INHOMOGENEOUS FUNCTIONALS AND APPROXIMATIONS OF INVARIANT DISTRIBUTIONS OF ERGODIC DIFFUSIONS: CENTRAL LIMIT THEOREM AND MODERATE DEVIATION ASYMPTOTICS

Arnab Ganguly* and P. Sundar
Department of Mathematics
Louisiana State University†

December 5, 2020

Abstract

The paper studies asymptotics of inhomogeneous integral functionals of an ergodic diffusion process under the effect of discretization. Convergence to the corresponding functionals of the invariant distribution is shown for suitably chosen discretization steps, and the fluctuations are analyzed through central limit theorem and moderate deviation principle. The results will be particularly useful for understanding accuracy of an Euler discretization based numerical scheme for approximating functionals of invariant distribution of an ergodic diffusion. This is an infinite-time horizon problem, and the accuracy of numerical schemes in this context are comparatively much less studied than the ones used for generating approximate trajectories of diffusions over finite time intervals. The potential applications of these results also extend to other areas including mathematical physics, parameter inference of ergodic diffusions and analysis of multiscale dynamical systems with averaging.

AMS 2010 subject classifications: 60F05, 60F10, 60H10, 60H35, 65C30.

Keywords: estimation of invariant distributions, ergodic diffusions, Euler approximation, central limit theorem, large deviations, moderate deviations, stochastic algorithms.

1 Introduction.

Consider the stochastic differential equation (SDE) driven by Brownian motion B

$$X(t) = x_0 + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dB(s), \qquad x_0 \in \mathbb{R}^d, \tag{1.1}$$

which we assume to be ergodic with invariant distribution π. Suppose one is interested in estimation of π. Of course, π satisfies the stationary Kolmogorv forward equation, $\mathcal{L}^*\pi = 0$, in the weak sense, where $\mathcal{L}^*$ is the adjoint of the generator $\mathcal{L}$ of X given by

$$\mathcal{L}g(x) = \sum_i b_i(x)\partial_i g(x) + \frac{1}{2}\sum_{ij} a_{ij}(x)\partial_{ij} g(x), \qquad g \in C^2(\mathbb{R}^d, \mathbb{R}). \tag{1.2}$$

Here $a = \sigma\sigma^T$. But since the above partial differential equation (PDE) is almost always difficult to solve in closed form or even numerically (when $d \geqslant 3$), a probabilistic approach is often the

*Corresponding author. Research supported in part by NSF DMS - 1855788 and Louisiana Board of Regents through the Board of Regents Support Fund (contract number: LEQSF(2016-19)-RD-A-04)
†Email: aganguly@lsu.edu (A. Ganguly), psundar@lsu.edu (P. Sundar)

most efficient one. This requires the ergodic theorem, which, under some standard conditions states that $\varepsilon \int_0^{1/\varepsilon} f(X(s))ds \stackrel{\varepsilon\to 0}{\longrightarrow} \pi(f) \doteq \int_{\mathbb{R}^d} f(x)d\pi(x) \quad a.s$, and a disctretization scheme, most commonly Euler-Maruyama method. Specifically, the integral $\pi(f)$ is approximated by a Riemann sum, $\hat{\pi}_\varepsilon^{\hat{\Delta}}(f) \doteq \frac{1}{N_\varepsilon}\sum_{k=1}^{N_\varepsilon} f(\hat{Z}(\hat{t}_k))$, where $N_\varepsilon = [(\varepsilon\hat{\Delta})^{-1}]$ and

$$\hat{Z}(\hat{t}_{k+1}) = \hat{Z}(\hat{t}_k) + b(\hat{Z}(\hat{t}_k))\hat{\Delta} + \sigma(\hat{Z}(\hat{t}_k))(W(\hat{t}_{k+1}) - W(\hat{t}_k)), \quad \hat{t}_{k+1} - \hat{t}_k = \hat{\Delta}.$$

Obviously, for such a scheme to be accurate, N_ε and $\hat{\Delta}$ have to be large and small, respectively. But the right choices of $\hat{\Delta}$ and N_ε(or equivalently, ε) are often not obvious for many models.

Euler-Maruyama schemes for simulating trajectories of X and estimates for weak and strong error over *finite time intervals* have been extensively studied, and we mention only a few comprehensive surveys and books for references [18, 31, 16, 12] (also see [1] for error analysis of Euler approximation for density-dependent jump Markov process). In comparison, much less is available on theoretical error analysis of its use in approximation of invariant measure for ergodic diffusions. To understand the issues here, note that although the error between X and $\hat{Z}$ over a fixed time interval $[0, T]$ is typically $O(\hat{\Delta})$ (weak error order), for many stochastic models, the constant involved grows with T. Thus, for these types of infinite-time horizon problems such estimates can lead to useless infinite error bound for the error, $\hat{\pi}_\varepsilon^{\hat{\Delta}}(f) - \pi(f)$, for a *fixed discretization step* $\hat{\Delta}$!

This shows that much care has to be taken for a rigorous error analysis, and important early results in this context were obtained by Talay [33, 34, 32]. The discretized chain $\{\hat{Z} \equiv \hat{Z}^{\hat{\Delta}}(t_k)\}$, under some regularity conditions, will have an invariant distribution $\pi^{\hat{\Delta}}$, and the total error can be split as

$$\hat{\pi}_\varepsilon^{\hat{\Delta}}(f) - \pi(f) = \left(\hat{\pi}_\varepsilon^{\hat{\Delta}}(f) - \pi^{\hat{\Delta}}(f)\right) + \left(\pi^{\hat{\Delta}}(f) - \pi(f)\right), \quad \hat{\pi}_\varepsilon^{\hat{\Delta}}(f) \doteq \frac{1}{N_\varepsilon}\sum_{k=1}^{N_\varepsilon} f(\hat{Z}(\hat{t}_k)) \stackrel{\varepsilon\to 0}{\longrightarrow} \pi^{\hat{\Delta}}(f).$$

The second error is 'purely' due to the discretization step, while the first depends on the integration time interval $[0, 1/\varepsilon]$. Talay provides L^1-type estimates on the second error in terms of Δ in [33] and [34], and notes that the first term is quite hard to estimate (also see [32]). But even the estimate on the second error term is given under some strict conditions, which in particular include boundedness of the derivatives (of all order) of coefficients. For many stochastic models, where the drift terms satisfy a recurrence condition, the boundedness assumption on the derivatives of drift could restrict applications of such a result. For SDEs on torus, Mattingly etal. [25] gives estimates on the L^p-error terms in terms of both N_ε (or equivalently, ε) and $\hat{\Delta}$ (also see [24] for some results in the case of additive noise), but the extension of these results to non-compact case is highly non-trivial.

The goal of the paper is to understand a proper scaling between ε (measuring inverse of time horizon) and the discretization step $\hat{\Delta} \equiv \hat{\Delta}(\varepsilon)$ for a complete error analysis, which will then lead to improved design of numerical schemes. Instead of looking at L^p-type error estimates, our study will be on asymptotics of the error probabilities of the form $\mathbb{P}\left(\frac{1}{\delta(\varepsilon)}|\hat{\pi}_\varepsilon^{\hat{\Delta}}(f) - \pi(f)| > \kappa\right)$ for different scaling regimes controlled by $\delta(\varepsilon)$. $\delta(\varepsilon) = \sqrt{\varepsilon}$ is of course the central limit theorem (CLT)-scaling, and if the estimator is good (due to the proper scaling of $\hat{\Delta} \equiv \hat{\Delta}(\varepsilon)$ and ε), then for regimes: $\delta(\varepsilon) \gg \sqrt{\varepsilon}$ (in the sense, $\sqrt{\varepsilon}/\delta(\varepsilon) \to 0$), we should expect exponential decay of these probabilities. The latter regime falls under the purview of moderate deviation principle (MDP). The use of large or moderate deviations in error analysis of these estimators, which we consider to be an interesting feature of our work, provides more insight than typical L^p-error bounds. It should be noted that exponential decay of error probabilities (in the right regimes) is not possible to deduce from L^p-error bounds. More importantly, in contrast to these error bounds (which can be suboptimal), the presence of both upper and lower bounds in a moderate deviation principle implies that the decay rate is optimal (for the given numerical scheme). In fact, large and moderate deviation analyses yield a precise expression of the rate of exponential decay, as opposed to L^p-error bounds which involve unknown constants. Thus they can potentially be powerful methods to compare between different numerical schemes.

