldsymbol{Q}$ *as defined in Proposition* 1.1.

Proof. We take a dual perspective and show that the above specified Gaussian measure π_h is the unique stationary solution of the corresponding forward equation; i.e.,

$$\left(\hat{\mathcal{P}}_h^{\text{gle-BAOAB}}\right)^\dagger \pi_h = \pi_h, \tag{3.9}$$

where $(\hat{\mathcal{P}}_h^{\text{gle-BAOAB}})^\dagger = \exp(\frac{h}{2}\mathcal{L}_B^\dagger)\exp(\frac{h}{2}\mathcal{L}_A^\dagger)\exp(h\mathcal{L}_O^\dagger)\exp(\frac{h}{2}\mathcal{L}_A^\dagger)\exp(\frac{h}{2}\mathcal{L}_B^\dagger)$, is the forward operator of the gle-BAOAB scheme, which by construction is simply the concatenation of the forward operators corresponding to the respective B/A/O-steps in the order given by the splitting scheme. The action of these forward operators when applied to a multivariate Gaussian measure

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{V})\mathrm{d}\boldsymbol{x} \propto \exp\left(-(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)\mathrm{d}\boldsymbol{x}$$

is found to be

$$\exp\left(\frac{h}{2}\mathcal{L}_A^\dagger\right)\mathcal{N}(\,\cdot\,\,|\boldsymbol{\mu},\boldsymbol{V}) = \mathcal{N}(\,\cdot\,\mid \boldsymbol{\Psi}_A\boldsymbol{\mu}, \boldsymbol{\Psi}_A\boldsymbol{V}\boldsymbol{\Psi}_A^T),$$
$$\exp\left(\frac{h}{2}\mathcal{L}_B^\dagger\right)\mathcal{N}(\,\cdot\,\,|\boldsymbol{\mu},\boldsymbol{V}) = \mathcal{N}(\,\cdot\,\mid \boldsymbol{\Psi}_B\boldsymbol{\mu}, \boldsymbol{\Psi}_B\boldsymbol{V}\boldsymbol{\Psi}_B^T),$$

where

$$\boldsymbol{\Psi}_A = \begin{bmatrix} \boldsymbol{I}_n & \frac{h}{2}\boldsymbol{M}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_m \end{bmatrix}, \quad \boldsymbol{\Psi}_B = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} & \boldsymbol{0} \\ -\frac{h}{2}\boldsymbol{\Omega} & \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_m \end{bmatrix},$$

and

$$\exp(h\mathcal{L}_O^\dagger)\mathcal{N}(\,\cdot\,\,|\boldsymbol{\mu},\boldsymbol{V}) = \mathcal{N}(\,\cdot\,\,|\widetilde{\boldsymbol{F}_h}\boldsymbol{\mu}, \widetilde{\boldsymbol{F}_h}\boldsymbol{V}\widetilde{\boldsymbol{F}_h}^T + \widetilde{\boldsymbol{S}_h}\widetilde{\boldsymbol{S}_h}^T),$$

where

$$\widetilde{\boldsymbol{F}_h} = \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_h \end{pmatrix}, \quad \widetilde{\boldsymbol{S}_h} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_h \end{pmatrix}. \tag{3.10}$$

Thus, the Gaussian measure with density $\mathcal{N}(\,\cdot\,\,|\boldsymbol{\mu},\boldsymbol{V})$ is invariant under the action of $(\hat{\mathcal{P}}_h^{\text{gle-BAOAB}})^\dagger$ precisely when

$$\boldsymbol{\mu} = \boldsymbol{\Psi}_B\boldsymbol{\Psi}_A\widetilde{\boldsymbol{F}_h}\boldsymbol{\Psi}_A\boldsymbol{\Psi}_B\boldsymbol{\mu},$$
$$\boldsymbol{V} = \boldsymbol{\Psi}_B\boldsymbol{\Psi}_A\widetilde{\boldsymbol{F}_h}\boldsymbol{\Psi}_A\boldsymbol{\Psi}_B\boldsymbol{V}\boldsymbol{\Psi}_B^T\boldsymbol{\Psi}_A^T\widetilde{\boldsymbol{F}_h}^T\boldsymbol{\Psi}_A^T\boldsymbol{\Psi}_B^T + \boldsymbol{\Psi}_B\boldsymbol{\Psi}_A\widetilde{\boldsymbol{S}_h}\widetilde{\boldsymbol{S}_h}^T\boldsymbol{\Psi}_A^T\boldsymbol{\Psi}_B^T,$$

which are satisfied if $\boldsymbol{\mu} = \boldsymbol{\mu}_h$ and $\boldsymbol{V} = \boldsymbol{V}_h$ with $\boldsymbol{\mu}_h = 0$ and $\boldsymbol{V}_h$ is as specified in the proposition. Since for sufficiently small $h > 0$ the discretized dynamics are ergodic (see subsection 3.1), this solution is also the unique solution of (3.9). ∎

Remark 3.5. With the same techniques as in the proof of Theorem 3.4, one may show that for the same quadratic potential function, the invariant measure of the Markov chain generated by gle-ABOBA is identical to the invariant measure of gle-BAOAB and that the unique invariant measure of the Markov chains generated by gle-OBABO and gle-OABAO is the Gaussian measure $\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_h,\boldsymbol{V}_h)\mathrm{d}\boldsymbol{x}$ with $\boldsymbol{\mu}_h = \boldsymbol{0}$ and $\boldsymbol{V}_h = \beta^{-1}\,\mathrm{diag}((1-h^2/4)\boldsymbol{\Omega}^{-1},\ \boldsymbol{M},\ \boldsymbol{Q})$.

3.2.2. Superconvergence of gle-BAOAB in the overdamped limit. The gle-BAOAB scheme possesses a superconvergence property in the discrete time version of the overdamped limit (see subsection 4.2). That is, for observables $\varphi \in L^2(\pi)$ which are purely functions of the position variable $\boldsymbol{q}$, the incurred discretization bias of the corresponding ergodic average when computed using the gle-BAOAB scheme applied to rescaled process (QGLE-scaled) with $\lambda = 1$ and $\mu_1 = \mu_2 = \varepsilon^{-1}$ behaves as

$$\left|\int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{q})\pi_h(\mathrm{d}\boldsymbol{x}) - \mathbb{E}_\pi\varphi\right| = O(\epsilon h^2) + O(h^4) \tag{3.11}$$

as $\varepsilon \to 0$ and $h \to 0$. For sufficiently small values of ε, the magnitude of the leading order term of the discretization bias decreases linearly in ε. In particular, in the limit $\varepsilon = 0$, the leading error term $O(\epsilon h^2)$ in (3.11) vanishes. This results in the discretization bias to decrease at fourth order in h (instead of second order as one would expect by construction)—a property which we refer to as "superconvergence."

