# Contraction and Convergence Rates for Discretized Kinetic Langevin Dynamics

# CONTRACTION AND CONVERGENCE RATES FOR DISCRETIZED KINETIC LANGEVIN DYNAMICS

BENEDICT LEIMKUHLER*, DANIEL PAULIN*, AND PETER A. WHALLEY*

**Abstract.** We provide a framework to prove convergence rates for discretizations of kinetic Langevin dynamics for $M$-$\nabla$Lipschitz $m$-log-concave densities. Our approach provides convergence rates of $\mathcal{O}(m/M)$, with explicit stepsize restrictions, which are of the same order as the stability threshold for Gaussian targets and are valid for a large interval of the friction parameter. We apply this methodology to various integration methods which are popular in the molecular dynamics and machine learning communities. Finally we introduce the property "$\gamma$-limit convergent" (GLC) to characterise underdamped Langevin schemes that converge to overdamped dynamics in the high friction limit and which have stepsize restrictions that are independent of the friction parameter; we show that this property is not generic by exhibiting methods from both the class and its complement.

**Key words.** Contractive numerical method, Wasserstein convergence, kinetic Langevin dynamics, underdamped Langevin dynamics, MCMC sampling.

**AMS subject classifications.** 65C05, 65C30, 65C40

## 1. Introduction.

In this article, we study the following form of the Langevin dynamics equations ("kinetic Langevin dynamics") :

$$
\begin{aligned}
dX_t &= V_t dt, \\
dV_t &= -\nabla U(X_t)dt - \gamma V_t dt + \sqrt{2\gamma} dW_t,
\end{aligned}
\tag{1.1}
$$

where $U$ is the potential energy, $\gamma > 0$ is a friction parameter, and $W_t$ is $d$-dimensional standard Brownian motion. It can be shown under mild conditions that this process has invariant measure with density proportional to $\exp(-U(X) - ||V||^2/2)$ [38]. Normally, Langevin dynamics is developed in the physical setting with additional parameters representing temperature and mass. However, our primary aim in using (1.1) is, ultimately, the computation of statistical averages involving only the position $X$, and in such situations both parameters can be neglected without any loss of generality, or alternatively incorporated into our results through suitable rescalings of time and potential energy. In this article we focus on the properties of (1.1) in relation to numerical discretization and variation of the friction coefficient.

Taking the limit as $\gamma \to \infty$ in (1.1), and introducing a suitable time-rescaling ($t' = \gamma t$) results in the overdamped Langevin dynamics given by (see [38][Sec 6.5])

$$
dX_t = -\nabla U(X_t)dt + \sqrt{2} dW_t. \tag{1.2}
$$

This equation has, again, a unique invariant measure with density proportional to $\exp(-U(x))$. Under the assumption of a Poincaré inequality, convergence rate guarantees can be established for the continuous dynamics [2]. In the case of kinetic dynamics a more delicate argument is needed to establish exponential convergence, due to the hypoelliptic nature of the SDE (see [11, 45, 21, 20, 1, 3, 4]).

Langevin dynamics, in its kinetic and overdamped forms, is the basis of many widely used sampling algorithms in machine learning and statistics [14, 48, 47]. In sampling, Langevin dynamics is discretized and the individual timesteps generated by integration are viewed as approximate draws from the target distribution, however

*Department of Mathematics, University of Edinburgh., Edinburgh, EH9 3FD, UK. (b.leimkuhler@ed.ac.uk, dpaulin@ed.ac.uk, p.a.whalley@sms.ed.ac.uk)

there is an inherent bias due to the finite difference numerical approximation. This bias is usually addressed by choosing a sufficiently small stepsize, or by adding bias correction by use of methods like Metropolis-Hastings adjustment. The choice of the discretization method has a significant effect on the quality of the samples and also on the computational cost of producing accurate samples, through stability properties, convergence rates and asymptotic bias.

Overdamped Langevin dynamics has been heavily studied both in the continuous and the discretized settings, with popular integrators being the Euler-Maruyama and the limit method of the BAOAB scheme [32]. The kinetic Langevin system has been extensively studied in the continuous case, but there are still many open questions around the design of the numerical integrator. A metric that is typically used to quantify the performance of a sampling scheme is the number of steps required to reach a certain level of accuracy in Wasserstein distance. Non-asymptotic bounds in Wasserstein distance reflect computational complexity, convergence rate and accuracy. Achieving such bounds relies on two steps: (1) determining explicit convergence rates of the process to its invariant measure and (2) proving non-asymptotic bias estimates for the invariant measure. The focus of the current article is the convergence of the time-discrete system to its invariant measure.

The approach that we use to obtain convergence rates is based on proving contraction for a synchronous coupling, as in [36, 18]. Proving contraction of a coupling has been a popular method for establishing convergence both in the continuous time setting and for the discretization for Langevin dynamics and Hamiltonian Monte Carlo ([27, 8, 19, 7, 6, 40, 42]), since a consequence of such a contraction is convergence in Wasserstein distance (viewed as the infimum over all possible couplings with respect to some norm). Synchronous coupling has been a popular means of achieving explicit convergence rates for discretizations [35, 37] due to its simplicity.

There has been other recent work aimed at providing convergence rates for kinetic Langevin dynamics under explicit restrictions on the parameters ([14, 18, 36, 37]), but these guarantees are valid only with sharp restrictions on stepsize. There has also been the work of [41] which considers a slightly different version of the SDE (1.1), where time is rescaled depending on $M$ and $m$ to optimize contraction rates and bias. We have included their results in Table 1 after converting them into our framework using [18][Lemma 1]. The results of [41] rely on a stepsize restriction of $\mathcal{O}(1/\gamma)$, but their analysis does not provide the stepsize threshold [41][Example 9], and the class of schemes considered is different, with only the stochastic Euler scheme in common. Other works on contraction of kinetic Langevin and its discretization include [29, 17].

In the current article, we apply direct convergence analysis to various popular integration methods, and provide a general framework for establishing convergence rates of kinetic Langevin dynamics with tight explicit stepsize restrictions of $\mathcal{O}(1/\gamma)$ or $\mathcal{O}(1/\sqrt{M})$ (depending on the scheme). As a consequence we improve the contraction rates significantly for many of the available algorithms (see Table 1). For a specific class of schemes, we establish explicit bounds on the convergence rate for stepsizes of $\mathcal{O}(1/\sqrt{M})$. In the limit of large friction, we distinguish two types of integrators – those that converge to overdamped dynamics ("$\gamma$-limit-convergent") and those that do not. We demonstrate with examples that this property is not universal: some seemingly reasonable methods have the property that the convergence rate falls to zero in the $\gamma \to \infty$ limit. This is verified numerically and analytically for an anisotropic Gaussian target.

The remainder of this article is structured as follows. We first introduce overdamped Langevin dynamics, the Euler-Maruyama (EM) and the high friction limit

| Algorithm | stepsize restriction | optimal one-step contraction rate |
|---|---|---|
| EM | $\mathcal{O}(1/\gamma)$ | $\mathcal{O}(m/M)$ |
| BAO, OBA, AOB | $\mathcal{O}(1/\sqrt{M})$ | $\mathcal{O}(m/M)$ |
| OAB, ABO, BOA | $\mathcal{O}(1/\gamma)$ | $\mathcal{O}(m/M)$ |
| BAOAB | $\mathcal{O}(1/\sqrt{M})$ | $\mathcal{O}(m/M)$ |
| OBABO | $\mathcal{O}(1/\sqrt{M})$ | $\mathcal{O}(m/M)$ |
| SES | $\mathcal{O}(1/\gamma)$ | $\mathcal{O}(m/M)$ |

| Algorithm | previous stepsize restriction | previous explicit best rate |
|---|---|---|
| OBABO | $\mathcal{O}(m/\gamma^3)$ | $\mathcal{O}(m^2/M^2)$ [36] |
| SES | $\mathcal{O}(1/\gamma)$ | $\mathcal{O}(m/M)$ [41] |

Table 1: The first table provides our stepsize restrictions and optimal contraction rates of the discretized kinetic Langevin dynamics. The second provides previous best results. Further there are no previous results regarding the EM scheme, the first order splittings and BAOAB to the best of our knowledge.

of BAOAB (LM) and discuss their convergence guarantees. Next, we introduce kinetic Langevin and describe various popular discretizations, and give our results on convergence guarantees with mild stepsize assumptions. These schemes include first and second order splittings and the stochastic Euler scheme (SES). Further we compare the results of overdamped Langevin and kinetic Langevin and show how schemes like BAOAB and OBABO exhibit the positive qualities of both cases with the GLC property, whereas schemes like EM and SES do not perform well for a large range of $\gamma$.

## 2. Assumptions and definitions.

**2.1. Assumptions on $U$.** We will make the following assumptions on the target measure $\exp(-U)$ to obtain convergence rates. We assume that the potential is $M$-smooth and $m$-convex:

*Assumption* 2.1 ($M$-$\nabla$Lipschitz). There exists a $M > 0$ such that for all $X, Y \in \mathbb{R}^d$

$$|\nabla U(X) - \nabla U(Y)| \leq M |X - Y|.$$

*Assumption* 2.2 ($m$-convexity). There exists a $m > 0$ such that for all $X, Y \in \mathbb{R}^d$

$$\langle \nabla U(X) - \nabla U(Y), X - Y \rangle \geq m |X - Y|^2.$$

The two assumptions are popular conditions used to obtain explicit convergence rates, see [16, 18] for example. It is worth mentioning that these assumptions can also produce explicit convergence rates for gradient descent [9].

**2.2. Modified Euclidean Norms.** For kinetic Langevin dynamics it is not possible to prove convergence with respect to the standard Euclidean norm due to the fact that the generator is hypoelliptic. We therefore work with a modified Euclidean norm as in [36]. For $z = (x, v) \in \mathbb{R}^{2d}$ we introduce the weighted Euclidean norm

$$||z||^2_{a,b} = ||x||^2 + 2b \langle x, v \rangle + a ||v||^2,$$

for $a, b > 0$ which is equivalent norm as long as $b^2 < a$. More precisely we have

$$\frac{1}{2}||z||_{a,0}^2 \leq ||z||_{a,b}^2 \leq \frac{3}{2}||z||_{a,0}^2.$$

**2.3. Wasserstein Distance.** We define $\mathcal{P}_p\left(\mathbb{R}^{2d}\right)$ to be the set of probability measures which have finite $p$-th moment, then for $p \in [0,\infty)$ we define the $p$-Wasserstein distance on this space. Let $\mu$ and $\nu$ be two probability measures. We define the $p$-Wasserstein distance between $\mu$ and $\nu$ with respect to the norm $||\cdot||_{a,b}$ (introduced in Sec. 2.2) to be

$$\mathcal{W}_{p,a,b}\left(\nu,\mu\right) = \left(\inf_{\xi\in\Gamma(\nu,\mu)} \int_{\mathbb{R}^{2d}} ||z_1 - z_2||_{a,b}^p d\xi\left(z_1,z_2\right)\right)^{1/p},$$

where $\Gamma\left(\mu,\nu\right)$ is the set of measures with marginals $\mu$ and $\nu$ (the set of all couplings between $\mu$ and $\nu$).

It is well known that the existence of couplings with a contractive property implies convergence in Wasserstein distance (which can be interpreted as the infimum over all such couplings). The simplest such coupling is to consider simulations with common noise, this is known as synchronous coupling, therefore if one can show contraction of two simulations which share noise increments with a explicit contraction rate. Then one has convergence in Wasserstein distance with the same rate. With all the constants and conditions derived for all the schemes for contraction, we have convergence in Wasserstein distance by the following proposition:

Proposition 2.3. *Assume a numerical scheme for kinetic Langevin dynamics with a $m$-strongly convex $M$-$\nabla$Lipschitz potential $U$ and transition kernel $P_h$. If any two synchronously coupled chains $(x_n, v_n)$ and $(\tilde{x}_n, \tilde{v}_n)$ of the numerical scheme have the contraction property*

$$||(x_n - \tilde{x}_n, v_n - \tilde{v}_n)||_{a,b}^2 \leq C(1 - c\left(h\right))^n ||(x_0 - \tilde{x}_0, v_0 - \tilde{v}_0)||_{a,b}^2, \tag{2.1}$$

*for $\gamma^2 \geq C_\gamma M$ and $h \leq C_h\left(\gamma, \sqrt{M}\right)$ for some $a, b > 0$ such that $b^2 > a$. Then we have that for all $\gamma^2 \geq C_\gamma M$, $h \leq C_h\left(\gamma, \sqrt{M}\right)$, $1 \leq p \leq \infty$ and all $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^{2d})$, and all $n \in \mathbb{N}$,*

$$\mathcal{W}_p^2\left(\nu P_h^n, \mu P_h^n\right) \leq 3C \max\left\{a, \frac{1}{a}\right\}\left(1 - c\left(h\right)\right)^n \mathcal{W}_p^2\left(\nu,\mu\right).$$

*Further to this, $P_h$ has a unique invariant measure which depends on the stepsize, $\pi_h$, where $\pi_h \in \mathcal{P}_p(\mathbb{R}^{2d})$ for all $1 \leq p \leq \infty$.*

*Proof.* The proof is given in [36][Corollary 20], which relies on [46][Corollary 5.22, Theorem 6.18].

