Langevin equations in the small-mass limit: Higher-order approximations

Jeremiah Birrell and Jan Wehr

Abstract. We study the small-mass (overdamped) limit of Langevin equations for a particle in a potential and/or magnetic field with matrix-valued and state-dependent drift and diffusion. We utilize a bootstrapping argument to derive a hierarchy of approximate equations for the position degrees of freedom that are able to achieve accuracy of order $m^{\ell/2}$ over compact time intervals for any $\ell \in \mathbb{Z}^+$. This generalizes prior derivations of the homogenized equation for the position degrees of freedom in the $m \to 0$ limit, which result in order $m^{1/2}$ approximations. Our results cover bounded forces, for which we prove convergence in L^p norms, and unbounded forces, in which case we prove convergence in probability.

1. Introduction

Langevin equations provide models of a diffusing particle; a simple example, illustrating several typical ingredients, is the system of stochastic differential equations (SDE)

$$dq_t^m = v_t^m dt, \ \ mdv_t^m = -\gamma v_t^m dt + \sigma dW_t, \tag{1}$$

where m is the mass of the particle (here and in the sequel we use a superscript to denote the m dependence), γ and σ are the dissipation (or: drag) and diffusion coefficients respectively and W_t is a Wiener process. Pioneering work, including investigation of the small-mass limit, was done by Smoluchowski [1] and Kramers [2]. A detailed discussion of the early literature can be found in [3].

Works studying the small-mass limit of (various generalizations of) Eq. (1) have rigorously established convergence of the position degrees of freedom, q_t^m, as $m \to 0$ to the solution, q_t, of a limiting SDE [4, 5, 6, 7, 8, 9].

Such problems fall under the umbrella of homogenization (see, for example, the recent sources [10, 11]) and so we refer to the SDE for q_t as the homogenized equation. Moving beyond the homogenized equation, results have been proven regarding the small-mass limit in the sense of rough-paths [12], the limit of the joint distribution of position and (scaled) velocity [13], the limit of the invariant measures [14], and the limit of the entropy production [15].

In this paper we build on the above small-mass limit results, specifically those in [9], where it was proven that q_t approximates q_t^m with $O(m^{1/2})$ error over compact time intervals (see also the summary in Section 1.2 below). Our main result is the derivation of a hierarchy of SDEs whose solutions are higher-order approximations to q_t^m; this hierarchy is able to achieve $O(m^{\ell/2})$ error over compact time intervals for any $\ell \in \mathbb{Z}^+$. We derive these approximations and prove the claimed error bounds via a bootstrapping argument. Related techniques are commonly used in many different contexts, such as in proving regularity of solutions to various classes of equations (see page 20 in [16] and page 489 in [17]), to derive error estimates for numerical methods [18, 19] and in predictor-corrector methods (see Chapter 15.5 in [20]), and in homogenization (see Chapters 16 and 20 in [11]) and perturbation theory (see Chapter 3.2 in [21] and Chapter 6.2 - 6.3 in [22]).

1.1. Langevin Equation with State-Dependent Drag and Noise

In this work, we study generalizations of Eq. (1) that allow for time- and state-dependent drag, noise, and external forcing:

$$dq_t^m = v_t^m dt, \qquad (2)$$

$$md(v_t^m)_i = \left(-\tilde{\gamma}_{ik}(t, q_t^m)(v_t^m)^k + F_i(t, q_t^m) \right) dt + \sigma_{i\rho}(t, q_t^m)dW_t^\rho, \qquad (3)$$

where q_t^m and v_t^m are $\mathbb{R}^n$-valued processes. The matrix-valued function $\tilde{\gamma}$ will have a symmetric part, the drag matrix, and is allowed to have an antisymmetric part, coming from a possible magnetic field; see Eq. (5) below for details. We again alert the reader that here, and elsewhere, the superscript m on vector or matrix-valued quantities denotes the value of the mass and not a component or a power.

Except in the simplest cases, the system Eq. (2) - Eq. (3) cannot be solved explicitly, and it is difficult to study numerically, especially for small values of m, since the velocity process v_t^m diverges as $m \to 0$. As discussed above, solutions of the homogenized SDE can serve as approximations to (the position components of) solutions of the original system, as long as the value of m is sufficiently small. The effectiveness of this has been confirmed numerically and experimentally for physically relevant values of m [5]. However, this approximate solution is independent of m. The present work improves on this, by deriving approximate position processes which are sensitive to the variation of m while still not requiring one to solve the full system. In addition, these m-dependent approximations are free from the type of singularity that makes the original system Eq. (2)-Eq. (3) difficult to work with when m is small.

More specifically, we obtain a hierarchy of approximations $q_t^{\ell,m}$, $\ell \in \mathbb{Z}^+$, starting with $q_t^{1,m} \equiv q_t$, where $q_t^{\ell,m}$ approximates q_t^m with $O(m^{\ell/2})$ error over compact time intervals. These processes will be constructed inductively (on ℓ) as solutions to SDEs of the form

$$dq_t^{\ell,m} = \tilde{b}(t, q_t^{\ell,m})dt + \tilde{\sigma}(t, q_t^{\ell,m})dW_t + \sqrt{m}dR_t^{\ell-1,m}. \qquad (4)$$

Here and in the following, SDEs are defined in the Itô sense.

The leading order terms in Eq. (4) are given by the same drift, $\tilde{b}$, (including the noise-induced drift from [6]) and diffusion, $\tilde{\sigma}$, that appear in the homogenized SDE for q_t (see Eq. (10) below). The corrections are captured by the semimartingale term $R_t^{\ell-1,m}$. The appropriate form of $R_t^{\ell-1,m}$ will be motivated by comparing the SDE for q_t^m to the SDE for the homogenized process, q_t, and extracting the error terms.

What makes the hierarchy particularly simple is that $R_t^{\ell-1,m}$ does *not* depend on $q_t^{\ell,m}$, but rather is an external driving semimartingale, constructed from the approximation at the previous step, $q_t^{\ell-1,m}$ (with $R_t^{0,m} \equiv 0$). This means that the singular nature of the $m \to 0$ limit does not complicate the limiting drift and diffusion, even for higher-order approximations. Moreover, the presence of m in the correction process, $R_t^{\ell-1,m}$, is rather benign; it primarily serves to exponentially damp out contributions from the past history of $q_t^{\ell-1,m}$.

In Section 1.2 we summarize the prior results that will be needed in this paper. Section 1.3 gives a summary of the new results that will be established. Section 2 contains an outline of our proof strategy, in a simplified setting, in order to highlight the key ideas. Our new results are fully developed in Sections 3 and 4. The former covers Langevin equations driven by bounded forces and the latter covers the extension to unbounded forces.

1.2. Homogenized Equation in the $m \to 0$ Limit: Established Results

Here we recall several previously proven results, pertaining to the Langevin equation Eq. (2)-Eq. (3), that will be needed going forward. Before doing so, we need to be a bit more specific about the objects appearing in Eq. (3). We will assume:

1. $\tilde{\gamma}$ is constructed from a continuous, positive definite matrix-valued drag, γ, and an antisymmetric part generated by a C^2 vector potential, ψ, as follows:

$$\tilde{\gamma}_{ik}(t, q) \equiv \gamma_{ik}(t, q) + \partial_{q^k}\psi_i(t, q) - \partial_{q^i}\psi_k(t, q). \qquad (5)$$

2. The diffusion, $\sigma : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^{n \times k}$ is continuous.
3. W is an $\mathbb{R}^k$-valued Wiener process on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, a filtered probability space satisfying the usual conditions [23].
4. The total forcing is

$$F(t, q) = -\partial_t \psi(t, q) - \nabla_q V(t, q) + \tilde{F}(t, q), \qquad (6)$$

where the C^2 function V represents an (electrostatic) potential and $\tilde{F}$ is a continuous external forcing.

Our usage of electromagnetic language is due to us viewing the antisymmetric part of the drag-matrix, $\tilde{\gamma}$, as begin generated by the vector-potential of an electromagnetic field, ψ. If one is not interested in such a term, then our framework still allows for consideration of quite general gradient and non-gradient forces. More assumptions on these objects will be required as we proceed; in particular the L^p-convergence result of Theorem 3.8 will apply only to bounded forces but Theorem 4.1 will prove convergence in probability for a large class of unbound forces.

Next, define the (kinematic) momentum

$$u_t^m = m v_t^m. \tag{7}$$

We showed in [9] that, under appropriate assumptions, there exist unique solutions (q_t^m, u_t^m), $t \in [0, \infty)$, that converge to $(q_t, 0)$ as $m \to 0$; q_t is the solution of a homogenized limiting SDE. The precise nature of this convergence, and the form of the SDE for q_t, are given below. These results provide the foundation that we build upon in order to derive higher-order approximations.

Summary of Previous Results:

Under the assumptions listed in Appendix A, one has the following convergence results (see [9] for a detailed proof): For any $T > 0$, $p > 0$, $\epsilon > 0$ we have

$$E\left[\sup_{t \in [0,T]} \|q_t^m - q_t\|^p\right]^{1/p} = O(m^{1/2-\epsilon}), \quad \sup_{t \in [0,T]} E\left[\|q_t^m - q_t\|^p\right]^{1/p} = O(m^{1/2}), \tag{8}$$

$$E\left[\sup_{t \in [0,T]} \|u_t^m\|^p\right]^{1/p} = O(m^{1/2-\epsilon}), \quad \sup_{t \in [0,T]} E\left[\|u_t^m\|^p\right]^{1/p} = O(m^{1/2}) \tag{9}$$

as $m \to 0$, where q_t is the solution to the SDE

$$dq_t = \tilde{\gamma}^{-1}(t, q_t) F(t, q_t) dt + S(t, q_t) dt + \tilde{\gamma}^{-1}(t, q_t) \sigma(t, q_t) dW_t. \tag{10}$$

$S(t, q)$ is called the noise-induced drift, *see [6, 9], and is given by (employing the summation convention on repeated indices):*

1. *$S^i(t, q) \equiv \partial_{q^k}(\tilde{\gamma}^{-1})^{ij}(t, q) \delta^{kl} G_{jl}^{rs}(t, q) \Sigma_{rs}(t, q)$,*
2. *$G_{ij}^{kl}(t, q) \equiv \delta^{rk} \delta^{sl} \int_0^\infty (e^{-\zeta \tilde{\gamma}(t,q)})_{ir} (e^{-\zeta \tilde{\gamma}(t,q)})_{js} d\zeta$,*
3. *$\Sigma_{ij} \equiv \sigma_{i\rho} \sigma_{j\xi} \delta^{\rho\xi}$.*

Here δ^{kl} denotes the Kronecker delta.

The initial conditions are assumed to satisfy $E[\|q_0^m\|^p] < \infty$, $E[\|q_0\|^p] < \infty$, and $E[\|q_0^m - q_0\|^p]^{1/p} = O(m^{1/2})$ for all $p > 0$.

The following bounds on q_t^m and q_t were also shown:

$$E\left[\sup_{t\in[0,T]} \|q_t^m\|^p\right] < \infty, \quad E\left[\sup_{t\in[0,T]} \|q_t\|^p\right] < \infty \tag{11}$$

for all $m > 0$, $T > 0$, $p > 0$.

Note that $u_t^m = O(m^{1/2})$ translates into $v_t^m = O(m^{-1/2})$. Also, in the above, we have defined the index placement on $\tilde{\gamma}^{-1}$ so that

$$(\tilde{\gamma}^{-1})^{ij}\tilde{\gamma}_{jk} = \delta_k^i, \tag{12}$$

and for any v_i we define the contraction $(\tilde{\gamma}^{-1}v)^i = (\tilde{\gamma}^{-1})^{ij}v_j$. Finally, note that what we call $\tilde{F}$ here was simply called F in [9], whereas here we use F to refer to Eq. (6).

As stated previously, a comprehensive list of assumptions that guarantee the above convergence and boundedness properties can be found in Appendix A. Of particular significance, we assume that the symmetric part, γ, of $\tilde{\gamma}$ is positive definite; this ensures that the term involving γ is dissipative, and is also sufficient to ensure that $\tilde{\gamma}$ is invertible.

1.3. Summary of New Results

The main content of the present work is the derivation of a hierarchy of approximating equations for the position degrees of freedom, generalizing the $O(m^{1/2})$-accurate Eq. (10), that is capable of approximating q_t^m to order $O(m^{\ell/2})$ for any ℓ. This is done in Section 3 under appropriate boundedness assumptions on the coefficients of the equation. Specifically, in Theorem 3.8 we show that for each $\ell \in \mathbb{Z}^+$ there is a family of $\mathbb{R}^n$-valued semimartingales, $R_t^{\ell-1,m}$ such that the solutions to the SDEs

$$dq_t^{\ell,m} = \tilde{\gamma}^{-1}(t, q_t^{\ell,m})F(t, q_t^{\ell,m})dt + S(t, q_t^{\ell,m})dt \tag{13}$$
$$+ \tilde{\gamma}^{-1}(t, q_t^{\ell,m})\sigma(t, q_t^{\ell,m})dW_t + \sqrt{m}dR_t^{\ell-1,m}$$

satisfy

$$\sup_{t\in[0,T]} E[\|q_t^m - q_t^{\ell,m}\|^p]^{1/p} = O(m^{\ell/2}), \tag{14}$$

$$E\left[\sup_{t\in[0,T]} \|q_t^m - q_t^{\ell,m}\|^p\right]^{1/p} = O(m^{\ell/2-\epsilon})$$

for all $T > 0$, $p > 0$, $\epsilon > 0$. We call $q_t^{\ell,m}$ the solution at the ℓ'th level of the hierarchy.

In Section 4 we will use the technique developed in [7] to significantly relax the assumption of bounded forcing, while still obtaining convergence in probability:

$$\lim_{m\to 0} P\left(\frac{\sup_{t\in[0,T]}\|q_t^m - q_t^{\ell,m}\|}{m^{\ell/2-\epsilon}} > \delta\right) = 0 \tag{15}$$

for all $T > 0$, $\delta > 0$, $\epsilon > 0$, $\ell \in \mathbb{Z}^+$; see Theorem 4.1.

The hierarchy begins with $R_t^{0,m} \equiv 0$, $q_t^{1,m} \equiv q_t$, the solution to the homogenized SDE, Eq. (10). We emphasize that $R_t^{\ell-1,m}$ acts as an external forcing semimartingale, and is *not* dependent on $q_t^{\ell,m}$. See Chapter V in [24] and Appendix D below for the general theory of SDEs that include forcing terms of this type.

Each $R_t^{\ell-1,m}$ will be defined in terms of $q^{\ell-1,m}|_{[0,t]}$, the approximation at the $\ell - 1$st step up to time t. In fact, it will be useful to think of the $R_t^{\ell,m}$ as functions of a continuous semimartingale. Thought of this way, they will satisfy $R_t^{\ell-1,m} = R_t^m[q^{\ell-1,m}]$. The mapping R_t^m is constructed by comparing the SDE for q_t^m with the homogenized SDE, Eq. (10), for q_t and extracting the error (i.e., remainder) terms. This will be carried out in Section 3.1; see Definition 3.7 for the precise definition of $R_t^m[y]$.

2. Outline of the Proof in a Simplified Setting

The convergence rates in Eq. (14) will be obtained by showing that, for an appropriate class of continuous semimartingales, y, the remainder terms $R_t^m[y]$ are Lipschitz transformations of $y|_{[0,t]}$ (Lipschitz with respect to pairs of norms that will be specified below), and then inductively using a Gronwall's inequality argument.

The proof of these Lipschitz properties is quite technical, and so we first provide an outline of our argument in the following simplified setting: Here we work in $n = 1$ dimensions, and consider the SDE

$$dq_t^m = v_t^m \, dt, \qquad\qquad (16)$$
$$m \, dv_t^m = F(q_t^m) \, dt - \gamma v_t^m \, dt + \sigma \, dW_t,$$

where γ, σ are positive constants or, in terms of $u_t^m \equiv m v_t^m$:

$$dq_t^m = \frac{1}{m} u_t^m \, dt, \qquad\qquad (17)$$
$$du_t^m = F(q_t^m) \, dt - \frac{1}{m} \gamma u_t^m \, dt + \sigma \, dW_t.$$

Solving the second equation for $\frac{1}{m}\gamma u_t^m \, dt$, substituting into the first, and rewriting in terms of

$$z_t^m \equiv \sqrt{m} v_t^m = u_t^m / \sqrt{m} \qquad\qquad (18)$$

(the velocity, normalized to be of order 1), one finds

$$dq_t^m = \gamma^{-1} F(q_t^m) \, dt + \gamma^{-1} \sigma \, dW_t - \sqrt{m} \gamma^{-1} \, dz_t^m. \qquad\qquad (19)$$

The last term is of order $\sqrt{m}$, so it vanishes in the limit $m \to 0$. Indeed, it can be shown that the limit q_t of the processes q_t^m satisfies

$$dq_t = \gamma^{-1} F(q_t) \, dt + \gamma^{-1} \sigma \, dW_t \qquad\qquad (20)$$

(see [3]). This is a considerably simpler equation than the original system and in many situations its solution q_t furnishes a good approximation of q_t^m. Here we are interested in a more accurate result, approximating q_t^m by an

m-dependent process that can still be obtained as a solution to a first-order SDE, albeit a somewhat more complicated one. In fact, we will obtain a hierarchy of such equations, whose solutions will approximate q_t^m to within an arbitrary power of m. To implement it, we do not neglect the remainder

$$\sqrt{m}\gamma^{-1} dz_t^m. \tag{21}$$

Instead, we rewrite it by solving the equation for u_t^m as an inhomogeneous linear equation (and multiplying the result by $\frac{1}{\sqrt{m}}$ to obtain z_t^m):

$$z_t^m = \frac{1}{\sqrt{m}}e^{-\frac{\gamma}{m}t}\left(\sqrt{m}z_0^m + \int_0^t e^{\frac{\gamma}{m}s}F\left(q_s^m\right) ds + \int_0^t e^{\frac{\gamma}{m}s}\sigma\, dW_s\right). \tag{22}$$

Substituting this expression into the remainder term, we obtain the delay equation

$$dq_t^m = \gamma^{-1}F\left(q_t\right) dt + \gamma^{-1}\sigma\, dW_t + \sqrt{m}\, dR_t^m[q^m], \tag{23}$$

where, for an arbitrary continuous semimartingale y_t, we define

$$dR_t^m[y] = \gamma^{-1}\, dz_t^m[y] \tag{24}$$

with

$$z_t^m[y] = \frac{1}{\sqrt{m}}e^{-\frac{\gamma}{m}t}\left(\sqrt{m}\tilde{z}_0 + \int_0^t e^{\frac{\gamma}{m}s}F\left(y_t\right) ds + \int_0^t e^{\frac{\gamma}{m}s}\sigma\, dW_s\right). \tag{25}$$

The equation for q_t^m can now be thought of as a fixed point problem. We will solve it iteratively, defining a sequence $q_t^{\ell,m}$ of approximate solutions, starting from $q_t^{1,m} = q_t$—the solution of the homogenized equation in the $m \to 0$ limit, and, given $q_t^{\ell-1,m}$, defining $q_t^{\ell,m}$ as the solution of the SDE

$$dq_t^{\ell,m} = \gamma^{-1}F\left(q_t^{\ell,m}\right) dt + \gamma^{-1}\sigma\, dW_t + \sqrt{m}dR_t^m\left[q^{\ell-1,m}\right]. \tag{26}$$

Note that the the first two terms are the same as in the SDE Eq. (20), while the last term is a fixed semimartingale forcing term, i.e., it does not depend on the process, $q_t^{\ell,m}$, that one is solving for. As we will see, these two features are maintained in the general case.