We formally show this result for a particle of unit mass in a one-dimensional positional domain and memory kernel corresponding to a matrix of the $\mathbf{\Gamma}$ of the generic form

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma_{1,1} & \Gamma_{1,2} \\ \Gamma_{2,1} & \Gamma_{2,2} \end{pmatrix} \in \mathbb{R}^2,$$

which is assumed to satisfy Assumption 1. Our derivation can be extended to more general forms of (QGLE), but we refrain from doing so in order to keep notation simple. As a starting point of the derivation we consider again a Taylor expansion of the evolution operator

$$\hat{\mathcal{P}}_h^{\text{gle-BAOAB}} = \mathrm{Id} + h\mathcal{A}_1 + h^3\mathcal{A}_3 + O(h^5), \tag{3.12}$$

where $\mathcal{A}_1 = \mathcal{L}_{\text{GLE}}$ and $\mathcal{A}_3$ is as defined in (2.5) with $X = \text{B}, Y = \text{A}$, and $Z = \text{O}$. By [32, Theorem 3.3] and under suitable regularity conditions on the generator $\mathcal{L}_{\text{GLE}}$ and on the operators $\mathcal{A}_k, k \geq 1$ (see Remark 3.6), there exists $h^* > 0$ so that the expectation of test functions $\varphi \in L^2(\pi)$ with respect to the perturbed invariant measure π_h can be expanded as

$$\int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{x})\pi_h(\mathrm{d}\boldsymbol{x}) = \mathbb{E}_\pi\varphi + h^2 \int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{x}) f_3(\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x}) + h^4 R_{\varphi,h} \tag{3.13}$$

with $|R_{\varphi,h}|$ uniformly bounded for $h \in (0, h^*]$. The correction term f_3 is obtained as the solution of

$$\mathcal{L}_{\text{GLE}}^* f_3 = \mathcal{A}_3^* \mathbf{1}, \tag{3.14}$$

where the explicit form of the right-hand side can be computed as

$$\begin{aligned} \mathcal{A}_3^*\mathbf{1} = &-\frac{1}{4}\beta\left(\boldsymbol{p}^2\Gamma_{1,1} + \boldsymbol{p}\,\boldsymbol{s}\Gamma_{2,1} - \beta^{-1}\Gamma_{1,1}\right)U''(\boldsymbol{q}) \\ &- \frac{1}{12}\beta\boldsymbol{p}^3U^{(3)}(\boldsymbol{q}) + \frac{1}{4}\beta\boldsymbol{p}U'(\boldsymbol{q})U''(\boldsymbol{q}). \end{aligned}$$

Here, and below, we denote $L^2(\pi)$-adjoint of an operator $\mathcal{A}$ by $\mathcal{A}^*$ so that $\langle \mathcal{A}g, f\rangle_{L^2(\pi)} = \langle g, \mathcal{A}^* f\rangle_{L^2(\pi)}$ for all $g, f \in L^2(\pi)$, where $\langle f, g\rangle_{L^2(\pi)} := \int fg \,\mathrm{d}\pi$. By virtue of the Fredholm alternative equation (3.14) possesses a solution iff $\langle g, \mathcal{A}_3^*\mathbf{1}\rangle_{L^2(\pi)} = 0$ for all functions g contained in the null space of $\mathcal{L}_{\text{GLE}}$. Since the SDE associated with the generator $\mathcal{L}_{\text{GLE}}$ is by assumption ergodic, the null space of $\mathcal{L}_{\text{GLE}}$ only contains constant functions for which $\langle g, \mathcal{A}_3^*\mathbf{1}\rangle_{L^2(\pi)} \propto \langle \mathbf{1}, \mathcal{A}_3^*\mathbf{1}\rangle_{L^2(\pi)} = 0$ is indeed true.

Finding a closed form solution of the PDE (3.14) is still intractable for general potentials. Instead, we employ a singular perturbation approach. Under the scaling $\lambda = 1, \mu_1 = \mu_2 = \varepsilon^{-1}$ the generator decomposes as $\mathcal{L}_{\text{GLE}} = \varepsilon^{-1}\mathcal{L}_{\text{O}} + \mathcal{L}_{\text{H}}$, and we can expand the solution f_3 in ε as $f_3 = f_{3,0} + \varepsilon f_{3,1} + \varepsilon^2 f_{3,2} + O(\varepsilon^3)$. By plugging this into (3.14) we get

$$\left(\frac{1}{\varepsilon}\mathcal{L}_{\text{O}}^* + \mathcal{L}_{\text{H}}^*\right)\left(f_{3,0} + \varepsilon f_{3,1} + \varepsilon^2 f_{3,2} + O(\varepsilon^3)\right) = \mathcal{A}_3^*\mathbf{1}, \tag{3.15}$$

from which we obtain the following collection of PDEs by equating powers of ε:

$$\mathcal{L}_{\text{O}}^* f_{3,0} = \frac{1}{4}\beta\boldsymbol{p}^2\Gamma_{1,1}U''(\boldsymbol{q}) - \frac{1}{4}\Gamma_{1,1}U''(\boldsymbol{q}) + \frac{1}{4}\beta s^T\Gamma_{2,1}\boldsymbol{p}U''(\boldsymbol{q}), \tag{3.16}$$

$$\mathcal{L}_{\text{H}}^* f_{3,0} + \mathcal{L}_{\text{O}}^* f_{3,1} = \frac{1}{12}\beta\boldsymbol{p}^3U^{(3)}(\boldsymbol{q}) - \frac{1}{4}\beta\boldsymbol{p}U'(\boldsymbol{q})U''(\boldsymbol{q}), \tag{3.17}$$

$$\mathcal{L}_{\text{H}}^* f_{3,1} + \mathcal{L}_{\text{O}}^* f_{3,2} = 0, \tag{3.18}$$

$$\mathcal{L}_{\text{H}}^* f_{3,i} + \mathcal{L}_{\text{O}}^* f_{3,i+1} = 0,\ i \geq 2. \tag{3.19}$$

Solving this system iteratively, we find (see section SM4 for details)

$$f_{3,0}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = -\frac{1}{8}\beta\boldsymbol{p}^2U''(\boldsymbol{q}) + \frac{1}{8}U''(\boldsymbol{q}),$$

which can be verified to satisfy $\int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{q}) f_{3,0}(\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x}) = 0$ for any observable $\varphi \in \mathcal{C}_P^\infty(\Omega_{\boldsymbol{x}},\mathbb{R})$ which is purely a function of $\boldsymbol{q}$. Thus, for such φ, (3.13) can be written as

$$\int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{q})\pi_h(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \mathbb{E}_\pi\varphi + \epsilon h^2 \int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{q}) f_{3,1}(\boldsymbol{q})\pi(\mathrm{d}\boldsymbol{x}) + O(\epsilon^2h^2) + O(h^4)$$

as $\varepsilon \to 0$ and $h \to 0$, which is the desired statement.