We now briefly describe the results in the paper and make some comments about the mathematical technicalities. In this paper, we actually work under the transformation $t \to t/\varepsilon$. A simple change of variable formula shows that the dynamics of $X(\cdot/\varepsilon)$ is given by the SDE (2.2), in the sense that its distribution is same as that of, X^ε, the solution of (2.2). Consequently, $\int_0^1 f(X^\varepsilon(s))ds \to \pi(f)$ as $\varepsilon \to 0$. Letting Z^ε denote the (continuous) Euler approximation of X^ε (see (2.3)) corresponding to the discretization step $\Delta \equiv \Delta(\varepsilon)$, we study the asymptotics of $\frac{1}{\delta(\varepsilon)}\left(\int_0^1 f(Z^\varepsilon(s)) - \pi(f)\right)$. It should be noted that the transformation $t \to t/\varepsilon$ is used for certain technical conveniences only. The two formulations are same mathematically, and so are any numerical schemes based on discretization of (1.1) or (2.2). Indeed, $\hat{\Delta}$, introduced previously, is related to Δ by $\hat{\Delta} = \Delta/\varepsilon$ and $Z^\varepsilon(\cdot) \overset{dist}{=} \hat{Z}(\cdot/\varepsilon)$. Our paper actually addresses the problem in more generality by considering (a) inhomogeneous integral functionals of the form $\int_0^\cdot f(s, Z^\varepsilon(s))ds$ (that is, we allow f to depend explicitly on time t as well), and (b) proving the CLT and MDP for $\int_0^\cdot f(s, Z^\varepsilon(s))ds$ at a process level (c.f. Theorem 2.7 and Theorem 2.9). Inhomogeneous functionals are more difficult to handle, especially, when differentiability is not assumed in the time variable, but they arise naturally in various applications including statistical inference of SDEs and averaging of certain multiscale systems with a fast diffusion component.

The MDP is proved by the weak convergence approach [7, 3, 4, 5, 6, 8] which in particular helped us to avoid some complicated exponential probability estimates which are particularly hard to obtain for our Euler approximation problem. This approach requires careful study of the tightness of certain associated controlled processes. Similar versions of many estimates that have been developed for studying the above tightness problem, are also used in the simpler uncontrolled setting for proving the CLT result. The latter proofs are much simpler and are therefore omitted with only the important changes being been pointed out. A crucial role in the tightness problem is played by the solution of the Poisson equation $\mathcal{L}u = -f$, and its regularity properties. Many of the results which provide sufficient conditions for this required regularity properties can be found in the work of Pardoux and Veretennikov [29] (also see [30]). However, we do note that, although not explicitly mentioned in [29], the proof of the estimate on the growth rate of the derivative of the solution of the Poisson equation requires the drift b to be bounded – a condition which is restrictive for ergodic diffusions (e.g. even for Ornstein-Uhlenbeck process $b(x) \sim -x$) - see Remark A.1. In the appendix we note how a slightly modified version of these results cover the case for b having some growth properties.

A different kind of numerical scheme and related error analysis for approximation of invariant measure has been studied in a series of papers [21, 22, 26, 27, 28]. There, a weighted estimator of the form $\sum_{k=1}^N w_k f(Y_k)/\sum_{k=1}^N w_k$ is considered where $\{Y_k\}$ is a Markov chain obtained by discretizing the SDE (1.1) with decreasing time step Δ_k such that $\Delta_k \to 0$ as $k \to \infty$, $\sum_{k=1}^N \Delta_k \to \infty$, $\sum_{k=1}^N w_k \to \infty$ as $N \to \infty$. In contrast, our Δ does not change with iteration step k, but is suitably scaled with N ($\equiv N_\varepsilon$, as per our notations). Although the convergence of the numerical scheme is shown for a broad class of functions (like our paper), a CLT for the error is proved for a smaller class of test functions of the form $\mathcal{L}\varphi$, with φ satisfying several conditions including requirement of bounded derivatives up to second or higher order. Moderate deviation analysis has not been undertaken in any of these papers, and all the results are only for homogeneous functionals.

Interestingly, but not surprisingly, the machineries which we develop here (actually, in their much simplified versions) also prove an MDP of the inhomogeneous integral functionals of the original process X^ε (see Theorem 2.11). This, by itself, is an interesting problem, homogeneous version of which has been studied in quite a few papers [23, 14] using different methods. For the inhomogeneous case, to the best of our knowledge there exist only one paper [15] on moderate deviation problem, which assumes that f is bounded (also see [13]). The weak convergence approach allows us to lift some of the restrictive conditions including boundedness of f in [15] and stronger ergodicity conditions in [14].

Although we motivated the usefulness of these results in terms of approximations of functionals of the invariant distribution, π, when π is unknown or complicated, these results will also be potentially useful in many other contexts including mathematical physics, multiscale systems and statistical inference of SDEs (where estimators of parameters are often functions of certain integral functionals - see[2, 20]). The rest of the paper is organized as follows. In Section 2.1, we give the mathematical formulation of our model and the statements of our main results. The variational representation and the controlled process underlying the weak convergence approach to LDP have been described in Section 2.2. Section 3 gives equivalent forms of the MDP rate functions which are useful in proving upper and lower bounds, and which are proved, respectively, in Section 5.2 and Section 6. Estimates and related tightness results required for these proofs are discussed in Section 4 and the beginning of Section 5. The proof of CLT is given in Section 5.1. Finally, the Appendix collects some necessary technical lemmas.

Notation: The following mathematical notation and conventions will be used in the paper. $a \vee b$ and $a \wedge b$ will respectively denote $\max\{a, b\}$ and $\min\{a, b\}$. For a Polish space S, we denote by $\mathcal{P}(S)$ (resp. $\mathcal{M}_F(S)$) the space of probability measures (resp. finite measures) on S equipped with the topology of weak convergence. We denote by $C_b(S)$ the space of real continuous and bounded functions on S, and by $C_b^1(S)$ the space of bounded Lipschitz continuous functions on S. The space of continuous functions from $[0, T]$ to S, equipped with the uniform topology, will be denoted as $C([0, T] : S)$. For a bounded $\mathbb{R}^d$ valued function g on S, we define $\|g\|_\infty = \sup_{x\in S} \|g(x)\|$. For a measure ν on S, and an integrable function $g : S \to \mathbb{R}^k$, $\nu(g) = \int_S g(x)\nu(dx)$. For $x \in \mathbb{R}^k$, $\|x\|$ will denote its Euclidean norm. For a matrix M, $\|M\|$ will denote some appropriate matrix norm. Since we are working in finite-dimensions, and all norms are equivalent, we will not explicitly mention which norms are used, unless it is required. For $g : \mathbb{R}^d \to \mathbb{R}^k$, Dg will denote its derivative matrix, that is, the l-th row is given by $(Dg)_{l*} = \nabla g_l$. D^2g will denote its second derivative, that is, $(D^2g)_{lij} = \partial^2_{ij} g_l$. The *big O* and *little o* notations will be used sometimes. That is $f(x) = O(g(x))$ as $x \to a$ if $|f(x)| \leqslant C|g(x)|$ for $|x - a| \leqslant \kappa$ for some constants C and κ, or if $a = \infty$, then for $x > B$ for some constant B (or equivalently, $\limsup_{x\to a} |f(x)/g(x)| < \infty$). Similarly, $f(x) = o(g(x))$ as $x \to a$ if $|f(x)/g(x)| \to 0$, as $x \to a$. These notations will be used mostly for the limiting regimes $x \to \infty$ and $\varepsilon \to 0$, and the regime intended for such a use of big O or little o notation will be clear from the context. Sometimes, $f(x) \sim g(x)$ will be used to mean that f and g have same rate of growth, that is, $f(x) = O(g(x))$ and $g(x) = O(f(x))$. This symbol will only be used informally for illustration purposes.