We prove the claimed $O(m^{\ell/2})$ difference between $q_t^{\ell,m}$ and q_t^m under a variety of norms, by using Gronwall's inequality and a bootstrapping argument. For specificity, here we outline the argument for the norm

$$\|y - \tilde{y}\|_{2,T} \equiv \sup_{0 \leq t \leq T} E[|y_t - \tilde{y}_t|^2]^{1/2}. \tag{27}$$

Subtracting Eq. (23) and Eq. (26), using the triangle inequality, and the simple bound $(a + b)^2 \leq 2(a^2 + b^2)$ for $a, b \geq 0$, we find

$$\|q^m - q^{\ell,m}\|_{2,T}^2 \tag{28}$$

$$\leq 2\|\int_0^t \gamma^{-1}F\left(q_s^m\right) - \gamma^{-1}F\left(q_s^{\ell,m}\right) ds\|_{2,T}^2 + 2m\|R^m[q^m] - R^m[q^{\ell-1,m}]\|_{2,T}^2.$$

Suppose now that F is Lipschitz and $R^m[y]$ is Lipschitz in y for the norm $\|\cdot\|_{2,T}$ (the latter is an oversimplification, but will allow us to illustrate the main idea without additional technical complications). Estimating the

first term in Eq. (28) by using the Lipschitz property for F, along with the Cauchy-Schwarz inequality on the Lebesgue integral, and the second, using the Lipschitz property of R^m; one arrives at

$$\|q^m - q^{\ell,m}\|_{2,T}^2 \le C \int_0^T \|q^m - q^{\ell,m}\|_{2,t}^2 dt + Cm\|q^m - q^{\ell-1,m}]\|_{2,T}^2 \qquad (29)$$

for some constant $C > 0$ (independent of m). Gronwall's inequality then yields

$$\|q^m - q^{\ell,m}\|_{2,T}^2 \le Cm\|q^m - q^{\ell-1,m}]\|_{2,T}^2 \qquad (30)$$

for (a different) $C > 0$. Taking the square root, one finds that the iteration from $q^{\ell-1,m}$ to $q^{\ell,m}$ has improved the error by a factor of $\sqrt{m}$. Starting with the base case $\|q^m - q^{1,m}\|_{2,T} = \|q^m - q\|_{2,T} = O(m^{1/2})$ (this is the error bound for the overdamped limit, proven in [9]) one obtains the claimed error bounds $\|q^m - q^{\ell,m}\|_{2,T} = O(m^{\ell/2})$ for all $\ell \in \mathbb{Z}^+$.

The proof in the general case follows the above outline, but introduces several technical complications:

1. Working in arbitrary dimension $n \ge 1$, and with state-dependent, matrix-valued γ and σ, complicates the derivation of the overdamped limit Eq. (20), as well as the remainder term in Eq. (23). The required computations are found in [9], but we outline them in Section 3.1 for completeness.
2. We prove below that $R^m[y]$ is Lipschitz under pairs of related norms, but not with respect to a single norm as was assumed above; see Lemmas 3.5 and 3.6. This constitutes the greatest technical hurdle in this paper, and requires repeated use of the inequalities collected in Appendix C.
3. The computations in Eq. (28) - Eq. (30) must be generalized beyond the $\|\cdot\|_{2,T}$-norm, to accommodate the usage of the norm-pairs mentioned in item (2). In addition, the stochastic integral terms will no longer cancel, due to state-dependence of γ and σ. These generalizations require further use of the inequalities from Appendix C; see Theorem 3.8.

3. Derivation of the Approximation Hierarchy for Bounded Forcing

Having outlined our argument, we now begin a detailed derivation of the hierarchy of approximating equations, in the general setting laid out in Section 1.2 and under the assumptions from Appendix A; in particular, for bounded forcing. The next two subsections lay the analytical groundwork, while the definition of the approximating hierarchy and the convergence proof are found in Section 3.3.

3.1. Identifying the Remainder Terms

It is convenient to rewrite the system Eq. (2) - Eq. (3) in terms of u_t^m (see Eq. (7)):

$$dq_t^m = \frac{1}{m} u_t^m dt, \tag{31}$$

$$d(u_t^m)_i = \left(-\frac{1}{m} \tilde{\gamma}_{ik}(t, q_t^m)(u_t^m)^k + F_i(t, q_t^m) \right) dt + \sigma_{i\rho}(t, q_t^m) dW_t^\rho. \tag{32}$$

The next step is to combine the SDEs for q_t^m and u_t^m and decompose the result into two pieces: one that becomes the homogenized SDE, Eq. (10), in the $m \to 0$ limit, and a remainder term that will motivate the definition of $R_t^m[y]$.

Eq. (32) is a linear equation for u_t^m, so the pathwise solution to

$$\frac{d}{dt} \Phi_t^m = -\frac{1}{m} \tilde{\gamma}(t, q_t^m) \Phi_t^m, \quad \Phi_0^m = I, \tag{33}$$

(i.e. the fundamental-solution process; see Appendix B) furnishes us with an explicit formula for u_t^m in terms of q_t^m:

$$u_t^m = \Phi_t^m \left(u_0^m + \int_0^t (\Phi_s^m)^{-1} F(s, q_s^m) ds + \int_0^t (\Phi_s^m)^{-1} \sigma(s, q_s^m) dW_s \right). \tag{34}$$

In principle, the above formula for u_t^m allows one to formulate a delay equation for q_t^m (i.e. with the right hand side depending on $q^m|_{[0,t]}$) by substituting Eq. (34) into Eq. (31). However, doing so in this form does little to shed light on the behavior in the singular $m \to 0$ limit. Nevertheless, by first rewriting the equation for q_t^m in an equivalent form we can turn this into a fruitful idea.

We begin by mimicking the convergence proof of q_t^m to q_t, as found in [9], and separating the terms that survive in the $m \to 0$ limit from the remaining $O(\sqrt{m})$ error terms, which will then be used to define $R_t^m[y]$. To make this section more self-contained, we will repeat a portion of that derivation here:

First solve Eq. (32) for $\frac{1}{m} u_t^m dt$ and substitute into Eq. (31) to obtain

$$d(q_t^m)^i = (\tilde{\gamma}^{-1})^{ij}(t, q_t^m) F_j(t, q_t^m) dt + (\tilde{\gamma}^{-1})^{ij}(t, q_t^m) \sigma_{j\rho}(t, q_t^m) dW_t^\rho \tag{35}$$
$$- (\tilde{\gamma}^{-1})^{ij}(t, q_t^m) d(u_t^m)_j.$$

Integrating the last term by parts results in

$$- (\tilde{\gamma}^{-1})^{ij}(t, q_t^m) d(u_t^m)_j = -d((\tilde{\gamma}^{-1})^{ij}(t, q_t^m)(u_t^m)_j) \tag{36}$$
$$+ (u_t^m)_j \partial_t (\tilde{\gamma}^{-1})^{ij}(t, q_t^m) dt + \frac{1}{m}(u_t^m)_j(u_t^m)_k \delta^{kl} \partial_{q^l} (\tilde{\gamma}^{-1})^{ij}(t, q_t^m) dt.$$

From Eq. (9), we see that $u_t^m / \sqrt{m}$ is $O(1)$ as $m \to 0$, so the last term above is $O(1)$ as $m \to 0$ and must be further decomposed to identify the q-dependent

piece that survives in the limit. To do this, use Eq. (32) to compute

$$d((u_t^m)_i(u_t^m)_j) = (u_t^m)_i d(u_t^m)_j + (u_t^m)_j d(u_t^m)_i + d[u_i^m, u_j^m]_t \qquad (37)$$

$$=\frac{1}{m}(-(u_t^m)_i\tilde{\gamma}_{jk}(t, q_t^m) - (u_t^m)_j\tilde{\gamma}_{ik}(t, q_t^m))(u_t^m)_l\delta^{kl}dt$$
$$+ ((u_t^m)_i F_j(t, q_t^m) + (u_t^m)_j F_i(t, q_t^m))\, dt$$
$$+ (u_t^m)_i\sigma_{j\rho}(t, q_t^m)dW_t^\rho + (u_t^m)_j\sigma_{i\rho}(t, q_t^m)dW_t^\rho + \Sigma_{ij}(t, q_t^m)dt.$$

We wish to solve for $\frac{1}{m}(u_t^m)_j(u_t^m)_k dt$, hence we rewrite this as

$$\frac{1}{m}(\tilde{\gamma}_{jk}(t, q_t^m)(u_t^m)_l(u_t^m)_i + \tilde{\gamma}_{ik}(t, q_t^m)(u_t^m)_l(u_t^m)_j)\delta^{kl}dt \qquad (38)$$

$$= - d((u_t^m)_i(u_t^m)_j) + ((u_t^m)_i F_j(t, q_t^m) + (u_t^m)_j F_i(t, q_t^m))\, dt$$
$$+ (u_t^m)_i\sigma_{j\rho}(t, q_t^m)dW_t^\rho + (u_t^m)_j\sigma_{i\rho}(t, q_t^m)dW_t^\rho + \Sigma_{ij}(t, q_t^m)dt.$$

If $\tilde{\gamma}$ is scalar-valued we can immediately solve for $\frac{1}{m}(u_t^m)_j(u_t^m)_k dt$. In general, one must solve a Lyapunov equation (see Eq.(4.15) and surrounding material in [9] for details). Doing so, and substituting back into Eq. (35) results in

$$d(q_t^m)^i =(\tilde{\gamma}^{-1})^{ij}(t, q_t^m)F_j(t, q_t^m)dt + Q^{ikl}(t, q_t^m)J_{kl}(t, q_t^m)dt \qquad (39)$$
$$+ (\tilde{\gamma}^{-1})^{ij}(t, q_t^m)\sigma_{j\rho}(t, q_t^m)dW_t^\rho + \sqrt{m}d(R_t^m)^i,$$

where

$$J_{ij}(t, q) \equiv G_{ij}^{kl}(t, q)\Sigma_{kl}(t, q), \qquad (40)$$

$$G_{ij}^{kl}(t, q) \equiv \delta^{rk}\delta^{sl}\int_0^\infty (e^{-\zeta\tilde{\gamma}(t,q)})_{ir}(e^{-\zeta\tilde{\gamma}(t,q)})_{js}d\zeta, \qquad (41)$$

$$Q^{ijl}(t, q) \equiv \partial_{q^k}(\tilde{\gamma}^{-1})^{ij}(t, q)\delta^{kl}, \qquad (42)$$

and, defining the $O(1)$ processes

$$z_t^m \equiv u_t^m/\sqrt{m}, \qquad (43)$$

$$d(R_t^m)^i \equiv - d((\tilde{\gamma}^{-1})^{ij}(t, q_t^m)(z_t^m)_j) + (z_t^m)_j\partial_t(\tilde{\gamma}^{-1})^{ij}(t, q_t^m)dt \qquad (44)$$
$$+ Q^{ikl}(t, q_t^m)G_{kl}^{ab}(t, q_t^m)((z_t^m)_a F_b(t, q_t^m) + (z_t^m)_b F_a(t, q_t^m))\, dt$$
$$+ (z_t^m)_a(z_t^m)_b(z_t^m)^c\partial_{q^c}(Q^{ikl}G_{kl}^{ab})(t, q_t^m)dt$$
$$+ Q^{ikl}(t, q_t^m)G_{kl}^{ab}(t, q_t^m)((z_t^m)_a\sigma_{b\rho}(t, q_t^m) + (z_t^m)_b\sigma_{a\rho}(t, q_t^m))\, dW_t^\rho$$
$$- \sqrt{m}d(Q^{ikl}(t, q_t^m)G_{kl}^{ab}(t, q_t^m)(z_t^m)_a(z_t^m)_b)$$
$$+ \sqrt{m}(z_t^m)_a(z_t^m)_b\partial_t(Q^{ikl}G_{kl}^{ab})(t, q_t^m)dt,$$

with $R_0^m = 0$. Though the precise form of R_t^m is less than intuitive, we emphasize that it simply constitutes all of the terms in the equation for q_t^m that are shown in [9] to not contribute in the $m \to 0$ limit. These terms will of course contribute to the higher-order approximations.

We have written R_t^m in terms of z_t^m (which is $O(1)$) so that the order in m of each term is more obvious. Note that the above definition of the remainder R_t^m differs from that in [9] by the factor of $\sqrt{m}$ that we have

explicitly pulled out in Eq. (39). Also, in obtaining Eq. (44) we have integrated the term $-Q^{ikl}(t,q_t^m)G_{kl}^{ab}(t,q_t^m)d((u_t^m)_a(u_t^m)_b)$ from [9] by parts.

One can then use Eq. (34) to write z_t^m in terms of q_t^m and substitute into Eq. (44). In this manner, we can view Eq. (39) as a delay equation for q_t^m, which will be the basis for the rest of the derivation.

3.2. Lipschitz and Boundedness Properties

As discussed above, we are viewing z_t^m as defined in terms of $q^m|_{[0,t]}$ via Eq. (34) and Eq. (43), and similarly for R_t^m, by substituting the expression for z_t^m into Eq. (44).

It will be useful to view both z_t^m and R_t^m as functions of an arbitrary continuous semimartingale, y, as follows:

Definition 3.1. *For y a continuous semimartingale, define*

$$z_t^m[y] \tag{45}$$
$$\equiv \frac{1}{\sqrt{m}}\Phi_t^m[y]\left(\sqrt{m}z_0^m + \int_0^t (\Phi_s^m[y])^{-1}F(s,y_s)ds + \int_0^t (\Phi_s^m[y])^{-1}\sigma(s,y_s)dW_s\right),$$

where $\Phi_t^m[y]$ is defined pathwise as the solution to

$$\frac{d}{dt}\Phi_t^m = -\frac{1}{m}\tilde{\gamma}(t,y_t)\Phi_t^m, \quad \Phi_0^m = I. \tag{46}$$

Using this, now define

$$d(R_t^m[y])^i \equiv -d((\tilde{\gamma}^{-1})^{ij}(t,y_t)(z_t^m[y])_j) + (z_t^m[y])_j\partial_t(\tilde{\gamma}^{-1})^{ij}(t,y_t)dt \tag{47}$$
$$+ Q^{ikl}(t,y_t)G_{kl}^{ab}(t,y_t)\left((z_t^m[y])_aF_b(t,y_t) + (z_t^m[y])_bF_a(t,y_t)\right)dt$$
$$+ (z_t^m[y])_a(z_t^m[y])_b(z_t^m[y])^c\partial_{q^c}(Q^{ikl}G_{kl}^{ab})(t,y_t)dt$$
$$+ Q^{ikl}(t,y_t)G_{kl}^{ab}(t,y_t)\left((z_t^m[y])_a\sigma_{b\rho}(t,y_t) + (z_t^m[y])_b\sigma_{a\rho}(t,y_t)\right)dW_t^\rho$$
$$- \sqrt{m}d(Q^{ikl}(t,y_t)G_{kl}^{ab}(t,y_t)(z_t^m[y])_a(z_t^m[y])_b)$$
$$+ \sqrt{m}(z_t^m[y])_a(z_t^m[y])_b\partial_t(Q^{ikl}G_{kl}^{ab})(t,y_t)dt$$

with $R_0^m[y] = 0$.

For any such y, $\Phi_t^m[y]$ is a pathwise-C^1 semimartingale, and $z_t^m[y]$ and $R_t^m[y]$ are continuous semimartingales. In terms of these maps, the processes entering the delay equation for q_t^m, Eq. (39) (see also Eq. (44)), are given by $R_t^m = R_t^m[q^m]$ and $z_t^m = z_t^m[q^m]$, as the notation suggests.

We will denote by Y the space of continuous semimartingales with

$$E\left[\sup_{t\in[0,T]}\|y_t\|^p\right] < \infty \text{ for all } T > 0, p > 0. \tag{48}$$

We will now show that $\Phi_t^m[y]$, $z_t^m[y]$, and $R_t^m[y]$ satisfy several Lipschitz and boundedness properties for $y \in Y$.

Lemma 3.2. *Let* $y, \tilde{y} \in Y$. *Then for any* $T > 0$ *there exists* $L_T > 0$ *such that for all* $0 \leq s \leq t \leq T$ *we have the pathwise bound*

$$\|\Phi_t^m[y](\Phi_s^m[y])^{-1} - \Phi_t^m[\tilde{y}](\Phi_s^m[\tilde{y}])^{-1}\| \tag{49}$$

$$\leq \frac{L_T}{m} \int_s^t \|y_r - \tilde{y}_r\| dr e^{-\lambda(t-s)/m}.$$

Proof. For $0 \leq r \leq t - s$ define $\Phi_1(r) = \Phi_{r+s}^m[y](\Phi_s^m[y])^{-1}$ and $\Phi_2(r) = \Phi_{r+s}^m[\tilde{y}](\Phi_s^m[\tilde{y}])^{-1}$. Lemma B1 applies to the Φ_i, where $B_1(r) = -m^{-1}\tilde{\gamma}(r + s, y_{r+s})$, $B_2(r) = -m^{-1}\tilde{\gamma}(r+s, \tilde{y}_{r+s})$, and $\mu = -\lambda/m$, and gives the pathwise bound

$$\|\Phi_t^m[y](\Phi_s^m[y])^{-1} - \Phi_t^m[\tilde{y}](\Phi_s^m[\tilde{y}])^{-1}\| \tag{50}$$

$$\leq \frac{1}{m} \int_s^t \|\tilde{\gamma}(r, y_r) - \tilde{\gamma}(r, \tilde{y}_r)\| dr e^{-\lambda(t-s)/m}.$$

The claimed result then follows from the fact that $\tilde{\gamma}$ is C^1 with bounded derivative on $[0, T] \times \mathbb{R}^n$ (see the assumptions in Appendix A).   □

Next we show Lipschitz and boundedness properties for $z_t^m[y]$ and $R_t^m[y]$ under various norms. These are the key technical results in this work; once they are established, the claimed convergence rates in Eq. (14) follow from a rather straightforward application of Gronwall's inequality and a bootstrapping argument.

In the following, we will repeatedly use the technique of bounding a difference of products by rewriting it as

$$ab - \tilde{a}\tilde{b} = (a - \tilde{a})b + \tilde{a}(b - \tilde{b}), \tag{51}$$

$$abc - \tilde{a}\tilde{b}\tilde{c} = (a - \tilde{a})bc + \tilde{a}(b - \tilde{b})c + \tilde{a}\tilde{b}(c - \tilde{c}), \quad \text{etc},$$

and then employing bounds on each of the terms and their differences. Bounds will be obtained by using the properties in Appendix A along with repeated usage of the following inequalities: Hölder's inequality (H), Hölder's inequality for finite sums (HFS), Minkowski's inequality for integrals (MI), the L^p-triangle inequality (T), and the the Burkholder-Davis-Gundy inequality (BDG). These inequalities are commonly used in the literature, but we have restated them in Appendix C for convenience, along with textbook citations. We will use the above abbreviations to indicate where the various inequalities are used. When multiple inequalities are used to obtain a particular line, we list the inequalities in the order they were applied.

Lemma 3.3. *Let* $y, \tilde{y} \in Y$. *Then for any* $m_0 > 0$, $T > 0$, $q > p \geq 2$ *there exist* C, L *such that. for* $0 < m \leq m_0$, $0 \leq t \leq T$:

$$\sup_{s \in [0,t]} E\left[\|z_s^m[y] - z_s^m[\tilde{y}]\|^p\right]^{1/p} \leq L \sup_{s \in [0,t]} E\left[\|y_s - \tilde{y}_s\|^q\right]^{1/q} \tag{52}$$

and

$$\sup_{t \in [0,T]} E\left[\|z_t^m[y]\|^p\right] \leq C. \tag{53}$$

We emphasize that the C and L are independent of m.