Remark 3.6. The formal error analysis can be made rigorous by showing that the remainder terms in expansions (3.13) and (3.15) are uniformly bounded for sufficient small h and ε, respectively. For the expansion (3.13) it would be sufficient to show that the conditions of [32, Theorem 3.3] are indeed satisfied. In particular, this would entail showing that the function set $\mathcal{C}_{P,0}^\infty(\Omega_{\boldsymbol{x}},\mathbb{R}) := \left\{\varphi \in \mathcal{C}_P^\infty(\Omega_{\boldsymbol{x}},\mathbb{R}) : \int \varphi \,\mathrm{d}\pi = 0\right\}$ is invariant under application of the operators $\mathcal{L}_{\mathrm{GLE}}^{-1}$ and $(\mathcal{L}_{\mathrm{GLE}}^*)^{-1}$, as well as that $\mathcal{C}_P^\infty(\Omega_{\boldsymbol{x}},\mathbb{R})$ is invariant under application of the operators $\mathcal{A}_k, k \in \mathbb{N}$. Analogous estimates have been shown in [24, 25, 47] for the generators of the overdamped Langevin equation, the underdamped Langevin equation, and Langevin equations with generalized kinetic energies, respectively. In order to make the expansion of (3.15) rigorous one would need to show—in analogy to a result shown in [30] for the underdamped Langevin equation—a uniform hypocoercivity property of the form: there is a $K > 0$ such that $\|(\varepsilon^{-1}\mathcal{L}_{\mathrm{O}} + \mathcal{L}_{\mathrm{H}})^{-1}\varphi\|_{H^1(\pi)} \le K\|\varphi\|_{H^1(\pi)}$ for any $\varepsilon > 0$ and all test functions φ contained in the weighted Sobolev space $H^1(\pi)$ which are such that for almost all $\boldsymbol{q}$ the mean with respect to the marginal measure $\pi(\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{s})$ vanishes.

4. White noise and overdamped limit of the gle-BAOAB method. In this section we analyze the behavior of the gle-BAOAB splitting method in the overdamped and white noise limit discussed in subsection 1.5. For this purpose consider the stochastic flow map of the gle-BAOAB method when applied to the rescaled process (QGLE-scaled),

$$\hat{\Phi}_{h,\mu}^{\text{gle-BAOAB}} = \Phi_{h/2}^{\mathrm{B}} \circ \Phi_{h/2}^{\mathrm{A}} \circ \Phi_{h,\mu}^{\mathrm{O}} \circ \Phi_{h/2}^{\mathrm{A}} \circ \Phi_{h/2}^{\mathrm{B}}, \tag{4.1}$$

where

$$\Phi_{h,\mu}^{\mathrm{O}} : (\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) \mapsto (\boldsymbol{q}, \boldsymbol{F}_h^\mu\,(\boldsymbol{p},\boldsymbol{s})^T + \boldsymbol{S}_h^\mu\mathscr{R}),\ \mathscr{R} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n+m}), \tag{4.2}$$

with

$$\boldsymbol{F}_h^\mu := \exp(-h\boldsymbol{\Gamma}^\mu), \quad (\boldsymbol{S}_h^\mu)^T\,\boldsymbol{S}_h^\mu = \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} - \boldsymbol{F}_h^\mu \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} \boldsymbol{F}_h^{\mu T},$$

and $\boldsymbol{\Gamma}^\mu$ as defined in (1.11). In both limits gle-BAOAB converges to state-of-the-art numerical integration schemes for the corresponding limiting dynamics which have been shown exhibit particularly low discretization bias (see [28, 30]).

4.1. White noise limit. The gle-BAOAB integration scheme, when applied to the rescaled process (QGLE-scaled) with $\lambda = 1, \mu_1 = \epsilon^{-1}$, and $\mu_2 = \epsilon^{-2}$, reduces to the ld-BAOAB discretization of an underdamped Langevin equation in the white noise limit $\epsilon \to 0$. More precisely, we have the following result.

THEOREM 4.1 (white noise limit of gle-BAOAB). *Let* $(\boldsymbol{q}_k, \boldsymbol{p}_k, \boldsymbol{s}_k)_{k\in\mathbb{N}}$ *be the Markov chain obtained with the* gle-BAOAB *method* $\hat{\Phi}^{\text{gle-BOAOB}}_{h,(\epsilon^{-1},\epsilon^{-2})}$. *Let* $(\hat{\boldsymbol{q}}_k, \hat{\boldsymbol{p}}_k)_{k\in\mathbb{N}}$ *denote the Markov chain generated by the* ld-BAOAB *method of* [28] *(see also Algorithm* SM2.1*) when applied to* (LD) *with friction tensor* $\widehat{\boldsymbol{\Gamma}} = \boldsymbol{D}_a^2\boldsymbol{D}_b^{-1}$, *diffusion tensor* $\widehat{\boldsymbol{\Sigma}} = \sqrt{2}\boldsymbol{D}_a\boldsymbol{D}_b^{-1/2}$, *and stepsize* h. *Then, for all* $N \in \mathbb{N}$, *we have*

$$(\boldsymbol{q}_k, \boldsymbol{p}_k)_{0\le k\le N} \xrightarrow[\varepsilon\to 0]{\text{law}} (\hat{\boldsymbol{q}}_k, \hat{\boldsymbol{p}}_k)_{0\le k\le N}.$$

Proof. Since $(\hat{\boldsymbol{q}}_k, \hat{\boldsymbol{p}}_k)_{k\in\mathbb{N}}$ is a Markov process, it is sufficient to show that the transition probabilities converge appropriately; i.e.,

$$\Pi\left(\hat{\Phi}^{\text{gle-BAOAB}}_{h,(\varepsilon^{-1},\varepsilon^{-2})}(q,p,s)\right) \xrightarrow[\varepsilon\to 0]{\text{law}} \hat{\Phi}^{\text{ld-BAOAB}}_{h}(q,p)$$

for all $(q,p,s) \in \Omega_{\boldsymbol{x}}$, where $\hat{\Phi}^{\text{ld-BAOAB}}_{h,\mu}$ denotes the stochastic flow map of the ld-BAOAB splitting scheme and $\Pi : (q,p,s) \mapsto (q,p)$ denotes the projection operator on the position and momentum component.