Convention: If $p_0 \geqslant 0$, and a function $g : \mathbb{R}^d \to \mathbb{R}^n$ satisfies $\|g(x)\| \leqslant C(1 + \|x\|)^{p_0}$, then by a slight abuse of notation, we will use the same constant C to write, when needed, $\|g(x)\| \leqslant C(1 + \|x\|^{p_0})$.

2 Mathematical framework and some prerequisites

2.1 Formulation and main results

Throughout, we will assume that (a) the SDE (1.1) admits a (pathwise) unique strong solution X^ε, and (b) the solution X of (1.1) has a unique stationary / invariant distribution π.

Existence and uniqueness of solutions, of course holds under a variety of conditions on the coefficients, the most common being Lipschitz continuity. Existence of unique stationary distribution, for example, holds under a recurrence condition like Condition 2.1-(i) and uniform ellipticity and boundedness of $a(x) = \sigma(x)\sigma^T(x)$. In fact, in this case X is geometrically ergodic [35, 29]; specifically,

$$\|P_t(x, \cdot) - \pi\|_{TV} \leqslant \Theta \exp(\theta_1\|x\|) \exp(-\theta_2 t), \qquad \int_{\mathbb{R}^d} \exp(\theta_1\|x\|)\pi(dx) < \infty,$$

for some constants $\Theta, \theta_1, \theta_2$. Here $P_t(x, \cdot)$ denotes the transition probability kernel and $\|\cdot\|_{TV}$ denotes the total variation norm.

Although the assumption of uniform ellipticity and boundedness of $a = \sigma\sigma^T$ is typical and quite commonly found in the literature for results involving stationary distribution of SDEs, we do not impose such conditions on the diffusion coefficient σ, and directly work under the assumption that a unique stationary distribution exists. This allows us to present our results in a bit more general framework and increases their applicability to potentially bigger class of ergodic SDEs where such restrictions might not necessarily hold (for example, Cox-Ingersoll model - see Example A.5).

The following conditions on the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ will be assumed.

Condition 2.1. *The coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ has the following properties:*

(i) there exist constants $\gamma > 0, \alpha \geqslant 0$ and $B \geqslant 0$ such that
$$\langle x, b(x)\rangle \leqslant -\gamma\|x\|^{1+\alpha}, \quad \text{for } \|x\| > B;$$

(ii) $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are Hölder continuous functions with respective Hölder exponents $\nu_b, \nu_\sigma \in [0, 1]$ and Holder constants, L_b and L_σ, respectively, that is,
$$L_b = \sup_{x\neq x'} \frac{\|b(x) - b(x')\|}{\|x - x'\|^{\nu_b}}, \quad L_\sigma = \sup_{x\neq x'} \frac{\|\sigma(x) - \sigma(x')\|}{\|x - x'\|^{\nu_\sigma}}; \tag{2.1}$$

(iii) there exists a constant $\mathcal{B}$ such that $\|b(x)\| \leqslant \mathcal{B}(1 + \|x\|)^{\bar\alpha}$, for $\bar\alpha \leqslant \alpha \wedge 1$;

(iv) there exists a constant $\bar{\mathcal{B}}$ such that $\|\sigma(x)\| \leqslant \bar{\mathcal{B}}(1 + \|x\|)^{\lambda}$, for $\lambda \leqslant \alpha/2$;

(v) the Hölder exponent of σ satisfies $\nu_\sigma \leqslant 1 - \lambda/\alpha$.

Remark 2.2. Under (i), (iii) and (iv) of Condition 2.1, the stationary distribution π (which is assumed to exist) has finite moments of any order, that is, $\int_{\mathbb{R}^d} \|x\|^q \pi(dx) < \infty$ for any $q \geqslant 0$. For justification of this fact, see Remark 5.9.

Appropriate assumptions on the moduli of continuity of the coefficients, which in this paper is in terms of Hölder continuity (and thus, of course, covering the case of Lipschitz continuous coefficients), is needed to analyze the discretized process. However we anticipate that parts of these assumptions could be sufficiently weakened to cover more general stochastic equations, as long as existence and uniqueness of solutions are guaranteed. Also we do note that such conditions are not needed for the MDP result of the original process X^ε (see Theorem 2.11).

Consider the scaled version of (1.1) obtained through the scaling $t \to t/\varepsilon$:
$$X^\varepsilon(t) = x_0 + \frac{1}{\varepsilon}\int_0^t b(X^\varepsilon(s))ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t \sigma(X^\varepsilon(s))dW(s). \tag{2.2}$$
We next consider an appropriate Euler-Maruyama discretization of scheme for X^ε. Let $\{t_k\}$ be a partition of $[0, T]$ such that $\Delta \equiv \Delta(\varepsilon) = t_k - t_{k-1}$, and let Z^ε denote the (continuous) Euler approximation of X^ε. In other words, let $\varrho_\varepsilon(s) = t_k$ for $t_k \leqslant s < t_{k+1}$, and Z^ε the solution to the SDE:
$$Z^\varepsilon(t) = x_0 + \frac{1}{\varepsilon}\int_0^t b(Z^\varepsilon(\varrho_\varepsilon(s)))ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t \sigma(Z^\varepsilon(\varrho_\varepsilon(s)))dW(s). \tag{2.3}$$
Let Ξ_ε, defined by $\Xi_\varepsilon(A \times [0, t]) = \int_0^t 1_{\{Z^\varepsilon(s)\in A\}}ds$, denote the occupation measure of the process Z^ε, and, as standard, for $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}^n$, $\Xi_\varepsilon(f)(t)$ will denote the following:
$$\Xi_\varepsilon(f)(t) = \int_{\mathbb{R}^d\times[0,t]} f(x, s)\Xi_\varepsilon(dx \times ds) \doteq \int_0^t f(Z^\varepsilon(s))ds.$$
As mentioned in the introduction, the objective of the paper is to study the asymptotics of
$$\Upsilon_\varepsilon(f) \doteq \frac{1}{\delta(\varepsilon)}\left(\Xi_\varepsilon(f)(\cdot) - \int_0^\cdot \pi(f(s, \cdot))ds\right)$$
in $C([0, T], \mathbb{R}^n)$, under a suitable scaling between the discretization step $\Delta \equiv \Delta(\varepsilon)$ and ε in the following regimes:

- *Central limit scaling:* $\delta(\varepsilon) = \varepsilon^{1/2}$.