Proof. Decomposing the stochastic convolution as in Eq. (140),

$$\Phi_t^m[y] \int_0^t (\Phi_s^m[y])^{-1}\sigma(s,y_s)dW_s \tag{54}$$

$$=\Phi_t^m[y] \int_0^t \sigma(s,y_s)dW_s + \frac{1}{m}\Phi_t^m[y] \int_0^t (\Phi_s^m[y])^{-1}\tilde{\gamma}(s,y_s) \int_s^t \sigma(r,y_r)dW_r ds,$$

we can write

$$z_t^m[y] \tag{55}$$

$$=\Phi_t^m[y]z_0^m + \frac{1}{\sqrt{m}} \int_0^t \Phi_t^m[y](\Phi_s^m[y])^{-1}F(s,y_s)ds + \frac{1}{\sqrt{m}}\Phi_t^m[y] \int_0^t \sigma(s,y_s)dW_s$$

$$+ \frac{1}{m^{3/2}} \int_0^t \Phi_t^m[y](\Phi_s^m[y])^{-1}\tilde{\gamma}(s,y_s) \int_s^t \sigma(r,y_r)dW_r ds.$$

Let $0 \le t \le T$. First, use the technique of Eq. (51) to bound the norm of the difference as follows:

$$\|z_t^m[y] - z_t^m[\tilde{y}]\| \le \|\Phi_t^m[y] - \Phi_t^m[\tilde{y}]\|\|z_0^m\| \tag{56}$$

$$+ \frac{1}{\sqrt{m}} \int_0^t \|\Phi_t^m[y](\Phi_s^m[y])^{-1} - \Phi_t^m[\tilde{y}](\Phi_s^m[\tilde{y}])^{-1}\|\|F(s,y_s)\|ds$$

$$+ \frac{1}{\sqrt{m}} \int_0^t \|\Phi_t^m[\tilde{y}](\Phi_s^m[\tilde{y}])^{-1}\|\|F(s,y_s) - F(s,\tilde{y}_s)\|ds$$

$$+ \frac{1}{\sqrt{m}}\|\Phi_t^m[y]\|\|\int_0^t \sigma(s,y_s) - \sigma(s,\tilde{y}_s)dW_s\|$$

$$+ \frac{1}{\sqrt{m}}\|\Phi_t^m[y] - \Phi_t^m[\tilde{y}]\|\|\int_0^t \sigma(s,\tilde{y}_s)dW_s\|$$

$$+ \frac{1}{m^{3/2}} \int_0^t \|\Phi_t^m[y](\Phi_s^m[y])^{-1} - \Phi_t^m[\tilde{y}](\Phi_s^m[\tilde{y}])^{-1}\|\|\tilde{\gamma}(s,y_s)\|\|\int_s^t \sigma(r,y_r)dW_r\|ds$$

$$+ \frac{1}{m^{3/2}} \int_0^t \|\Phi_t^m[\tilde{y}](\Phi_s^m[\tilde{y}])^{-1}\|\|\tilde{\gamma}(s,y_s)\|\|\int_s^t \sigma(r,y_r) - \sigma(r,\tilde{y}_r)dW_r\|ds$$

$$+ \frac{1}{m^{3/2}} \int_0^t \|\Phi_t^m[\tilde{y}](\Phi_s^m[\tilde{y}])^{-1}\|\|\tilde{\gamma}(s,y_s) - \tilde{\gamma}(s,\tilde{y}_s)\|\|\int_s^t \sigma(r,\tilde{y}_r)dW_r\|ds.$$

Let L denote a constant, independent of m, that may change from line to line. Using Eq. (49), Eq. (136), and the bounds and Lipschitz properties from Appendix A, we obtain

$$\|z_t^m[y] - z_t^m[\tilde{y}]\| \tag{57}$$

$$\le \frac{LL_T}{m} \int_0^t \|y_r - \tilde{y}_r\|dre^{-\lambda t/m} + \frac{L_T\|F\|_\infty}{m^{3/2}} \int_0^t \int_s^t \|y_r - \tilde{y}_r\|dre^{-\lambda(t-s)/m}ds$$

$$+ \frac{L}{\sqrt{m}} \int_0^t e^{-\lambda(t-s)/m} \|y_s - \tilde{y}_s\| ds + \frac{e^{-\lambda t/m}}{\sqrt{m}} \| \int_0^t \sigma(s, y_s) - \sigma(s, \tilde{y}_s) dW_s \|$$

$$+ \frac{L_T}{m^{3/2}} e^{-\lambda t/m} \int_0^t \|y_r - \tilde{y}_r\| \| \int_0^t \sigma(s, \tilde{y}_s) dW_s \| dr$$

$$+ \frac{L_T \|\tilde{\gamma}\|_\infty}{m^{5/2}} \int_0^t \int_s^t \|y_r - \tilde{y}_r\| dr e^{-\lambda(t-s)/m} \| \int_s^t \sigma(r, y_r) dW_r \| ds$$

$$+ \frac{\|\tilde{\gamma}\|_\infty}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m} \| \int_s^t \sigma(r, y_r) - \sigma(r, \tilde{y}_r) dW_r \| ds$$

$$+ \frac{L}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m} \|y_s - \tilde{y}_s\| \| \int_s^t \sigma(r, \tilde{y}_r) dW_r \| ds.$$

Here, $\|g\|_\infty \equiv \sup_{t \in [0,T], q \in \mathbb{R}^n} \|g(t, x)\|$.

Next we compute an L^p bound by using the inequalities from Appendix C (indicated by the abbreviations in parentheses; if more than one inequality is used, we list them in order of usage): Let $q > p \geq 2$ and define $\tilde{p} = pq/(q - p)$. First use the triangle inequality (T) to bound the L^p-norm of $\|z_t^m[y] - z_t^m[\tilde{y}]\|$ by the sum of L^p-norms of each of the terms on the right-hand-side in Eq. (57), and then bound each term as follows:

$$E[\|z_t^m[y] - z_t^m[\tilde{y}]\|^p]^{1/p} \tag{58}$$

$$\leq \frac{L L_T}{m} e^{-\lambda t/m} \int_0^t E[\|y_r - \tilde{y}_r\|^p]^{1/p} dr \tag{MI}$$

$$+ \frac{L_T \|F\|_\infty}{m^{3/2}} \int_0^t \int_s^t e^{-\lambda(t-s)/m} E[\|y_r - \tilde{y}_r\|^p]^{1/p} dr ds \tag{MI}$$

$$+ \frac{L}{\sqrt{m}} \int_0^t e^{-\lambda(t-s)/m} E[\|y_s - \tilde{y}_s\|^p]^{1/p} ds \tag{MI}$$

$$+ \frac{1}{\sqrt{m}} e^{-\lambda t/m} E\left[\| \int_0^t \sigma(s, y_s) - \sigma(s, \tilde{y}_s) dW_s \|^p\right]^{1/p}$$

$$+ \frac{L_T}{m^{3/2}} e^{-\lambda t/m} \int_0^t E[\|y_r - \tilde{y}_r\|^q]^{1/q} E\left[\| \int_0^t \sigma(s, \tilde{y}_s) dW_s \|^{\tilde{p}}\right]^{1/\tilde{p}} dr \tag{MI,H}$$

$$+ \frac{L_T \|\tilde{\gamma}\|_\infty}{m^{5/2}} \int_0^t \int_s^t e^{-\lambda(t-s)/m} E[\|y_r - \tilde{y}_r\|^q]^{1/q} \tag{MI,H}$$

$$\times E\left[\| \int_s^t \sigma(\tilde{r}, y_{\tilde{r}}) dW_{\tilde{r}} \|^{\tilde{p}}\right]^{1/\tilde{p}} dr ds$$

$$+ \frac{\|\tilde{\gamma}\|_\infty}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m} E\left[\| \int_s^t \sigma(r, y_r) - \sigma(r, \tilde{y}_r) dW_r \|^p\right]^{1/p} ds \tag{MI}$$

$$+ \frac{L}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m} E[\|y_s - \tilde{y}_s\|^q]^{1/q} E\left[\| \int_s^t \sigma(r, \tilde{y}_r) dW_r \|^{\tilde{p}}\right]^{1/\tilde{p}} ds. \tag{MI,H}$$

The uses of Hölder's inequality are all with the conjugate exponents q/p and $q/(q - p)$.

Bounding the stochastic integrals via the Burkholder-Davis-Gundy inequality, using the bounds and Lipschitz properties form Appendix A, and extracting the powers of m from the Lebesgue integrals gives

$$E[\|z_t^m[y] - z_t^m[\tilde{y}]\|^p]^{1/p} \leq L\frac{t}{m}e^{-\lambda t/m} \sup_{r \in [0,t]} E\left[\|y_r - \tilde{y}_r\|^p\right]^{1/p} \tag{59}$$

$$+ Lm^{1/2} \int_0^{t/m} ue^{-\lambda u}du \sup_{r \in [0,t]} E\left[\|y_r - \tilde{y}_r\|^p\right]^{1/p}$$

$$+ Lm^{1/2} \int_0^{t/m} e^{-\lambda u}du \sup_{s \in [0,t]} E\left[\|y_s - \tilde{y}_s\|^p\right]^{1/p}$$

$$+ \frac{1}{\sqrt{m}}e^{-\lambda t/m}E\left[\left(\int_0^t \|y_s - \tilde{y}_s\|^2 ds\right)^{p/2}\right]^{1/p} \tag{BDG}$$

$$+ L(t/m)^{3/2}e^{-\lambda t/m} \sup_{r \in [0,t]} E\left[\|y_r - \tilde{y}_r\|^q\right]^{1/q} \tag{BDG}$$

$$+ L\int_0^{t/m} u^{3/2}e^{-\lambda u}du \sup_{r \in [0,t]} E\left[\|y_r - \tilde{y}_r\|^q\right]^{1/q} \tag{BDG}$$

$$+ \frac{L}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m}E\left[\left(\int_s^t \|y_r - \tilde{y}_r\|^2 dr\right)^{p/2}\right]^{1/p} ds \tag{BDG}$$

$$+ L\int_0^{t/m} e^{-\lambda u}u^{1/2}du \sup_{s \in [0,t]} E\left[\|y_s - \tilde{y}_s\|^q\right]^{1/q} \tag{BDG}$$

$$\leq L \sup_{s \in [0,t]} E\left[\|y_s - \tilde{y}_s\|^q\right]^{1/q}. \tag{H}$$

We have used several times the fact that $\sup_{m>0} \sup_{t\geq 0} (t/m)^k e^{-\lambda t/m} < \infty$ for all $k \geq 0$. To obtain the last line, we used Hölder's inequality to bound all expectations by the L^q-norm; the condition $p \geq 2$ was needed to use Hölder's inequality on the term in the third-to-last line. Taking a supremum over t on the left hand side gives the claimed Lipschitz bound.

The bound on z_t^m proceeds using the same tools. First we bound

$$\|z_t^m[y]\| \tag{60}$$

$$\leq Ce^{-\lambda t/m} + \frac{\|F\|_\infty}{\sqrt{m}} \int_0^t e^{-\lambda(t-s)/m}ds + \frac{1}{\sqrt{m}}e^{-\lambda t/m}\|\int_0^t \sigma(s, y_s)dW_s\|$$

$$+ \frac{\|\tilde{\gamma}\|_\infty}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m}\|\int_s^t \sigma(r, y_r)dW_r\|ds.$$

Hence (letting C vary from line to line) for $p \geq 2$ we obtain

$$E[\|z_t^m[y]\|^p]^{1/p} \tag{61}$$

$$\leq C e^{-\lambda t/m} + \|F\|_\infty \sqrt{m} \int_0^{t/m} e^{-\lambda u} du \tag{T}$$

$$+ \frac{1}{\sqrt{m}} e^{-\lambda t/m} E\left[\|\int_0^t \sigma(s, y_s) dW_s\|^p\right]^{1/p} \tag{T}$$

$$+ \frac{\|\tilde{\gamma}\|_\infty}{m^{3/2}} E\left[\left(\int_0^t e^{-\lambda(t-s)/m}\|\int_s^t \sigma(r, y_r) dW_r\| ds\right)^p\right]^{1/p} \tag{T}$$

$$\leq C + \frac{C}{\sqrt{m}} e^{-\lambda t/m} E\left[\left(\int_0^t \|\sigma(s, y_s)\|^2 ds\right)^{p/2}\right]^{1/p} \tag{BDG}$$

$$+ \frac{C}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m} E\left[\left(\int_s^t \|\sigma(r, y_r)\|^2 dr\right)^{p/2}\right]^{1/p} ds. \tag{MI,BDG}$$

We have assumed σ is bounded, therefore

$$E[\|z_t^m[y]\|^p]^{1/p} \tag{62}$$

$$\leq C + C(t/m)^{1/2} e^{-\lambda t/m} + \frac{C}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m} (t-s)^{1/2} ds$$

$$\leq C + C(t/m)^{1/2} e^{-\lambda t/m} + C \int_0^{t/m} e^{-\lambda u} u^{1/2} du.$$

Taking the supremum over $t \in [0, T]$, we arrive at the claimed bound Eq. (53).

$\square$

Similarly to the previous lemma, but this time also employing Lemma B3, we have:

Lemma 3.4. *Let $y, \tilde{y} \in Y$. Then for any $m_0 > 0$, $T > 0$, $p \geq 2$, $\epsilon > 0$ there exist $q > p$, C, L such that for $0 < m \leq m_0$, $0 \leq t \leq T$:*

$$E\left[\sup_{s \in [0,t]} \|z_s^m[y] - z_s^m[\tilde{y}]\|^p\right]^{1/p} \leq \frac{L}{m^\epsilon} E\left[\sup_{s \in [0,t]} \|y_s - \tilde{y}_s\|^q\right]^{1/q} \tag{63}$$

and

$$E\left[\sup_{t \in [0,T]} \|z_t^m[y]\|^p\right]^{1/p} \leq C/m^\epsilon. \tag{64}$$

Again, C and L are independent of m. Also note that, in contrast with the previous result, q depends on ϵ and can't be chosen arbitrarily.

Proof. Let C be a constant that varies from line to line. To derive Eq. (64) we start from Eq. (60) and use the triangle inequality (T) to compute

$$E\left[\sup_{t\in[0,T]}\|z_t^m[y]\|^p\right]^{1/p}\leq C+\frac{1}{\sqrt{m}}E\left[\left(\sup_{t\in[0,T]}e^{-\lambda t/m}\|\int_0^t\sigma(s,y_s)dW_s\|\right)^p\right]^{1/p} \tag{65}$$

$$+\frac{C}{m^{3/2}}E\left[\left(\sup_{t\in[0,T]}\int_0^t e^{-\lambda(t-s)/m}\|\int_s^t\sigma(r,y_r)dW_r\|ds\right)^p\right]^{1/p}.$$

Here, and in the following, we will need to bound expected values of the following types:

$$E_{1,j}\equiv E\left[\left(\sup_{t\in[0,T]}t^j e^{-\lambda t/m}\|\int_0^t V_s dW_s\|\right)^p\right]^{1/p} \tag{66}$$

and

$$E_{2,j}\equiv E\left[\left(\sup_{t\in[0,T]}\int_0^t(t-s)^j e^{-\lambda(t-s)/m}\|\int_s^t V_r dW_r\|ds\right)^p\right]^{1/p}, \tag{67}$$

where $j\in\mathbb{Z}_0$ and V_s is a continuous adapted $\mathbb{R}^{n\times k}$-valued process.

First, for any $T>0$, $p\geq 2$, use integration by parts to write

$$E_{1,j}=E\left[\left(\sup_{t\in[0,T]}\|\int_0^t s^j e^{-\lambda s/m}V_s dW_s+\int_0^t\int_0^s V_r dW_r\partial_s(s^j e^{-\lambda s/m})ds\|\right)^p\right]^{1/p} \tag{68}$$

and then use the inequalities from Appendix C to obtain

$$E_{1,j}\leq E\left[\left(\sup_{t\in[0,T]}\|\int_0^t s^j e^{-\lambda s/m}V_s dW_s\|\right)^p\right]^{1/p} \tag{T}$$

$$+\int_0^T E\left[\|\int_0^s V_r dW_r\|^p\right]^{1/p}h_{j,m}(s)ds \tag{T,MI}$$

$$\leq C_{p,n}^{1/p}\left(\int_0^T E\left[\|V_s\|^p\right]^{2/p}s^{2j}e^{-2\lambda s/m}ds\right)^{1/2} \tag{BDG,MI}$$

$$+C_{p,n}^{1/p}\int_0^T\left(\int_0^s E[\|V_r\|^p]^{2/p}dr\right)^{1/2}h_{j,m}(s)ds \tag{BDG,MI}$$

$$\leq C_{1,p,n,j,\lambda}m^{j+1/2}\sup_{s\in[0,T]}E\left[\|V_s\|^p\right]^{1/p}, \tag{69}$$

$$h_{j,m}(s)\equiv js^{j-1}e^{-\lambda s/m}+s^j e^{-\lambda s/m}\lambda/m,$$

$$C_{1,p,n,j,\lambda}\equiv C_{p,n}^{1/p}\left(\left(\int_0^\infty u^{2j}e^{-2\lambda u}du\right)^{1/2}+\int_0^\infty u^{1/2}(ju^{j-1}e^{-\lambda u}+\lambda u^j e^{-\lambda u})du\right).$$

Here, $C_{p,n}$ denotes the constant from (the vector version of) the Burkholder-Davis-Gundy inequality. Note that $p \geq 2$ was needed for the second and third usage of Minkowski's inequality for integrals.

Eq. (67) can be bounded similarly, but this time we also need to employ the bound from Lemma B3. Using this, and then the triangle inequality (T), for any $T > 0$, $\delta > 0$, $p \geq 2$ we have

$$E_{2,j} \leq \frac{C_j m^{j+1}}{\lambda^{j+1}} \left(E\left[\max_{\ell=1,\dots,N} \sup_{\tau \in [t_{\ell-1}, T_{\ell+1}]} \| \int_{t_{\ell-1}}^{\tau} V_r dW_r \|^p \right]^{1/p} \right. \tag{70}$$
$$\left. + e^{-\lambda \delta/(2m)} E\left[\sup_{\tau \in [0,T]} \| \int_0^{\tau} V_r dW_r \|^p \right]^{1/p} \right),$$

where $N \equiv \min\{\ell \in \mathbb{Z} : \ell\delta \geq T\}$, $t_{\ell-1} \equiv (\ell-1)\delta$, $T_{\ell+1} \equiv \min\{(\ell+1)\delta, T\}$, and C_j depends only on j.

For any $\tilde{q} \geq 1$ we can bound the maximum of an N-term sequence by its $\ell^{\tilde{q}}$ norm. This, along with the inequalities from Appendix C, yields

$$E_{2,j} \tag{71}$$

$$\leq \frac{C_j m^{j+1}}{\lambda^{j+1}} \left(E\left[\left(\sum_{\ell=1}^{N} \sup_{\tau \in [t_{\ell-1}, T_{\ell+1}]} \| \int_{t_{\ell-1}}^{\tau} V_r dW_r \|^{p\tilde{q}} \right)^{1/\tilde{q}} \right]^{1/p} \right.$$

$$\left. + C_{p,n}^{1/p} e^{-\lambda\delta/(2m)} E\left[\left(\int_0^T \|V_r\|^2 dr \right)^{p/2} \right]^{1/p} \right) \tag{BDG}$$

$$\leq \frac{C_j m^{j+1}}{\lambda^{j+1}} \left(\left(\sum_{\ell=1}^{N} E\left[\sup_{\tau \in [t_{\ell-1}, T_{\ell+1}]} \| \int_{t_{\ell-1}}^{\tau} V_r dW_r \|^{p\tilde{q}} \right] \right)^{1/(p\tilde{q})} \right. \tag{H}$$

$$\left. + C_{p,n}^{1/p} e^{-\lambda\delta/(2m)} \left(\int_0^T E\left[\|V_r\|^p \right]^{2/p} dr \right)^{1/2} \right) \tag{MI}$$

$$\leq \frac{C_j m^{j+1}}{\lambda^{j+1}} \left(C_{p\tilde{q},n}^{1/(p\tilde{q})} \left(\sum_{\ell=1}^{N} E\left[\left(\int_{t_{\ell-1}}^{T_{\ell+1}} \|V_r\|^2 dr \right)^{p\tilde{q}/2} \right] \right)^{1/(p\tilde{q})} \right. \tag{BDG}$$

$$\left. + T^{1/2} C_{p,n}^{1/p} e^{-\lambda\delta/(2m)} \sup_{r \in [0,T]} E\left[\|V_r\|^p \right]^{1/p} \right)$$

$$\leq \frac{C_j m^{j+1}}{\lambda^{j+1}} \left(C_{p\tilde{q},n}^{1/(p\tilde{q})} \left((1 + T/\delta)(2\delta)^{p\tilde{q}/2} \right)^{1/(p\tilde{q})} + T^{1/2} C_{p,n}^{1/p} e^{-\lambda\delta/(2m)} \right) \tag{MI,H}$$

$$\times \sup_{r \in [0,T]} E\left[\|V_r\|^{p\tilde{q}} \right]^{1/(p\tilde{q})}.$$

In obtaining the last inequality, we used $N \leq 1 + T/\delta$.