The two methods only differ in terms of their respective O-steps. It is therefore sufficient to show that in the limit $\varepsilon \to 0$, these become identical in distribution, which is exactly the case if

$$\lim_{\epsilon\to 0} \exp\left(-h\boldsymbol{\Gamma}^{(\epsilon^{-1},\epsilon^{-2})}\right) = \begin{pmatrix} \exp(-h\boldsymbol{D}_a^2\boldsymbol{D}_b^{-1}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix},$$

where

$$\boldsymbol{\Gamma}^{(\epsilon^{-1},\epsilon^{-2})} = \begin{pmatrix} \boldsymbol{0} & -\epsilon^{-1}\boldsymbol{D}_a \\ \epsilon^{-1}\boldsymbol{D}_a & \epsilon^{-2}\boldsymbol{D}_b \end{pmatrix}.$$

We show this by applying a suitable similarity transformation: without loss of generality let $h = 1$, and consider the orthogonal matrix

$$\boldsymbol{O} = \hat{\boldsymbol{I}}^{(2n)}_{2n-1,2n}\hat{\boldsymbol{I}}^{(2n)}_{2n-3,2n-1}\cdots\hat{\boldsymbol{I}}^{(2n)}_{5,n+3}\hat{\boldsymbol{I}}^{(2n)}_{3,n+2}\hat{\boldsymbol{I}}^{(2n)}_{1,n+1},$$

where $\hat{\boldsymbol{I}}^{(2n)}_{i,j}$ denotes the elementary matrix whose action when multiplied from the left to a matrix $\boldsymbol{A} \in \mathbb{R}^{2n\times 2n}$ corresponds to a swap of ith and jth rows of $\boldsymbol{A}$ so that

$$\boldsymbol{\Gamma}^{(\epsilon^{-1},\epsilon^{-2})} = \boldsymbol{O}^T \operatorname{diag}(A_1^\epsilon, A_2^\epsilon, \ldots, A_n^\epsilon)\,\boldsymbol{O}, \quad \text{with } A_i^\epsilon = \begin{pmatrix} 0 & -\epsilon^{-1}a_i \\ \epsilon^{-1}a_i & \epsilon^{-2}b_i \end{pmatrix},\ 1 \le i \le n.$$

By Lemma SM3.1 we have $\lim_{\epsilon\to 0} A_i^\epsilon = B_i$ with $B_i = \left(\begin{smallmatrix} e^{-a_i^2/b_i} & 0 \\ 0 & 0 \end{smallmatrix}\right)$. Thus,

$$\begin{aligned}
\lim_{\epsilon\to 0} \exp\left(-\boldsymbol{\Gamma}^{(\epsilon^{-1},\epsilon^{-2})}\right) &= \lim_{\epsilon\to 0} \boldsymbol{O}^T \operatorname{diag}(\exp(-A_1^\epsilon), \ldots, \exp(-A_n^\epsilon))\,\boldsymbol{O} \\
&= \boldsymbol{O}^T \operatorname{diag}\left(\lim_{\epsilon\to 0}\exp(-A_1^\epsilon), \ldots, \lim_{\epsilon\to 0}\exp(-A_n^\epsilon)\right)\boldsymbol{O} \\
&= \boldsymbol{O}^T \operatorname{diag}(B_1, \ldots, B_n)\,\boldsymbol{O} = \begin{pmatrix} \exp(-h\boldsymbol{D}_a^2\boldsymbol{D}_b^{-1}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}. \qquad \square
\end{aligned}$$

4.2. Overdamped limit. When applied to the rescaled process (QGLE-scaled) with $\lambda = 1, \mu_2 = \epsilon^{-1}$, and $\mu_3 = \epsilon^{-1}$ the gle-BAOAB method reduces to the BAOAB-limit method (the "Leimkuhler–Matthews method") of [28] in the asymptotic limit $\epsilon \to 0$ as shown in the following.

THEOREM 4.2. *Let* $(\boldsymbol{q}_k, \boldsymbol{p}_k, \boldsymbol{s}_k)_{k\in\mathbb{N}}$ *be the Markov chain obtained by with the* gle-BAOAB *method* $\hat{\Phi}^{\text{gle-BOAOB}}_{h,(\epsilon^{-1},\epsilon^{-1})}$ *with* $\boldsymbol{p}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$. *Let* $(\widetilde{\boldsymbol{q}}_k)_{k\in\mathbb{N}}$ *denote the Markov chain generated by the* BAOAB*-limit method,*

$$\widetilde{\boldsymbol{q}}_{k+1} \leftarrow \widetilde{\boldsymbol{q}}_k - \tilde{h}\Lambda\nabla U(\boldsymbol{q}_k) + \sqrt{2\tilde{h}\Lambda}\frac{1}{2}(\widetilde{\mathscr{R}}_k + \widetilde{\mathscr{R}}_{k+1}), \tag{4.3}$$

with $\widetilde{\boldsymbol{q}}_0$ *and* $\boldsymbol{q}_0$ *being identically distributed,* $\widetilde{\mathscr{R}}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$, $k \in \mathbb{N}$ *independent, stepsize* $\tilde{h} = h^2/2$, *and* $\Lambda = \boldsymbol{I}_n$. *Then, for all* $N \in \mathbb{N}$, *we have*

$$(\boldsymbol{q}_k)_{0\le k\le N} \xrightarrow[\varepsilon\to 0]{\text{law}} (\widetilde{\boldsymbol{q}}_k)_{0\le k\le N}.$$

Proof. For $\mu = (\varepsilon, \varepsilon^{-1})$, we have $\boldsymbol{\Gamma}^\mu = \varepsilon^{-1}\left(\begin{smallmatrix} \boldsymbol{0} & -\boldsymbol{D}_a \\ \boldsymbol{D}_a & \boldsymbol{D}_b \end{smallmatrix}\right)$; thus,

$$\boldsymbol{F}_h = \exp\left(-\frac{h}{\epsilon}\begin{pmatrix} \boldsymbol{0} & -\boldsymbol{D}_a \\ \boldsymbol{D}_a & \boldsymbol{D}_b \end{pmatrix}\right) \to \boldsymbol{0} \quad \text{as} \quad \epsilon \to 0,$$

and therefore also $\boldsymbol{S}_h \to \beta^{-1/2}\boldsymbol{I}_{n+m}$ as $\epsilon \to 0$. Thus,

$$\lim_{\epsilon\to 0} \Phi^{\text{O}}_{h,(\epsilon^{-1},\epsilon^{-1})}(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s}) = (\boldsymbol{q}, \beta^{-1/2}\mathscr{R}),\ \mathscr{R} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n+m}),$$

which removes any coupling between the auxiliary variable $\boldsymbol{s}$ and $(\boldsymbol{q}, \boldsymbol{p})$. Consequently, in the limit of $\varepsilon \to 0$ we can disregard the $\boldsymbol{s}$-component in the corresponding updating sequence of the positions and momenta. Moreover, since the momentum variables are independently resampled at every iteration, we can eliminate the momentum component from the updating sequence to obtain (4.3) with $\Lambda = \boldsymbol{I}_n$, $\tilde{h} = h^2/2$, and $\mathscr{R}_{-1} = \boldsymbol{p}_0$. ∎

5. Numerical experiments. In this section we assess the performance of the splitting methods which we introduced in section 2 in numerical experiments.