- *Moderate deviation scaling:* $\varepsilon \to 0, \quad \delta(\varepsilon) \to 0, \quad \beta(\varepsilon) \equiv \varepsilon/\delta^2(\varepsilon) \to 0.$

The case $\delta(\varepsilon) = 1$ requires investigating large deviation asymptotics which we do not undertake in this paper; large deviation analysis requires some different estimates and deserves a separate paper-long treatment. In this paper, the notation $\delta(\varepsilon)$ will only be used in the moderate deviation scaling regime.

Our results will be proved for the class of functions f of the form $f = -\mathcal{L}u$, with $u \in W^{2,p}_{loc}$ satisfying some additional conditions, which we state shortly. In other words, the MDP results are applicable to a *subset of the following class of functions*: $\{\mathcal{L}u : u \in W^{2,p}_{loc}\}$.

Assumption 2.3. *For each $t \geqslant 0$, $\pi(f(t,\cdot)) \equiv \int f(t,x)\pi(dx) = 0$, that is, f is centralized. Furthermore, there exist exponents $p_0, q_0 \in \mathbb{R}$ and a constant $\mathcal{C}(T)$ such that*

(i) $\sup_{t\leqslant T} \|f(t,x)\| \leqslant \mathcal{C}(T)(1+\|x\|)^{p_0}$;

(ii) $\omega_f(\Delta, x) \leqslant \mathcal{C}(T)\mathfrak{r}(\Delta)(1+\|x\|)^{q_0}$, *where* $\omega_f(\Delta,x) \doteq \sup_{|t-s|\leqslant\Delta, 0\leqslant s,t\leqslant T} \|f(x,t) - f(x,s)\|$ *is the modulus of continuity of f.*

(iii) $f = -\mathcal{L}u$, *with* $u \in \cap_{p>0} W^{2,p}_{loc}$, *where* $u = (u_1, u_2, \ldots, u_n)$ *satisfies the following conditions:*

(a) $\sup_{t\leqslant T} \|u_l(t,x)\| \leqslant \mathcal{C}_1(T)\,(1+\|x\|)^{p_1}$,

(b) $\sup_{t\leqslant T} \|\nabla u_l(t,x)\| \leqslant \mathcal{C}_1(T)\,(1+\|x\|)^{p_2}$,

(c) $\omega_{u_l}(\Delta,x) \doteq \sup_{\{|t-s|\leqslant\Delta,\ 0\leqslant s,t\leqslant T\}} \|u_l(t,x) - u_l(s,x)\| \leqslant \mathcal{C}_1(T)\mathfrak{r}(\Delta)(1+\|x\|)^{q_1}$,

(d) $\omega_{\nabla u_l}(\Delta,x) \doteq \sup_{\{|t-s|\leqslant\Delta,\ 0\leqslant s,t\leqslant T\}} \|\nabla u_l(t,x) - \nabla u_l(s,x)\| \leqslant \mathcal{C}_1(T)\mathfrak{r}(\Delta)(1+\|x\|)^{q_2}$,

(e) $\sup_{t\leqslant T} \|D^2 u_l(t,x)\| \leqslant \mathcal{C}_1(T)\,(1+\|x\|)^{p_3}$.

Remark 2.4. *(Discussion of Assumption 2.3)* Observe that class of f satisfying Assumption 2.3 is certainly a rich class of functions, as one can always pick a u satisfying Assumption 2.3-(iii), and define $f = -\mathcal{L}u$ (whose expression can be explicitly computed).

On the other hand, it is natural ask if a given f (which is centralized) satisfying some standard conditions falls in this class. This question can be answered by studying existence and regularity of solutions of *Poisson equation:* $\mathcal{L}u = -f$. For some models and certain f, the solution u can be computed directly and the required assumptions can be directly checked. For example, consider an one-dimensional SDE with $xb(x) = -|x|^{1+\alpha}$ and $\sigma(x) \equiv 1$. Notice that $\int_{\mathbb{R}} b(x)\pi(dx) = 0$. For $f(x) = -b(x)$, the coresponding $u(x) = x$ and clearly Assumption 2.8 below holds for $\alpha \geqslant 1$.

However, in most models, a closed form expression of the Poisson equation is not available, but required existence and regularity results can still be studied theoretically in certain important cases. For example, when $a = \sigma\sigma^T$ is uniformly elliptic and bounded, and f has polynomial growth, [29, Theorem 2] (also see Proposition A.2) gurantees existence and uniqueness of solution of the corresponding Poisson equation satisfying Assumption 2.3 -(iii). Thus, when the diffusion term of (1.1) is uniformly elliptic and bounded, our asymptotic results hold (essentially) for function classes of the form

$$\{f : [0,T]\times\mathbb{R}^d \to \mathbb{R}^n : \|f(t,x)\| \leqslant \mathcal{C}(T)(1+\|x\|)^{p_0},\ \text{for } 0\leqslant t\leqslant T,\ x\in\mathbb{R}^d\}.$$

The appendix contains more detailed discussion of some of the relevant results on Poisson equation and their connection to Assumption 2.3 -(iii). But the advantage of presenting the theorems directly for a subclass of $\{\mathcal{L}u : u \in W^{2,p}\}$ allows us to avoid some extra restrictions on the coefficients of SDE (1.1), which might only be needed for existence and desired regularity properties of the solutions of Poisson equations, but are not required directly for proof of Theorem 2.7 and Theorem 2.7.

For implementation, it might be even more practical and convenient to use the Riemann sum,

$$\Xi_\varepsilon^R(f)(t) = \sum_{i=1}^{[t/\Delta(\varepsilon)]} f(t_i, Z^\varepsilon(t_i))\Delta(\varepsilon) = \int_0^t f(\varrho_\varepsilon(s), Z^\varepsilon(\varrho_\varepsilon(s)))ds \tag{2.4}$$

as the estimator (the superscript R stands for Riemann sum). The associated limit theorems could be proved under either one of the following additional conditions on f.

Assumption 2.5. *Either*

(A) f is Hölder continuous with Hölder exponent $\nu_f \in (0,1]$; or

(B) f is differentiable and $\sup_{t\leqslant T}\|Df(t,x)\| \leqslant \mathcal{C}(T)(1+\|x\|)^{p_0'}$, for some $p_0' \geqslant 0$.

The following result guarantees the convergence of our scheme.

Theorem 2.6. *Let $f : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^n$ satisfy Assumption 2.3, with $\mathfrak{r}(\Delta) = O(\sqrt{\Delta})$. Let Z^ε be defined by (2.3), where the the step size $\Delta(\varepsilon)$ is such that $\Delta(\varepsilon)/\varepsilon \to 0$, as $\varepsilon \to 0$. Then under Condition 2.1, for $T > 0$, there exists a constant $\mathcal{K}(T)$ such that*

$$\mathbb{E}\left[\sup_{t\leqslant T}\|\Xi_\varepsilon(f)(t)\|\right] \leqslant \mathcal{K}(T)\sqrt{\varepsilon}.$$

In particular, $\Xi_\varepsilon(f) \to 0$ in probability in $C([0,T],\mathbb{R}^n)$ as $\varepsilon \to 0$. (Recall that f is already centralized). If, in addition, Assumption 2.5 holds, the above assertion is also true for $\Xi_\varepsilon^R(f)$.

The proof of this theorem follows easily from the proof of the CLT (stated below) which is given in Section 5.1. Indeed, multiplying (5.15) by $\sqrt{\varepsilon}$, one uses similar estimates (actually simpler versions) used in Section 5.1 and the proof of Theorem 5.8. In fact by Markov's inequality and Borel-Cantelli lemma, the subsequences along which the convergence is almost sure can be constructed.