For the current purposes, it is useful to let $\delta = m^\kappa$ for $\kappa \in (0,1)$. Hence we have shown there is a constant $C_{T,m_0,p,\kappa,\tilde{q},n,\lambda,j}$, with the indicated dependencies, such that

$$
E_{2,j} \leq \frac{C_j m^{j+1}}{\lambda^{j+1}} \left(2^{1/2} C_{p\tilde{q},n}^{1/(p\tilde{q})} (m^\kappa + T)^{1/(p\tilde{q})} m^{\kappa(1/2 - 1/(p\tilde{q}))} \right.
$$
$$
\left. + T^{1/2} C_{p,n}^{1/p} e^{-\lambda/(2m^{1-\kappa})} \right) \sup_{r \in [0,T]} E\left[\|V_r\|^{p\tilde{q}} \right]^{1/(p\tilde{q})}
$$
$$
\leq C_{T,m_0,p,\kappa,\tilde{q},n,\lambda,j} m^{j+1+\kappa(1/2 - 1/(p\tilde{q}))} \sup_{r \in [0,T]} E\left[\|V_r\|^{p\tilde{q}} \right]^{1/(p\tilde{q})}. \qquad (72)
$$

We note that the constant can be chosen to be increasing in T.

Applying the bounds on $E_{1,j}$ and $E_{2,j}$ to Eq. (65) implies that, for any $\tilde{q} \geq 1$, $\kappa \in (0,1)$, we have

$$
E\left[\sup_{t \in [0,T]} \|z_t^m[y]\|^p \right]^{1/p} \leq C + C_{1,p,n,0,\lambda} \sup_{s \in [0,T]} E[\|\sigma(s,y_s)\|^p]^{1/p} \qquad (73)
$$
$$
+ C C_{T,m_0,p,\kappa,\tilde{q},n,\lambda,0} m^{-(1/2 - \kappa(1/2 - 1/(p\tilde{q})))} \sup_{r \in [0,T]} E\left[\|\sigma(r,y_r)\|^{p\tilde{q}} \right]^{1/(p\tilde{q})}.
$$

σ is bounded, so for any $\epsilon > 0$ we can fix κ and $\tilde{q}$ and find a C so that

$$
E\left[\sup_{t \in [0,T]} \|z_t^m[y]\|^p \right]^{1/p} \leq C/m^\epsilon \qquad (74)
$$

as claimed.

For the Lipschitz bound we go back to Eq. (57) and compute

$$
\|z_t^m[y] - z_t^m[\tilde{y}]\| \qquad (75)
$$
$$
\leq \left(\frac{LL_T}{m} t e^{-\lambda t/m} + \frac{L_T \|F\|_\infty}{m^{3/2}} \int_0^t (t-s) e^{-\lambda(t-s)/m} ds + \frac{L}{\sqrt{m}} \int_0^t e^{-\lambda(t-s)/m} ds \right.
$$
$$
+ \frac{L_T}{m^{3/2}} t e^{-\lambda t/m} \| \int_0^t \sigma(s,\tilde{y}_s) dW_s \|
$$
$$
+ \frac{L_T \|\tilde{\gamma}\|_\infty}{m^{5/2}} \int_0^t (t-s) e^{-\lambda(t-s)/m} \| \int_s^t \sigma(r,y_r) dW_r \| ds
$$
$$
\left. + \frac{L}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m} \| \int_s^t \sigma(r,\tilde{y}_r) dW_r \| ds \right) \sup_{r \in [0,t]} \|y_r - \tilde{y}_r\|
$$
$$
+ \frac{\|\tilde{\gamma}\|_\infty}{m^{3/2}} \int_0^t e^{-\lambda(t-s)/m} \| \int_s^t \sigma(r,y_r) - \sigma(r,\tilde{y}_r) dW_r \| ds
$$
$$
+ \frac{1}{\sqrt{m}} e^{-\lambda t/m} \| \int_0^t \sigma(s,y_s) - \sigma(s,\tilde{y}_s) dW_s \|.
$$

Let $\tilde{t} \leq T$, take the supremum $[0, \tilde{t}]$ and use the triangle inequality (T) to bound the L^p-norm of the sum. Next, let $q > p$, define $\tilde{p} = pq/(q-p)$, and

use Hölder's inequality (H) with exponents q/p and $q/(q-p)$ followed by the triangle inequality (T) with exponent $\tilde{p}$ (again, letting L vary line to line):

$$E\left[\sup_{t\in[0,\tilde{t}]}\|z_t^m[y]-z_t^m[\tilde{y}]\|^p\right]^{1/p} \tag{76}$$

$$\leq\left(L+\frac{L}{m^{3/2}}E\left[\left(\sup_{t\in[0,\tilde{t}]}te^{-\lambda t/m}\|\int_0^t\sigma(s,\tilde{y}_s)dW_s\|\right)^{\tilde{p}}\right]^{1/\tilde{p}}\right.\tag{H,T}$$

$$+\frac{L}{m^{5/2}}E\left[\left(\sup_{t\in[0,\tilde{t}]}\int_0^t(t-s)e^{-\lambda(t-s)/m}\|\int_s^t\sigma(r,y_r)dW_r\|ds\right)^{\tilde{p}}\right]^{1/\tilde{p}}\tag{H,T}$$

$$+\frac{L}{m^{3/2}}E\left[\left(\sup_{t\in[0,\tilde{t}]}\int_0^t e^{-\lambda(t-s)/m}\|\int_s^t\sigma(r,\tilde{y}_r)dW_r\|ds\right)^{\tilde{p}}\right]^{1/\tilde{p}}\right)\tag{H,T}$$

$$\times E\left[\sup_{r\in[0,\tilde{t}]}\|y_r-\tilde{y}_r\|^q\right]^{1/q}\tag{H}$$

$$+\frac{L}{m^{3/2}}E\left[\left(\sup_{t\in[0,\tilde{t}]}\int_0^t e^{-\lambda(t-s)/m}\|\int_s^t\sigma(r,y_r)-\sigma(r,\tilde{y}_r)dW_r\|ds\right)^p\right]^{1/p}$$

$$+\frac{1}{\sqrt{m}}E\left[\left(\sup_{t\in[0,\tilde{t}]}e^{-\lambda t/m}\|\int_0^t\sigma(s,y_s)-\sigma(s,\tilde{y}_s)dW_s\|\right)^p\right]^{1/p}.$$

Each of the expected values involving σ is of the form Eq. (66) or Eq. (67). Applying the corresponding bounds, Eq. (69) and Eq. (72), to the first term results in

$$\frac{1}{m^{3/2}}E\left[\left(\sup_{t\in[0,\tilde{t}]}te^{-\lambda t/m}\|\int_0^t\sigma(s,\tilde{y}_s)dW_s\|\right)^{\tilde{p}}\right]^{1/\tilde{p}}\tag{77}$$

$$+\frac{1}{m^{5/2}}E\left[\left(\sup_{t\in[0,\tilde{t}]}\int_0^t(t-s)e^{-\lambda(t-s)/m}\|\int_s^t\sigma(r,y_r)dW_r\|ds\right)^{\tilde{p}}\right]^{1/\tilde{p}}$$

$$+\frac{1}{m^{3/2}}E\left[\left(\sup_{t\in[0,\tilde{t}]}\int_0^t e^{-\lambda(t-s)/m}\|\int_s^t\sigma(r,\tilde{y}_r)dW_r\|ds\right)^{\tilde{p}}\right]^{1/\tilde{p}}$$

$$\leq\|\sigma\|_\infty C_{1,pq/(q-p),n,1,\lambda}\tag{78}$$

$$+\|\sigma\|_\infty C_{T,m_0,\tilde{p},\kappa,\tilde{q},n,\lambda,1}m^{-5/2+2+\kappa(1/2-(q-p)/(\tilde{q}pq))}$$

$$+\|\sigma\|_\infty C_{T,m_0,\tilde{p},\kappa,\tilde{q},n,\lambda,0}m^{-3/2+1+\kappa(1/2-(q-p)/(\tilde{q}pq))}$$

for any $\tilde{q}\geq 1$, $\kappa\in(0,1)$.

Therefore, given $\epsilon > 0$ we can choose $\tilde{q}$, κ

$$E\left[\sup_{t\in[0,\tilde{t}]} \|z_t^m[y] - z_t^m[\tilde{y}]\|^p\right]^{1/p} \le \frac{L}{m^\epsilon} E\left[\sup_{r\in[0,\tilde{t}]} \|y_r - \tilde{y}_r\|^q\right]^{1/q} \tag{79}$$

$$+ \frac{L}{m^{3/2}} E\left[\left(\sup_{t\in[0,\tilde{t}]} \int_0^t e^{-\lambda(t-s)/m}\| \int_s^t \sigma(r, y_r) - \sigma(r, \tilde{y}_r) dW_r\| ds\right)^p\right]^{1/p}$$

$$+ \frac{1}{\sqrt{m}} E\left[\left(\sup_{t\in[0,\tilde{t}]} e^{-\lambda t/m}\| \int_0^t \sigma(s, y_s) - \sigma(s, \tilde{y}_s) dW_s\|\right)^p\right]^{1/p}$$

for some $L > 0$.

We can similarly use Eq. (69) and Eq. (72) to bound the last two terms:

$$E\left[\sup_{t\in[0,\tilde{t}]} \|z_t^m[y] - z_t^m[\tilde{y}]\|^p\right]^{1/p} \le \frac{L}{m^\epsilon} E\left[\sup_{r\in[0,\tilde{t}]} \|y_r - \tilde{y}_r\|^q\right]^{1/q} \tag{80}$$

$$+ LC_{T,m_0,p,\kappa,\tilde{q},n,\lambda,0} m^{-3/2+1+\kappa(1/2-1/(p\tilde{q}))}$$

$$\times \sup_{r\in[0,\tilde{t}]} E\left[\|\sigma(r, y_r) - \sigma(r, \tilde{y}_r)\|^{p\tilde{q}}\right]^{1/(p\tilde{q})}$$

$$+ C_{1,p,n,0,\lambda} \sup_{s\in[0,\tilde{t}]} E\left[\|\sigma(s, y_s) - \sigma(s, \tilde{y}_s)\|^p\right]^{1/p}.$$

Using the assumption that σ is Lipschitz in its second argument, for appropriate choices of $\kappa \in (0,1)$ and $\tilde{q} \ge 1$ we finally obtain

$$E\left[\sup_{t\in[0,\tilde{t}]} \|z_t^m[y] - z_t^m[\tilde{y}]\|^p\right]^{1/p} \le \frac{L}{m^\epsilon} E\left[\sup_{s\in[0,\tilde{t}]} \|y_s - \tilde{y}_s\|^{p\tilde{q}}\right]^{1/(p\tilde{q})}. \tag{81}$$

This completes the proof. □

Next we prove analogous results for $R_t^m[y]$.

Lemma 3.5. *Let $y, \tilde{y} \in Y$. Then for any $m_0 > 0$, $T > 0$, $q > p \ge 2$, there exist C, L such that for $0 < m \le m_0$, $0 \le t \le T$:*

$$\sup_{s\in[0,t]} E\left[\|R_s^m[y] - R_s^m[\tilde{y}]\|^p\right]^{1/p} \le L \sup_{s\in[0,t]} E\left[\|y_s - \tilde{y}_s\|^q\right]^{1/q}. \tag{82}$$

and

$$\sup_{t\in[0,T]} E\left[\|R_t^m[y]\|^p\right] \le C. \tag{83}$$

Once again, C and L are independent of m.

Proof. We group the terms in $R_t^m[y]$ of similar form as follows:

$$(R_t^m[y])^i \tag{84}$$

$$= -\left((\tilde{\gamma}^{-1})^{ij}(t, y_t)(z_t^m[y])_j - (\tilde{\gamma}^{-1})^{ij}(0, y_0)(z_0^m[y])_j\right)$$

$$-\sqrt{m}\left((Q^{ikl}G_{kl}^{ab})(t, y_t)(z_t^m[y])_a(z_t^m[y])_b - (Q^{ikl}G_{kl}^{ab})(0, y_0)(z_0^m[y])_a(z_0^m[y])_b\right)$$

$$+\int_0^t (z_s^m[y])_j \left(\partial_s(\tilde{\gamma}^{-1})^{ij} + Q^{ikl}G_{kl}^{jb}F_b + Q^{ikl}G_{kl}^{aj}F_a\right)(s, y_s)ds$$

$$+\int_0^t (z_s^m[y])_a(z_s^m[y])_b(z_s^m[y])_c\partial_{q^c}(Q^{ikl}G_{kl}^{ab})(s, y_s)ds$$

$$+\sqrt{m}\int_0^t (z_s^m[y])_a(z_s^m[y])_b\partial_s(Q^{ikl}G_{kl}^{ab})(s, y_s)dt$$

$$+\int_0^t (z_s^m[y])_j \left(Q^{ikl}G_{kl}^{jb}\sigma_{b\rho} + Q^{ikl}G_{kl}^{aj}\sigma_{a\rho}\right)(s, y_s)dW_s^\rho.$$

We will show that each of these terms satisfies the claimed Lipschitz property. The computations are all similar, and use the same tools as the previous lemmas, so we illustrate the main ideas while omitting some details:

We will again make repeated use of the Eq. (51). The estimates of the first two terms are similar, and the derivations rely on the fact that $\tilde{\gamma}^{-1}$, $Q^{ikl}(t, q)$, and $G_{kl}^{ab}(t, q)$ are bounded and Lipschitz in q, uniformly in $t \in [0, T]$, and $z_0^m[y] = z_0^m$ is independent of y and uniformly bounded (these properties follow from the assumptions in Appendix A; boundedness of G_{kl}^{ab} follows from the lower bound on the eigenvalues of γ). We illustrate the computation with the first sub-term of the second term. Again, let L denote a constant that may vary from line to line and use the triangle inequality (T) to bound the L^p-norm as follows:

$$\sup_{t\in[0,\tilde{t}]} E\left[\|\sqrt{m}(Q^{ikl}G_{kl}^{ab})(t, y_t)(z_t^m[y])_a(z_t^m[y])_b\right. \tag{85}$$

$$\left. -\sqrt{m}(Q^{ikl}G_{kl}^{ab})(t, \tilde{y}_t)(z_t^m[\tilde{y}])_a(z_t^m[\tilde{y}])_b\|^p\right]^{1/p}$$

$$\leq \sqrt{m} \sup_{t\in[0,\tilde{t}]} E\left[\|(Q^{ikl}G_{kl}^{ab})(t, y_t)((z_t^m[y])_a(z_t^m[y])_b - (z_t^m[\tilde{y}])_a(z_t^m[\tilde{y}])_b)\|^p\right]^{1/p}$$

$$+\sqrt{m} \sup_{t\in[0,\tilde{t}]} E\left[\|((Q^{ikl}G_{kl}^{ab})(t, y_t) - (Q^{ikl}G_{kl}^{ab})(t, \tilde{y}_t))(z_t^m[\tilde{y}])_a(z_t^m[\tilde{y}])_b\|^p\right]^{1/p}$$

$$\leq L\left(\sup_{t\in[0,\tilde{t}]} E\left[\|z_t^m[y]\|^p\|z_t^m[y] - z_t^m[\tilde{y}]\|^p\right]^{1/p}\right.$$

$$\left. +\sup_{t\in[0,\tilde{t}]} E\left[\|z_t^m[y] - z_t^m[\tilde{y}]\|^p\|z_t^m[\tilde{y}]\|^p\right]^{1/p} + \sup_{t\in[0,\tilde{t}]} E\left[\|y_t - \tilde{y}_t\|^p\|z_t^m[\tilde{y}]\|^{2p}\right]^{1/p}\right).$$

Here and in the following, $\|x^i\|$ denotes the ℓ^2 norm of the vector with components x^i. The other, paired, indices still indicate summations.

Let $p < \tilde{q} < q$. Using Hölder's inequality (H) gives

$$\sup_{t\in[0,\tilde{t}]} E\left[\|\sqrt{m}(Q^{ikl}G^{ab}_{kl})(t, y_t)(z^m_t[y])_a(z^m_t[y])_b\right. \tag{86}$$

$$\left.-\sqrt{m}(Q^{ikl}G^{ab}_{kl})(t, \tilde{y}_t)(z^m_t[\tilde{y}])_a(z^m_t[\tilde{y}])_b\|^p\right]^{1/p}$$

$$\leq L\left(\sup_{t\in[0,\tilde{t}]} E\left[\|z^m_t[y]\|^{p\tilde{q}/(\tilde{q}-p)}\right]^{(\tilde{q}-p)/(p\tilde{q})} \sup_{t\in[0,\tilde{t}]} E\left[\|z^m_t[y] - z^m_t[\tilde{y}]\|^{\tilde{q}}\right]^{1/\tilde{q}} \quad \text{(H)}$$

$$+ \sup_{t\in[0,\tilde{t}]} E\left[\|z^m_t[\tilde{y}]\|^{p\tilde{q}/(\tilde{q}-p)}\right]^{(\tilde{q}-p)/(p\tilde{q})} \sup_{t\in[0,\tilde{t}]} E\left[\|z^m_t[y] - z^m_t[\tilde{y}]\|^{\tilde{q}}\right]^{1/\tilde{q}} \quad \text{(H)}$$

$$+ \sup_{t\in[0,\tilde{t}]} E\left[\|y_t - \tilde{y}_t\|^q\right]^{1/q} \sup_{t\in[0,\tilde{t}]} E\left[\|z^m_t[\tilde{y}]\|^{2pq/(q-p)}\right]^{(q-p)/(qp)}\Bigg) \quad \text{(H)}$$

$$\leq L \sup_{t\in[0,\tilde{t}]} E\left[\|y_t - \tilde{y}_t\|^q\right]^{1/q}.$$

To obtain the last line, we used Lemma 3.3.

The third, fourth and fifth terms in Eq. (84) are bounded similarly, using also the facts that F and $\partial_s(\tilde{\gamma}^{-1})^{ij}$, $\partial_{q^c}(Q^{ikl}G^{ab}_{kl})$, $\partial_s(Q^{ikl}G^{ab}_{kl})$ are bounded and Lipschitz in q (as are implied by the assumptions in Appendix A). For example, defining

$$H^j(t,q) \equiv \left(\partial_s(\tilde{\gamma}^{-1})^{ij} + Q^{ikl}G^{jb}_{kl}F_b + Q^{ikl}G^{aj}_{kl}F_a\right)(t,q), \tag{87}$$

the estimate of the third term is derived by first using Minkowski's inequality for integrals (MI) to write

$$\sup_{t\in[0,\tilde{t}]} E\left[\|\int_0^t (z^m_s[y])_j H^j(s, y_s)ds - \int_0^t (z^m_s[\tilde{y}])_j H^j(s, \tilde{y}_s)ds\|^p\right]^{1/p} \tag{88}$$

$$\leq \int_0^{\tilde{t}} E\left[\|(z^m_s[y])_j H^j(s, y_s) - (z^m_s[\tilde{y}])_j H^j(s, \tilde{y}_s)\|^p\right]^{1/p} ds.$$

The rest of the derivation mimics that of the first two terms.