The focus of this article is to prove contractions of the form (2.1), and hence to achieve Wasserstein convergence rates by Prop. 2.3. With convergence to the invariant measure of the discretizations of kinetic Langevin dynamics considered here it will be possible to combine our results with estimates of the bias of each scheme as in [18], [36], [41] and [14] to obtain non-asymptotic estimates.

**3. Overdamped Langevin discretizations and contraction.** We first consider two discretizations of the SDE (1.2), namely the Euler-Maruyama discretization and the high friction limit of the popular kinetic Langevin dynamics scheme BAOAB [32]. The simplest discretization of overdamped Langevin dynamics is using the Euler-Maruyama (EM) method which is defined by the update rule

$$X_{n+1} = X_n - h\nabla U\left(X_n\right) + \sqrt{2h}\xi_{n+1}. \tag{3.1}$$

This scheme is combined with Metropolization in the popular MALA algorithm.

An alternative method is the BAOAB limit method of Leimkuhler and Matthews (LM)([32], [34]) which is defined by the update rule

$$X_{n+1} = X_n - h\nabla U\left(X_n\right) + \sqrt{2h}\frac{\xi_{n+1} + \xi_n}{2}.$$

The advantage of this method is that it gains a weak order of accuracy asymptotically.

**3.1. Convergence guarantees.** The convergence guarantees of overdamped Langevin dynamics and its discretizations have been extensively studied under the assumptions presented (see [16, 24, 13, 15, 25, 23, 26]). We use synchronous coupling as a proof strategy to obtain convergence rates as in [16]. We first consider two chains $x_n$ and $y_n$ with shared noise such that

$$x_{n+1} = y_n - h\nabla U(x_n) + \sqrt{2h}\xi_{n+1}, \quad y_{n+1} = y_n - h\nabla U(y_n) + \sqrt{2h}\xi_{n+1}.$$

Then we have that

$$\begin{aligned}
&||x_{n+1} - y_{n+1}||^2 = ||x_n - y_n + (-\nabla U(x_n) - (-\nabla U(y_n))||^2 \\
&= ||x_n - y_n||^2 - 2h\langle \nabla U(x_n) - \nabla U(y_n), x_n - y_n\rangle + h^2||\nabla U(x_n) - \nabla U(y_n)||^2 \\
&= ||x_n - y_n||^2 - 2h\langle x_n - y_n, Q(x_n - y_n)\rangle + h^2\langle x_n - y_n, Q^2(x_n - y_n)\rangle,
\end{aligned}$$

where $Q = \int_{t=0}^{1} \nabla^2 U(x_n + t(y_n - x_n))dt$. $Q$ has eigenvalues which are bounded between $m$ and $M$, so $Q^2 \preceq MQ$, and hence

$$h^2\langle x_n - y_n, Q^2(x_n - y_n)\rangle \leq h^2 M\langle x_n - y_n, Q(x_n - y_n)\rangle.$$

Therefore

$$\begin{aligned}
||x_{n+1} - y_{n+1}||^2 &\leq ||x_n - y_n||^2 - h(2 - hM)\langle x_n - y_n, Q(x_n - y_n)\rangle \\
&\leq ||x_n - y_n||^2(1 - hm(2 - hM)),
\end{aligned}$$

assuming that $h \leq \frac{2}{M}$. We have a contraction and

$$||x_n - y_n|| \leq (1 - hm(2 - hM))^{n/2}||x_0 - y_0||.$$

A consequence of this contraction result is that we have convergence in Wasserstein distance to the invariant measure with rate $hm\left(2 - hM\right)$, under the imposed assumptions on $h$ (as discussed in Sec. 2.3)[36, 46].

Note that this argument is exactly the same for the LM discretization of overdamped Langevin dynamics as all the noise components are shared. The stepsize assumption for convergence of overdamped Langevin dynamics in this setting is weak and is the same assumption as is needed to guarantee convergence of gradient descent in optimisation [9][Eq. (9.18)].

**4. Kinetic Langevin Dynamics.** We now consider many discretizations of the SDE (1.1) using a framework established in Sec. 4.1, where we construct an alternative Euclidean norm in which we can prove contraction (it is not possible to prove contraction in the standard Euclidean norm). Essentially, we convert the problem of proving contraction to the problem of showing that certain matrices are positive definite.

**4.1. Proof Strategy.** We will consider a modified Euclidean norm as defined in Sec. 2.2 for some choice of $a$ and $b$. Our aim is to construct an equivalent Euclidean norm such that contraction occurs for two Markov chains simulated by the same discretization $z_n = (x_n, v_n) \in \mathbb{R}^{2d}$ and $\tilde{z}_n = (\tilde{x}_n, \tilde{v}_n) \in \mathbb{R}^{2d}$ that are syncronously coupled. That is, for some choice of $a$ and $b$ such that $a, b > 0$ and $b^2 < a$

$$||\tilde{z}_{k+1} - z_{k+1}||^2_{a,b} < (1 - c(h)) ||\tilde{z}_k - z_k||^2_{a,b}, \tag{4.1}$$

where $a$ and $b$ are chosen to provide reasonable explicit assumptions on the stepsize $h$ and friction parameter $\gamma$. Our initial choices of $a$ and $b$ for simple schemes are motivated by [36], and are derived by considering contraction of the continuous dynamics. Let $\overline{z}_j = \tilde{z}_j - z_j$ for $j \in \mathbb{N}$, then (4.1) is equivalent to showing that

$$\overline{z}_k^T \left((1 - c(h)) M - P^T M P\right) \overline{z}_k > 0, \quad \text{where} \quad M = \begin{pmatrix} 1 & b \\ b & a \end{pmatrix}, \tag{4.2}$$

and $\overline{z}_{k+1} = P\overline{z}_k$ ($P$ depends on $z_k$ and $\tilde{z}_k$, but we omit this in the notation).

*Example* 4.1. As an example we have for the Euler-Maruyama scheme the update rule for $\overline{z}_k$

$$\overline{x}_{k+1} = \overline{x}_k + h\overline{v}_k, \qquad \overline{v}_{k+1} = \overline{v}_k - \gamma h \overline{v}_k - hQ\overline{x}_k,$$

where by mean value theorem we can define $Q = \int_{t=0}^{1} \nabla^2 U(\tilde{x}_k + t(x_k - \tilde{x}_k))dt$, then $\nabla U(\tilde{x}_k) - \nabla U(x_k) = Q\overline{x}$. One can show that in the notation of equation (4.2) we have

$$P = \begin{pmatrix} I & hI \\ -hQ & (1 - \gamma h) I \end{pmatrix}. \tag{4.3}$$

Proving contraction for a general scheme is equivalent to showing that the matrix $\mathcal{H} := (1 - c(h)) M - P^T M P \succ 0$ is positive definite. The matrix $\mathcal{H}$ is symmetric and hence of the form

$$\mathcal{H} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}, \tag{4.4}$$

we can show that $\mathcal{H}$ is positive definite by applying the following Prop. 4.2.

PROPOSITION 4.2. *Let $\mathcal{H}$ be a symmetric matrix of the form (4.4), then $\mathcal{H}$ is positive definite if and only if $A \succ 0$ and $C - BA^{-1}B \succ 0$. Further if $A$, $B$ and $C$ commute then $\mathcal{H}$ is positive definite if and only if $A \succ 0$ and $AC - B^2 \succ 0$.*

*Proof.* The proof of the first result is given in [31]. To establish the second statement, observe from [30] that if two matrices are positive definite and they commute then the product is positive definite. Also if $A \succ 0$ then $A^{-1} \succ 0$ (as $A$ is symmetric positive definite). Further $A, B$ and $C$ commute and hence $B$, $C$ and $A^{-1}$ commute. Therefore by applying the first result we have that $A \succ 0$ and

$$A^{-1}\left(AC - B^2\right) = C - BA^{-1}B \succ 0,$$

hence $\mathcal{H}$ is positive definite. If $\mathcal{H}$ is positive definite then $A \succ 0$ and $C - BA^{-1}B \succ 0$ by the first result. Thus as $A$, $B$ and $C$ commute we have $AC - B^2 \succ 0$. ◻

*Remark* 4.3. An equivalent condition for a symmetric matrix $\mathcal{H}$ of the form (4.4) to be positive definite is $C \succ 0$ and $AC - B^2 \succ 0$ when $A$, $B$ and $C$ commute. One could equivalently prove that $C \succ 0$ instead of $A \succ 0$ if it is more convenient.

Our general approach to prove contraction of some popular kinetic Langevin dynamics schemes is to prove the conditions of Prop. 4.2 are satisfied to establish contraction. We will use the notation laid out in this section in the proofs given in the appendix.

**4.2. Euler-Maruyama discretization.** We define the EM chain with initial condition $(x_0, v_0)$ by $(x_n, v_n, \xi_n)$ where the $(\xi_n)_{n\in\mathbb{N}}$ are independent $\mathcal{N}(0,1)$ draws and $(x_n, v_n)$ are updated according to:

$$x_{k+1} = x_k + hv_k, \tag{4.5}$$

$$v_{k+1} = v_k - h\nabla U(x_k) - h\gamma v_k + \sqrt{2\gamma h}\xi_{k+1}. \tag{4.6}$$

Theorem 4.4. *Assume $U$ is a $m$-strongly convex and $M$-$\nabla$Lipschitz potential. When $\gamma^2 \geq 4M$ and $h < \frac{1}{2\gamma}$, we have that, for all initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, and for any sequence of standard normal random variables $(\xi_n)_{n\in\mathbb{N}}$, the corresponding EM chains $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$ and $(\tilde{x}_n, \tilde{v}_n, \xi_n)_{n\in\mathbb{N}}$ with initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, respectively, satisfy*

$$||(x_k - \tilde{x}_k, v_k - \tilde{v}_k)||_{a,b} \leq (1 - c(h))^{\frac{k}{2}} ||(x_0 - \tilde{x}_0, v_0 - \tilde{v}_0)||_{a,b},$$

*where $a = \frac{1}{M}$, $b = \frac{1}{\gamma}$ and*

$$c(h) = \frac{mh}{2\gamma}.$$

*Example* 4.5. *An example to illustrate the tightness of the restrictions on the step-size $h$ and the restriction on the friction parameter $\gamma$.* We consider the anisotropic Gaussian distribution on $\mathbb{R}^2$ with potential $U : \mathbb{R}^2 \mapsto \mathbb{R}$ given by $U(x,y) = \frac{1}{2}mx^2 + \frac{1}{2}My^2$. This potential satisfies the assumptions 2.1 with constants $M$ and $m$ respectively. By computing the eigenvalues of the transition matrix $P$ (for contraction) we can see for what values of $h$ contraction occurs. For EM we have that

$$P = \begin{pmatrix} I & hI \\ -hQ & (1-\gamma h) I \end{pmatrix}, \text{ where } Q = \begin{pmatrix} m & 0 \\ 0 & M \end{pmatrix},$$

with eigenvalues

$$\frac{1}{2}\left(2 - \gamma h \pm h\sqrt{\gamma^2 - 4\lambda}\right),$$

for $\lambda = m, M$. For stability and contraction we require that

$$\frac{1}{2}\left(2 - \gamma h - h\sqrt{\gamma^2 - 4m}\right) > 0, \quad \text{and} \quad \frac{1}{2}\left(2 - \gamma h + h\sqrt{\gamma^2 - 4m}\right) < 1.$$

The second condition is equivalent to $\gamma > \sqrt{\gamma^2 - 4m}$ which trivially holds and the first condition is equivalent to $h \leq 2/(\gamma + \sqrt{\gamma^2 - 4m}) \approx 1/\gamma$.