For the CLT and the MDP results, we first define the matrix $M_f(t)$ by

$$\begin{aligned}(M_f(t))_{i,j} &= \int_{\mathbb{R}^d} Du(t,x)a(x)(Du(t,x))^T\pi(dx)\\ &= \int_{\mathbb{R}^d}\int_0^\infty [f_i(t,x)P_sf_j(t,\cdot)(x) + f_j(t,\cdot)P_sf_i(t,\cdot)(x)]\,dsd\pi(x),\end{aligned} \tag{2.5}$$

where, by a slight abuse of notation, we used $\{P_t\}$ to denote the semigroup corresponding to the transition probability kernels $\{P_t\}$ of X; in other words, $P_tg(x) = \int_{\mathbb{R}^d} g(y)P_t(x,dy)$. Note that the above quantity is finite by Remark 2.2. The second equality in (2.5) holds by Lemma 3.1.

Theorem 2.7. *Let $f : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^n$ satisfy Assumption 2.3, with $\mathfrak{r}(\Delta) = o(\sqrt{\Delta})$, and let $\nu = \nu_b \wedge \nu_\sigma$. Let Z^ε be defined by (2.3), where the the step size $\Delta(\varepsilon)$ is such that $(\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\sqrt{\varepsilon} \to 0$, as $\varepsilon \to 0$. Then under Condition 2.1,*

$$\varepsilon^{-1/2}\Xi_\varepsilon(f) \Rightarrow \int_0^\cdot M_f^{1/2}(s)dW(s),$$

in $C([0,T],\mathbb{R}^n)$ as $\varepsilon \to 0$. Moreover the above assertion is also true for $\varepsilon^{-1/2}\Xi_\varepsilon^R(f)$ if either one of the two conditions in Assumption 2.5 holds.

Finally, we state our MDP result, which we deem to be the most important contribution of the present paper. This requires additional restrictions on the exponents appearing in Assumption 2.3.

Assumption 2.8. *The exponents in Assumption 2.3 satisfy the following bounds:*

(i) $p_1 \leqslant (1+\alpha-2\lambda)/2$, *(ii)* $p_2 < \alpha - 2\lambda$ *if* $\alpha - 2\lambda \leqslant 1$, *and* $p_2 \leqslant (1+\alpha-2\lambda)/2$ *o.w* ,
(iii) $q_0 \leqslant 2(\alpha-\lambda)$, *(iv)* $q_2 \leqslant \alpha - 2\lambda$, *(v)* $q_1 \leqslant 2(\alpha-\lambda)\wedge(1+\alpha-2\lambda)$, *(vi)* $p_3 \leqslant \alpha - 2\lambda$.

Theorem 2.9. *Let $f : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^n$ satisfy Assumption 2.3 with $\mathfrak{r}(\Delta) = O(\sqrt{\Delta})$, and let $\nu = \nu_b \wedge \nu_\sigma$. Moreover suppose Assumption 2.8 holds. Let Z^ε be defined by* (2.3)*, where the the step size $\Delta(\varepsilon)$ is such that $(\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\sqrt{\varepsilon} \to 0$, as $\varepsilon \to 0$. Then under Condition 2.1 (with $\alpha > 0$), as $\varepsilon \to 0$, $\{\Upsilon_\varepsilon(f)\}$ satisfies a LDP on $C([0,T], \mathbb{R}^n)$ with speed $\beta(\varepsilon) \equiv \varepsilon/\delta^2(\varepsilon)$ and rate function I_f given by*

$$I_f(\xi) = \begin{cases} \frac{1}{2}\int_0^T (\dot{\xi}(s))^T M_f(s)^{-1}\dot{\xi}(s)ds, & \xi \text{ is absolutely continuous;} \\ \infty, & \text{otherwise.} \end{cases} \tag{2.6}$$

That is,

(i) $\liminf_{\varepsilon\to 0} \beta(\varepsilon) \log \mathbb{P}(\Upsilon_\varepsilon(f) \in O) \geqslant -I_f(O)$, *for every open set* $O \in C([0,T], \mathbb{R}^n)$*;*

(ii) $\limsup_{\varepsilon\to 0} \beta(\varepsilon) \log \mathbb{P}(\Upsilon_\varepsilon(f) \in C) \leqslant -I_f(C)$, *for every closed set* $C \in C([0,T], \mathbb{R}^n)$.

Moreover the above assertion is also true for $\varepsilon^{-1/2}\Xi^R_\varepsilon(f)$ if Assumption 2.5-(A) or Assumption 2.5-(B) holds with $\mathfrak{p}_0' \leqslant \alpha - 2\lambda$. Here for a set A, $I_f(A) = \inf_{x\in A} I_f(x)$.

Remark 2.10. If $\sigma(x) \equiv \sigma$ (a constant), then $\nu \equiv \nu_b$, and for the MDP result to hold, we only need $(\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\delta(\varepsilon) \to 0$. Thus the discretization steps can be chosen slightly bigger. Also, in this case, the growth assumptions of D^2u (Assumption 2.3-(iii)-(e) and Assumption 2.8-(iv)) are not needed.

As mentioned, not surprisingly, the same techniques prove an MDP of the inhomogeneous functionals of the original process X^ε under less restrictive conditions. Indeed, some of the estimates that are essential for study of MDP for $\Xi_\varepsilon(f)$ do not come up while considering the case of $\Gamma_\varepsilon(f)$, defined by $\Gamma_\varepsilon(f) = \int_{\mathbb{R}^d\times[0,\cdot]} f(s,x)\Gamma_\varepsilon(dx \times ds) = \int_0^\cdot f(s, X^\varepsilon(s))ds$. Some assumptions can be removed (including Hölder continuity of b and σ, provided existence and uniqeness of solution X are available), and some complex arguments could be simplified as a result.

Theorem 2.11. *Let $f : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^n$ satisfy (i) - (iii)(d) of Assumption 2.3 with $\mathfrak{r}(\Delta) = O(\sqrt{\Delta})$. Let X^ε be the unique solution to* (2.2)*. Then under (i), (iv) and (v) of Condition 2.1 (with $\alpha > 0$), and (i) - (iii) of Assumption 2.8, as $\varepsilon \to 0$, $\left\{U_\varepsilon(f) \equiv \frac{1}{\delta(\varepsilon)}\Gamma_\varepsilon(f) = \frac{1}{\delta(\varepsilon)}\int_0^\cdot f(s, X^\varepsilon(s))ds\right\}$ satisfies a LDP on $C([0,T], \mathbb{R}^n)$ with speed $\beta(\varepsilon) \equiv \varepsilon/\delta^2(\varepsilon)$ and rate function I_f given by* (2.6)*.*

2.2 Laplace principle, variational representation and controlled processes

To establish Theorem 2.9, we will actually prove the *Laplace principle*, which is equivalent to proving LDP [7, Section 1.2]. In other words, we will show that for all $F \in C_b^1(C([0,T] : \mathbb{R}^n))$

$$\lim_{\varepsilon\to 0} \beta(\varepsilon) \ln \mathbb{E}\left[\exp\Big(-F(\Upsilon_\varepsilon(f))/\beta(\varepsilon)\Big)\right] = -\inf_{\xi\in C([0,T],\mathbb{R}^d)} [I(\xi) + F(\xi)]. \tag{2.7}$$

This is the weak convergence approach to large deviation asymptotics. The first step in this approach requires variational representation of the prelimit of the left side of (2.7), that is, of expectation of exponential functionals of $\Upsilon_\varepsilon(f)$. We briefly describe the steps below.