Finally, for the last term define

$$L^j(t,q) = \left(Q^{ikl}G^{jb}_{kl}\sigma_{b\rho} + Q^{ikl}G^{aj}_{kl}\sigma_{a\rho}\right)(t,q). \tag{89}$$

This is bounded and Lipschitz, and the Burkholder-Davis-Gundy inequality (BDG) followed by Minkowski's inequality for integrals (MI) give

$$\sup_{t\in[0,\tilde{t}]} E\left[\|\int_0^t (z^m_s[y])_j L^j(s, y_s)dW^\rho_s - \int_0^t (z^m_s[\tilde{y}])_j L^j(s, \tilde{y}_s)dW^\rho_s\|^p\right]^{1/p}$$
$$\tag{90}$$

$$\leq L\left(\int_0^{\tilde{t}} E\left[\|(z^m_s[y])_j L^j(s, y_s) - (z^m_s[\tilde{y}])_j L^j(s, \tilde{y}_s)\|^p\right]^{2/p} ds\right)^{1/2}.$$

The rest of the proof proceeds similarly to the previous cases.

This completes the proof of the Lipschitz property. The proof of the inequality

$$\sup_{t\in[0,T]} E\left[\|R_t^m[y]\|^p\right] \leq C \tag{91}$$

with C independent of m follows from Eq. (53) using similar techniques. We omit the details.   □

We end this section with Lipschitz and boundedness properties for R_t^m, corresponding to the norms from Lemma 3.4.

Lemma 3.6. *Let $y, \tilde{y} \in Y$. Then for any $m_0 > 0$, $T > 0$, $p \geq 2$, $\epsilon > 0$ there exist $q > p$, C, L such that for $0 < m \leq m_0$, $0 \leq t \leq T$:*

$$E\left[\sup_{s\in[0,t]} \|R_s^m[y] - R_s^m[\tilde{y}]\|^p\right]^{1/p} \leq \frac{L}{m^\epsilon} E\left[\sup_{s\in[0,t]} \|y_s - \tilde{y}_s\|^q\right]^{1/q} \tag{92}$$

and

$$E\left[\sup_{t\in[0,T]} \|R_t^m[y]\|^p\right]^{1/p} \leq C/m^\epsilon. \tag{93}$$

C and L are independent of m.

Proof. The proof is very similar to that of Lemma 3.5, with the bounds and Lipschitz constants for $z_t^m[y]$ coming from Lemma 3.4. We omit the details.   □

3.3. Hierarchy of Approximations and the Convergence Proof

With the notation of the previous subsection, q_t^m solves the delay equation

$$dq_t^m = \tilde{\gamma}^{-1}(t, q_t^m)F(t, q_t^m)dt + S(t, q_t^m)dt \tag{94}$$
$$+ \tilde{\gamma}^{-1}(t, q_t^m)\sigma(t, q_t^m)dW_t + \sqrt{m}dR_t^m[q^m].$$

We use this form of the equation to motivate the definition of a hierarchy of approximating processes, $q_t^{\ell,m}$, and prove the claimed convergence result, Eq. (14). We first recall the following definitions for convenience:

Definition 3.7. *For y a continuous martingale, define*

$$d(R_t^m[y])^i \equiv -d((\tilde{\gamma}^{-1})^{ij}(t, y_t)(z_t^m[y])_j) + (z_t^m[y])_j\partial_t(\tilde{\gamma}^{-1})^{ij}(t, y_t)dt \tag{95}$$
$$+ Q^{ikl}(t, y_t)G_{kl}^{ab}(t, y_t)\left((z_t^m[y])_a F_b(t, y_t) + (z_t^m[y])_b F_a(t, y_t)\right)dt$$
$$+ (z_t^m[y])_a(z_t^m[y])_b(z_t^m[y])^c\partial_{q^c}(Q^{ikl}G_{kl}^{ab})(t, y_t)dt$$
$$+ Q^{ikl}(t, y_t)G_{kl}^{ab}(t, y_t)\left((z_t^m[y])_a\sigma_{b\rho}(t, y_t) + (z_t^m[y])_b\sigma_{a\rho}(t, y_t)\right)dW_t^\rho$$
$$- \sqrt{m}d(Q^{ikl}(t, y_t)G_{kl}^{ab}(t, y_t)(z_t^m[y])_a(z_t^m[y])_b)$$
$$+ \sqrt{m}(z_t^m[y])_a(z_t^m[y])_b\partial_t(Q^{ikl}G_{kl}^{ab})(t, y_t)dt,$$

$R_0^m[y] \equiv 0$, where

$$z_t^m[y] \equiv \Phi_t^m[y]z_0^m + \frac{1}{\sqrt{m}}\Phi_t^m[y]\int_0^t (\Phi_s^m[y])^{-1}F(s,y_s)ds \qquad (96)$$

$$+ \frac{1}{\sqrt{m}}\Phi_t^m[y]\int_0^t (\Phi_s^m[y])^{-1}\sigma(s,y_s)dW_s,$$

$\Phi_t^m[y]$ is defined pathwise as the solution to

$$\frac{d}{dt}\Phi_t^m = -\frac{1}{m}\tilde{\gamma}(t,y_t)\Phi_t^m, \quad \Phi_0^m = I, \qquad (97)$$

and

1. $\tilde{\gamma}_{ik}(t,q) \equiv \gamma_{ik}(t,q) + \partial_{q^k}\psi_i(t,q) - \partial_{q^i}\psi_k(t,q)$,
2. $Q^{ijl}(t,q) \equiv \partial_{q^k}(\tilde{\gamma}^{-1})^{ij}(t,q)\delta^{kl}$,
3. $S^i(t,q) \equiv \partial_{q^k}(\tilde{\gamma}^{-1})^{ij}(t,q)\delta^{kl}G_{jl}^{rs}(t,q)\Sigma_{rs}(t,q)$,
4. $G_{ij}^{kl}(t,q) \equiv \delta^{rk}\delta^{sl}\int_0^\infty (e^{-\zeta\tilde{\gamma}(t,q)})_{ir}(e^{-\zeta\tilde{\gamma}(t,q)})_{js}d\zeta$,
5. $\Sigma_{ij} \equiv \sigma_{i\rho}\sigma_{j\xi}\delta^{\rho\xi}$,
6. $F(t,q) = -\partial_t\psi(t,q) - \nabla_q V(t,q) + \tilde{F}(t,q)$.

Theorem 3.8. *Assume the conditions in Appendix A hold. Fix an initial condition q_0 such that $E[\|q_0\|^p] < \infty$ for all $p > 0$ and let q_t^m, q_t be the solutions to the original SDE (Eq. (31)) and the homogenized SDE (Eq. (10)) respectively, all with the same initial position, q_0.*

With the notation from Definition 3.7, define the continuous semimartingales $q_t^{\ell,m}$, $\ell \in \mathbb{Z}^+$, by setting $q_t^{1,m} \equiv q_t$ and, for $\ell > 1$, inductively defining $q_t^{\ell,m}$ to be the solution to

$$q_t^{\ell,m} = q_0 + \int_0^t \tilde{\gamma}^{-1}(s,q_s^{\ell,m})F(s,q_s^{\ell,m})ds + \int_0^t S(s,q_s^{\ell,m})ds \qquad (98)$$

$$+ \int_0^t \tilde{\gamma}^{-1}(s,q_s^{\ell,m})\sigma(s,q_s^{\ell,m})dW_s + \sqrt{m}R_t^m[q^{\ell-1,m}].$$

(Note that this also holds for $\ell = 1$ if one defines $R_t^m[q^{0,m}] \equiv 0$.)

Then for any $T > 0$, $p > 0$, $\epsilon > 0$, $\ell \in \mathbb{Z}^+$ we have

$$\sup_{t\in[0,T]} E[\|q_t^m - q_t^{\ell,m}\|^p]^{1/p} = O(m^{\ell/2}), \qquad (99)$$

$$E\left[\sup_{t\in[0,T]} \|q_t^m - q_t^{\ell,m}\|^p\right]^{1/p} = O(m^{\ell/2-\epsilon}).$$

Proof. Given $y \in Y$ (recall Eq. (48)), Theorem D1 and Corollary D2 give a unique solution to

$$x_t = q_0 + \int_0^t \tilde{\gamma}^{-1}(s,x_s)F(s,x_s)ds + \int_0^t S(s,x_s)ds \qquad (100)$$

$$+ \int_0^t \tilde{\gamma}^{-1}(s,x_s)\sigma(s,x_s)dW_s + \sqrt{m}R_t^m[y]$$

defined for all $t \geq 0$. The solution is a continuous semimartingale and bound-edness of the drift and diffusion, together with Eq. (93), imply

$$E\left[\sup_{t\in[0,T]} \|x_t\|^p\right] < \infty \text{ for all } T > 0,\, p > 0, \tag{101}$$

i.e., $x \in Y$ as well.

Recalling Eq. (11), we see that $q_t^{1,m} = q_t \in Y$. Therefore, the inductive definition of the $q_t^{\ell,m}$ results in a sequence of processes in Y.

To prove Eq. (99) by induction, first note that the $\ell = 0$ result is the content of Eq. (8), proven in [9]. Supposing that Eq. (99) holds for $\ell \in \mathbb{Z}^+$ and any $T > 0$, $p > 0$, $\epsilon > 0$, we now show it holds for $\ell + 1$:

Eq. (11) implies $q_t^m \in Y$ for all $m > 0$ as well. From Eq. (94) and Eq. (98), we have

$$q_t^m - q_t^{\ell+1,m} = \int_0^t (\tilde{\gamma}^{-1}F + S)(s, q_s^m) - (\tilde{\gamma}^{-1}F + S)(s, q_s^{\ell+1,m})ds \tag{102}$$

$$+ \int_0^t (\tilde{\gamma}^{-1}\sigma)(s, q_s^m) - (\tilde{\gamma}^{-1}\sigma)(s, q_s^{\ell+1,m})dW_s + \sqrt{m}\left(R_t^m[q^m] - R_t^m[q^{\ell,m}]\right).$$

Note that $(\tilde{\gamma}^{-1}F + S)(t, q)$ is Lipschitz in q, uniformly in $t \in [0, T]$. Let L denote a constant that may vary from line to line and $0 \leq \tilde{t} \leq T$, $q > p \geq 2$. The inequalities in Appendix C give

$$\sup_{t\in[0,\tilde{t}]} E[\|q_t^m - q_t^{\ell+1,m}\|^p] \tag{103}$$

$$\leq 3^{p-1}\left(L^p \sup_{t\in[0,\tilde{t}]} E\left[\left(\int_0^t \|q_s^m - q_s^{\ell+1,m}\|ds\right)^p\right]\right. \tag{HFS}$$

$$+ \sup_{t\in[0,\tilde{t}]} E\left[\|\int_0^t (\tilde{\gamma}^{-1}\sigma)(s, q_s^m) - (\tilde{\gamma}^{-1}\sigma)(s, q_s^{\ell+1,m})dW_s\|^p\right] \tag{HFS}$$

$$\left. + m^{p/2} \sup_{t\in[0,\tilde{t}]} E\left[\|R_t^m[q^m] - R_t^m[q^{\ell,m}]\|^p\right]\right) \tag{HFS}$$

$$\leq 3^{p-1}\left(L^p \sup_{t\in[0,\tilde{t}]} E\left[\left(\int_0^t \|q_s^m - q_s^{\ell+1,m}\|ds\right)^p\right]\right.$$

$$+ \sup_{t\in[0,\tilde{t}]} E\left[\left(\int_0^t \|(\tilde{\gamma}^{-1}\sigma)(s, q_s^m) - (\tilde{\gamma}^{-1}\sigma)(s, q_s^{\ell+1,m})\|^2 ds\right)^{p/2}\right] \tag{BDG}$$

$$\left. + m^{p/2}L^p \sup_{t\in[0,\tilde{t}]} E\left[\|q_t^m - q_t^{\ell,m}\|^q\right]^{p/q}\right).$$

Lemma 3.5 was used to obtain the last line.

$\tilde{\gamma}^{-1}\sigma$ is Lipschitz in q, uniformly in $t \in [0,T]$. This, together with Hölder's inequality and the induction hypothesis yields

$$\sup_{t\in[0,\tilde{t}]} E[\|q_t^m - q_t^{\ell+1,m}\|^p] \tag{104}$$

$$\leq 3^{p-1}\left(T^{p-1}L^p \int_0^{\tilde{t}} E\left[\|q_s^m - q_s^{\ell+1,m}\|^p\right]ds\right. \tag{H}$$

$$\left. + T^{p/2-1}L^p \int_0^{\tilde{t}} E\left[\|q_s^m - q_s^{\ell+1,m}\|^p\right]ds + O(m^{p(\ell+1)/2})\right) \tag{H}$$

$$\leq C \int_0^{\tilde{t}} \sup_{r\in[0,s]} E\left[\|q_r^m - q_r^{\ell+1,m}\|^p\right]ds + O(m^{p(\ell+1)/2})$$

for some constant $C > 0$.

The integrand is L^1 because q_t^m, $q_t^{\ell+1,m} \in Y$, hence Gronwall's inequality gives

$$\sup_{t\in[0,T]} E[\|q_t^m - q_t^{\ell+1,m}\|^p] \leq O(m^{p(\ell+1)/2})e^{CT} = O(m^{p(\ell+1)/2}). \tag{105}$$

We have proven the desired bound for $p \geq 2$, but it follows for all $p > 0$ by Hölder's inequality (H). This completes the proof of the first part of Eq. (99).

The proof of the second part is similar. Again starting from Eq. (102), for $p \geq 2$ we have

$$E\left[\sup_{t\in[0,\tilde{t}]} \|q_t^m - q_t^{\ell+1,m}\|^p\right] \tag{106}$$

$$\leq 3^{p-1}\left(L^p E\left[\left(\int_0^{\tilde{t}} \|q_s^m - q_s^{\ell+1,m}\|ds\right)^p\right]\right. \tag{HFS}$$

$$+ E\left[\sup_{t\in[0,\tilde{t}]} \|\int_0^t (\tilde{\gamma}^{-1}\sigma)(s,q_s^m) - (\tilde{\gamma}^{-1}\sigma)(s,q_s^{\ell+1,m})dW_s\|^p\right] \tag{HFS}$$

$$\left. + m^{p/2}E\left[\sup_{t\in[0,\tilde{t}]} \|R_t^m[q^m] - R_t^m[q^{\ell,m}]\|^p\right]\right) \tag{HFS}$$

$$\leq 3^{p-1}\left(L^p T^{p-1} \int_0^{\tilde{t}} E\left[\sup_{r\in[0,s]} \|q_r^m - q_r^{\ell+1,m}\|^p\right]ds\right. \tag{H}$$

$$+ T^{p/2-1}L^p \int_0^{\tilde{t}} E\left[\sup_{r\in[0,s]} \|q_r^m - q_r^{\ell+1,m}\|^p\right]ds \tag{BDG,H}$$

$$\left. + m^{p/2}E\left[\sup_{t\in[0,\tilde{t}]} \|R_t^m[q^m] - R_t^m[q^{\ell,m}]\|^p\right]\right).$$

By Eq. (92) and the induction hypothesis, given $\epsilon > 0$ there exist $L > 0$, $q > 0$ such that

$$E\left[\sup_{t\in[0,\tilde{t}]}\|R_t^m[q^m] - R_t^m[q^{\ell,m}]\|^p\right] \leq \frac{L^p}{m^{p\epsilon/2}}E\left[\sup_{t\in[0,\tilde{t}]}\|q^m - q^{\ell,m}\|^q\right]^{p/q}$$

(107)

$$\leq \frac{L^p}{m^{p\epsilon/2}}O(m^{p(\ell-\epsilon)/2}) = O(m^{p(\ell/2-\epsilon)}).$$

Therefore

$$E\left[\sup_{t\in[0,\tilde{t}]}\|q_t^m - q_t^{\ell+1,m}\|^p\right]$$

(108)

$$\leq C\int_0^{\tilde{t}} E\left[\sup_{r\in[0,s]}\|q_r^m - q_r^{\ell+1,m}\|^p\right]ds + O(m^{p((\ell+1)/2-\epsilon)})$$

for some $C > 0$.

The integrand is again in L^1 because q_t^m, $q_t^{\ell+1,m} \in Y$, hence Gronwall's inequality similarly gives

$$E\left[\sup_{t\in[0,T]}\|q_t^m - q_t^{\ell+1,m}\|^p\right] \leq O(m^{p((\ell+1)/2-\epsilon)}).$$

(109)

The bound for arbitrary $p > 0$ again follows from Hölder's inequality (H), and so the proof of the second half of Eq. (99) is also complete.

$\square$

Remark 3.9. *By introducing auxiliary variables $z_t^{\ell-1,m} \equiv z_t^m[q^{\ell-1,m}]$, noting that they satisfy*

$$dz_t^{\ell-1,m} = -\frac{1}{m}\tilde{\gamma}(s, q_s^{\ell-1,m})z_t^{\ell-1,m}dt + \frac{1}{\sqrt{m}}F(t, q_t^{\ell-1,m})dt$$

(110)

$$+ \frac{1}{\sqrt{m}}\sigma(t, q_t^{\ell-1,m})dW_t,$$

and using Itô's formula on the terms

$$-d((\tilde{\gamma}^{-1})^{ij}(t, y_t)(z_t^m[y])_j) \text{ and } -\sqrt{m}d(Q^{ikl}(t, y_t)G_{kl}^{ab}(t, y_t)(z_t^m[y])_a(z_t^m[y])_b)$$

in $R_t^m[q^{\ell-1,m}]$, the hierarchy Eq. (98) can be embedded in a system of SDEs. However, for our purposes the resulting form is much less convenient to work with than the hierarchy Eq. (98), largely due to singular nature of the $1/m$ and $1/\sqrt{m}$ factors in Eq. (110). In contrast, the m-dependence of the integral formula for $z_t^m[q^{\ell-1,m}]$ (see Eq. (96) and Eq. (97), especially once combined with Eq. (136) and Eq. (140)) is manifestly more well-behaved. Similarly, the m-dependence of the formula for $R_t^m[q^{\ell-1,m}]$ (see Eq. (95)) and of the equations for $q_t^{\ell,m}$ (see Eq. (98)) presents no additional trouble. These facts play a crucial role in our proofs.