**5. First order splittings.** A common discretization strategy for kinetic Langevin dynamics is based on splitting up the dynamics into parts which can be integrated exactly, in the weak sense. An increasingly popular splitting choice used in molecular dynamics modelling is to divide the SDE into deterministic parts corresponding to linear positional drift and an impulse due to the force and a dissipative-stochastic term corresponding to an Ornstein-Uhlenbeck equation [10]. These parts are denoted by $\mathcal{B}$, $\mathcal{A}$ and $\mathcal{O}$ with update rules given by

$$\begin{aligned}&\mathcal{B}: v \to v - h\nabla U(x),\\&\mathcal{A}: x \to x + hv,\\&\mathcal{O}: v \to \eta v + \sqrt{1-\eta^2}\xi,\end{aligned} \tag{5.1}$$

where

$$\eta := \exp(-\gamma h).$$

The reasoning for such a splitting is based on the fact that the infinitesimal generator of the SDE (1.1) can be split as $\mathcal{L} = \mathcal{L}_{\mathcal{A}} + \mathcal{L}_{\mathcal{B}} + \gamma\mathcal{L}_{\mathcal{O}}$, where

$$\mathcal{L}_{\mathcal{A}} = \langle v, \nabla_x \rangle, \qquad \mathcal{L}_{\mathcal{B}} = -\langle \nabla U(x), \nabla_v \rangle, \qquad \mathcal{L}_{\mathcal{O}} = -\langle v, \nabla_v \rangle + \Delta_v.$$

The dynamics associated to $\mathcal{L}_{\mathcal{A}}$ and $\mathcal{L}_{\mathcal{B}}$ are the deteriministic dynamics corresponding to $\mathcal{A}$ and $\mathcal{B}$. The dynamics associated to $\gamma\mathcal{L}_{\mathcal{O}}$ is the Ornstein-Uhlenbeck process, which can be solved exactly, in the sense of distributions. This corresponds to the $\mathcal{O}$ step. We use the convention that one applies the operators left to right.

The BAO method would first apply $\mathcal{B}$ then $\mathcal{A}$ and lastly $\mathcal{O}$. For more details on these splittings we refer the reader to [33].

We will now consider contraction for all first order splitting (permutations of the $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{O}$ pieces), which are schemes with weak order 1. We first consider BAO, where we define a BAO chain with initial condition $(x_0, v_0) \in \mathbb{R}^{2d}$ by $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$, using the update $\mathcal{BAO}$ (5.1) and $(\xi_n)_{n\in\mathbb{N}}$ are vectors of standard normal random variables.

THEOREM 5.1 (BAO). *Assume $U$ is a $m$-strongly convex and $M$-$\nabla$Lipschitz potential. When $h < \frac{1-\eta}{\sqrt{6M}}$, we have that for all initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, and for any sequence of standard normal random variables $(\xi_n)_{n\in\mathbb{N}}$ the BAO chains $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$ and $(\tilde{x}_n, \tilde{v}_n, \xi_n)_{n\in\mathbb{N}}$ with initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, respectively, satisfy*

$$||(x_k - \tilde{x}_k, v_k - \tilde{v}_k)||_{a,b} \leq (1 - c(h))^{\frac{k}{2}} ||(x_0 - \tilde{x}_0, v_0 - \tilde{v}_0)||_{a,b},$$

*where $a = \frac{1}{M}$ and $b = \frac{h}{(1-\eta)}$ and*

$$c(h) = \frac{h^2 m}{4(1-\eta)}.$$

*Remark* 5.2. The modified Euclidean norm has now been chosen to be stepsize dependent and is needed to eliminate the corresponding dependency of the stepsize on the strong convexity constant $m$. We note that that simply choosing $b = 1/\gamma$ does not result in a norm which guarantees a stepsize restriction which is independent of $m$, as is clear from the motivation of the construction of our choice of $b$. When $b \neq h/(1-\eta)$ one can always choose $m$ small enough such that $AC - B^2$ is not positive definite. We also point out that the stepsize restriction implicitly implies that $\gamma^2$ is larger than some constant factor multiplied by $M$. Further, for large $\gamma$ (for example $\gamma \geq 5\sqrt{M}$) we have convergence for stepsizes independent of the size of $\gamma$ (for example $h < 1/8\sqrt{M}$), which improves on the results of [41].

*Example* 5.3. *An example to illustrate the tightness of the restrictions on the stepsize $h$ and the restriction on the friction parameter $\gamma$.* We consider the anisotropic Gaussian distribution on $\mathbb{R}^2$ with potential $U : \mathbb{R}^2 \mapsto \mathbb{R}$ given by $U(x,y) = \frac{1}{2}mx^2 + \frac{1}{2}My^2$. By computing the eigenvalues of the transition matrix $P$ (for contraction) we can see for what values of $h$ contraction occurs. For BAO we have that

$$P = \begin{pmatrix} I - h^2 Q & hI \\ -h\eta Q & \eta I \end{pmatrix}, \text{ where } Q = \begin{pmatrix} m & 0 \\ 0 & M \end{pmatrix},$$

with eigenvalues

$$\frac{1}{2}\left(1 + \eta - h^2\lambda \pm \sqrt{-4\eta + (-1 - \eta + h^2\lambda)^2}\right),$$

for $\lambda = m, M$. For stability and contraction it is necessary and sufficient that

$$\left(1 + \eta - h^2 M\right) > 0, \quad \text{and} \quad \frac{1}{2}\left(1 + \eta - h^2\lambda + \sqrt{-4\eta + (-1 - \eta + h^2\lambda)^2}\right) < 1.$$

The first condition requires $h < \sqrt{\frac{1+\eta}{M}}$, where $\frac{1}{\sqrt{M}} < \sqrt{\frac{1+\eta}{M}} < \frac{2}{\sqrt{M}}$. The second condition holds when

$$1 - \eta + h^2\lambda > \sqrt{-4\eta + (-1 - \eta + h^2\lambda)^2},$$

which is equivalent to $4h^2\lambda > 0$, which trivially holds. Due to these stability conditions the best contraction rate possible is $\mathcal{O}\left(\frac{m}{M}\right)$, which coincides with our results. Further we have that the contraction rate is precisely $1 - \lambda_{max}$ which simplifies to

$$c_{\mathcal{N}} = 1 - \eta + h^2 m - \sqrt{(1 - \eta + h^2 m)^2 - 4h^2 m}.$$

Moreover, it can be shown that $4c(h) > c_{\mathcal{N}}$ for $h < 1/\sqrt{22m}$ and $\gamma \geq 4\sqrt{m}$. It is shown in [35][Proposition 4] that for the continuous dynamics this condition on $\gamma$ is necessary.

THEOREM 5.4 (OAB). *Assume $U$ is a $m$-strongly convex and $M$-$\nabla$Lipschitz potential. When $h < \min\{\frac{1}{4\gamma}, \frac{1-\eta}{\sqrt{6M}}\}$, we have that for all initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, and for any sequence of standard normal random variables $(\xi_n)_{n\in\mathbb{N}}$ the OAB chains $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$ and $(\tilde{x}_n, \tilde{v}_n, \xi_n)_{n\in\mathbb{N}}$ with initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, respectively, satisfy*

$$||(x_k - \tilde{x}_k, v_k - \tilde{v}_k)||_{a,b} \leq (1 - c(h))^{\frac{k}{2}} ||(x_0 - \tilde{x}_0, v_0 - \tilde{v}_0)||_{a,b},$$

*where $a = \frac{1}{M}$, $b = \frac{\eta h}{(1-\eta)}$ and $c(h) = \frac{\eta h^2 m}{4(1-\eta)}$.*

Considering other splittings one could use the same techniques as above or we can use the contractions results of BAO and OAB to achieve a contraction result for the remaining permutations by writing

$$(\mathcal{ABO})^n = \mathcal{AB}(\mathcal{OAB})^{n-1}\mathcal{O}, \quad (\mathcal{BOA})^n = \mathcal{B}(\mathcal{OAB})^{n-1}\mathcal{OA}$$
$$(\mathcal{OBA})^n = \mathcal{O}(\mathcal{BAO})^{n-1}\mathcal{BA}, \quad (\mathcal{AOB})^n = \mathcal{AO}(\mathcal{BAO})^{n-1}\mathcal{B}$$

However by applying direct arguments as done for OAB and BAO one would achieve better preconstants. Let $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$ and $(x_0, v_0) \in \mathbb{R}^{2d}$ be two initial conditions for a synchronous coupling of sample paths of the ABO splitting and $\overline{x}_0 := \tilde{x}_0 - x_0$, $\overline{v}_0 := \tilde{v}_0 - v_0$. In the following argument we let $Q$ be such that $\nabla U(\tilde{x}_0 + h\tilde{v}_0) - \nabla U(x_0 + hv_0) = Q(\overline{x}_0 + h\overline{v}_0)$ by the mean value theorem. Using the notation $\Psi_{\text{ABO}}$ to denote the one step map of the ABO discretization we have that for $h < \min\{\frac{1}{4\gamma}, \frac{1-\eta}{\sqrt{6M}}\}$

$$\begin{aligned}
&||\Psi_{\text{ABO}}(\tilde{x}_k) - \Psi_{\text{ABO}}(x_k)||^2_{a,b} = ||(\Psi_{\text{ABO}})^k(\tilde{x}_0) - (\Psi_{\text{ABO}})^k(x_0)||^2_{a,b} \\
&= ||\Psi_{\text{O}} \circ (\Psi_{\text{OAB}})^{k-1} \circ \Psi_{\text{AB}}(\tilde{x}_0) - \Psi_{\text{O}} \circ (\Psi_{\text{OAB}})^{k-1} \circ \Psi_{\text{AB}}(x_0)||^2_{a,b} \\
&\leq 3(1 - c(h))^{k-1} ||\Psi_{\text{AB}}(\tilde{x}_0, \tilde{v}_0) - \Psi_{\text{AB}}(x_0, v_0)||^2_{a,b} \\
&\leq 9(1 - c(h))^{k-1} \left(\left(1 + 2h^2M^2a\right)||\overline{x}_0||^2 + \left(h^2 + a + 2h^4M^2a\right)||\overline{v}_0||^2\right) \\
&\leq 27(1 - c(h))^{k-1} ||(\overline{x}_0, \overline{v}_0)||^2_{a,b},
\end{aligned}$$

where we have used the norm equivalence introduced in Sec. 2.2. The same method of argument can be used for the other first order splittings.

**6. Higher order splittings.** We now consider higher order schemes which are obtained by the splittings introduced in Sec. 5. These schemes are weak order two and they are symmetric in the order of the operators, with repeated operators corresponding to multiple steps with half the stepsize. We will focus our attention to two popular splittings which are BAOAB and ABOBA (or OBABO) as in [32]. Due to the fact that the modified Euclidean norms developed in the previous section are different for different first order splittings we aren't able to simply compose the results of say OBA and ABO to obtain contraction of OBABO. First we consider the BAOAB discretization, where we denote a BAOAB chain with initial condition $(x_0, v_0) \in \mathbb{R}^{2d}$ by $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$, which are defined by the update $\mathcal{BAOAB}$ (5.1) and $(\xi_n)_{n\in\mathbb{N}}$ are independent Gaussian random variables.