Let $\mathcal{P}$ denote the predictable σ-field on $[0,T] \times \Omega$ associated with the filtration $\{\mathcal{F}_t : 0 \leqslant t \leqslant T\}$, and let $P_2^M \equiv \{h : [0,T] \to \mathbb{R}^m : \int_0^T \|h(s)\|^2 ds \leqslant M\}$, and

$$\mathcal{P}_2^M \equiv \left\{\psi : \psi \text{ is } \mathcal{P}\backslash\mathcal{B}(\mathbb{R}^m) \text{ measurable and } \psi \in P_2^M, \text{ a.s. } \mathbb{P}\right\}, \quad \mathcal{P}_2 \doteq \cup_{M=1}^\infty \mathcal{P}_2^M,$$

Then by the variational representation and an application of Girsanov's theorem [3, 4],

$$-\beta(\varepsilon) \ln \mathbb{E}\left[\exp\Big(-F(\Upsilon_\varepsilon(f))/\beta(\varepsilon)\Big)\right] = \inf_{\psi\in\mathcal{P}_2} \mathbb{E}\left\{\frac{1}{2}\int_0^T \|\psi(s)\|^2 ds + F(\bar{\Upsilon}_\varepsilon^\psi(f))\right\}, \tag{2.8}$$

where $\bar{\Upsilon}^{\psi}_{\varepsilon}(f)(t) = \frac{1}{\delta(\varepsilon)}\int_0^t f(s, \bar{Z}^{\psi}_{\varepsilon}(s))ds$ and $\bar{Z}^{\psi}_{\varepsilon}$ solves the controlled stochastic equation:

$$\bar{Z}^{\psi}_{\varepsilon}(t) = x_0 + \frac{1}{\varepsilon}\int_0^t b(\bar{Z}^{\psi}_{\varepsilon}(\varrho_\varepsilon(s)))ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t \sigma(\bar{Z}^{\psi}_{\varepsilon}(\varrho_\varepsilon(s)))dW(s) + \frac{\delta(\varepsilon)}{\varepsilon}\int_0^t \sigma(\bar{Z}^{\psi}_{\varepsilon}(\varrho_\varepsilon(s)))\psi(s)ds. \tag{2.9}$$

Similarly, defining $\bar{\Xi}^{R,\psi}_{\varepsilon}(f)(t) \equiv \int_0^t f(\varrho_\varepsilon(s), \bar{Z}^{\psi}_{\varepsilon}(\varrho_\varepsilon(s)))ds$,

$$-\beta(\varepsilon)\ln\mathbb{E}\left[\exp\Big(-F(\Xi^{R,\psi}_{\varepsilon}(f)/\delta(\varepsilon))/\beta(\varepsilon)\Big)\right] = \inf_{\psi}\mathbb{E}\left\{\frac{1}{2}\int_0^T \|\psi(s)\|^2 ds + F(\bar{\Xi}^{R}_{\varepsilon}(f)/\delta(\varepsilon))\right\},$$

Since P_2^M is a closed ball in $L^2([0,T])$, it is compact under the weak topology, which is metrizable, and throughout the paper, this topology will be used on P_2^M.
Notational convention: The overbar on a process will denote its controlled version. For convenience, superscripts like ψ will mostly be dropped from the notation of the controlled process.

3 Equivalent forms of the rate function

In this section we describe two equivalent forms of the rate function I_f that will be convenient to work with in the proof of upper and lower bounds of Laplace principle.

Let λ_T denote the Lebesgue measure on $[0,T]$. Let $\mathbb{B}_T = [0,T]\times\mathbb{R}^d\times\mathbb{R}^m$, and let $\mathcal{M}_1(\mathbb{B}_T)$ be the space of finite measures R on $\mathbb{B}_T$ such that $R_{(1)} = \lambda_T$ and $R_{(2,3|1)}$ is a probability measure on $\mathbb{R}^d\times\mathbb{R}^m$. Here for $i=1,2,3$, $R_{(i)}$ denotes the i-th marginal of R and $R_{(i,j|k)}$ denotes the conditional distribution of i-th and j-th coordinate given the k-th coordinate.

For each $\xi \in C([0,T],\mathbb{R})$, let $\mathcal{R}_\xi$ denote the family of measures $R\in\mathcal{M}_1(\mathbb{B}_T)$ such that

$$\int_{\mathbb{B}_T}\|z\|^2 R(d\mathbf{y}) < \infty; \tag{3.1}$$

$$\xi(t) = \int_{\mathbb{B}_t} Du(s,x)\sigma(x)zR(d\mathbf{y}); \tag{3.2}$$

$$\int_{\mathbb{B}_t}\mathcal{L}g(x)R(d\mathbf{y}) = 0, \qquad \text{for all } t\in[0,T],\ g\in C_b^2(\mathbb{R}^d,\mathbb{R}), \tag{3.3}$$

where the l-th row of the derivative matrix Du is given by

$$(Du(s,x))_{l*} = \nabla^T u_l(s,x) = (\partial_1 u_l(s,x), \partial_2 u_l(s,x), \ldots, \partial_d u_l(s,x))$$

and a typical tuple $(s,x,z)\in\mathbb{B}_T$ is denoted by $\mathbf{y}$. Define $\bar{I}_f : C([0,T],\mathbb{R}^d)\to[0,\infty]$ by

$$\bar{I}_f(\xi) = \inf_{R\in\mathcal{R}_\xi}\left\{\frac{1}{2}\int_{\mathbb{B}_T}\|z\|^2 R(d\mathbf{y})\right\}. \tag{3.4}$$

Next, let $\mathcal{A}_\xi$ denote the space of $\phi\in L^2(\mathbb{R}^d\times[0,T],\pi\times\lambda_T)$ such that

$$\xi(t) = \int_{\mathbb{R}^d\times[0,t]} Du(s,x)\sigma(x)\phi(x,s)\pi(dx)ds.$$

Define $\hat{I}_f : C([0,T],\mathbb{R}^d)\to[0,\infty]$ by

$$\hat{I}_f(\xi) = \inf_{\phi\in\mathcal{A}_\xi}\left\{\frac{1}{2}\int_{\mathbb{R}^d\times[0,T]}\|\phi(x,s)\|^2\pi(dx)ds\right\}. \tag{3.5}$$

Lemma 3.1. *$M_f(t) = \int_{\mathbb{R}^d} Du(t,x)a(x)(Du(t,x))^T\pi(dx)$, where $a=\sigma\sigma^T$, u is defined by* (A.1) *and M_f is defined by* (2.5).

Proof. Fix $t > 0$. By Itô's lemma, we have

$$\begin{aligned}u_i(t, X(r)) &= u_i(t, X(0)) + \int_0^r \mathcal{L}u_i(t, \cdot)(X(s))ds + \int_0^r \nabla^T u_i(t, X(s))\sigma(X(s))dB(s)\\ &= u_i(t, X(0)) - \int_0^r f_i(t_0, X(s))ds + \int_0^r \nabla^T u_i(t, X(s))\sigma(X(s))dB(s).\end{aligned}$$

Then by integration by parts and observing that the last term on the right side is a martingale, we have, for any $t > 0$, after taking expectation with $X(0)$ distributed as π

$$\begin{aligned}\mathbb{E}_\pi\left(u_i(t, X(r))u_j(t, X(r))\right) = \; & \mathbb{E}_\pi\left(u_i(t, X(0))u_j(t, X(0))\right) - \int_0^r \mathbb{E}_\pi\left(u_i(t, X(s))f_j(t, X(s))\right)ds\\ &- \int_0^r \mathbb{E}_\pi\left(u_j(t, X(s))f_i(t, X(s))\right)ds\\ &+ \int_0^r \mathbb{E}_\pi\left(\nabla^T u_i(t, X(s))\sigma(X(s))\sigma^T(X(s))\nabla^T u_j(t, X(s))\right)ds.\end{aligned}$$