5.1. Comparison of proposed splitting schemes. We first compare the performance of the methods discussed in this article against each other. For this purpose we consider a simple QGLE on a one-dimensional positional domain with potential function

$$U_{DW}(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^2 + \sin(1/4 + 2\boldsymbol{q}), \tag{5.1}$$

which is an uneven double-well. We evaluate the performance in terms of the incurred stepsize-dependent discretization bias for observables which are purely functions of the position variable. For the parameterization of the noise process in the GLE we consider the memory kernels

$$\boldsymbol{K}(t) = 2^r K(t2^r), \quad K(t) := \frac{5}{2}\exp(-t/4) + \frac{1}{2}\exp(-t/8), \tag{5.2}$$

where we let r take values in $\{0, 1, 2\}$. The r-dependent parameterization of the memory kernels is chosen such that in the limit of $r \to \infty$, the corresponding GLE approaches an underdamped Langevin equation.

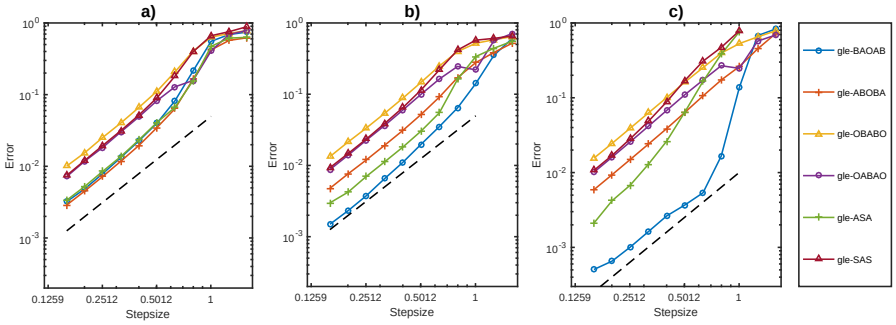

FIG. 2. *Log-log plot of the mean approximate error of numerical integrators proposed in this article. Results for the GLE with potential function* (5.1) *and memory kernel* (5.2) *with* $r = 0, 1, 2$ *are shown in panels* (a), (b), *and* (c), *respectively. Any missing error value indicates numerical instability of the respective method for the corresponding stepsize. The dashed black line corresponds to a second order decay. Details on the integrators* gle-ASA *and* gle-SAS *can be found in subsection* SM5.3.

We consider as an error measure

$$\mathcal{E}_{\mathrm{MAE}}((\boldsymbol{q}_k)_{1\le k\le N}) = n_B^{-1}\sum_{i=1}^{n_B}\left|\left(\frac{1}{N}\sum_{k=0}^{N-1}\mathbb{1}_{B_i}(\boldsymbol{q}_k) - \int_{\Omega_{\boldsymbol{x}}}\mathbb{1}_{B_i}(\boldsymbol{q})\pi(\mathrm{d}\boldsymbol{x})\right)\right|,$$

where the equal-sized bins $B_i \subset \mathbb{R}, i = 1, \dots, n_B$, are chosen such that they form a partition of an interval $[a, b] \subset \mathbb{R}$, which contains 99.99% of the probability mass of the Gibbs measure associated with $\mathbb{R}$. The quantity $\mathcal{E}_{\mathrm{MAE}}((\boldsymbol{q}_k)_{1\le k\le N})$ may be considered as the mean approximate error (MAE) of the discretization bias incurred for the observables $\varphi_{B_i} : \boldsymbol{q} \mapsto \mathbb{1}_{B_i}(\boldsymbol{q}), i = 1, \dots, n_B$, or as an estimate of the total variation distance between the perturbed invariant measure π_h and the exact target measure π.

In total, 100 trajectories, all initialized in accordance with the exact equilibrium distribution π, were simulated over a physical time period of length $T = hN = 10^7$ to obtain the statistics.

Figure 2 shows $\mathcal{E}_{\mathrm{MAE}}$ for the splitting schemes discussed in section 2. All methods displayed are by construction second order. Differences in performance are thus measured in terms of the magnitude of the corresponding prefactors of the leading error term. We find that the discretization error incurred in gle-OBABO and gle-OABAO is comparable and is not noticeably affected by the parameterization of the memory kernel. In comparison to that, the discretization error of gle-BAOAB and gle-ABOBA is smaller, and the accuracy of gle-BAOAB improves significantly with increasing value of r in the parameterization of the memory kernel.

5.2. Comparison with previously proposed GLE schemes. We next compare the performance of gle-BAOAB with methods previously proposed in the literature using the same setup as in subsection 5.1. We compare the gle-BAOAB method with the methods proposed in [5] (BB-BAOB, BB-BACOCAB), [51] (KLS-OBOAB), and [12] (gle-OBABO). These methods are all constructed as weak second order schemes.

For moderate variance and slowly decaying autocorrelation of the noise process (that is, $r = 0, 1$) we observe that the error incurred by the methods BB-BACOCAB and KLS-OBABO is very similar to the error of the gle-BAOAB method (Figure 3(a) and (b)). For all choices of the memory kernel, the error in gle-OBABO

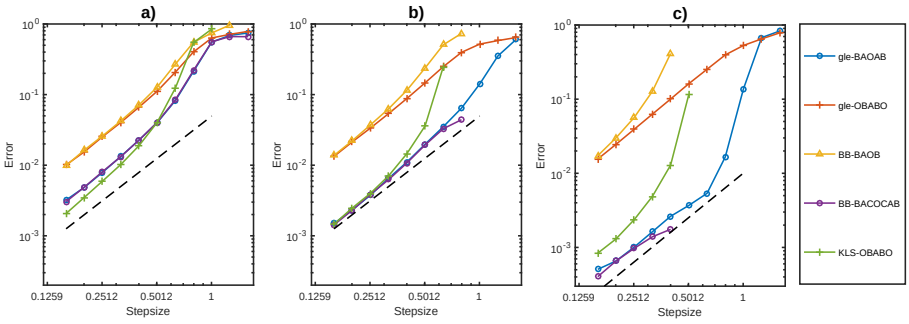

FIG. 3. *Different integrators but otherwise same setup as in Figure* 2.

and BB-BAOB is at least by a factor of 10 higher than the error of gle-BAOAB, and this factor increases further with an increasing value of r. Similarly, with an increasing value of r, the accuracy of the KLS-OBABO method decreases in comparison to that of gle-BAOAB. The maximum admissible stepsize of BB-BAOB and BB-BACOCAB decreases significantly, while the maximum admissible stepsize for schemes discussed in section 2 is not affected. The high accuracy of BB-BACOCAB is not surprising, as Baczewski and Bond in [5] specifically design this method for the sampling of accurate configurational averages. Interestingly, the KLS-OBABO method exhibits comparable accuracy even though the construction of this numerical scheme is not based on a systematic analysis of the discretization error in configurational averages. It is important to note that the scope of the memory kernels to which the methods proposed in [5] are applicable is very limited in comparison to the class of memory kernels which can be simulated using gle-BAOAB and gle-OBABO. Similarly, the scope of applications for which KLS-OBABO is designed is different from the sampling applications which are the focus of this article. As such the KLS-OBABO scheme is not applicable to GLEs which are parameterized with the type of kernels proposed in [12, 54].