THEOREM 6.1 (BAOAB). *Assume $U$ is a $m$-strongly convex and $M$-$\nabla$Lipschitz potential. When $h \leq \frac{1-\eta}{2\sqrt{M}}$, we have that for all initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, and for any sequence of standard normal random variables $(\xi_n)_{n\in\mathbb{N}}$ the BAOAB chains $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$ and $(\tilde{x}_n, \tilde{v}_n, \xi_n)_{n\in\mathbb{N}}$ with initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, respectively, satisfy*

$$||(x_k - \tilde{x}_k, v_k - \tilde{v}_k)||_{a,b} \leq 7(1 - c(h))^{\frac{k-1}{2}} ||(x_0 - \tilde{x}_0, v_0 - \tilde{v}_0)||_{a,b},$$

*where $a = \frac{1}{M}$ and $b = \frac{h}{(1-\eta)}$ and*

$$c(h) = \frac{1}{4}\left(\frac{\eta h^2 m}{(1-\eta)} + h^2 m\right) = \frac{h^2 m}{4(1-\eta)}.$$

Next we consider the OBABO discretization which has been studied in the recent work [36]. In [37] they analyse Hamiltonian Monte Carlo as $\mathcal{O}(\mathcal{ABA})^L\mathcal{O}$ for $L$ leapfrog steps. In [37] a similar norm is used to study Hamiltonian Monte Carlo, however they obtain stepsize restrictions of at least $\mathcal{O}\left(m/L^{3/2}\right)$. We note that the OABAO scheme can also be analysed in our framework. We denote a OBABO chain with initial condition $(x_0, v_0) \in \mathbb{R}^{2d}$ by $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$, which are defined by the update $\mathcal{OBABO}$ (5.1) and $(\xi_n)_{n\in\mathbb{N}}$ are independent Gaussian random variables.

THEOREM 6.2 (OBABO). *Assume $U$ is a $m$-strongly convex and $M$-$\nabla$Lipschitz potential. When $h < \frac{1-\eta}{\sqrt{4M}}$, we have that for all initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, and for any sequence of standard normal random variables $(\xi_n)_{n\in\mathbb{N}}$ the OBABO chains $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$ and $(\tilde{x}_n, \tilde{v}_n, \xi_n)_{n\in\mathbb{N}}$ with initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, respectively, satisfy*

$$|| (x_k - \tilde{x}_k, v_k - \tilde{v}_k) ||_{a,b} \leq 7 (1 - c(h))^{\frac{k-1}{2}} || (x_0 - \tilde{x}_0, v_0 - \tilde{v}_0) ||_{a,b},$$

*where $a = \frac{1}{M}$, $b = \frac{h}{(1-\eta)}$ and*

$$c(h) = \frac{h^2 m}{4(1-\eta)}.$$

*Remark* 6.3. In [18] it is shown that the continuous dynamics converges with a rate of $\mathcal{O}(m/\gamma)$. There is a major difference in terms of contraction rate for large $\gamma$ between the rates achieved by BAOAB and OBABO and the continuous dynamics. As $\gamma \to \infty$ for BAOAB and OBABO you have convergence rates of $\mathcal{O}(h^2 m)$, whereas the contraction rate of the continuous dynamics converges to zero.

*Remark* 6.4. In Theorem 6.1 and Theorem 6.2 we have a prefactor of 7 due to the fact that we have converted the problem of contraction into proving a simpler problem with one gradient evaluation. More specifically for BAOAB using the relation $(\mathcal{BAOAB})^n = \mathcal{BAO}(\mathcal{ABAO})^{n-1}\mathcal{AB}$ and proving contraction for $\mathcal{ABAO}$ and similarly for OBABO. The prefactor comes from the remaining terms $\mathcal{BAO}$ and $\mathcal{AB}$.

**7. Stochastic exponential Euler scheme.** See [22] for an introduction to the Stochastic exponential Euler scheme and a derivation, based on keeping the gradient constant and analytically integrating the OU process with this constant gradient by combining the $\mathcal{B}$ and the $\mathcal{O}$ steps in the previous splitting. This scheme is the one considered in [14, 18] and has gained a lot of attention in the machine learning community and we can apply our methods to this scheme. Similar schemes have also been considered in [12, 28, 44] and it has been analysed in [22, 43]. The scheme in the notation we have used is given by the update rule

$$\begin{aligned} X_{k+1} &= X_k + \frac{1-\eta}{\gamma}V_k - \frac{\gamma h + \eta - 1}{\gamma^2}\nabla U(X_k) + \zeta_{k+1}, \\ V_{k+1} &= \eta V_k - \frac{1-\eta}{\gamma}\nabla U(X_k) + \omega_{k+1}, \end{aligned} \tag{7.1}$$

where

$$\zeta_{k+1} = \sqrt{2\gamma}\int_0^h e^{-\gamma(h-s)}dW_{h\gamma+s}, \qquad \omega_{k+1} = \sqrt{2\gamma}\int_0^h \frac{1-e^{-\gamma(h-s)}}{\gamma}dW_{h\gamma+s}.$$

$(\zeta_k, \omega_k)_{k\in\mathbb{N}}$ are Gaussian random vectors with covariances matrices which are stated in [22]. Now we can couple two trajectories which have common noise $(\zeta_k, \omega_k)_{k\in\mathbb{N}}$ then we can obtain contraction rates by the previously introduced methods. For the SES discretization where we denote a SES chain with initial condition $(x_0, v_0) \in \mathbb{R}^{2d}$ by $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$, which are defined by the update SES (7.1) and $(\xi_n)_{n\in\mathbb{N}}$ are independent Gaussian random variables.

THEOREM 7.1 (Stochastic Euler Scheme). *Assume $U$ is a $m$-strongly convex and $M$-$\nabla$Lipschitz potential. When $\gamma \geq 5\sqrt{M}$ and $h \leq \frac{1}{2\gamma}$, we have that for all initial conditions $(x_0, v_0) \in \mathbb{R}^{2d}$ and $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$, and for any sequence of standard* *normal random variables* $(\xi_n)_{n\in\mathbb{N}}$ *the SES chains* $(x_n, v_n, \xi_n)_{n\in\mathbb{N}}$ *and* $(\tilde{x}_n, \tilde{v}_n, \xi_n)_{n\in\mathbb{N}}$ *with initial conditions* $(x_0, v_0) \in \mathbb{R}^{2d}$ *and* $(\tilde{x}_0, \tilde{v}_0) \in \mathbb{R}^{2d}$*, respectively, satisfy*

$$||\left(x_k - \tilde{x}_k, v_k - \tilde{v}_k\right)||_{a,b} \leq \left(1 - c\left(h\right)\right)^{\frac{k}{2}} ||\left(x_0 - \tilde{x}_0, v_0 - \tilde{v}_0\right)||_{a,b},$$

*where* $a = \frac{1}{M}$*,* $b = \frac{1}{\gamma}$ *and*

$$c\left(h\right) = \frac{mh}{4\gamma}.$$

**8. Overdamped Limit.** We will now compare and analyze how the different schemes behave in the high friction limit. Starting with the first order schemes. It is a desirable property that the high friction limit is a discretization of the overdamped dynamics, therefore if a user of such a scheme sets a friction parameter $\gamma$ too large, they will not suffer from the $\mathcal{O}(1/\gamma)$ scaling of the convergence rate. We will call schemes with this desirable property $\gamma$-limit convergent (GLC), out of the schemes we have analysed it is only BAOAB and OBABO which are GLC.

**8.1. BAO.** If we consider the update rule of the BAO scheme

$$x_{k+1} = x_k + h\left(v_k - h\nabla U(x_k)\right), \quad v_{k+1} = \eta v_k - h\eta\nabla U(x_k) + \sqrt{1-\eta^2}\xi_{k+1},$$

and take the limit as $\gamma \to \infty$ we obtain

$$x_{k+1} = x_k - h^2\nabla U(x_k) + h\xi_k,$$

which is simply the Euler-Maruyama scheme with stepsize $h^2$ for potential $\tilde{U} := 4U$, which imposes stepsize restrictions $h^2 \leq \frac{2}{4M}$ and hence consistent with our analysis. Further if we take the limit of the contraction rate and the modified Euclidean norm we have

$$\lim_{\gamma\to\infty} c\left(h\right) = \frac{h^2 m}{4}, \qquad \lim_{\gamma\to\infty} ||x||^2 + 2b\langle x, v\rangle + a||v||^2 = ||x||^2 + 2h\langle x, v\rangle + \frac{1}{M}||v||^2,$$

which is again consistent with the convergence rates achieved in Sec. 3.1 and the norm is essentially the Euclidean norm when considered on the overdamped process as $\overline{v} = 0$. Due to the fact that the potential is rescaled in the limit, this is not a discretization of the overdamped dynamics.

**8.2. OAB.** If we consider the update rule of the OAB scheme

$$\begin{aligned} x_{k+1} &= x_k + h\eta v_k + h\sqrt{1-\eta^2}\xi_{k+1}, \\ v_{k+1} &= \eta v_k\sqrt{1-\eta^2}\xi_{k+1} - h\eta\nabla U(x_k + h\eta v_k + h\sqrt{1-\eta^2}\xi_{k+1}), \end{aligned}$$

and take the limit as $\gamma \to \infty$ we obtain the update rule $x_{k+1} = x_k + h\xi_{k+1}$, therefore the overdamped limit is not inherited by the scheme and further we do not expect contraction. This is consistent with our analysis of OAB and our contraction rate which tends towards 0 in the high friction limit.

**8.3. BAOAB.** If we consider the update rule of the BAOAB scheme

$$\begin{aligned} x_{k+1} &= x_k + \frac{h}{2}\left(1+\eta\right)v_k - \frac{h^2}{4}\left(1+\eta\right)\nabla U(x_k) + \frac{h}{2}\sqrt{1-\eta^2}\xi_{k+1}, \\ v_{k+1} &= \eta\left(v_k - \frac{h}{2}\nabla U(x_k)\right) + \sqrt{1-\eta^2}\xi_{k+1} - \frac{h}{2}\nabla U(x_{k+1}), \end{aligned}$$

and take the limit as $\gamma \to \infty$ we obtain

$$x_{k+1} = x_k - \frac{h^2}{2}\nabla U(x_k) + \frac{h}{2}\left(\xi_k + \xi_{k+1}\right),$$

which is simply the LM scheme with stepsize $h^2/2$ (as originally noted in [32]), which imposes stepsize restrictions $h^2 \leq 2/M$ and hence consistent with our analysis. Further if we take the limit of the contraction rate and the modified Euclidean norm we have

$$\lim_{\gamma\to\infty} c\left(h\right) = \frac{h^2 m}{4}, \qquad \lim_{\gamma\to\infty} ||x||^2 + 2b\langle x, v\rangle + a||v||^2 = ||x||^2 + 2h\langle x, v\rangle + \frac{1}{M}||v||^2,$$

which is again consistent with the convergence rates achieved in Sec. 3.1 and the modified Euclidean norm is essentially the Euclidean norm when considered on the overdamped process as $\overline{v} = 0$.

**8.4. OBABO.** If we consider the update rule of the OBABO scheme

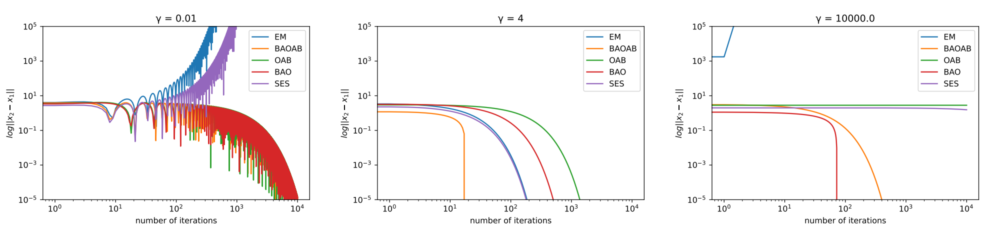
$$x_{k+1} = x_k + h\eta v_k + h\sqrt{1-\eta^2}\xi_{1,k+1} - \frac{h^2}{2}\nabla U(x_k),$$
$$v_{k+1} = \eta\left(\eta v + \sqrt{1-\eta^2}\xi_{1,k+1} - \frac{h}{2}\nabla U(x_k) - \frac{h}{2}\nabla U(x_{k+1})\right) + \sqrt{1-\eta^2}\xi_{2,k+1},$$

where $(\eta = \exp\left(-\gamma h/2\right))$ and for ease of notation in the above scheme and we have labelled the two noises of one step $\xi_1$ and $\xi_2$ . Now we take the limit as $\gamma \to \infty$ we obtain

$$x_{k+1} = x_k - \frac{h^2}{2}\nabla U(x_k) + h\xi_{k+1},$$

which is the Euler-Maruyama scheme for overdamped Langevin with stepsize $h^2/2$, which has convergence rate $\mathcal{O}\left(h^2 m\right)$. Hence consistent with our analysis of OBABO and our contraction rate which tends towards $h^2 m/4$ in the high friction limit.