3.4. Special Cases

We end this section by presenting simplified formulas for the hierarchy in several important special cases, all of which are direct consequences of the theorems stated earlier. First we recall the noise-induced drift in the fluctuation-dissipation case:

Corollary 3.10. *Suppose that $\psi = 0$ and a fluctuation-dissipation relation holds,*

$$\Sigma_{ij}(t, q) = 2k_B T(t, q)\gamma_{ij}(t, q), \tag{111}$$

for a time and position dependent 'temperature' $T(t, q)$. Then the noise-induced drift has the following simplified form:

$$S^i(t, q) = k_B T(t, q)\partial_{q^j}(\gamma^{-1})^{ij}(t, q). \tag{112}$$

While Eq. (112) greatly simplifies the first approximation, the full hierarchy is still quite complicated. Things simplify further in the scalar case:

Corollary 3.11. *Suppose that $\psi = 0$ and γ and σ are scalar-valued. Note that this automatically gives a fluctuation dissipation relation with*

$$T(t, q) = \frac{\sigma^2(t, q)}{2k_B \gamma(t, q)}. \tag{113}$$

Under these conditions, the approximating hierarchy, Eq. (98), takes the following form for $\ell > 1$:

$$q_t^{\ell, m} = q_0 + \int_0^t \gamma^{-1}(s, q_s^{\ell, m})F(s, q_s^{\ell, m})ds + \int_0^t k_B T(s, q_s^{\ell, m})\nabla_q(\gamma^{-1})(s, q_s^{\ell, m})ds \tag{114}$$

$$+ \int_0^t \gamma^{-1}(s, q_s^{\ell, m})\sigma(s, q_s^{\ell, m})dW_s + \sqrt{m}R_t^m[q^{\ell-1, m}],$$

where

$$
\begin{aligned}
d(R_t^m[y])^i =& - d((\gamma^{-1})^{ij}(t, y_t)(z_t^m[y])_j) + (z_t^m[y])_j\partial_t(\gamma^{-1})^{ij}(t, y_t)dt \quad (115) \\
&+ Y^{ikl}(t, y_t)\left((z_t^m[y])_k F_l(t, y_t) + (z_t^m[y])_l F_k(t, y_t)\right)dt \\
&+ (z_t^m[y])_k(z_t^m[y])_l(z_t^m[y])^j\partial_{q^j}(Y^{ikl})(t, y_t)dt \\
&+ Y^{ikl}(t, y_t)\left((z_t^m[y])_k\sigma_{l\rho}(t, y_t) + (z_t^m[y])_l\sigma_{k\rho}(t, y_t)\right)dW_t^\rho \\
&- \sqrt{m}d(Y^{ikl}(t, y_t)(z_t^m[y])_k(z_t^m[y])_l) \\
&+ \sqrt{m}(z_t^m[y])_k(z_t^m[y])_l\partial_t(Y^{ikl})(t, y_t)dt,
\end{aligned}
$$

$$R_0^m[y] = 0,$$

$$z_t^m[y] = \exp\left(-\frac{1}{m}\int_0^t \gamma(r, y_r)dr\right)z_0^m \tag{116}$$

$$+ \frac{1}{\sqrt{m}}\int_0^t \exp\left(-\frac{1}{m}\int_s^t \gamma(r, y_r)dr\right)F(s, y_s)ds$$

$$+ \frac{1}{\sqrt{m}}\exp\left(-\frac{1}{m}\int_0^t \gamma(r, y_r)dr\right)\int_0^t \exp\left(\frac{1}{m}\int_0^s \gamma(r, y_r)dr\right)\sigma(s, y_s)dW_s,$$

$$Y^{ikl}(t, q) \equiv \frac{1}{2}\gamma^{-1}(t, q)\partial_{q^j}\gamma^{-1}(t, q)\delta^{ik}\delta^{jl}, \tag{117}$$

$$F(t, q) = -\nabla_q V(t, q) + \tilde{F}(t, q). \tag{118}$$

Finally, instead of a fluctuation-dissipation relation, suppose that γ is state-independent:

Corollary 3.12. *Suppose that $\psi = 0$ and γ is independent of q. Then the approximating hierarchy, Eq. (98), takes the following form for $\ell > 1$:*

$$q_t^{\ell,m} = q_0 + \int_0^t \gamma^{-1}(s)F(s, q_s^{\ell,m})ds + \int_0^t \gamma^{-1}(s)\sigma(s, q_s^{\ell,m})dW_s \tag{119}$$

$$+ \sqrt{m}R_t^m[q^{\ell-1,m}],$$

where $F(t, q) = -\nabla_q V(t, q) + \tilde{F}(t, q)$,

$$R_t^m[q^{\ell-1,m}] \tag{120}$$

$$= \int_0^t \partial_t(\gamma^{-1})(s)z_s^m[q^{\ell-1,m}]ds - \left(\gamma^{-1}(t)z_t^m[q^{\ell-1,m}] - \gamma^{-1}(0)z_0^m\right),$$

$$z_t^m[y] \tag{121}$$

$$= \Phi_t^m z_0^m + \frac{1}{\sqrt{m}}\Phi_t^m\left(\int_0^t (\Phi_s^m)^{-1}F(s, y_s)ds + \int_0^t (\Phi_s^m)^{-1}\sigma(s, y_s)dW_s\right),$$

and Φ_t^m is the (non-random) matrix-valued function that solves

$$\frac{d}{dt}\Phi_t^m = -\frac{1}{m}\gamma(t)\Phi_t^m, \quad \Phi_0^m = I. \tag{122}$$

4. Extension to Unbounded Forcing

The boundedness assumptions in Theorem 3.8 can be relaxed by using the technique developed in [7], and similarly used in [9], at the cost of weakening the convergence mode to convergence in probability. Specifically, here we will focus on accommodating unbounded forces, $\tilde{F}$ and $\nabla_q V$, where V is sufficiently confining. In this section, we will no longer be working under the assumptions from Appendix A, but rather:

Assumption 4.1. *Assume that:*

1. $V(t,q)$ is C^2 and there exist $a \geq 0, b \geq 0$ such that

$$\tilde{V}(t,q) \equiv a + b\|q\|^2 + V(t,q) \tag{123}$$

   is non-negative.
2. $\psi(t,q)$ is C^4 and $\nabla_q \psi$ is bounded.
3. $\gamma(t,q)$ is a bounded, C^3 function valued in the symmetric $n \times n$ real matrices with eigenvalues bounded below by some $\lambda > 0$.
4. $\sigma(t,q)$ is bounded, continuous and Lipschitz in q with the Lipschitz constant uniform on compact time intervals.
5. $\tilde{F}(t,q)$ is continuous and locally Lipschitz in q with the Lipschitz constant uniform on compact time intervals.
6. There exist $C > 0$, $M > 0$ such that

$$|\partial_t V(t,q)| \leq M + C(\|q\|^2 + \tilde{V}(t,q)), \tag{124}$$

$$\| - \partial_t \psi(t,q) + \tilde{F}(t,q)\|^2 \leq M + C\left(\|q\|^2 + \tilde{V}(t,q)\right), \tag{125}$$

$$\|\partial_{q^i} \tilde{\gamma}(t,q)\|^2 \leq M + C\left(\|q\|^2 + \tilde{V}(t,q)\right), \quad i = 1, ..., n, \tag{126}$$

$$\|\nabla_q V(t,q)\| \leq M + C(\|q\|^2 + \tilde{V}(t,q)), \tag{127}$$

   and

$$\left(\sum_{i,j} |\partial_{q^i} \partial_{q^j} V(t,q)|^2\right)^{1/2} \leq M + C\left(\|q\|^2 + \tilde{V}(t,q)\right). \tag{128}$$

7. We have $\mathbb{R}^n$-valued initial conditions ($\mathcal{F}_0$-measurable random variables) (q_0, u_0^m) that satisfy the following:
   $E[\|q_0\|^p] < \infty$ for all $p > 0$ and there exists $C > 0$ such that $\|u_0^m\|^2 \leq Cm$ for all $m > 0$ and all $\omega \in \Omega$.

These assumptions are not of utmost generality, but they are still quite general, are commonly satisfied, and are relatively convenient to work with. These assumptions are similar, but not identical, to those of Theorem 6.1 in [9] (recall that what we call $\tilde{F}$ here was called F in [9]); we will comment on the differences below. Under the above assumptions, we are able to prove the following theorem.

Theorem 4.1. *Under Assumption 4.1, let (q_t^m, u_t^m) be the solutions to Eq. (31)-Eq. (32) and q_t to Eq. (10). Define the continuous semimartingales $q_t^{\ell,m}$, $\ell \in \mathbb{Z}^+$, by setting $q_t^{1,m} \equiv q_t$ and inductively defining $q_t^{\ell,m}$ to be the unique maximal solution to*

$$q_t^{\ell,m} = q_0 + \int_0^t \tilde{\gamma}^{-1}(s, q_s^{\ell,m}) F(s, q_s^{\ell,m}) ds + \int_0^t S(s, q_s^{\ell,m}) ds \tag{129}$$

$$+ \int_0^t \tilde{\gamma}^{-1}(s, q_s^{\ell,m}) \sigma(s, q_s^{\ell,m}) dW_s + \sqrt{m} R_t^m[q^{\ell-1,m}].$$

Then all q_t^m, $q_t^{\ell,m}$ are continuous semimartingales, they are defined for all $t \geq 0$, and

$$\lim_{m \to 0} P \left(\frac{\sup_{t \in [0,T]} \|q_t^m - q_t^{\ell,m}\|}{m^{\ell/2 - \epsilon}} > \delta \right) = 0 \qquad (130)$$

for all $T > 0$, $\delta > 0$, $\epsilon > 0$, $\ell \in \mathbb{Z}^+$.

The method of proof is very similar to that of Theorem 6.1 in [9], though there are some additional technical complications, primarily arising from the need to prove non-explosion of solutions to various SDEs with unbounded coefficients and semimartingale external-forcing terms (see Appendix D for the relevant tools). The main ideas are outlined below and, for completeness, a full proof is included in Appendix E:

The majority of the items in Assumption 4.1 are there to ensure non-explosion of solutions to the Langevin equation, as well as the approximation hierarchy, i.e., so that we can apply the Liapunov function method from Theorem D3 (see the assumptions therein). These requirements differ somewhat from those in [9], as here we must consider SDEs with external semimartingale forcing terms. Our method for handling such systems requires the additional bound Eq. (127) (compare with item (4) in Theorem D3). On the other hand, we combined Eq. (6.8)-(6.9) from Theorem 6.1 of [9] into the single condition Eq. (125) simply for efficiency, as it does not change the proof.

To prove the convergence result Eq. (130), one begins by defining a family of cutoff systems: Let $\chi : \mathbb{R}^n \to [0,1]$ be a C^∞ bump function, equal to 1 on $\overline{B_1(0)} \equiv \{\|q\| \leq 1\}$ and zero outside $\overline{B_2(0)}$. Given $r > 0$ let $\chi_r(q) = \chi(q/r)$ and define

$$V_r(t,q) = \chi_r(q)V(t,q), \quad \tilde{F}_r(t,q) = \chi_r(q)\tilde{F}(t,q), \quad \psi_r(t,q) = \chi_r(q)\psi(t,q),$$
$$\gamma_r(t,q) = \chi_r(q)\gamma(t,q) + (1 - \chi_r(q))\lambda I. \qquad (131)$$

For each $r > 0$, replacing V with V_r, F with F_r etc., we arrive at an SDE satisfying the hypotheses of Theorem 3.8; the regularity conditions in Eq. (4.1) were chosen for this exact purpose. Note that this also accounts for the stronger assumptions here, as opposed to in Theorem 6.1 from [9]; see also the discussion at the end of Appendix A.

Let $(q_t^{r,m}, u_t^{r,m})$ be the solutions to the cutoff system, q_t^r the solution to the corresponding homogenized equation, and $q_t^{r,\ell,m}$ the solutions to the corresponding approximating hierarchy, all using the same initial conditions as the system without the cutoff. Corresponding solutions to the original and cutoff systems agree up until the first exit time, denoted $\sigma_r^{\ell,m}$, of any position processes (Langevin, homogenized, or hierarchy up to level ℓ) from the ball

of radius r. Therefore

$$P\left(\frac{\sup_{t\in[0,T]}\|q_t^m - q_t^{\ell,m}\|}{m^{\ell/2-\epsilon}} > \delta\right) \tag{132}$$

$$\leq P\left(\sigma_r^{\ell,m} > T, \frac{\sup_{t\in[0,T]}\|q_t^{r,m} - q_t^{r,\ell,m}\|}{m^{\ell/2-\epsilon}} > \delta\right) + P\left(\sigma_r^{\ell,m} \leq T\right).$$

The first term involves the cutoff processes, to which Theorem 3.8 applies, and hence it converges to zero by Markov's inequality. For the second, one must show that it converges to zero as $r \to \infty$, uniformly in m, i.e., the probability that any of the position processes exits the r-ball goes to zero as $r \to \infty$, uniformly in m. This is done in Appendix E, starting with Eq. (181).

5. Discussion

In this paper we have shown how a bootstrapping method can be used to derive higher-order approximations to the position degrees-of-freedom of Langevin dynamics, in the small-mass limit. We obtain a hierarchy of approximations $q_t^{\ell,m}$ (m denotes the mass), where the ℓ'th level is an $O(m^{\ell/2})$ approximation to the Langevin position degrees-of-freedom, q_t^m, over compact time intervals. The equations for the $q_t^{\ell,m}$'s (see Theorem 3.8) consist of the standard overdamped equation (i.e., the $\ell = 1$ equation) with an added semimartingale correction term, which is independent of $q_t^{\ell,m}$; the correction term (for $\ell \geq 2$) is constructed from the solution, $q_t^{\ell-1,m}$, at the previous level. This work naturally leads to the following two questions.

First, can the hierarchy of approximations derived here be used as the basis for efficient numerical methods with higher-order-in-m accuracy? This is a question for future work, but the form of hierarchy derived here suggests that one should be able to avoid the difficulty inherent in the $O(m^{-1/2})$ divergence of the velocity degrees of freedom in the underdamped Langevin equation. The SDE for $q_t^{\ell,m}$, Eq. (98), consists of the standard overdamped approximation with an explicit semimartingale correction term. Methods for simulating the overdamped Langevin equation are well studied and much used, so the question is whether methods can be devised to efficiently incorporate this correction term. The fact that the correction term does not depend on the variables, $q_t^{\ell,m}$, that one is solving for is promising.

Secondly, can the method employed here be adapted to study the singular limit of other SDEs, moving beyond the small-mass limit of Langevin dynamics? We do not have an answer at this time, but we note that [25] generalizes the technique of [9] to derive a homogenized SDE for a larger class of systems, with convergence in the same L^p-sense over compact time intervals, and with explicit remainder terms; in fact, [9] itself studies more general noisy, dissipative Hamiltonian systems than just Langevin dynamics. Such convergence results for the Langevin equation were the starting point for the bootstrapping method used in this paper, and so [25] might provide a

starting point for higher-order approximations to more general singular limits of SDEs. We do anticipate that proving the required Lipschitz properties of the remainder terms will be more difficult; here we made heavy use of the formula Eq. (45) for the fast degrees-of-freedom, z_t^m, but a similar expression is not available in general.

Appendix A. Assumptions Implying Homogenization as $m \to 0$

In this appendix, we give a list of properties that, as shown in [9], are sufficient to guarantee that the solutions to the SDE Eq. (31)-Eq. (32) satisfy the properties Eq. (8), Eq. (9), and Eq. (11) (note that what we call $\tilde{F}$ here was simply called F in [9]). Some of the assumptions below are strengthened, as compared to [9], in order to meet the needs of the current paper; we remark on this further below.

We assume that

1. $\gamma : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^{n \times n}$ is C^3.
    (a) The values of γ are symmetric matrices.
    (b) The eigenvalues of γ are uniformly bounded below by some $\lambda > 0$.
    (c) γ is bounded.
    (d) For all $T > 0$ and all multi-indices α with $1 \leq |\alpha| \leq 3$, $\partial_{q^\alpha} \gamma$ is bounded uniformly for $(t, q) \in [0, T] \times \mathbb{R}^n$.
    (e) For all $T > 0$ and all multi-indices α with $0 \leq |\alpha| \leq 2$, $\partial_{q^\alpha} \partial_t \gamma$ is bounded uniformly for $(t, q) \in [0, T] \times \mathbb{R}^n$.
2. $\psi : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ is C^4.
    (a) For all $T > 0$ and all multi-indices α with $1 \leq |\alpha| \leq 4$, $\partial_{q^\alpha} \psi$ is bounded uniformly for $(t, q) \in [0, T] \times \mathbb{R}^n$.
    (b) For all $T > 0$ and all multi-indices α with $0 \leq |\alpha| \leq 3$, $\partial_{q^\alpha} \partial_t \psi$ is bounded uniformly for $(t, q) \in [0, T] \times \mathbb{R}^n$.
3. $\tilde{F} : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ is continuous.
    (a) $\tilde{F}$ is bounded.
    (b) $\tilde{F}$ is Lipschitz in q uniformly in t.
4. $\sigma : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^{n \times k}$ is continuous.
    (a) σ is bounded.
    (b) σ is Lipschitz in q uniformly in t.
5. $V : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ is C^2.
    (a) $\nabla_q V$ is Lipschitz in q uniformly in t.
    (b) For all $T > 0$, $\nabla_q V$ is bounded uniformly for $(t, q) \in [0, T] \times \mathbb{R}^n$.
    (c) There exist $a, b \geq 0$ such that $\tilde{V}(t, q) \equiv a + b\|q\|^2 + V(t, q)$ is non-negative for all t, q.
    (d) There exist $M, C > 0$ such that

$$|\partial_t V(t, q)| \leq M + C(\|q\|^2 + \tilde{V}(t, q)) \tag{133}$$

    and

$$\| - \partial_t \psi(t, q) + \tilde{F}(t, q)\|^2 \leq M + C(\|q\|^2 + \tilde{V}(t, q)) \tag{134}$$

for all t, q.

6. There exists $C > 0$ such that the (random) initial conditions satisfy $\|u_0^m\|^2 \leq Cm$ for all $m > 0$ and all $\omega \in \Omega$ and $E[\|q_0^m\|^p] < \infty$, $E[\|q_0\|^p] < \infty$, and $E[\|q_0^m - q_0\|^p]^{1/p} = O(m^{1/2})$ for all $p > 0$.

The various global-in-time properties are used to prove non-explosion of solutions, while the properties over compact time intervals are used to prove convergence to the homogenized SDE in [9]. The reason we needed to strengthen certain regularity properties here, as compared to [9], is so we can prove the required Lipschitz properties of the remainder terms, Eq. (47), on compact time intervals; this is in contrast to [9], where one only had to show that the remainder terms converge to zero as $m \to 0$. For example, the third line of Eq. (95) includes a $\partial_{q^c} Q^{ikl}$ term, which in turn involves $\partial_{q^c} \partial_{q^b} \tilde{\gamma}^{-1}$. To ensure this is Lipschitz in q, we have assumed that $\tilde{\gamma}$ is C^3 with third derivative being bounded on compact time intervals; more precisely, we have assumed this of both γ and $\partial_j \psi$, as these are used to construct $\tilde{\gamma}$. This is why we require conditions on the third derivative of γ and the fourth derivative of ψ, as opposed to [9] where we only required conditions on derivatives up to order two and three respectively. Similar remarks apply to the other objects.

Appendix B. Properties of the Fundamental Solution

Our derivations will require the use of several properties of the fundamental solution of a linear ordinary differential equation (ODE). Specifically, we need to consider the process obtained by pathwise solving the linear ODE

$$\frac{d}{dt} \Phi_t^m = -\frac{1}{m} \tilde{\gamma}(t, y_t) \Phi_t^m, \quad \Phi_0^m = I, \tag{135}$$

where y is a continuous semimartingale. The process Φ_t^m is adapted and pathwise C^1; we will call it the fundamental-solution process, as each of its paths is the fundamental solution to a linear ODE.

The symmetric part of $\tilde{\gamma}$, denoted by γ, is assumed to have eigenvalues bounded below by $\lambda > 0$ (see Appendix A). This implies the following crucial bound

$$\|\Phi_t^m (\Phi_s^m)^{-1}\| \leq e^{-\lambda(t-s)/m} \quad \text{for all } t \geq s \tag{136}$$

(see, for example, p.86 of [26]). Note that while the left hand side is random, the upper bound is not. As we have stated it, this bound requires the use of the the ℓ^2 operator norm. Otherwise, there is an additional constant multiplying the exponential.

We will also need the following bound on the difference between the fundamental solutions corresponding to two linear ODEs. See the Appendix to [13].

Lemma B1. *Let $B_i : [0, T] \to \mathbb{R}^{n \times n}$; $i = 1, 2$, be continuous and suppose their symmetric parts have eigenvalues bounded above by μ, uniformly in t.*

Consider the fundamental solutions, $\Phi_i(t)$, satisfying

$$\frac{d}{dt}\Phi_i(t) = B_i(t)\Phi_i(t), \ \Phi_i(0) = I. \tag{137}$$

Then for any $0 \leq t \leq T$ we have the bound

$$\|\Phi_1(t) - \Phi_2(t)\| \leq e^{\mu t}\int_0^t \|B_1(s) - B_2(s)\|ds. \tag{138}$$

We will need the following lemma concerning stochastic convolutions, adapted from Lemma 5.1 in [8]:

Lemma B2. *Let B_s be a continuous adapted $\mathbb{R}^{n\times n}$-valued processes. Let $\Phi(t)$ be the fundamental-solution process, pathwise satisfying*

$$\frac{d}{dt}\Phi(t) = B(t)\Phi(t), \ \Phi(0) = I. \tag{139}$$

Let V_s be a continuous adapted $\mathbb{R}^{n\times k}$-valued processes. Then we have the P-a.s. equality

$$\Phi(t)\int_0^t \Phi^{-1}(s)V_s dW_s \tag{140}$$

$$=\Phi(t)\int_0^t V_s dW_s - \Phi(t)\int_0^t \Phi^{-1}(s)B(s)\left(\int_s^t V_r dW_r\right)ds \text{ for all } t.$$

The following lemma will assist us in bounding processes having the form of the last term in Eq. (140). The proof is very similar to that of Lemma 5.1 in [8], but we provide it for completeness.