5.3. Parameter-dependent accuracy of gle-BAOAB. In order to support the results derived by the singular perturbation ansatz in subsection 3.2.2, we evaluate the sampling accuracy of gle-BAOAB when applied to a GLE with a simple exponentially decaying memory kernel, i.e.,

$$\boldsymbol{K}(t) = \gamma e^{-t/\tau}, \gamma > 0, \tau > 0, \tag{5.3}$$

and the potential function (5.1). As predicted we find that the discretization bias decreases as the overdamped limit is approached (see Figure 4(b). Moreover, for parameter values $\lambda = 128$ and $\tau = 1/16$, we find the predicted fourth order decay of the discretization bias as h tends to 0. For the chosen range of parameter values we further observe (i) a decrease of the MAE in the white noise limit (Figure 4(a)), (ii) a decrease of the MAE for fixed decay rate $\tau = 1$ as the prefactor γ increases (Figure 4(d)), and (iii) no systematic change of the magnitude of the MAE for fixed prefactor $\gamma = 4$ and varying decay rate τ (Figure 4(c)).

5.4. Application to Bayesian posterior sampling. In this section, we consider the Bayesian parameterization of a Gaussian mixture model as a simple application of the discussed sampling methods, and we use this application to demonstrate how the gle-BAOAB method when used in combination with the GLE-dynamics developed in [12] results in a sampling scheme which has drastically improved sampling

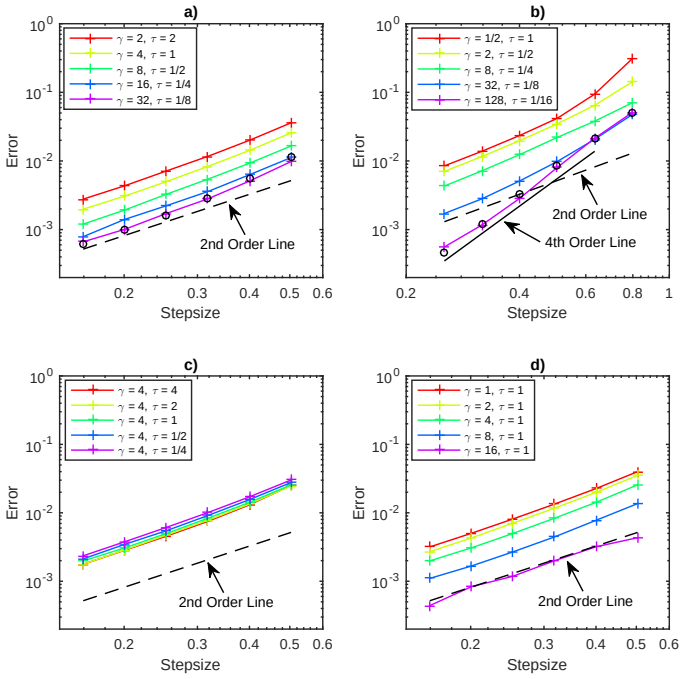

FIG. 4. *Log-Log plot of stepsize versus MAE of* gle-BAOAB *applied to the GLE with potential function* (5.1) *and memory kernel* $K(t) = \gamma e^{-t/\tau}$. *Black circles show the observed MAE of the respective limiting dynamics.*

properties in comparison to BAOAB discretizations of the underdamped Langevin equation as well as in comparison to the sampling scheme proposed in the above-mentioned reference, which in the language used in this article corresponds to the gle-OBABO method.

As a benchmark system we consider a Bayesian Gaussian mixture model applied to the Hidalgo stamp data set [21], which consists of the measurements $\{x_i\}_{i=1}^N \subset \mathbb{R}$ of the thicknesses of $N = 482$ postage stamps from a certain Mexican issue of postage stamps from the year 1872. We parameterize the model as described in [14] (see also [49] and [22]). Specifically we choose the number of components to be $N_c = 3$ and assume isotropic Gaussian components resulting in a parameter vector $\boldsymbol{q} = ((\boldsymbol{w}_k)_{1\le k\le 3}, (\boldsymbol{\mu}_k, \boldsymbol{\lambda}_k)_{1\le k\le 3}, \beta) \in \boldsymbol{\Delta}^3 \times \mathbb{R}^6 \times \mathbb{R}$, where $\boldsymbol{\Delta}^3$ denotes the standard simplex in $\mathbb{R}^3$, $\boldsymbol{w}_k$ is the weight parameter, $\boldsymbol{\mu}_k, \boldsymbol{\lambda}_k$ are the mean and precision of the kth Gaussian component, respectively, and $\beta \in \mathbb{R}$ denotes an additional hyper-parameter of the prior distribution. The resulting target distribution is then given as the Gibbs measure of the corresponding negative log-posterior function

$$U(\boldsymbol{q}) = -\sum_{i=1}^{N} \log p\,(x_i \mid \boldsymbol{q}) - \log p_{\text{prior}}(\boldsymbol{q}), \tag{5.4}$$

where the exact form of the likelihood function $p(x_i \mid \boldsymbol{q})$ and the prior $p_{\text{prior}}(\boldsymbol{q})$ are both specified in subsection SM5.1 of the supplementary material.

We parameterize both the gle-BAOAB scheme and the gle-OBABO scheme with the preoptimized memory kernel kv-8-8 obtained from the website GLE4MD [9] (see also subsection SM5.2). Even in this very modest example, the computational cost incurred by simulating the additional 8 auxiliary variables per parameter in the GLE

dynamics is negligible in comparison to the computational cost of the evaluation of the gradient of the log-likelihood over the whole data set. Thus, computational costs per time step for the considered GLE schemes and discretizations of the underdamped Langevin equation are very similar. In the cases with more complicated gradients or larger data sets the differences would only further decrease. We compare the performance of the sampling schemes:

(i) in terms of the observed discretization bias which we measure by the relative error incurred for the variable specific configurational temperatures

$$\varphi_{CT,i}(\boldsymbol{q}) = \boldsymbol{q}_i \partial_{\boldsymbol{q}_i} U(\boldsymbol{q}), \quad i = 1, \ldots, 9, \textit{and}$$

(ii) in terms of mixing which we measure by estimates of the integrated autocorrelation times

$$\tau_i = \int_0^\infty \mathbb{E}\left[(\boldsymbol{q}_i(t) - \mu_{\boldsymbol{q}_i})(\boldsymbol{q}_i(0) - \mu_{\boldsymbol{q}_i})\right] \mathrm{d}t, \quad i = 1, \ldots, 9,$$

where $\boldsymbol{q}(0) \sim \pi$ and $\mathbb{E}[\cdot]$ is the expectation with respect to $\boldsymbol{q}(0)$ and the Wiener process $\boldsymbol{W}$ in (QGLE).