**8.5. SES.** If we consider the limit as $\gamma \to \infty$ of the scheme (7.1) we obtain the update rule $x_{k+1} = x_k$ and therefore the overdamped limit is not inherited by the scheme and further we do not expect contraction. Hence consistent with our analysis of the stochastic Euler scheme as the contraction rate tends to 0 in the high friction limit.




(a) Low friction (b) (c) High friction

Fig. 1: Contraction of two kinetic Langevin trajectories $x_1$ and $x_2$ with initial conditions $[-1,-1]$ and $[1,1]$ for a 2-dimensional standard Gaussian with stepsize $h = 0.25 = 1/4\sqrt{M}$.

**9. Discussion.** We tested our observations numerically in Fig. 1 with a 2-dimensional standard Gaussian. Fig. 1 is consistent with our analysis that all schemes are stable when $\gamma \approx 4\sqrt{M}$ and in the high friction regime EM, OAB and SES behave poorly compared to BAOAB and BAO. In the low friction regime again EM and SES perform poorly compared to the other schemes.

In [18] it is shown that the optimal convergence rate for the continuous time dynamics is $\mathcal{O}(m/\gamma)$, therefore our contraction rates are consistent up to a constant for the discretizations, however for some of the schemes considered for example BAOAB and OBABO we have that the scheme inherits convergence to the overdamped Langevin dynamics (without time rescaling) and this is reflected in our convergence rate estimates. Therefore for MCMC applications it does not suffer from the scaling of $1/\gamma$ on the convergence rate, if the user picks a friction parameter which is too high. This robustness with respect to the friction parameter is shown in Fig. 1.

The constants in our arguments can be improved by sharper bounds and a more careful analysis, but the restriction on $\gamma$ is consistent with other works on synchronous coupling for the continuous time Langevin diffusions [5, 14, 18, 19, 49]. Further it is shown in [35][Proposition 4] that the continuous time process yields Wasserstein contraction of synchronous coupling for all $M$-$\nabla$Lipschitz and $m$-strongly convex potentials $U$ if and only if $M - m < \gamma(\sqrt{M} + \sqrt{m})$ for the norms that we considered. This condition when $M$ is much larger than $m$ is $\mathcal{O}(\sqrt{M})$. It may be possible to achieve convergence rates for small $\gamma$, by using a more sophisticated argument like that of [27]. Using a different Lyapunov function or techniques may lead to being able to extend these results to all $\gamma > 0$ [22, 39], following results for the continuous case [27], but this is beyond the scope of this paper.

The restrictions on the stepsize $h$ are tight for the optimal contraction rate for EM and BAO and hence result in stability conditions of $\mathcal{O}(1/\gamma)$ for EM and SES. Also we have shown BAO, OBA, AOB, BAOAB and OBABO have convergence guarantees for stepsizes $\mathcal{O}(1/\sqrt{M})$ and BAOAB and OBABO have the desirable GLC property which is not common amongst the schemes we studied. For the choice of parameters which achieve optimal contraction rate we derive $\mathcal{O}(m/M)$ rates of contraction, which are sharp up to a constant and we achieve this for every scheme that we studied.

**Acknowledgments.** The authors would like to thank Kostas Zygalakis for helpful comments on this work. The authors acknowledge the support of the Engineering and Physical Sciences Research Council Grant EP/S023291/1 (MAC-MIGS Centre for Doctoral Training).

REFERENCES

[1] D. Bakry and M. Émery, *Diffusions hypercontractives*, in Séminaire de Probabilités XIX 1983/84: Proceedings, Springer, 2006, pp. 177–206.

[2] D. Bakry, I. Gentil, M. Ledoux, et al., *Analysis and geometry of Markov diffusion operators*, vol. 103, Springer, 2014.

[3] F. Baudoin, *Wasserstein contraction properties for hypoelliptic diffusions*, arXiv preprint arXiv:1602.04177, (2016).

[4] F. Baudoin, *Bakry–émery meet villani*, Journal of functional analysis, 273 (2017), pp. 2275–2291.

[5] F. Bolley, A. Guillin, and F. Malrieu, *Trend to equilibrium and particle approximation for a weakly selfconsistent vlasov-fokker-planck equation*, ESAIM: Mathematical Modelling and Numerical Analysis, 44 (2010), pp. 867–884.

[6] N. Bou-Rabee and A. Eberle, *Couplings for andersen dynamics*, in Annales de l'Institut Henri Poincare (B) Probabilites et statistiques, vol. 58, Institut Henri Poincaré, 2022,

pp. 916–944.
[7] N. Bou-Rabee and A. Eberle, *Mixing time guarantees for unadjusted hamiltonian monte carlo*, Bernoulli, 29 (2023), pp. 75–104.
[8] N. Bou-Rabee, A. Eberle, and R. Zimmer, *Coupling and convergence for hamiltonian monte carlo*, The Annals of applied probability, 30 (2020), pp. 1209–1250.
[9] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.
[10] G. Bussi and M. Parrinello, *Accurate sampling using langevin dynamics*, Phys. Rev. E, 75 (2007), p. 056707.
[11] Y. Cao, J. Lu, and L. Wang, *On explicit $l^2$-convergence rate estimate for underdamped langevin dynamics*, arXiv preprint arXiv:1908.04746, (2019).
[12] S. Chandrasekhar, *Stochastic problems in physics and astronomy*, Reviews of modern physics, 15 (1943), p. 1.
[13] X. Cheng and P. Bartlett, *Convergence of langevin mcmc in kl-divergence*, in Algorithmic Learning Theory, PMLR, 2018, pp. 186–211.
[14] X. Cheng, N. S. Chatterji, P. L. Bartlett, and M. I. Jordan, *Underdamped langevin mcmc: A non-asymptotic analysis*, in Conference on learning theory, PMLR, 2018, pp. 300–323.
[15] A. Dalalyan, *Further and stronger analogy between sampling and optimization: Langevin monte carlo and gradient descent*, in Conference on Learning Theory, PMLR, 2017, pp. 678–689.
[16] A. S. Dalalyan, *Theoretical guarantees for approximate sampling from smooth and log-concave densities*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79 (2017), pp. 651–676.
[17] A. S. Dalalyan, A. Karagulyan, and L. Riou-Durand, *Bounding the error of discretized langevin algorithms for non-strongly log-concave targets*, Journal of Machine Learning Research, 23 (2022), pp. 1–38.
[18] A. S. Dalalyan and L. Riou-Durand, *On sampling from a log-concave density using kinetic langevin diffusions*, Bernoulli, 26 (2020), pp. 1956–1988.
[19] G. Deligiannidis, D. Paulin, A. Bouchard-Côté, and A. Doucet, *Randomized hamiltonian monte carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates*, The Annals of Applied Probability, 31 (2021), pp. 2612–2662.
[20] J. Dolbeault, C. Mouhot, and C. Schmeiser, *Hypocoercivity for kinetic equations with linear relaxation terms*, Comptes Rendus Mathematique, 347 (2009), pp. 511–516.
[21] J. Dolbeault, C. Mouhot, and C. Schmeiser, *Hypocoercivity for linear kinetic equations conserving mass*, Transactions of the American Mathematical Society, 367 (2015), pp. 3807–3828.
[22] A. Durmus, A. Enfroy, É. Moulines, and G. Stoltz, *Uniform minorization condition and convergence bounds for discretizations of kinetic langevin dynamics*, arXiv preprint arXiv:2107.14542, (2021).
[23] A. Durmus, S. Majewski, and B. Miasojedow, *Analysis of langevin monte carlo via convex optimization*, The Journal of Machine Learning Research, 20 (2019), pp. 2666–2711.
[24] A. Durmus and E. Moulines, *Nonasymptotic convergence analysis for the unadjusted langevin algorithm*, The Annals of Applied Probability, 27 (2017), pp. 1551–1587.
[25] A. Durmus and E. Moulines, *High-dimensional bayesian inference via the unadjusted langevin algorithm*, Bernoulli, 25 (2019), pp. 2854–2882.
[26] R. Dwivedi, Y. Chen, M. J. Wainwright, and B. Yu, *Log-concave sampling: Metropolis-hastings algorithms are fast!*, in Conference on learning theory, PMLR, 2018, pp. 793–797.
[27] A. Eberle, A. Guillin, and R. Zimmer, *Couplings and quantitative contraction rates for langevin dynamics*, The Annals of Probability, 47 (2019), pp. 1982–2010.
[28] D. L. Ermak and H. Buckholz, *Numerical integration of the langevin equation: Monte carlo simulation*, Journal of Computational Physics, 35 (1980), pp. 169–182.
[29] J. Foster, T. Lyons, and H. Oberhauser, *The shifted ode method for underdamped langevin mcmc*, arXiv preprint arXiv:2101.03446, (2021).
[30] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge university press, 2012.
[31] R. A. Horn and F. Zhang, *Basic properties of the schur complement*, The Schur Complement and Its Applications, (2005), pp. 17–46.
[32] B. Leimkuhler and C. Matthews, *Rational construction of stochastic numerical methods for molecular sampling*, Applied Mathematics Research eXpress, 2013 (2013), pp. 34–56.
[33] B. Leimkuhler, C. Matthews, and G. Stoltz, *The computation of averages from equilibrium and nonequilibrium langevin molecular dynamics*, IMA Journal of Numerical Analysis, 36 (2016), pp. 13–79.

[34] B. Leimkuhler, C. Matthews, and M. Tretyakov, *On the long-time integration of stochastic gradient systems*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 470 (2014), p. 20140120.
[35] P. Monmarché, *Almost sure contraction for diffusions on rd. application to generalised langevin diffusions*, arXiv preprint arXiv:2009.10828, (2020).
[36] P. Monmarché, *High-dimensional mcmc with a standard splitting scheme for the underdamped langevin diffusion.*, Electronic Journal of Statistics, 15 (2021), pp. 4117–4166.
[37] P. Monmarché, *Hmc and langevin united in the unadjusted and convex case*, arXiv preprint arXiv:2202.00977, (2022).
[38] G. A. Pavliotis, *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, vol. 60, Springer, 2014.
[39] Q. Qin and J. P. Hobert, *Geometric convergence bounds for markov chains in wasserstein distance based on generalized drift and contraction conditions*, in Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, vol. 58, 2022, pp. 872–889.
[40] L. Riou-Durand and J. Vogrinc, *Metropolis adjusted langevin trajectories: a robust alternative to hamiltonian monte carlo*, arXiv preprint arXiv:2202.13230, (2022).
[41] J. M. Sanz-Serna and K. C. Zygalakis, *Wasserstein distance estimates for the distributions of numerical approximations to ergodic stochastic differential equations.*, J. Mach. Learn. Res., 22 (2021), pp. 242–1.
[42] K. Schuh, *Global contractivity for langevin dynamics with distribution-dependent forces and uniform in time propagation of chaos*, arXiv preprint arXiv:2206.03082, (2022).
[43] C. Shi, Y. Xiao, and C. Zhang, *The convergence and ms stability of exponential euler method for semilinear stochastic differential equations*, in Abstract and Applied Analysis, vol. 2012, Hindawi, 2012.
[44] R. D. Skeel and J. A. Izaguirre, *An impulse integrator for langevin dynamics*, Molecular Physics, 100 (2002), pp. 3885–3891.
[45] C. Villani, *Hypocoercivity*, (2009).
[46] C. Villani, *Optimal transport: old and new*, vol. 338, Springer, 2009.
[47] S. J. Vollmer, K. C. Zygalakis, and Y. W. Teh, *Exploration of the (non-) asymptotic bias and variance of stochastic gradient langevin dynamics*, The Journal of Machine Learning Research, 17 (2016), pp. 5504–5548.
[48] M. Welling and Y. W. Teh, *Bayesian learning via stochastic gradient langevin dynamics*, in Proceedings of the 28th international conference on machine learning (ICML-11), 2011, pp. 681–688.
[49] T. Zajic, *Non-asymptotic error bounds for scaled underdamped langevin mcmc*, arXiv preprint arXiv:1912.03154, (2019).