The result now easily follows from (A.2) and from the observation that the left side is equal to the first term on the right side as for all $r > 0$, $X(r)$ is distributed as π (π is the invariant measure). $\square$

Theorem 3.2. $\bar{I}_f = \hat{I}_f = I_f$. *(see* (3.4)*,* (3.5) *and* (2.6)*, for their definitions).*

Proof. We first show that $\bar{I}_f(\xi) = \hat{I}_f(\xi)$. Fix $\kappa > 0$. Let $R \in \mathcal{R}_\xi$ be such that

$$\frac{1}{2}\int_{\mathbb{B}_T} \|z\|^2 R(d\mathbf{y}) \leqslant \bar{I}_f(\xi) + \kappa. \tag{3.6}$$

Writing $R(d\mathbf{y}) = R_{(2,3|1)}(dx \times dz|s)ds$ and using (3.3), for any $g \in C_b^2(\mathbb{R}^d, \mathbb{R})$, we have

$$0 = \int_{\mathbb{R}^d\times\mathbb{R}^m} \mathcal{L}g(x)R_{(2,3|1)}(dx \times dz|s) = \int_{\mathbb{R}^d} \mathcal{L}g(x)R_{(2|1)}(dx|s), \quad \text{for } a.a\ s \in [0, T]$$

By the uniqueness of π, we have $R_{(2|1)}(dx|s) = \pi(dx)$ for a.a $s \in [0, T]$ and thus we have

$$R(d\mathbf{y}) = R_{(3|1,2)}(dz|x, s)R_{(2|1)}(dx|s)ds = R_{(3|1,2)}(dz|x, s)\pi(dx)ds.$$

Define $\phi(x, s) = \int_{\mathbb{R}^m} zR_{(3|2,1)}(dz|x, s)$. Clearly, by Cauchy-Schwarz inequality,

$$\int_{\mathbb{R}^d\times[0,T]} \|\phi(x, s)\|^2\pi(dx)ds \leqslant \int_{\mathbb{R}^m\times\mathbb{R}^d\times[0,T]} \|z\|^2 R_{(3|2,1)}(dz|x, s)\pi(dx)ds = \int_{\mathbb{B}_T} \|z\|^2 R(d\mathbf{y}).$$

Also,

$$\begin{aligned}\xi(t) = &\int_{\mathbb{B}_t} Du(s, x)\sigma(x)zR(d\mathbf{y}) = \int_{\mathbb{B}_t} Du(s, x)\sigma(x)zR_{(3|2,1)}(dz|x, s)\pi(dx)ds\\ = &\int_{\mathbb{R}^d\times[0,t]} Du(s, x)\sigma(x)\phi(s, x)\pi(dx)ds.\end{aligned}$$

Hence $\phi \in \mathcal{A}_\xi$.

$$\hat{I}_f(\xi) \leqslant \frac{1}{2}\int_{\mathbb{R}^d\times[0,T]} \|\phi(x, s)\|^2\pi(dx)ds \leqslant \frac{1}{2}\int_{\mathbb{B}_T} \|z\|^2 R(d\mathbf{y}) \leqslant \bar{I}_f(\xi) + \kappa.$$

Since this is true for all κ, $\hat{I}_f(\xi) \leqslant \bar{I}_f(\xi)$.

Conversely, for a fixed $\kappa > 0$, let $\phi \in \mathcal{A}_\xi$ be such that

$$\frac{1}{2}\int_{\mathbb{R}^d\times[0,T]} \|\phi(x, s)\|^2\pi(dx)ds \leqslant \hat{I}_f(\xi) + \kappa. \tag{3.7}$$

Define the measure R on $\mathbb{B}_t$ by

$$R([0, t] \times A \times B) = \int_{A\times[0,t]} 1_{\{\phi(x,s)\in B\}}\pi(dx)ds.$$

Clearly, by the definition of R,

$$\int_{\mathbb{B}_T} \|z\|^2 R(d\mathbf{y}) = \int_{\mathbb{R}^d \times [0,T]} \|\phi(x,s)\|^2 \pi(dx)ds, \quad \text{and}$$

$$\xi(t) = \int_{\mathbb{R}^d \times [0,t]} Du(s,x)\sigma(x)\phi(x,s)\pi(dx)ds = \int_{\mathbb{B}_t} Du(s,x)\sigma(x) z R(d\mathbf{y}).$$

Thus

$$\bar{I}_f(\xi) \leqslant \frac{1}{2}\int_{\mathbb{B}_T} \|z\|^2 R(d\mathbf{y}) = \frac{1}{2}\int_{\mathbb{R}^d \times [0,T]} \|\phi(x,s)\|^2 \pi(dx)ds \leqslant \hat{I}_f(\xi) + \kappa.$$

Consequently, $\bar{I}_f(\xi) \leqslant \hat{I}_f(\xi)$.

We next show that $\hat{I}_f(\xi) = I_f(\xi)$. Let $\kappa > 0$ and let $\phi \in \mathcal{A}_\xi$ be such that (3.7) holds. Notice that

$$\dot{\xi}(s) = \int_{\mathbb{R}^d} Du(s,x)\sigma(x)\phi(x,s)\pi(dx).$$

By Lemma B.5 (taking $(\Omega, \mathbb{P}) = (\mathbb{R}^d, \pi)$, $H(s,x) = Du(s,x)\sigma(x)$, $b = \dot{\xi}(s)$) for a.a s

$$(\dot{\xi}(s))^T M_f(s)^{-1} \dot{\xi}(s) \leqslant \int_{\mathbb{R}^d} \|\phi(x,s)\|^2 \pi(dx),$$

where we used the fact that by Lemma 3.1,

$$M_f(s) = \int_{\mathbb{R}^d} H(s,x)H(s,x)^T \pi(dx) = \int_{\mathbb{R}^d} Du(s,x)a(x)(Du(s,x))^T \pi(dx).$$

It now readily follows that $I_f(\xi) \leqslant \hat{I}_f(\xi) + \kappa$, and since this is true for all $\kappa > 0$, we have $I_f(\xi) \leqslant \hat{I}_f(\xi)$. Conversely, for an absolutely continuous ξ, define $\phi(x,s) = H^T(s,x)M_f(s)^{-1}\dot{\xi}(s)$. Clearly, $\phi \in \mathcal{A}_\xi$, and $\frac{1}{2}\int_{\mathbb{R}^d \times [0,T]} \|\phi(x,s)\|^2 \pi(dx)ds = \frac{1}{2}\int_{[0,T]} (\dot{\xi}(s))^T M_f(s)^{-1}\dot{\xi}(s)ds$. It follows that $I_f(\xi) \geqslant \hat{I}_f(\xi)$. □

4 Some estimates

We begin by making the following simple observation. Let $\{\tilde{t}_k\}$ be a partition of $[0,t]$ such that $\tilde{t}_k - \tilde{t}_{k-1} = \tilde{\Delta}$. Let $\eta(s) = \tilde{t}_k$, if $\tilde{t}_k \leqslant s < \tilde{t}_{k+1}$. Then for any locally integrable function h, by changing the order of integration, we get

$$\begin{aligned}\int_0^t \int_{\eta(s)}^s |h(r)|dr\, ds &= \sum_k \int_{\tilde{t}_k}^{\tilde{t}_{k+1}} \int_{\tilde{t}_k}^s |h(r)|dr\, ds = \sum_k \int_{\tilde{t}_k}^{\tilde{t}_{k+1}} \int_r^{\tilde{t}_{k+1}} |h(r)|ds\, dr \\ &\leqslant \tilde{\Delta} \sum_k \int_{\tilde{t}_k}^{\tilde{t}_{k+1}} |h(r)|dr = \tilde{\Delta}\int_0^t |h(r)|dr. \end{aligned} \tag{4.1}$$

The various constants that will appear in this and subsequent sections will only depend on the parameters of the system like $\mathcal{B}, \bar{\mathcal{B}}, \nu_b, \nu_\sigma, \bar{\alpha}, \lambda$ etc., and possibly on T. The explicit dependence will not be stated again, but can easily be inferred from the context.