For ld-BAOAB we considered the commonly used parameterization with a single scalar friction coefficient, i.e., $\widehat{\boldsymbol{\Gamma}} = \gamma \boldsymbol{I}_n$. The simulation run corresponding to the parameter values $\gamma = 1.0$ and $h = .01$ was obtained as the result of minimizing the integrated autocorrelation time for the slowest parameters by varying the stepsize after fixing the friction coefficient to $\gamma = 1.0$. The simulation run corresponding to the parameter values $\gamma = .1$ with $h = .01$ was obtained as the result of minimizing the integrated autocorrelation time for the "slowest parameter" (i.e., the parameter with the largest associated integrated autocorrelation time) by simultaneously optimizing both the stepsize as well as the friction coefficient γ. The results reported for gle-BAOAB and gle-OBABO use a stepsize $h = .02$, which was determined approximately as the maximum admissible stepsize with a few (short) test runs. We find that, in terms of sampling efficiency which we measured in terms of the integrated autocorrelation time of the "slowest" sampled parameter λ_1, the GLE schemes clearly outperform these as Figure 5 shows. Among the GLE schemes, we find that the discretization error in the

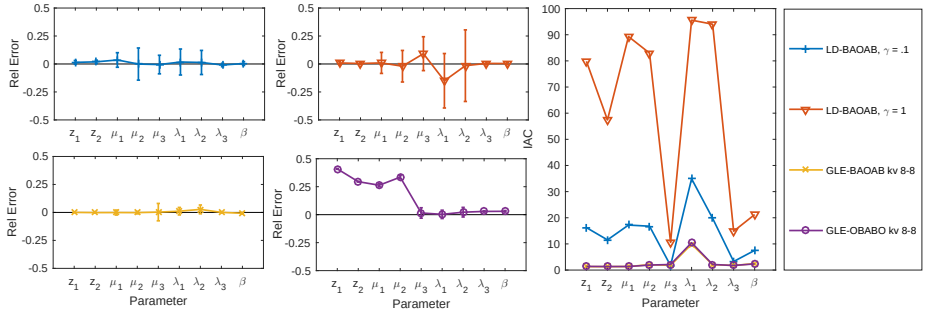

FIG. 5. *Sampling statistics for the numerical experiments performed on the Hidalgo stamp data set. The rightmost panel shows the integrated autocorrelation time (labeled as IAC) for each sampled parameter of the Gaussian mixture model. The remaining four panels on the left show the discretization bias incurred for the variable-specific configurational temperature. Note that since the IACs of gle-BAOAB and gle-OBABO are almost identical the corresponding graph of the former is covered by the graph of the latter.*

sample obtained from gle-BAOAB is significantly smaller than the discretization error in the sample obtained with gle-OBABO. The improvement in terms of the maximum admissible stable stepsize of the GLE methods in comparison to the Langevin schemes is an interesting feature. Presumably, this is due to resonance effects which occur in the discretized dynamics of the underdamped Langevin due to insufficient damping of fast frequency modes for the tuned value of the friction coefficient.

Acknowledgments. We thank Gabriel Stoltz for helpful discussions. We also wish to thank the referees for their generous assistance in refining the manuscript.

REFERENCES

[1] A. Abdulle, G. Vilmart, and K. C. Zygalakis, *High order numerical approximation of the invariant measure of ergodic SDES*, SIAM J. Numer. Anal., 52 (2014), pp. 1600–1622.
[2] A. Abdulle, G. Vilmart, and K. C. Zygalakis, *Long time accuracy of Lie-Trotter splitting methods for Langevin dynamics*, SIAM J. Numer. Anal., 53 (2015), pp. 1–16.
[3] S. Adelman and J. Doll, *Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids*, J. Chem. Phys., 64 (1976), pp. 2375–2388.
[4] T. Ando and Y. Sugita, *Algorithms for Brownian dynamics simuluation*, in Quantum Bio-Informatics IV: From Quantum Information to Bio-Informatics - Proceedings of Quantum Bio-Informatics 2014, World Scientific, River Edge, NJ, 2020, pp. 29–36.
[5] A. D. Baczewski and S. D. Bond, *Numerical integration of the extended variable generalized Langevin equation with a positive Prony representable memory kernel*, J. Chem. Phys., 139 (2013), 044107.
[6] L. R. Bellet, *Ergodic properties of Markov processes*, in Open Quantum Systems II, Springer, Berlin, 2006, pp. 1–39.
[7] R. N. Bhattacharya, *On the functional central limit theorem and the law of the iterated logarithm for Markov processes*, Z. Wahrscheinlichkeit., 60 (1982), pp. 185–201.
[8] N. Bou-Rabee and H. Owhadi, *Long-run accuracy of variational integrators in the stochastic context*, SIAM J. Numer. Anal., 48 (2010), pp. 278–297.
[9] M. Ceriotti, *Gle4md*, http://gle4md.org.
[10] M. Ceriotti, *A Novel Framework for Enhanced Molecular Dynamics Based on the Generalized Langevin Equation*, Doctoral thesis, ETH Zürich, 2010.
[11] M. Ceriotti, G. Bussi, and M. Parrinello, *Langevin equation with colored noise for constant-temperature molecular dynamics simulations*, Phys. Rev. Lett., 102 (2009), 020601.
[12] M. Ceriotti, G. Bussi, and M. Parrinello, *Colored-noise thermostats à la carte*, J. Chem. Theory Comput., 6 (2010), pp. 1170–1180.
[13] M. Chak, N. Kantas, and G. A. Pavliotis, *On the Generalised Langevin Equation for Simulated Annealing*, preprint, arXiv:2003.06448, 2020.
[14] N. Chopin, T. Lelièvre, and G. Stoltz, *Free energy methods for Bayesian inference: Efficient exploration of univariate Gaussian mixture posteriors*, Stat. Comput., 22 (2012), pp. 897–916.
[15] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet, *Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures*, Comm. Math. Phys., 201 (1999), pp. 657–697.
[16] J. Fricks, L. Yao, T. C. Elston, and M. G. Forest, *Time-domain methods for diffusive transport in soft matter*, SIAM J. Appl. Math., 69 (2009), pp. 1277–1308.
[17] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer, Berlin, 1994.
[18] D. Givon, R. Kupferman, and A. Stuart, *Extracting macroscopic dynamics: Model problems and algorithms*, Nonlinearity, 17 (2004), R55.
[19] F. Grogan, H. Lei, X. Li, and N. A. Baker, *Data-driven molecular modeling with the generalized langevin equation*, J. Comput. Phys., 418 (2020), 109633.
[20] M. Hairer and J. C. Mattingly, *Yet another look at Harris ergodic theorem for Markov chains*, in Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab. 63, Springer, Basel, Switzerland, 2011, pp. 109–117.
[21] A. J. Izenman and C. J. Sommer, *Philatelic mixtures and multimodal densities*, J. Amer. Statist. Assoc., 83 (1988), pp. 941–953.