## Appendix A. Proofs.

*Proof of Theorem 4.4.* We will denote two synchronous realisations of EM as $(x_j, v_j)$ and $(\tilde{x}_j, \tilde{v}_j)$ for $j \in \mathbb{N}$. Now we will denote $\overline{x} := (\tilde{x}_j - x_j)$, $\overline{v} = (\tilde{v}_j - v_j)$ and $\overline{z} = (\overline{x}, \overline{v})$, where $\overline{z}_j = (\overline{x}_j, \overline{v}_j)$ for $j = k, k+1$ for $k \in \mathbb{N}$. We have the following update rule for $\overline{z}_k$

$$\overline{x}_{k+1} = \overline{x}_k + h\overline{v}_k, \qquad \overline{v}_{k+1} = \overline{v}_k - \gamma h \overline{v}_k - hQ\overline{x}_k,$$

where by mean value theorem we define $Q = \int_{t=0}^{1} \nabla^2 U(\tilde{x}_k + t(x_k - \tilde{x}_k))dt$, then $\nabla U(\tilde{x}_k) - \nabla U(x_k) = Q\overline{x}$. One can show that in the notation of equation (4.2) we have

$$P = \begin{pmatrix} I & hI \\ -hQ & (1-\gamma h) I \end{pmatrix}, \tag{A.1}$$

and therefore for Euler-Maruyama (using the notation of equation (4.4))

$$\begin{aligned} &A = -c(h) + 2bhQ - h^2 aQ^2, \quad B = -c(h)\, b + h\left((b\gamma - 1) + (a + h(b - a\gamma))Q\right), \\ &C = -c(h)\, a + h\left(2a\gamma - 2b - h(1 - 2b\gamma + a\gamma^2)\right). \end{aligned}$$

We now invoke Prop. 4.2 as $A$, $B$ and $C$ commute as they are all polynomials in $Q$. $A$ is positive definite if and only if all it's eigenvalues are positive. We note that the eigenvalues of $A$ are precisely $P_A(\lambda) := -c(h) + h\left(2b\lambda - ha\lambda^2\right)$, where $\lambda$ are the eigenvalues of $Q$, where $m \leq \lambda \leq M$. We wish to show that $P_A(\lambda) > 0$ for all $\lambda \in [m, M]$. This is equivalent to

$$\frac{P_A(\lambda)}{h} = -\frac{m}{2\gamma} + \frac{2\lambda}{\gamma} - \frac{h\lambda^2}{M} \geq \lambda\left(-\frac{1}{2\gamma} + \frac{2}{\gamma} - h\right) > 0,$$

which are both satisfied when $h < 1/\gamma$. Hence we have that $A \succ 0$. Now it remains to prove that $AC - B^2 \succ 0$, where $AC - B^2$ is a polynomial of $Q$, which we denote $P_{AC-B^2}(Q)$. Hence has eigenvalues dictated by the eigenvalues $\lambda$ of $Q$. That is the eigenvalues of $AC - B^2$ are $P_{AC-B^2}(\lambda)$ for $\lambda$ an eigenvalue of $Q$. Considering $P_{AC-B^2}$ we have

$$\frac{P_{AC-B^2}(\lambda)}{h^2\lambda} = \frac{4}{M} - \frac{4}{\gamma^2} + \frac{m^2}{4\gamma^2 M\lambda} - \frac{m^2}{4\gamma^4\lambda} + \frac{m}{\gamma^2\lambda} - \frac{m}{M\lambda} - \frac{\lambda}{M^2} + h^2\left(\frac{\lambda}{M} - \frac{\lambda}{\gamma^2}\right)$$
$$+ h\left(\frac{2}{\gamma} - \frac{m}{\gamma M} - \frac{m}{2\gamma\lambda} + \frac{m}{\gamma^3} + \frac{m\lambda}{2\gamma M^2} + \frac{\gamma m}{2M\lambda} - \frac{2\gamma}{M}\right) > \frac{1}{M} - h\frac{2\gamma}{M} > 0, \qquad \square$$

where we have used the fact that $\gamma \geq 2\sqrt{M}$ and $h < \frac{1}{2\gamma}$ and hence $AC - B^2 \succ 0$.

*Proof of Theorem 5.1.* We will denote two synchronous realisations of BAO as $(x_j, v_j)$ and $(\tilde{x}_j, \tilde{v}_j)$ for $j \in \mathbb{N}$. Now we will denote $\overline{x} := (\tilde{x}_j - x_j)$, $\overline{v} = (\tilde{v}_j - v_j)$ and $\overline{z} = (\overline{x}, \overline{v})$, where $\overline{z}_j = (\overline{x}_j, \overline{v}_j)$ for $j = k, k+1$ for $k \in \mathbb{N}$. We have the following update rule for $\overline{z}_k$

$$\overline{x}_{k+1} = \overline{x}_k + h\left(\overline{v}_k - hQ\overline{x}_k\right), \quad \overline{v}_{k+1} = \eta(\overline{v}_k - hQ\overline{x}_k),$$

where by mean value theorem we define $Q = \int_{t=0}^{1} \nabla^2 U(\tilde{x}_k + t(x_k - \tilde{x}_k))dt$, then $\nabla U(\tilde{x}_k) - \nabla U(x_k) = Q\overline{x}_k$. In the notation of (4.4) we have that

$$A = -c(h) + 2(b\eta + h)hQ - \left(a\eta^2 + 2b\eta h + h^2\right)h^2Q^2,$$
$$B = b(1-\eta) - h - bc(h) + \left(a\eta^2 + 2b\eta h + h^2\right)hQ,$$
$$C = a\left(1 - \eta^2\right) - 2b\eta h - h^2 - ac(h),$$

where $\eta = \exp\{-\gamma h\}$. For our choice of $a$ and $b$, $B$ simplifies to

$$B = -bc(h) + \left(a\eta^2 + 2b\eta h + h^2\right)hQ.$$

This simplification motivates our choice of $b$ as we would like to factor out $m$ from $B$ in proving $AC - B^2 \succ 0$ in a later calculation and this factorisation is necessary to remove the dependence on $m$ in our stepsize estimates.

We will now apply Prop. 4.2. First considering $A$, we have that the eigenvalues for our choice of $a$, $b$ and $c(h)$ are precisely

$$P_A(\lambda) = -c(h) + 2b\lambda h - \left(a\eta^2 + 2b\eta h + h^2\right)h^2\lambda^2 > h\lambda\left(\frac{7b}{4} - \left(h + \frac{b}{2} + \frac{h}{4}\right)\right) > 0,$$

where $\lambda$ are the eigenvalues of $Q$ and we have used that $h < 1/(\sqrt{4M})$ and $b \geq h$. Hence we have that $A \succ 0$. Now it remains to prove that $AC - B^2 \succ 0$, where $AC - B^2$ is a polynomial of $Q$, which we denote $P_{AC-B^2}(Q)$ and hence has eigenvalues dictated by the eigenvalues $\lambda$ of $Q$. We have that

$$\begin{aligned}
\frac{P(\lambda)}{h\lambda} &= \frac{\left(\eta^2-1\right)c\left(h\right)}{hM\lambda} + h\left(\frac{2\left(1-\eta^2\right)}{\left(1-\eta\right)M} + \frac{c(h)^2}{h^2M\lambda} - \frac{\eta^2\lambda}{M^2}\right) \\
&+ \frac{c(h)}{h}h^2\left(\frac{1}{\lambda} + \frac{2}{(1-\eta)}(\frac{\eta}{\lambda} + \frac{\eta^2-1}{M}) + \frac{\eta^2\lambda}{M^2}\right) \\
&+ h^3\left(-\frac{4\eta}{(1-\eta)^2} - \frac{2}{(1-\eta)} - \frac{c(h)^2}{(1-\eta)^2h^2\lambda} - \frac{\lambda}{M} - \frac{2\eta\lambda}{(1-\eta)M}\right) \\
&+ \frac{c(h)}{h}h^4\left(\frac{4\eta}{(1-\eta)^2} + \frac{2}{(1-\eta)} + \frac{\lambda}{M} + \frac{2\eta\lambda}{(1-\eta)M}\right) \\
&> \left(\frac{7h\left(1+\eta\right)}{4M} - \frac{\eta^2h}{M} - \frac{h^2}{\sqrt{6M}}\right) + h^3\left(-\frac{1}{(1-\eta)^2}\left(4+\frac{1}{96}\right) - 1 - \frac{4}{(1-\eta)}\right) \\
&> h\left(\frac{7}{4M} - \frac{1}{6M} - \frac{2}{3M}\left(1+\frac{1}{384}\right) - \frac{1}{6M} - \frac{2}{3M}\right) > 0,
\end{aligned}$$

where we have used the fact that $h < \frac{1-\eta}{\sqrt{6M}}$. This holds for any $\lambda \in [m, M]$ and hence $AC - B^2 \succ 0$ and our contraction results hold. ∎

*Proof of Theorem 5.4.* We will denote two synchronous realisations of OAB as $(x_j, v_j)$ and $(\tilde{x}_j, \tilde{v}_j)$ for $j \in \mathbb{N}$. Now we will denote $\overline{x} := (\tilde{x}_j - x_j)$, $\overline{v} = (\tilde{v}_j - v_j)$ and $\overline{z} = (\overline{x}, \overline{v})$, where $\overline{z}_j = (\overline{x}_j, \overline{v}_j)$ for $j = k, k+1$ for $k \in \mathbb{N}$. We have the following update rule for $\overline{z}_k$

$$\overline{x}_{k+1} = \overline{x}_k + h\eta\overline{v}_k, \quad \overline{v}_{k+1} = \eta\overline{v}_k - hQ\overline{x}_{k+1},$$

where by mean value theorem we define $Q = \int_{t=0}^1 \nabla^2 U(\tilde{x}_{k+1} + t(x_{k+1} - \tilde{x}_{k+1}))dt$, then $\nabla U\left(\tilde{x}_{k+1}\right) - \nabla U\left(x_{k+1}\right) = Q\overline{x}_{k+1}$. In the notation of (4.4) we have that

$$\begin{aligned}
&A = -c\left(h\right) + 2bhQ - ah^2Q^2, \\
&B = b\left(1-\eta\right) - \eta h - bc\left(h\right) + (a\eta h + 2b\eta h^2)Q - a\eta h^3Q^2, \\
&C = a\left(1-\eta^2\right) - 2b\eta^2h - \eta^2h^2 - ac\left(h\right) + (2a\eta^2h^2 + 2b\eta^2h^3)Q - a\eta^2h^4Q^2,
\end{aligned}$$

where $\eta = \exp\{-\gamma h\}$. For our choice of $a$ and $b$, $B$ simplifies to

$$B = -bc\left(h\right) + (a\eta h + 2b\eta h^2)Q - a\eta h^3Q^2.$$

We will now apply Prop. 4.2, first considering $A$ we have that the eigenvalues are precisely

$$P_A(\lambda) = h\lambda\left(-\frac{c\left(h\right)}{h\lambda} + 2b - ah\lambda\right) \geq h\lambda\left(\frac{7b}{4} - h\right) > 0,$$

Hence we have that $A \succ 0$ and we note that this condition enforces a dependency of $h$ on $\gamma$ to ensure convergence. Now it remains to prove that $AC - B^2 \succ 0$, now we have that $AC - B^2$ is a polynomial of $Q$, which we denote $P_{AC-B^2}(Q)$ and hence has eigenvalues dictated by the eigenvalues $\lambda$ of $Q$.