Lemma B3. *Let V_s be a continuous adapted $\mathbb{R}^{n\times k}$-valued process and $\alpha > 0$.*
Then for every $j \in \mathbb{Z}_0$ there exists $C_j > 0$ such that for all $T > 0$, $\delta > 0$ we have the P-a.s. bound

$$\sup_{t\in[0,T]}\int_0^t (t-s)^j e^{-\alpha(t-s)}\|\int_s^t V_r dW_r\|ds \tag{141}$$

$$\leq \frac{C_j}{\alpha^{j+1}}\left(\max_{\ell=1,\ldots,N}\sup_{\tau\in[(\ell-1)\delta,\min\{(\ell+1)\delta,T\}]}\|\int_{(\ell-1)\delta}^{\tau}V_r dW_r\|\right.$$

$$\left. + e^{-\alpha\delta/2}\sup_{\tau\in[0,T]}\|\int_0^{\tau}V_r dW_r\|\right),$$

where $N = \min\{\ell \in \mathbb{Z} : \ell\delta \geq T\}$. We emphasize that C_j depends only on j.

Proof. Suppose $\delta < T$. First split

$$\sup_{t\in[0,T]}\int_0^t (t-s)^j e^{-\alpha(t-s)}\|\int_s^t V_r dW_r\|ds \tag{142}$$

$$\leq \sup_{t\in[0,\delta]}\int_0^t (t-s)^j e^{-\alpha(t-s)}\|\int_s^t V_r dW_r\|ds$$

$$+ \sup_{t\in[\delta,T]}\int_0^t (t-s)^j e^{-\alpha(t-s)}\|\int_s^t V_r dW_r\|ds.$$

The first term can be bounded as follows.

$$\sup_{t\in[0,\delta]} \int_0^t (t-s)^j e^{-\alpha(t-s)} \| \int_s^t V_r dW_r \| ds \qquad (143)$$

$$= \sup_{t\in[0,\delta]} \int_0^t (t-s)^j e^{-\alpha(t-s)} \| \int_0^t V_r dW_r - \int_0^s V_r dW_r \| ds$$

$$\leq \frac{2}{\alpha^{j+1}} \sup_{0\leq\tau\leq\delta} \| \int_0^\tau V_r dW_r \| \int_0^{\alpha\delta} u^j e^{-u} du.$$

In the second term we split the integral to obtain

$$\sup_{t\in[\delta,T]} \int_0^t (t-s)^j e^{-\alpha(t-s)} \| \int_s^t V_r dW_r \| ds \qquad (144)$$

$$\leq \sup_{t\in[\delta,T]} \int_0^{t-\delta} (t-s)^j e^{-\alpha(t-s)} \| \int_s^t V_r dW_r \| ds$$

$$+ \sup_{t\in[\delta,T]} \int_{t-\delta}^t (t-s)^j e^{-\alpha(t-s)} \| \int_s^t V_r dW_r \| ds$$

$$\leq \frac{2}{\alpha^{j+1}} \left(\int_0^\infty u^{2j} e^{-u} du \right)^{1/2} e^{-\alpha\delta/2} \sup_{\tau\in[0,T]} \| \int_0^\tau V_r dW_r \|$$

$$+ \sup_{t\in[\delta,T]} \int_{t-\delta}^t (t-s)^j e^{-\alpha(t-s)} \| \int_s^t V_r dW_r \| ds.$$

Let $N = \min\{\ell \in \mathbb{Z} : \ell\delta \geq T\}$. Then P-a.s.

$$\sup_{t\in[\delta,T]} \int_{t-\delta}^t (t-s)^j e^{-\alpha(t-s)} \| \int_s^t V_r dW_r \| ds \qquad (145)$$

$$\leq \max_{\ell=1,\dots,N-1} \sup_{t\in[\ell\delta,\min\{(\ell+1)\delta,T\}]} \int_{(\ell-1)\delta}^t (t-s)^j e^{-\alpha(t-s)} \| \int_s^t V_r dW_r \| ds$$

$$\leq \frac{2}{\alpha^{j+1}} \int_0^\infty u^j e^{-u} du \max_{\ell=1,\dots,N-1} \sup_{\tau\in[(\ell-1)\delta,\min\{(\ell+1)\delta,T\}]} \| \int_{(\ell-1)\delta}^\tau V_r dW_r \|.$$

Combining Eq. (143), Eq. (144), and Eq. (145) gives the P-a.s. bound

$$\sup_{t\in[0,T]} \int_0^t (t-s)^j e^{-\alpha(t-s)} \| \int_s^t V_r dW_r \| ds \qquad (146)$$

$$\leq \frac{C_j}{\alpha^{j+1}} \left(\max_{\ell=1,\dots,N-1} \sup_{\tau\in[(\ell-1)\delta,\min\{(\ell+1)\delta,T\}]} \| \int_{(\ell-1)\delta}^\tau V_r dW_r \| \right.$$

$$\left. + e^{-\alpha\delta/2} \sup_{\tau\in[0,T]} \| \int_0^\tau V_r dW_r \| \right).$$

The case $\delta \geq T$ is covered by bounding $\max_{\ell=1,\dots,N-1}$ by $\max_{\ell=1,\dots,N}$. $\square$

Appendix C. Frequently Used Inequalities

For the convenience of the reader, here we collect several inequalities that are repeatedly used in our proofs. In proofs, we will refer to them via the abbreviations given in parentheses.

Hölder's Inequality (see, for example, Theorem 6.2 in [27]):

Lemma C1 (H). *Let $(X, \mathcal{M}, \mu)$ be a measure space, $1 < p, q < \infty$ with $1/p + 1/q = 1$, and f, g be measurable functions on X. Then*

$$\int |fg| d\mu \leq \left(\int |f|^p d\mu \right)^{1/p} \left(\int |g|^q d\mu \right)^{1/q}. \tag{147}$$

When applied to counting measure on $\{1, ..., N\}$, with $g_i = 1$, Hölder's Inequality gives the following useful bound on finite sums (one can also obtain it from Jensen's inequality):

Lemma C2 (HFS). *Let $1 \leq p < \infty$ and $f_i \geq 0$, $i = 1, ..., N$. Then*

$$\left(\sum_{i=1}^{N} f_i \right)^p \leq N^{p-1} \sum_{i=1}^{N} f_i^p. \tag{148}$$

Minkowski's Inequality for Integrals (see Theorem 6.19 in [27]):

Lemma C3 (MI). *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be sigma-finite measure spaces, $1 \leq p < \infty$, and f be a product-measurable function on $X \times Y$ that satisfies one of the following two conditions:*

1. $f \geq 0$,
2. $f(\cdot, y) \in L^p(\mu)$ for ν-a.e. y and $y \to \|f(\cdot, y)\|_{L^p(\mu)}$ is in $L^1(\nu)$.

Then

$$\left(\left| \int f(x, y) \nu(dy) \right|^p \mu(dx) \right)^{1/p} \leq \int \left(\int |f(x, y)|^p \mu(dx) \right)^{1/p} \nu(dy). \tag{149}$$

L^p-Triangle Inequality (also known as Minkowski's inequality, see Theorem 6.5 in [27]):

Lemma C4 (T). *Let $(X, \mathcal{M}, \mu)$ be a measure space, $1 \leq p < \infty$, and f, g be measurable functions on X. Then*

$$\left(\int |f + g|^p d\mu \right)^{1/p} \leq \left(\int |f|^p d\mu \right)^{1/p} + \left(\int |g|^p d\mu \right)^{1/p}. \tag{150}$$

Burkholder-Davis-Gundy Inequality (see Theorem 3.28 in [23]):

Lemma C5 (BDG). *For every $p > 0$ there exists constants $k_p, K_p \in (0, \infty)$ such that for all $\mathbb{R}$-valued continuous local martingales, M, and all stopping times, T, we have*

$$E \left[\sup_{0 \leq s \leq T} |M_s|^p \right] \leq K_p E[\langle M \rangle_T^{p/2}], \tag{151}$$

where $\langle M \rangle$ denotes the quadratic variation of M.

Recall that the quadratic variation of an Itô integral of a $\mathbb{R}^k$-valued continuous, adapted process, a_t, with respect to an $\mathbb{R}^k$-valued Wiener process, W_t, is (using summation convention) given by

$$\left\langle \int_0^{(\cdot)} a_j(s)dW_s^j \right\rangle_T = \int_0^T \|a(s)\|^2 ds \qquad (152)$$

($\|\cdot\|$ denotes the ℓ^2 norm). If M is $\mathbb{R}^n$-valued then one can still use Lemma C5 to bound $E\left[\sup_{0\leq s\leq T} \|M_s\|^p\right]$ by first using

$$\sup_{0\leq s\leq T} \|M_s\|^p \leq D_{p,n} \sum_{j=1}^n \sup_{0\leq s\leq T} |M_s^j|^p, \qquad (153)$$

where $D_{p,n}$ is a constant, depending only on p and n.

Appendix D. SDEs with Semimartingale Forcing

Let W_t be an $\mathbb{R}^k$-valued Wiener process on $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$, a filtered probability space satisfying the usual conditions [23]. In this section, we give some of the background theory of SDEs of the form

$$X_t = N_t + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \qquad (154)$$

i.e., SDEs where the initial condition is generalized to a time-dependent, continuous semimartingale forcing term, N_t. Much of the following can be found in [24], with the generalization to SDEs with explosions adapted from [28]. Both of these references discuss the generalization where W_t is replaced by a more general driving semimartingale, but we do not need that extension here.

The main existence and uniqueness result for Eq. (154) mirrors that of the more standard SDE theory:

Theorem D1. *Let $U \subset \mathbb{R}^n$ be open and $\sigma : [0, \infty) \times U \to \mathbb{R}^{n\times k}$, $b : [0, \infty) \times U \to \mathbb{R}^n$ satisfy the following:*

1. *b, σ are measurable.*
2. *For every $T > 0$ and compact $C \subset U$ there exists $K_{T,C} > 0$ such that for all $t \in [0, T]$, $x, y \in C$ we have*

$$\sup_{t\in[0,T], x\in C} \|b(t,x)\| + \sup_{t\in[0,T], x\in C} \|\sigma(t,x)\| \leq K_{T,C}. \qquad (155)$$

3. *For every $T > 0$ and compact $C \subset U$ there exists $L_{T,C} > 0$ such that for all $t \in [0, T]$, $x, y \in C$ we have*

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \leq L_{T,C}\|x - y\|. \qquad (156)$$

   i.e. $b(t,x)$ and $\sigma(t,x)$ are locally Lipschitz in x, uniformly in t on compact intervals.

  Then for any continuous semimartingale N_t with N_0 valued in U, the
SDE

$$X_t = N_t + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s \tag{157}$$

has a unique (pathwise) maximal solution up to a stopping time, e, called the
explosion time. For every $\omega \in \Omega$, e satisfies one of the following:

1. $e(\omega) = \infty$,
2. There exists a subsequence $t_n \nearrow e(\omega)$ with $\lim_{n\to\infty} X_{t_n}(\omega) = \infty$,
3. There exists a subsequence $t_n \nearrow e(\omega)$ with $\lim_{n\to\infty} d(X_{t_n}(\omega), \partial U) = 0$.

  As with standard SDEs, non-explosion of solutions follows when the
drift and diffusion are linearly bounded:

Corollary D2. *Let* $\sigma : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^{n \times k}$, $b : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ *be contin-
uous and satisfy the local Lipschitz property Eq. (156). Suppose we also have
the following linear growth bound:*
For each $T > 0$ *there exists* $L_T > 0$ *such that*

$$\sup_{t\in[0,T]} (\|b(t, x)\| + \|\sigma(t, x)\|) \le L_T(1 + \|x\|). \tag{158}$$

  Then for any continuous semimartingale, N_t, *the SDE*

$$X_t = N_t + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s \tag{159}$$

has a unique maximal solution and it is defined for all $t \ge 0$ *i.e. its explosion
time is* $e = \infty$ *a.s.*

  We will also need a generalization of the theory of Lyapunov functions
to the current setting; it is needed to prove non-explosion for the hierarchy of
approximating equations when the assumption of bounded forcing is relaxed.

Theorem D3. *Let* $U \subset \mathbb{R}^n$ *be open,* W_t *be an* $\mathbb{R}^k$-*valued Wiener process.
Suppose* $b : [0, \infty) \times U \to \mathbb{R}^n$ *and* $\sigma : [0, \infty) \times U \to \mathbb{R}^{n \times k}$ *are continuous and
satisfy the local Lipschitz property Eq. (156).*
  Let X_0 *be an* $\mathcal{F}_0$-*measurable random variable valued in* U, $a : [0, \infty) \times
\Omega \to \mathbb{R}^n$ *and* $c : [0, \infty) \times \Omega \to \mathbb{R}^{n \times k}$ *be pathwise continuous, adapted processes,
and let* N_t *be the continuous semimartingale*

$$N_t = X_0 + \int_0^t a_s ds + \int_0^t c_s dW_s. \tag{160}$$

  Suppose we have a $C^{1,2}$ *function* $V : [0, \infty) \times U \to [0, \infty)$ *and measurable
functions* $C, M : [0, \infty) \to [0, \infty)$ *that satisfy:*

1. $M(t)$ *and* $C(t)$ *are integrable on compact subsets of* $[0, \infty)$.
2. *For any* t *and any* $R > 0$ *there exists a compact* $C \subset U$ *and* $\delta > 0$ *such
   that* $V(s, x) \ge R$ *for all* $(s, x) \in [t - \delta, t] \times C^c$.

3.

$$L[V](t,x) \equiv \partial_t V(t,x) + b^i(t,x)\partial_{x^i} V(t,x) + \frac{1}{2}\Sigma^{ij}(t,x)\partial_{x_i}\partial_{x_j} V(t,x)$$
$$\leq M(t) + C(t)V(t,x),$$

where $\Sigma^{ij} = \sum_\rho \sigma^i_\rho \sigma^j_\rho$,
4. $\|\nabla_x V(t,x)\| \leq M(t) + C(t)V(t,x)$,
5. $\|D^2_x V(t,x)\|(1 + \|\sigma(t,x)\|) \leq M(t) + C(t)V(t,x)$.

Then the unique maximal solution to the SDE

$$X_t = N_t + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s \tag{161}$$

has explosion time $e = \infty$ a.s. We call V a Lyapunov function for the SDE Eq. (161).

Proof. Existence of a solution, X_t, up to explosion time, e, follows from Theorem D1. Let U_n be precompact open sets with $\overline{U_n} \subset U_{n+1} \subset U$ and $\cup_n U_n = U$. By looking at the equation on the events $\{X_0 \in U_n \setminus U_{n-1}\}$ it suffices to suppose X_0 is contained in a compact subset of U (say, U_1).

Define $\eta_m = \inf\{t : \|a_t\| \geq m\} \wedge \inf\{t : c_t \geq m\}$. a_t and c_t are continuous and adapted, so η_m are stopping times. Since η_m increase to infinity, proving that there is no explosion with N_t replaced by $N^m_t \equiv N^{\eta_m}_t$ for each m will imply that $e = \infty$.

Therefore we can fix m and consider X, the solution to

$$X_t = N^m_t + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \tag{162}$$

with explosion time e.

Define the stopping times $\tau_n = \inf\{t : X_t \in U^c_n\} \wedge n$ and note that $\tau_n < e$ a.s and $\|X^{\tau_n}_t\| \leq \sup_{x \in \overline{U_n}} \|x\|$. The continuous semimartingales X^{τ_n} are solutions to

$$X^{\tau_n}_t = X_0 + \int_0^{t \wedge \tau_n} 1_{s \leq \eta_m} a_s ds + \int_0^{t \wedge \tau_n} 1_{s \leq \eta_m} c_s dW_s \tag{163}$$
$$+ \int_0^{t \wedge \tau_n} b(s, X^{\tau_n}_s)ds + \int_0^{t \wedge \tau_n} \sigma(s, X^{\tau_n}_s)dW_s,$$

hence Itô's Lemma implies

$$V(t \wedge \tau_n, X^{\tau_n}_t) - V(0, X_0) \tag{164}$$
$$= \int_0^{t \wedge \tau_n} \partial_s V(s, X^{\tau_n}_s)ds + \int_0^{t \wedge \tau_n} \partial_{x^i} V(s, X^{\tau_n}_s)d(X^{\tau_n})^i_s$$
$$+ \frac{1}{2} \int_0^{t \wedge \tau_n} \partial_{x^i}\partial_{x^j} V(s, X^{\tau_n}_s)d[(X^{\tau_n})^i, (X^{\tau_n})^j]_s$$

$$\leq \int_0^{t \wedge \tau_n} M(s)ds + \int_0^{t \wedge \tau_n} C(s)V(s, X_s^{\tau_n})ds$$

$$+ \int_0^{t \wedge \tau_n} \partial_{x^i} V(s, X_s^{\tau_n})(1_{s \leq \eta_m}(c_s)_j^i + \sigma_j^i(s, X_s^{\tau_n}))dB_s^j$$

$$+ \int_0^{t \wedge \tau_n} 1_{s \leq \eta_m} a_s^i \partial_{x^i} V(s, X_s^{\tau_n})ds$$

$$+ \frac{1}{2} \int_0^{t \wedge \tau_n} 1_{s \leq \eta_m} \partial_{x^i} \partial_{x^j} V(s, X_s^{\tau_n})((\sigma c)^{ij}(s, X_s^{\tau_n}) + (c\sigma)^{ij}(s, X_s^{\tau_n}) + C_s^{ij})ds.$$

Note that if $\eta_m > 0$ then $1_{s \leq \eta_m} \|a_s\| \leq m$, $1_{s \leq \eta_m} \|c_s\| \leq m$ and if $\eta_m = 0$ then the integrals involving $1_{s \leq \eta_m}$ are zero. Therefore

$$V(t \wedge \tau_n, X_t^{\tau_n}) - V(0, X_0) \tag{165}$$

$$\leq \int_0^{t \wedge \tau_n} M(s)ds + \int_0^{t \wedge \tau_n} C(s)V(s, X_s^{\tau_n})ds$$

$$+ \int_0^{t \wedge \tau_n} \partial_{x^i} V(s, X_s^{\tau_n})(1_{s \leq \eta_m}(c_s)_j^i + \sigma_j^i(s, X_s^{\tau_n}))dB_s^j$$

$$+ \frac{1}{2} \int_0^{t \wedge \tau_n} \|D^2 V(s, X_s^{\tau_n})\|(2m\|\sigma(s, X_s^{\tau_n})\| + m^2)ds$$

$$+ \int_0^{t \wedge \tau_n} m\|\nabla V(s, X_s^{\tau_n})\|ds$$

$$\leq \int_0^{t \wedge \tau_n} M(s)ds + \int_0^{t \wedge \tau_n} C(s)V(s, X_s^{\tau_n})ds$$

$$+ \int_0^{t \wedge \tau_n} \partial_{x^i} V(s, X_s^{\tau_n})(1_{s \leq \eta_m}(c_s)_j^i + \sigma_j^i(s, X_s^{\tau_n}))dW_s^j,$$

where we have absorbed constants into $M(s)$ and $C(s)$.