[22] A. Jasra, C. C. Holmes, and D. A. Stephens, *Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling*, Statist. Sci., 20 (2005), pp. 50–67.
[23] L. Kantorovich, *Generalized Langevin equation for solids.* I. *Rigorous derivation and main properties*, Phys. Rev. B, 78 (2008), 094304.
[24] M. Kopec, *Weak backward error analysis for overdamped Langevin processes*, IMA J. Numer. Anal., 35 (2014), pp. 583–614.
[25] M. Kopec, *Weak backward error analysis for Langevin process*, BIT, 55 (2015), pp. 1057–1103.
[26] H. S. Lee, S.-H. Ahn, and E. F. Darve, *The multi-dimensional generalized Langevin equation for conformational motion of proteins*, J. Chem. Phys., 150 (2019), 174113.
[27] H. Lei, N. A. Baker, and X. Li, *Data-driven parameterization of the generalized Langevin equation*, Proc. Natl. Acad. Sci. USA, 113 (2016), pp. 14183–14188.
[28] B. Leimkuhler and C. Matthews, *Robust and efficient configurational molecular sampling via Langevin dynamics*, J. Chem. Phys., 138 (2013), 05B601_1.
[29] B. Leimkuhler and C. Matthews, *Efficient molecular dynamics using geodesic integration and solvent–solute splitting*, Proc. A, 472 (2016), 20160138.
[30] B. Leimkuhler, C. Matthews, and G. Stoltz, *The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics*, IMA J. Numer. Anal., 36 (2015), pp. 13–79.
[31] B. Leimkuhler and M. Sachs, *Ergodic properties of quasi-Markovian generalized Langevin equations with configuration dependent noise and non-conservative force*, in Stochastic Dynamics Out of Equilibrium, Springer, Cham, Switzerland, 2017, pp. 282–330.
[32] T. Lelièvre and G. Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*, Acta Numer., 25 (2016), pp. 681–880.
[33] Z. Li, X. Bian, X. Li, and G. E. Karniadakis, *Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism*, J. Chem. Phys., 143 (2015), 243128.
[34] Z. Li, H. S. Lee, E. Darve, and G. E. Karniadakis, *Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: Application to polymer melts*, J. Chem. Phys., 146 (2017), 014104.
[35] S. H. Lim and J. Wehr, *Homogenization of a Class of Non-Markovian Langevin Equations with an Application to Thermophoresis*, preprint, arXiv:1704.00134, 2017.
[36] S. A. McKinley, L. Yao, and M. G. Forest, *Transient anomalous diffusion of tracer particles in soft matter*, J. Rheol., 53 (2009), pp. 1487–1506.
[37] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, London, 2012.
[38] G. Mil'shtein, *Weak approximation of solutions of systems of stochastic differential equations*, Theory Probab. Appl., 30 (1986), pp. 750–766.
[39] H. Mori, *A continued-fraction representation of the time-correlation functions*, Prog. Theor. Phys., 34 (1965), pp. 399–416.
[40] J. A. Morrone, T. E. Markland, M. Ceriotti, and B. Berne, *Efficient multiple time scale molecular dynamics: Using colored noise thermostats to stabilize resonances*, J. Chem. Phys., 134 (2011), 014103.
[41] W. Mou, Y.-A. Ma, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan, *High-Order Langevin Diffusion Yields an Accelerated MCMC Algorithm*, preprint, arXiv:1908.10859, 2019.
[42] H. Ness, L. Stella, C. Lorenz, and L. Kantorovich, *Applications of the generalized Langevin equation: Towards a realistic description of the baths*, Phys. Rev. B, 91 (2015), 014301.
[43] H. Ness, L. Stella, C. Lorenz, and L. Kantorovich, *Nonequilibrium generalised Langevin equation for the calculation of heat transport properties in model* $1D$ *atomic chains coupled for two* $3D$ *thermal baths*, J. Chem. Phys., 146 (2017), 164103, https://doi.org/10.1063/1.4981816.
[44] M. Ottobre and G. Pavliotis, *Asymptotic analysis for the generalized Langevin equation*, Nonlinearity, 24 (2011), p. 1629.
[45] G. A. Pavliotis, *Stochastic Processes and Applications*, Springer, New York, 2014.
[46] G. A. Pavliotis, G. Stoltz, and U. Vaes, *Scaling limits for the generalized Langevin equation*, J. Nonlinear Sci., 31 (2021), pp. 1–58.
[47] S. Redon, G. Stoltz, and Z. Trstanova, *Error analysis of modified Langevin dynamics*, J. Stat. Phys., 164 (2016), pp. 735–771.
[48] L. Rey-Bellet and L. E. Thomas, *Exponential convergence to non-equilibrium stationary states in classical statistical mechanics*, Comm. Math. Phys., 225 (2002), pp. 305–329.
[49] S. Richardson and P. J. Green, *On Bayesian analysis of mixtures with an unknown number of components (with discussion)*, J. R. Stat. Soc. Ser. B Stat. Methodol., 59 (1997), pp. 731–792.

[50] Z. Schuss, *Theory and Applications of Stochastic Processes*, Appl. Math. Sci. 170, Springer, New York, 2010.
[51] L. Stella, C. Lorenz, and L. Kantorovich, *Generalized Langevin equation: An efficient approach to nonequilibrium molecular dynamics of open systems*, Phys. Rev. B, 89 (2014), 134303.
[52] P. A. Vasquez and M. G. Forest, *Complex fluids and soft structures in the human body*, in Complex Fluids in Biological Systems, Springer, New York, 2015, pp. 53–110.
[53] G. Vilmart, *Postprocessed integrators for the high order integration of ergodic sdes*, SIAM J. Sci. Comput., 37 (2015), pp. A201–A220.
[54] X. Wu, B. R. Brooks, and E. Vanden-Eijnden, *Self-guided Langevin dynamics via generalized Langevin equation*, J. Comput. Chem., 37 (2016), pp. 595–601.
[55] R. Zwanzig, *Nonlinear generalized Langevin equations*, J. Stat. Phys., 9 (1973), pp. 215–220.