Using the fact that $h\gamma \geq 1 - \eta \geq \frac{h\gamma}{2}$ and hence $\eta \geq \frac{3}{4}$ for $h < \frac{1}{4\gamma}$. Also using that

$h \leq \frac{1-\eta}{\sqrt{6M}}$ we have that

$$\frac{P_{AC-B^2}(\lambda)}{h\lambda} = \frac{\left(\eta^2-1\right)c(h)}{hM\lambda} + h\left(\frac{2\eta\left(1-\eta^2\right)}{(1-\eta)M} + \frac{c(h)^2}{h^2M\lambda} - \frac{\lambda}{M^2}\right)$$

$$+ h^2\frac{c(h)}{h}\left(-\frac{2\eta}{(1-\eta)M} - \frac{2\eta^2}{M} + \frac{2\eta^2}{(1-\eta)M} + \frac{\eta^2}{\lambda} + \frac{2\eta^3}{(1-\eta)\lambda} + \frac{\lambda}{M^2}\right)$$

$$+ h^3\left(-\frac{4\eta^4}{(1-\eta)^2} - \frac{2\eta^3}{1-\eta} - \frac{\eta^2 c(h)^2}{(1-\eta)^2h^2\lambda} + \frac{\lambda\eta^2}{M} + \frac{2\eta^3\lambda}{(1-\eta)M}\right)$$

$$+ h^4\frac{c(h)}{h}\left(\frac{4\eta^3}{(1-\eta)^2} - \frac{2\eta^3}{(1-\eta)} + \frac{\eta^2\lambda}{M} - \frac{2\eta^2\lambda}{(1-\eta)M}\right)$$

$$\geq \frac{5h}{4M} + \left(-\frac{h}{6M}\right) + h^3\left(-\frac{4}{(1-\eta)^2} - \frac{1}{96(1-\eta)^2} - \frac{2}{(1-\eta)}\right) - \frac{h^3}{12} > 0.$$

Hence $AC - B^2 \succ 0$ and our contraction results hold. □

*Proof of Theorem 6.1.* We first note that $(\mathcal{BAOAB})^n = \mathcal{BAO}\left(\mathcal{ABAO}\right)^{n-1}\mathcal{AB}$. We will now focus our attention on proving contraction of $\mathcal{ABAO}$, by doing this we only have to deal with a single evaluation of the Hessian at each step. We will denote two synchronous realisations of ABAO as $(x_j, v_j)$ and $(\tilde{x}_j, \tilde{v}_j)$ for $j \in \mathbb{N}$. Now we will denote $\overline{x} := (\tilde{x}_j - x_j)$, $\overline{v} = (\tilde{v}_j - v_j)$ and $\overline{z} = (\overline{x}, \overline{v})$, where $\overline{z}_j = (\overline{x}_j, \overline{v}_j)$ for $j = k, k+1$ for $k \in \mathbb{N}$. We have the following update rule for $\overline{z}_k$

$$\overline{x}_{k+1} = \overline{x}_k + h\overline{v}_k - \frac{h^2}{2}Q\left(\overline{x} + \frac{h}{2}\overline{v}\right), \quad \overline{v}_{k+1} = \eta\overline{v}_k - h\eta Q\left(\overline{x} + \frac{h}{2}\overline{v}\right),$$

where by mean value theorem we define $Q = \int_{t=0}^{1} \nabla^2 U(\tilde{x}_k + \frac{h}{2}\tilde{v}_k + t(x_k - \tilde{x}_k + \frac{h}{2}(v_k - \tilde{v}_k)))dt$, then $\nabla U\left(\tilde{x}_k + \frac{h}{2}\tilde{v}_k\right) - \nabla U\left(x_k + \frac{h}{2}v_k\right) = Q\left(\overline{x} + \frac{h}{2}\overline{v}\right)$. In the notation of (4.4) we have that for this scheme

$$A = -c(h) + \left(2b\eta h + h^2\right)Q + \left(-a\eta^2h^2 - b\eta h^3 - \frac{1}{4}h^4\right)Q^2,$$

$$B = b(1-\eta) - h - bc(h) + \left(a\eta^2 h + 2b\eta h^2 + \frac{3}{4}h^3\right)Q$$

$$+ \left(-\frac{1}{2}a\eta^2h^3 - \frac{1}{2}b\eta h^4 - \frac{1}{8}h^5\right)Q^2,$$

$$C = a\left(1-\eta^2\right) - 2b\eta h - h^2 - ac(h) + \left(a\eta^2h^2 + \frac{3}{2}b\eta h^3 + \frac{1}{2}h^4\right)Q$$

$$+ \left(-\frac{1}{4}a\eta^2h^4 - \frac{1}{4}b\eta h^5 - \frac{1}{16}h^6\right)Q^2,$$

where $\eta = \exp\{-\gamma h\}$. For our choice of $a$ and $b$, $B$ simplifies to

$$B = -bc(h) + \left(a\eta^2h + 2b\eta h^2 + \frac{3}{4}h^3\right)Q + \left(-\frac{1}{2}a\eta^2h^3 - \frac{1}{2}b\eta h^4 - \frac{1}{8}h^5\right)Q^2,$$

Now it is sufficient to prove that $A \succ 0$ and that $C - BA^{-1}B \succ 0$, noting that $A$, $B$ and $C$ commute as they are all polynomials in $Q$; it is sufficient to prove that $A \succ 0$

and $AC - B^2 \succ 0$. First considering $A$ we have that $A$ is symmetric and hence it is positive definite if and only if all it's eigenvalues are positive. We note that the eigenvalues of $A$ are precisely

$$\begin{aligned}
P_A(\lambda) &= h\left(-\frac{c(h)}{h} + (2b\eta + h)\lambda + (-a\eta^2 h - b\eta h^2 - \frac{1}{4}h^3)\lambda^2\right) \\
&\geq h\lambda\left(\frac{7b\eta}{4} + \frac{3h}{4} - \eta^2 h - b\eta h^2 M - \frac{h^3 M}{4}\right) \\
&\geq h\lambda\left(\frac{5b\eta}{4} + \frac{3h}{4} - \frac{3\eta^2 h}{4} - \frac{h}{16}\right) \geq h\lambda\left(\frac{b\eta}{2} + \frac{3h}{4} - \frac{h}{16}\right) > 0,
\end{aligned}$$

where $\lambda$ are the eigenvalues of $Q$, where $m \leq \lambda \leq M$ and we have used the fact that $h < \frac{1-\eta}{\sqrt{4M}}$ and $b \geq h$. Hence we have that $A \succ 0$. Now it remains to prove that $AC - B^2 \succ 0$, now we have that $AC - B^2$ is a polynomial of $Q$, which we denote $P_{AC-B^2}(Q)$ and hence has eigenvalues dictated by the eigenvalues $\lambda$ of $Q$. Now considering

$$\begin{aligned}
\frac{P_{AC-B^2}(\lambda)}{h\lambda} &= \frac{(\eta^2-1)\,c(h)}{hM\lambda} + h\left(\frac{1-\eta^2}{M} - \frac{\eta^2\lambda}{M^2} + \frac{2\eta(1-\eta^2)}{(1-\eta)M} + \frac{c(h)^2}{h^2 M\lambda}\right) \\
&+ h^2\frac{c(h)}{h}\left(-\frac{(1+\eta)^2}{M} + \frac{1+\eta}{(1-\eta)\lambda} + \frac{\eta^2\lambda}{M^2}\right) \\
&+ h^3\left(-1 - \frac{4\eta}{(1-\eta)^2} - \frac{c(h)^2}{h^2(1-\eta)^2\lambda} - \frac{\lambda}{4M} + \frac{3\eta^2\lambda}{4M} - \frac{\eta\lambda(1-\eta^2)}{(1-\eta)M}\right) \\
&+ h^4\frac{c(h)}{h}\left(1 + \frac{4\eta}{(1-\eta)^2} + \frac{\lambda(1+\eta^2)}{4M} + \frac{\eta\lambda}{M}\right) + h^5\lambda\left(\frac{3}{16} + \frac{\eta}{(1-\eta)^2}\right) \\
&+ h^6\frac{c(h)}{h}\left(-\frac{3\lambda}{16} - \frac{\eta\lambda}{(1-\eta)^2}\right) \geq -\frac{(1+\eta)h}{4M} + h\left(\frac{1+2\eta}{M}\right) \\
&+ h^3\left(-\frac{5}{4} - \frac{4\eta}{(1-\eta)^2} - \frac{1}{64(1-\eta)^2} - \eta(1+\eta)\right) - \frac{h^3}{32} > 0,
\end{aligned}$$

where we have used the fact that $h < \frac{1-\eta}{\sqrt{4M}}$. Hence $AC - B^2 \succ 0$ and our contraction results hold. All computations can be checked using Mathematica. We can bound $\mathcal{AB}$ operator on $||\cdot||_{a,b}$ by

$$\begin{aligned}
&||\Phi_{\mathrm{AB}}(\tilde{x}_k, \tilde{v}_k) - \Phi_{\mathrm{AB}}(x_k, v_k)||^2_{a,b} \leq \\
&3\left(\left(1 + \frac{ah^2M^2}{2}\right)||\overline{x}_k||^2 + \left(a + \frac{h^2}{4} + \frac{ah^4M^2}{8}\right)||\overline{v}_k||^2\right) \leq 7||\overline{x}_k, \overline{v}_k||^2_{a,b},
\end{aligned}$$

where we have used the norm equivalence in Sec. 2.2. We can also bound

$$\begin{aligned}
&||\Phi_{\mathrm{BAO}}(\tilde{x}_k, \tilde{v}_k) - \Phi_{\mathrm{BAO}}(x_k, v_k)||^2_{a,b} \\
&\leq 3\left(\left(1 + \frac{ah^2M^2}{4} + \frac{h^4M^2}{8}\right)||\overline{x}_k||^2 + \left(\frac{h^2}{2} + a\right)||\overline{v}_k||^2\right) \leq 7||\overline{x}_k, \overline{v}_k||^2_{a,b}.
\end{aligned}$$

Combining these results we have the required result. ∎

*Proof of Theorem 6.2.* We first note that $(\mathcal{OBABO})^n = \mathcal{OB}(\mathcal{ABOB})^{n-1}\mathcal{ABO}$. We will now focus our attention on proving contraction of $\mathcal{ABOB}$. Note we are only have to deal with a single evaluation of the Hessian at each step as the position variable is not updated between gradient evaluations. We will denote two synchronous realisations of ABOB as $(x_j, v_j)$ and $(\tilde{x}_j, \tilde{v}_j)$ for $j \in \mathbb{N}$. Now we will denote $\overline{x} := (\tilde{x}_j - x_j)$, $\overline{v} = (\tilde{v}_j - v_j)$ and $\overline{z} = (\overline{x}, \overline{v})$, where $\overline{z}_j = (\overline{x}_j, \overline{v}_j)$ for $j = k, k+1$ for $k \in \mathbb{N}$. We have the following update rule for $\overline{z}_k$

$$\overline{x}_{k+1} = \overline{x}_k + h\overline{v}_k, \quad \overline{v}_{k+1} = \eta\overline{v}_k - \frac{h}{2}(\eta+1)Q\left(\overline{x} + h\overline{v}\right),$$

where by mean value theorem we define $Q = \int_{t=0}^{1} \nabla^2 U(\tilde{x}_k + h\tilde{v}_k + t(x_k - \tilde{x}_k + h(v_k - \tilde{v}_k)))dt$, then $\nabla U\left(\tilde{x}_k + h\tilde{v}_k\right) - \nabla U\left(x_k + hv_k\right) = Q\left(\overline{x} + h\overline{v}\right)$. In the notation of (4.4) we have that for this scheme

$$\begin{aligned}
A &= -c\left(h\right) + bh\left(1+\eta\right)Q - (1+\eta)^2\frac{ah^2Q^2}{4},\\
B &= b\left(1-\eta\right) - h - bc\left(h\right) + \left(\frac{1}{2}a\eta + bh\right)\left(\eta+1\right)hQ - \left(\eta+1\right)^2\frac{ah^3}{4}Q^2,\\
C &= a\left(1-\eta^2\right) - 2b\eta h - h^2 - ac\left(h\right) + \left(a\eta + bh\right)\left(\eta+1\right)h^2Q - a\left(\eta+1\right)^2\frac{h^4}{4}Q^2,
\end{aligned}$$

where $\eta = \exp\{-\gamma h\}$. This form motivates the choice $b = \frac{h}{1-\eta}$ and $a = \frac{1}{M}$ inspired by the continuous dynamics. For our choice of $a$ and $b$, $B$ simplifies to

$$B = -bc\left(h\right) + (\frac{1}{2}a\eta + bh)(\eta+1)hQ - (\eta+1)^2\frac{ah^3}{4}Q^2.$$