$X_s^{\tau_n}$ is valued in U_n, a precompact subset of U. Therefore continuity of V and $\partial_{x^i} V$ imply all of these terms have finite expectations. Also

$$E[\int_0^t |1_{s \leq t \wedge \tau_n} \partial_{x^i} V(s, X_s^{\tau_n})(1_{s \leq \eta_m}(c_s)_j^i + \sigma_j^i(s, X_s^{\tau_n})|^2 ds] < \infty \tag{166}$$

for all t, implying the stochastic integral is a martingale. Therefore

$$E[V(t \wedge \tau_n, X_t^{\tau_n})] \tag{167}$$

$$\leq E[V(0, X_0)] + \int_0^t M(s)ds + \int_0^t C(s)E[V(s \wedge \tau_n, X_s^{\tau_n})]ds.$$

The integrands are L^1, hence Gronwall's inequality implies

$$E[V(t \wedge \tau_n, X_t^{\tau_n})] \leq \left(E[V(0, X_0)] + \int_0^t M(s)ds \right) \exp \left(\int_0^t C(s)ds \right) \tag{168}$$

for all $t \geq 0$.

Taking $n \geq t$ and using Fatou's lemma gives

$$\left(E[V(0, X_0)] + \int_0^t M(s)ds\right) \exp\left(\int_0^t C(s)ds\right) \geq E[\liminf_{n \to \infty} V(\tau_n, X_{\tau_n})1_{e<t}]. \tag{169}$$

Now take $\omega \in \Omega$ with $e(\omega) < t$. Given $R > 0$ we have a compact $C \subset U$ and a $\delta > 0$ such that $V \geq R$ on $[e(\omega) - \delta, e(\omega)] \times C^c$. Noting that $\tau_n(\omega) \nearrow e(\omega)$ we can take N large enough that for $n \geq N$ we have $\tau_n(\omega) \in [e(\omega) - \delta, e(\omega)]$ and $C \subset U_n$. Therefore $V(\tau_n(\omega), X_{\tau_n}(\omega)) \geq R$ for $n \geq N$. So $\liminf_{n \to \infty} V(\tau_n(\omega), X_{\tau_n}(\omega)) \geq R$ i.e. $\liminf_{n \to \infty} V(\tau_n, X_{\tau_n})1_{e<t} = \infty 1_{e<t}$. But we have a finite upper bound Eq. (169) so we must have $P(e < t) = 0$. $t \geq 0$ was arbitrary and so $e = \infty$ a.s. $\qquad\square$

Appendix E. Proof of Theorem 4.1

In this section we provide a proof of Theorem 4.1, which extends Theorem 3.8 to unbounded forces, at the cost of weakening the convergence mode to convergence in probability. Recall that here, we are working under Assumption 4.1. First, we require several lemmas:

Assumption 4.1 is sufficient to ensure non-explosion of solutions to the Langevin equation. This can be shown by constructing Lyapunov functions:

Lemma E1. *Given Assumption 4.1, there exist unique global in time solutions (q_t^m, u_t^m) to Eq. (31)-Eq. (32) and q_t to Eq. (10).*

Proof. Despite the slightly different assumptions made here, the proof in Appendix C of [9] goes through essentially unchanged. We omit the details. $\qquad\square$

For y a continuous semimartingale, we define $z_t^m[y]$ and $R_t^m[y]$ as in Definition 3.7. The following two properties will be needed:

Lemma E2. *If η is a stopping time and $y, \tilde{y}$ are continuous semimartingales that satisfy $y_t^\eta = \tilde{y}_t^\eta$ then*

$$R_{t\wedge\eta}^m[y] = R_{t\wedge\eta}^m[\tilde{y}] \tag{170}$$

for all $t \geq 0$, P-a.s.

Proof. The proof is a straightforward use of the formulas in Definition 3.7. $\quad\square$

Lemma E3. *Define $\tilde{Y}$ to be the set of continuous semimartingales of the form*

$$y_t = y_0 + \int_0^t a_s ds + \int_0^t c_s dW_s \tag{171}$$

where y_0 is $\mathcal{F}_0$-measurable and $a : [0, \infty) \times \Omega \to \mathbb{R}^n$, $c : [0, \infty) \times \Omega \to \mathbb{R}^{n \times k}$ are pathwise continuous, adapted processes.
If $y \in \tilde{Y}$ then $z_t^m[y] \in \tilde{Y}$ and $R_t^m[y] \in \tilde{Y}$.

Proof. The set of semimartingales of the form Eq. (171) is a vector space and, using integration by parts, one can see that is closed under multiplication by $\mathbb{R}$-valued processes of the form Eq. (171) (i.e., with $n = 1$), and contains $z_t^m[y]$ for any continuous semimartingale y.

The result then follows for $R_t^m[y]$ by noting that Assumption 4.1 implies all of the integrands are pathwise continuous, adapted, and that $\tilde{\gamma}^{-1}(t, q)$, $Q^{ikl}(t, q)$, and $G_{kl}^{a,b}(t, q)$ are C^2. The latter allows Itô's Lemma to be applied to $\tilde{\gamma}^{-1}(t, y_t)$ etc., yielding terms in $\tilde{Y}$, provided that $y \in \tilde{Y}$. $\qquad\square$

We also need to know that solutions to the SDE defining the hierarchy exist under the current weakened assumptions:

Lemma E4. *Under Assumption* 4.1*, for any* $y \in \tilde{Y}$ *(defined in Lemma* E3*) there is a unique continuous semimartingale,* x_t*, defined for all* $t \geq 0$ *that solves*

$$x_t = q_0 + \int_0^t \tilde{\gamma}^{-1}(s, x_s) F(s, x_s) ds + \int_0^t S(s, x_s) ds \qquad (172)$$
$$+ \int_0^t \tilde{\gamma}^{-1}(s, x_s) \sigma(s, x_s) dW_s + \sqrt{m} R_t^m[y].$$

We also have $x \in \tilde{Y}$*.*

Proof. $\tilde{\gamma}^{-1}F + S$ and $\tilde{\gamma}^{-1}\sigma$ are continuous and satisfy the local Lipschitz property, Eq. (156). Lemma E3 implies $R^m[y]$ is a continuous semimartingale (in fact, $R^m[y] \in \tilde{Y}$). Therefore Theorem D1 shows a unique maximal solution exists up to explosion time.

One can check that the function

$$(t, q) \to \|q\|^2 + \tilde{V}(t, q), \qquad (173)$$

where $\tilde{V}$ was defined in Eq. (123), satisfies all the conditions required by Theorem Eq. (D3) to make it a Lyapunov function for the SDE Eq. (172), thereby proving x_t has explosion time $e = \infty$. $R^m[y] \in \tilde{Y}$ together with Eq. (172) shows that $x \in \tilde{Y}$ as well. $\qquad\square$

We are now ready to prove Theorem 4.1:

Proof. By Lemma E1, there exist unique global in time solutions (q_t^m, u_t^m) to Eq. (31)-Eq. (32) and $q_t \in \tilde{Y}$ ($\tilde{Y}$ was defined in Lemma E3) to Eq. (10), and by induction, Lemma E4 gives globally defined continuous semimartingale solutions to the approximation hierarchy, Eq. (172).

Let $\chi : \mathbb{R}^n \to [0, 1]$ be a C^∞ bump function, equal to 1 on $\overline{B_1(0)} \equiv \{\|q\| \leq 1\}$ and zero outside $\overline{B_2(0)}$. Given $r > 0$ let $\chi_r(q) = \chi(q/r)$. Define

$$V_r(t, q) = \chi_r(q) V(t, q), \quad \tilde{F}_r(t, q) = \chi_r(q) \tilde{F}(t, q), \quad \psi_r(t, q) = \chi_r(q) \psi(t, q),$$
$$\gamma_r(t, q) = \chi_r(q) \gamma(t, q) + (1 - \chi_r(q)) \lambda I. \qquad (174)$$

For each $r > 0$, replacing V with V_r, F with F_r etc., we arrive at an SDE satisfying the hypotheses of Theorem 3.8. We will call these the cutoff systems.

Let $R_t^{r,m}[y]$ denote Eq. (95), with V replaced by V_r, etc. All of these objects and their derivatives agree on $\overline{B_r(0)}$, so for any continuous semimartingale, y, if we let $\eta_r^y = \inf\{t : \|y_t\| \geq r\}$, we have

$$R_{t \wedge \eta_r^y}^m[y] = R_{t \wedge \eta_r^y}^{r,m}[y] \tag{175}$$

for all $t \geq 0$, P-a.s.

Let $(q_t^{r,m}, u_t^{r,m})$ be the solutions to the cutoff system, q_t^r the solution to the corresponding homogenized equation, and $q_t^{r,\ell,m}$ the solutions to the corresponding approximating hierarchy, all using the same initial conditions as the system without the cutoff.

For each $r > R$ define the stopping times

$$\eta_r^m = \inf\{t : \|q_t^m\| \geq r\}, \ \eta_r^{\ell,m} = \inf\{t : \|q_t^{\ell,m}\| \geq r\}, \tag{176}$$
$$\eta_r^{r,\ell,m} = \inf\{t : \|q_t^{r,\ell,m}\| \geq r\},$$

and

$$\tau_r^{\ell,m} = \eta_r^{\ell,m} \wedge \eta_r^{\ell-1,m} \wedge \dots \wedge \eta_r^{1,m}, \ \tau_r^{r,\ell,m} = \eta_r^{r,\ell,m} \wedge \eta_r^{r,\ell-1,m} \wedge \dots \wedge \eta_r^{r,1,m}. \tag{177}$$

Note that $\eta_r^{1,m} = \inf\{t : \|q_t\| \geq r\} \equiv \eta_r$ is independent of m. Finally, define $\sigma_r^{\ell,m} = \tau_r^{\ell,m} \wedge \eta_r^m$, the first exit time for any of the position processes up to level ℓ of the hierarchy.

The drifts and diffusions of the modified and unmodified SDEs agree on the ball $\{\|q\| \leq r\}$. Therefore, using induction on ℓ, Lemma E2, Eq. (175), and pathwise uniqueness of solutions, we see that the driving semimartingales of the hierarchy up to ℓ for both the original and cutoff systems agree up to the stopping time $\tau_r^{\ell,m}$ and

$$q_{t \wedge \eta_r^m}^m = q_{t \wedge \eta_r^m}^{r,m} \text{ for all } t \geq 0 \text{ a.s.}, \tag{178}$$

$$\tau_r^{\ell,m} = \tau_r^{r,\ell,m} \text{a.s., and } q_{t \wedge \tau_r^{\ell,m}}^{\ell,m} = q_{t \wedge \tau_r^{\ell,m}}^{r,\ell,m} \text{ for all } t \geq 0 \text{ a.s.} \tag{179}$$

Fixing $r > 0$ and using Eq. (178) and Eq. (179), for any $T > 0$, $\delta > 0$, $\epsilon > 0$, $\ell \in \mathbb{Z}^+$ we can calculate

$$P\left(\frac{\sup_{t \in [0,T]} \|q_t^m - q_t^{\ell,m}\|}{m^{\ell/2-\epsilon}} > \delta\right) \tag{180}$$

$$= P\left(\sigma_r^{\ell,m} > T, \frac{\sup_{t \in [0,T]} \|q_{t \wedge \eta_r^m}^{r,m} - q_{t \wedge \tau_r^{\ell,m}}^{r,\ell,m}\|}{m^{\ell/2-\epsilon}} > \delta\right)$$

$$+ P\left(\sigma_r^{\ell,m} \leq T, \frac{\sup_{t \in [0,T]} \|q_t^m - q_t^{\ell,m}\|}{m^{\ell/2-\epsilon}} > \delta\right)$$

$$\leq P\left(\frac{\sup_{t \in [0,T]} \|q_t^{r,m} - q_t^{r,\ell,m}\|}{m^{\ell/2-\epsilon}} > \delta\right) + P\left(\sigma_r^{\ell,m} \leq T\right).$$

The first term, involving the cutoff system, converges to zero as $m \to 0$ by Markov's inequality and the convergence result for bounded forces, Eq. (99). Hence we focus on the second term. We note that the only essential difference between the argument below and the similar computation in the proof of Theorem 6.1 from [9] is the need to consider all processes in the hierarchy up to level ℓ, and not just the processes, q_t^m and $q_t^{\ell,m}$, that were being compared in Eq. (180). This is due to the iterative construction of each level in the hierarchy from the solution at the previous level. The second term can be bounded as follows:

$$P\left(\sigma_r^{\ell,m} \leq T\right) \tag{181}$$

$$\leq P(\eta_r \leq T) + P\left(\eta_r^m \leq T, \eta_r > T, \sup_{t\in[0,T]} \|q_t^r - q_t^{r,m}\| \leq 1\right)$$

$$+ \sum_{k=2}^{\ell} P\left(\tau_r^{k-1,m} > T, \eta_r^{k,m} \leq T, \sup_{t\in[0,T]} \|q_t^r - q_t^{r,k,m}\| \leq 1\right)$$

$$+ P\left(\sup_{t\in[0,T]} \|q_t^r - q_t^{r,m}\| > 1\right) + \sum_{k=2}^{\ell} P\left(\sup_{t\in[0,T]} \|q_t^r - q_t^{r,k,m}\| > 1\right)$$

$$\leq P\left(\sup_{t\in[0,T]} \|q_t\| \geq r\right) + P\left(\eta_r^m \leq T, \|q_{T\wedge\eta_r^m} - q_{T\wedge\eta_r^m}^m\| \leq 1\right)$$

$$+ \sum_{k=2}^{\ell} P\left(\eta_r^{k,m} \leq T, \|q_{T\wedge\eta_r^{k,m}} - q_{T\wedge\eta_r^{k,m}}^{k,m}\| \leq 1\right)$$

$$+ E\left[\sup_{t\in[0,T]} \|q_t^r - q_t^{r,m}\|\right] + \sum_{k=2}^{\ell} E\left[\sup_{t\in[0,T]} \|q_t^r - q_t^{r,k,m}\|\right],$$

where we again used the uniqueness results, Eq. (178)-Eq. (179). The terms in the last line go to zero as $m \to 0$, as seen from the triangle inequality and Eq. (99).

On the event where $\eta_r^m \leq T$ and $\|q_{T\wedge\eta_r^m} - q_{T\wedge\eta_r^m}^m\| \leq 1$ we have $\|q_{\eta_r^m}^m\| \geq r$ and

$$\|q_{\eta_r^m}^m\| \geq \|q_{T\wedge\eta_r^m}^m\| - \|q_{T\wedge\eta_r^m} - q_{T\wedge\eta_r^m}^m\| \geq r - 1. \tag{182}$$

Hence $\sup_{t\in[0,T]} \|q_t\| \geq r - 1$ on this event. Similarly,

$$\left\{\eta_r^{k,m} \leq T, \|q_{T\wedge\eta_r^{k,m}} - q_{T\wedge\eta_r^{k,m}}^{k,m}\| \leq 1\right\} \subset \left\{\sup_{t\in[0,T]} \|q_t\| \geq r - 1\right\}. \tag{183}$$

Therefore we obtain

$$\limsup_{m\to 0} P\left(\frac{\sup_{t\in[0,T]}\|q_t^m - q_t^{\ell,m}\|}{m^{\ell/2-\epsilon}} > \delta\right) \tag{184}$$

$$\leq P\left(\sup_{t\in[0,T]}\|q_t\| \geq r\right) + \ell P\left(\sup_{t\in[0,T]}\|q_t\| \geq r-1\right)$$

$$\leq (\ell+1)P\left(\sup_{t\in[0,T]}\|q_t\| \geq r-1\right).$$

This holds for all $r > 0$ and non-explosion of q_t implies that

$$P\left(\sup_{t\in[0,T]}\|q_t\| \geq r-1\right) \to 0 \tag{185}$$

as $r \to \infty$, hence we have proven the claimed result. □

Acknowledgments

References

[1] M. Smoluchowski. Drei Vortrage uber Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Zeitschrift fur Physik*, 17:557–585, 1916.

[2] H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284 – 304, 1940.

[3] E. Nelson. *Dynamical Theories of Brownian Motion*. Mathematical Notes. Princeton University Press, 1967.

[4] J.M. Sancho, M. San Miguel, and D. Dürr. Adiabatic elimination for systems of Brownian particles with nonconstant damping coefficients. *Journal of Statistical Physics*, 28(2):291–305, 1982.

[5] G. Volpe, L. Helden, T. Brettschneider, J. Wehr, and C. Bechinger. Influence of noise on force measurements. *Physical Review Letters*, 104(17):170602, 2010.

[6] S. Hottovy, A. McDaniel, G. Volpe, and J. Wehr. The Smoluchowski-Kramers Limit of Stochastic Differential Equations with Arbitrary State-Dependent Friction. *Communications in Mathematical Physics*, 336(3):1259–1283, 2014.

[7] D.P. Herzog, S. Hottovy, and G. Volpe. The small-mass limit for Langevin dynamics with unbounded coefficients and positive friction. *Journal of Statistical Physics*, 163(3):659–673, 2016.

[8] J. Birrell, S. Hottovy, G. Volpe, and J. Wehr. Small mass limit of a Langevin equation on a manifold. *Annales Henri Poincaré*, 18(2):707–755, 2017.

[9] J. Birrell and J. Wehr. Homogenization of dissipative, noisy, Hamiltonian dynamics. *Stochastic Processes and their Applications*, 2017.

[10] J.P. Fouque, J. Garnier, G. Papanicolaou, and K. Solna. *Wave Propagation and Time Reversal in Randomly Layered Media*. Stochastic Modelling and Applied Probability. Springer New York, 2007.

[11] G.A. Pavliotis and A. Stuart. *Multiscale Methods: Averaging and Homogenization*. Texts in Applied Mathematics. Springer New York, 2008.

[12] P. Friz, P. Gassiat, and T. Lyons. Physical Brownian motion in a magnetic field as a rough path. *Trans. Amer. Math. Soc.*, 367:7939–7955, 2015.

[13] J. Birrell and J. Wehr. Phase space homogenization of noisy Hamiltonian systems. *Annales Henri Poincaré*, 19:1081, 2018.

[14] Wenqing Hu and Konstantinos Spiliopoulos. Hypoelliptic multiscale Langevin diffusions: large deviations, invariant measures and small mass asymptotics. *Electron. J. Probab.*, 22:38 pp., 2017.

[15] Jeremiah Birrell. Entropy anomaly in Langevin–Kramers dynamics with a temperature gradient, matrix drag, and magnetic field. *Journal of Statistical Physics*, 173(6):1549–1586, Dec 2018.

[16] T. Tao and Conference Board of the Mathematical Sciences. *Nonlinear Dispersive Equations: Local and Global Analysis*. Conference Board of the Mathematical Sciences. Regional conference series in mathematics. American Mathematical Society, 2006.

[17] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 2010.

[18] Xiaojie Wang and Siqing Gan. The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Difference Equations and Applications*, 19(3):466–490, 2013.

[19] Arnulf Jentzen and Primo Punik. Strong convergence rates for an explicit numerical approximation method for stochastic evolution equations with non-globally Lipschitz continuous nonlinearities. *IMA Journal of Numerical Analysis*, 04 2019.

[20] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2013.

[21] J.A. Murdock. *Perturbations: Theory and Methods*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1999.

[22] D.R. Smith. *Singular-Perturbation Theory: An Introduction with Applications*. Cambridge University Press, 1985.

[23] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 2014.

[24] P. Protter. *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2013.

[25] Jeremiah Birrell and Jan Wehr. A homogenization theorem for langevin systems with an application to hamiltonian dynamics. In Vladas Sidoravicius, editor, *Sojourns in Probability Theory and Statistical Physics - I*, pages 89–122, Singapore, 2019. Springer Singapore.

[26] G. Teschl. *Ordinary Differential Equations and Dynamical Systems*. Graduate studies in mathematics. American Mathematical Society, 2012.

[27] G.B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2013.

[28] E.P. Hsu. *Stochastic Analysis on Manifolds*. Contemporary Mathematics. American Mathematical Society, 2002.

Jeremiah Birrell
Department of Mathematics and Statistics
University of Massachusetts Amherst
Amherst, MA 01003, USA
e-mail: `birrell@math.umass.edu`

Jan Wehr
Department of Mathematics
Program in Applied Mathematics
University of Arizona
Tucson, AZ, 85721, USA
e-mail: `wehr@math.arizona.edu`