We will now apply Prop. 4.2, first considering $A$ we have that the eigenvalues are precisely

$$P_A\left(\lambda\right) = -c\left(h\right) + bh(1+\eta)\lambda - (1+\eta)^2\frac{ah^2\lambda^2}{4} \geq h\lambda\left(\frac{3b}{4} + b\eta - \frac{h}{4} - \frac{3b\eta}{4}\right) > 0,$$

where $\lambda$ are the eigenvalues of $Q$, where $m \leq \lambda \leq M$ and we have used the fact that $b \geq h$. Hence we have that $A \succ 0$. Now it remains to prove that $AC - B^2 \succ 0$, now we have that $AC - B^2$ is a polynomial of $Q$, which we denote $P_{AC-B^2}(Q)$ and hence has eigenvalues dictated by the eigenvalues $\lambda$ of $Q$. Now considering

$$\begin{aligned}
\frac{P_{AC-B^2}(\lambda)}{h\lambda} &= \frac{\left(\eta^2-1\right)c(h)}{hM\lambda} + h\left(\frac{(1+\eta)^2}{M} + \frac{c(h)^2}{h^2M\lambda} - \frac{(1+\eta)^2\lambda}{4M^2}\right) \\
&+ h^2\frac{c(h)}{h}\left(-\frac{(1+\eta)^2}{M} - \frac{1}{\lambda} + \frac{2}{(1-\eta)\lambda} + \frac{\lambda(1+\eta)^2}{4M^2}\right) \\
&+ h^3\left(-\frac{(1+\eta)^2}{(1-\eta)^2} - \frac{c(h)^2}{(1-\eta)^2h^2\lambda} - \frac{\lambda(1+\eta)^2}{4M} + \frac{\lambda(1+\eta)^2}{2(1-\eta)M}\right) \\
&+ h^4\frac{c(h)}{h}\left(\frac{(1+\eta)^2}{(1-\eta)^2} + \frac{\lambda(1+\eta)^2}{4M} - \frac{\lambda(1+\eta)^2}{2M(1-\eta)}\right) \\
&> -\frac{h(1+\eta)}{4M} + h\left(\frac{3(1+\eta)^2}{4M}\right) - h\left(\frac{3(1+\eta)^2}{64M}\right) \\
&+ h\left(-\frac{(1+\eta)^2}{4M} - \frac{1}{64M}\right) > 0,
\end{aligned}$$

where we have used the fact that $h < \frac{1-\eta}{\sqrt{4M}}$. Hence $AC - B^2 \succ 0$ and our contraction results hold. All computations can be checked using Mathematica. We can bound $\mathcal{ABO}$ operator on $||\cdot||_{a,b}$ by

$$\begin{aligned}
&||\Phi_{\mathrm{ABO}}(\tilde{x}_k,\tilde{v}_k) - \Phi_{\mathrm{ABO}}(x_k,v_k)||^2_{a,b} \\
&\leq 3\left(\left(1+\frac{ah^2M^2}{2}\right)||\overline{x}_k||^2 + \left(a+h^2+\frac{ah^4M^2}{2}\right)||\overline{v}_k||^2\right) \\
&\leq 8||\overline{x}_k,\overline{v}_k||^2_{a,b},
\end{aligned}$$

where we have used the norm equivalence in Sec. 2.2. We can also bound

$$\begin{aligned}
||\Phi_{\mathrm{OB}}(\tilde{x}_k,\tilde{v}_k) - \Phi_{\mathrm{OB}}(x_k,v_k)||^2_{a,b} &\leq 3\left(\left(\frac{1}{2}+\frac{ah^2M^2}{4}\right)||\overline{x}_k||^2 + a||\overline{v}_k||^2\right) \\
&\leq 6||\overline{x}_k,\overline{v}_k||^2_{a,b}.
\end{aligned}$$

Combining these results we have the required result. ∎

*Proof of Theorem 7.1.* We remark that synchronously coupling between two realisations of the stochastic Euler scheme results in a synchronous coupling of $(\zeta_k,\omega_k)_{k\in\mathbb{N}}$. Now we will denote $\overline{x} := (\tilde{x}_j - x_j)$, $\overline{v} = (\tilde{v}_j - v_j)$ and $\overline{z} = (\overline{x},\overline{v})$, where $\overline{z}_j = (\overline{x}_j,\overline{v}_j)$ for $j = k, k+1$ for $k \in \mathbb{N}$. We have the following update rule for $\overline{z}_k$

$$\overline{x}_{k+1} = \overline{x}_k + \frac{1-\eta}{\gamma}\overline{v}_k - \frac{\gamma h+\eta-1}{\gamma^2}Q\overline{x}_k, \quad \overline{v}_{k+1} = \eta\overline{v}_k - \frac{1-\eta}{\gamma}Q\overline{x}_k,$$

where by mean value theorem we define $Q = \int_{t=0}^1 \nabla^2 U(\tilde{x}_k + t(x_k - \tilde{x}_k))dt$, then $\nabla U(\tilde{x}_k) - \nabla U(x_k) = Q(\overline{x} + h\overline{v})$. In the notation of (4.4) we have that for this scheme

$$
\begin{aligned}
A &= -c\,(h) + 2\left(\frac{b\,(1-\eta)}{\gamma} + \frac{\eta - 1 + \gamma h}{\gamma^2}\right) Q \\
&\quad - \left(\frac{a\,(1-\eta)^2}{\gamma^2} + \frac{2b\,(1-\eta)\,(-1+\eta+\gamma h)}{\gamma^3} + \frac{(-1+\eta+\gamma h)^2}{\gamma^4}\right) Q^2, \\
B &= b\,(1-\eta) - \frac{(1-\eta)}{\gamma} - bc\,(h) \\
&\quad + \left(\frac{a\eta\,(1-\eta)}{\gamma} + \frac{b\,(1-\eta)^2}{\gamma^2} + \frac{b\eta\,(-1+\eta+\gamma h)}{\gamma^2} + \frac{(1-\eta)\,(-1+\eta+\gamma h)}{\gamma^3}\right) Q, \\
C &= a(1-\eta^2) - ac\,(h) - \frac{2b\eta\,(1-\eta)}{\gamma} - \frac{(1-\eta)^2}{\gamma^2},
\end{aligned}
$$

where $\eta = \exp\{-\gamma h\}$. This form motivates the choice $b = \frac{1}{\gamma}$ and $a = \frac{1}{M}$ inspired by the continuous dynamics. For our choice of $a$ and $b$, $B$ simplifies to $B = -bc\,(h) + \mathcal{O}\,(Q)$. We will now apply Prop. 4.2, first considering $A$ we wish to show that all it's eigenvalues are positive which are precisely

$$
\begin{aligned}
P_A(\lambda) &:= -c\,(h) + 2\left(\frac{b\,(1-\eta)}{\gamma} + \frac{\eta - 1 + \gamma h}{\gamma^2}\right)\lambda \\
&\quad - \left(\frac{a\,(1-\eta)^2}{\gamma^2} + \frac{2b\,(1-\eta)\,(-1+\eta+\gamma h)}{\gamma^3} + \frac{(-1+\eta+\gamma h)^2}{\gamma^4}\right)\lambda^2, \\
&\geq h\lambda\left(\frac{7}{4\gamma} - \left(h + \frac{h^2 M}{\gamma} + \frac{h^3 M}{4}\right)\right) > 0,
\end{aligned}
$$

where $\lambda$ are the eigenvalues of $Q$, where $m \leq \lambda \leq M$ and using the fact that, $\gamma^2 \geq 4M$, $1-\eta \leq h\gamma$, $h\gamma + \eta - 1 \leq \frac{(h\gamma)^2}{2}$ and $h < \frac{1}{2\gamma}$. Hence we have that $A \succ 0$. Now it remains to prove that $AC - B^2 \succ 0$, now we have that $AC - B^2$ is a polynomial of $Q$, which we denote $P_{AC-B^2}(Q)$ and hence has eigenvalues dictated by the eigenvalues $\lambda$ of $Q$. Due to the fact that the terms are more complicated than the previous discretizations we choose a convenient way of expanding the expression which can obtain positive definiteness. That is to expand the expression in terms of $a$. By using for example

Mathematica one can show that $P_{AC-B^2}(\lambda) = c_0 + c_1 a + c_2 a^2$, where

$$\begin{aligned}
c_1 + c_2 a &= \left(\eta^2 - 1\right) c(h) + c(h)^2 + 2\left(1-\eta^2\right)\left(\frac{b(1-\eta)}{\gamma} + \frac{-1+\eta+\gamma h}{\gamma^2}\right)\lambda \\
&+ \frac{2b(1-\eta)\eta c(h)\lambda}{\gamma} - 2\left(\frac{b(1-\eta)}{\gamma} + \frac{-1+\eta+\gamma h}{\gamma^2}\right)c(h)\lambda + \frac{(1-\eta)^4\lambda^2}{\gamma^4} \\
&- \frac{2\eta(1-\eta)^2(-1+\eta+\gamma h)\lambda^2}{\gamma^4} - \frac{2b(1-\eta)(-1+\eta+\gamma h)\lambda^2}{\gamma^3} \\
&- \frac{\left(1-\eta^2\right)(-1+\eta+\gamma h)^2\lambda^2}{\gamma^4} + \frac{2b(1-\eta)(-1+\eta+\gamma h)c(h)\lambda^2}{\gamma^3} \\
&+ \frac{(-1+\eta+\gamma h)^2 c(h)\lambda^2}{\gamma^4} + a\left(-\frac{(1-\eta)^2\lambda^2}{\gamma^2} + \frac{(1-\eta)^2 c(h)\lambda^2}{\gamma^2}\right) \\
&\geq c_1 - \frac{(1-\eta)^2\lambda}{\gamma^2} + \frac{(1-\eta)^2 c(h)\lambda}{\gamma^2} \\
&\geq \left(\eta^2-1\right)c(h) + \left(1-\eta^2\right)\left(\frac{2h}{\gamma} - \frac{1-\eta}{(1+\eta)\gamma^2}\right)\lambda + \ldots \\
&\geq \lambda\left(\left(-\frac{h^2}{2}\right) + h^2\left(2 - \frac{1}{1+\eta}\right) + \ldots\right) \\
&> \lambda\left(\frac{h^2}{2} - \frac{h^2}{16} - \frac{h^3\gamma}{16} - \frac{h^3\gamma}{8} - \frac{h^3\gamma}{16}\right) \geq \lambda\left(\frac{7h^2}{16} - \frac{h^3\gamma}{4}\right),
\end{aligned}$$

where $h < \frac{1}{2\gamma}$, $\gamma^2 \geq 8M \geq 8m$, $\frac{h\gamma}{2} \leq 1-\eta \leq h\gamma$ and $h\gamma + \eta - 1 \leq \frac{(h\gamma)^2}{2}$. Further we have that

$$\begin{aligned}
c_0 &= \frac{(1-\eta)^2 c(h)}{\gamma^2} + \frac{2b(1-\eta)\eta c(h)}{\gamma} - b^2 c(h)^2 - \frac{2b(1-\eta)^3\lambda}{\gamma^3} \\
&- \frac{2(1-\eta)^2(-1+\eta+\gamma h)\lambda}{\gamma^4} - \frac{4b^2\eta(1-\eta)^2\lambda}{\gamma^2} - \frac{4b\eta(1-\eta)(-1+\eta+\gamma h)\lambda}{\gamma^3} \\
&- \frac{b^2\eta^2\lambda^2(\eta+\gamma h-1)^2}{\gamma^4} + \frac{2b^2(1-\eta)^2 c(h)\lambda}{\gamma^2} + \frac{2b^2\eta c(h)\lambda(\eta+\gamma h-1)}{\gamma^2} \\
&+ \frac{2b^2\eta(1-\eta)^2\lambda^2(\eta+\gamma h-1)}{\gamma^4} - \frac{b^2(1-\eta)^4\lambda^2}{\gamma^4} + \frac{2b(1-\eta)c(h)\lambda(\eta+\gamma h-1)}{\gamma^3} \\
&> \lambda\left(-\frac{2(h\gamma)^3}{\gamma^4} - \frac{(h\gamma)^4}{\gamma^4} - \frac{4(h\gamma)^2}{\gamma^4} - \frac{2(h\gamma)^3}{\gamma^4} - \frac{(h\gamma)^4\lambda}{4\gamma^6} - \frac{(h\gamma)^4\lambda}{\gamma^6}\right) \\
&> \lambda\left(-\frac{4h^3}{\gamma} - \frac{4h^2}{\gamma^2} - 2h^4\right) > \lambda\left(-\frac{7h^2}{\gamma^2}\right),
\end{aligned}$$

now we can combine this with the previous estimate and we have

$$P_{AC-B^2}(\lambda) > \lambda\left(\frac{7h^2}{16M} - \frac{h^3\gamma}{4M} - \frac{7h^2}{\gamma^2}\right) > h^2\lambda\left(\frac{5}{16M} - \frac{7}{\gamma^2}\right) \geq 0,$$

which is true when $\gamma \geq 5\sqrt{M}$. Hence $AC - B^2 \succ 0$ and our contraction results hold. All computations can be checked using Mathematica. ∎