Lemma 4.1. *Let $\bar{Z}^\psi_\varepsilon$ as in (2.9), and assume that Condition 2.1 holds. Let $\Delta(\varepsilon)$ be such that $(\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\sqrt{\varepsilon} \to 0$ as $\varepsilon \to 0$. Then for any $M \geqslant 0$ and $m \geqslant 0$, there exist $\varepsilon_0 > 0$, and constants C^1, $C^2(T)$ such that for any $\psi \in \mathcal{P}_2^M$ and $\varepsilon \leqslant \varepsilon_0$*

(i) $\mathbb{E}\left[\|\bar{Z}^\psi_\varepsilon(s) - \bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|^m \Big| \mathcal{F}_{\varrho_\varepsilon(s)}\right] \leqslant C^1 \varsigma^m(\varepsilon)\left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{m(\bar{\alpha}\vee\lambda)}$,

(ii) $\mathbb{E}\left[\|\bar{Z}^\psi_\varepsilon(s) - \bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|^m\right] \leqslant C^1 \varsigma^m(\varepsilon)\mathbb{E}\left((1 + \|\bar{Z}^\psi_\varepsilon(s)\|)^{m(\bar{\alpha}\vee\lambda)}\right)$,

where $\varsigma(\varepsilon) = \delta(\varepsilon)\Delta^{1/2}(\varepsilon)/\varepsilon$. Furthermore, if $m \leqslant 2(1 - \lambda/\alpha)$, then *(iii)* $\displaystyle\int_0^T \mathbb{E}\left[\|\bar{Z}_\varepsilon^\psi(s) - \bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|^m\right] ds \leqslant C^2(T)\left(\frac{\Delta(\varepsilon)}{\varepsilon}\right)^{m/2} \mathbb{E}\int_0^T \left(1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|\right)^{2(\alpha-\lambda)} ds,$

(iv) $\displaystyle\int_0^T \mathbb{E}\left[\|\bar{Z}_\varepsilon^\psi(s) - \bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|^m\right] ds \leqslant C^2(T)\left(\frac{\Delta(\varepsilon)}{\varepsilon}\right)^{m/2} \mathbb{E}\int_0^T \left(1+\|\bar{Z}_\varepsilon^\psi(s)\|\right)^{2(\alpha-\lambda)} ds.$

Remark 4.2. Notice that the condition on the step size $\Delta(\varepsilon)$ implies that $\Delta^{1/2}(\varepsilon)/\varepsilon \to 0$, as $\varepsilon \to 0$. To see this simply observe that since $\nu \leqslant 1$,

$$\left(\Delta^{1/2}(\varepsilon)/\varepsilon\right)^\nu = \varepsilon^{(1-\nu)/2} \times (\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\sqrt{\varepsilon} \quad \stackrel{\varepsilon\to 0}{\longrightarrow} 0.$$

In particular, not only does this imply $\varsigma(\varepsilon) \to 0$, but also

$$\varsigma^\nu(\varepsilon)/\delta(\varepsilon) = \left(\frac{\sqrt{\varepsilon}}{\delta(\varepsilon)}\right)^{1-\nu} \times (\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\sqrt{\varepsilon} \stackrel{\varepsilon\to 0}{\longrightarrow} 0.$$

Proof. (Lemma 4.1) Observe that

$$\bar{Z}_\varepsilon^\psi(s) - \bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s)) = \frac{1}{\varepsilon} b(\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))(s-\varrho_\varepsilon(s)) + \frac{1}{\sqrt{\varepsilon}}\sigma(\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s)))(W(s)-W(\varrho_\varepsilon(s)) + \frac{\delta(\varepsilon)}{\varepsilon}\sigma(\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s)))\int_{\varrho_\varepsilon(s)}^s \psi(r)dr. \tag{4.2}$$

Now using (a) for any $m>0$, there exists a constant $\tilde{C}_m$ such that $\|x+y\|^m \leqslant \tilde{C}_m(\|x\|^m + \|y\|^m)$, (b) $\mathbb{E}(\|W(h)\|^m) = O(h^{m/2})$, (c) $\|b(x)\|^m \leqslant \mathcal{B}^m(1+\|x\|)^{m\bar{\alpha}}$, $\|\sigma(x)\|^m \leqslant \bar{\mathcal{B}}^m(1+\|x\|)^{m\lambda}$ and (d) $\|s-\varrho_\varepsilon(s)| \leqslant \Delta$, we can estimate $\mathcal{A}_0 = \mathbb{E}\left[\|\bar{Z}_\varepsilon^\psi(s) - \bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|^m \middle| \mathcal{F}_{\varrho_\varepsilon(s)}\right]$ as

$$\begin{aligned}
\mathcal{A}_0 \leqslant \tilde{C}^0 \Bigg[&(1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|)^{m\bar{\alpha}}\left(\frac{\Delta}{\varepsilon}\right)^m + (1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s)\|)^{m\lambda}\left(\frac{\Delta}{\varepsilon}\right)^{m/2} \\
&+ (1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|)^{m\lambda}\left(\frac{\delta(\varepsilon)}{\varepsilon}\right)^m \times \mathbb{E}\left(\left(\Delta\int_{\varrho_\varepsilon(s)}^s \|\psi(r)\|^2 dr\right)^{m/2} \middle| \mathcal{F}_{\varrho_\varepsilon(s)}\right)\Bigg] \\
\leqslant &\tilde{C}^1\left(1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|\right)^{m\bar{\alpha}\vee m\lambda}\left[\left(\frac{\Delta}{\varepsilon}\right)^m + \left(\frac{\Delta}{\varepsilon}\right)^{m/2} + M^{m/2}\left(\frac{\delta(\varepsilon)\Delta^{1/2}}{\varepsilon}\right)^m\right] \\
\leqslant &\tilde{C}^2\left(1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|\right)^{m(\bar{\alpha}\vee\lambda)}\varsigma^m(\varepsilon),
\end{aligned}$$

where $\varsigma(\varepsilon) = \delta(\varepsilon)\Delta^{1/2}/\varepsilon$. The last inequality follows because $(\Delta/\varepsilon)^{1/2} = \frac{\sqrt{\varepsilon}}{\delta(\varepsilon)}\varsigma(\varepsilon) \leqslant \varsigma(\varepsilon)$. Notice that since $\sqrt{\varepsilon}/\delta(\varepsilon) \to 0$ and $\varsigma(\varepsilon) \to 0$ as $\varepsilon \to 0$, we can assume that $\max\{\sqrt{\varepsilon}/\delta(\varepsilon), \Delta/\varepsilon\} \leqslant 1$ for $\varepsilon \leqslant 1$. This proves (i). It could be easily seen that (i) leads to (ii) for sufficiently small ε by using the inequality $\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\| \leqslant \|\bar{Z}_\varepsilon^\psi(s)\| + \|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s)) - \bar{Z}_\varepsilon^\psi(s)\|$ and the fact $\bar{\alpha} \leqslant 1$.

For (iii) and (iv) we first estimate $\mathcal{A}_1 \doteq \mathbb{E}\left[\int_0^T \left(1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|\right)^{m\lambda}\left(\int_{\varrho_\varepsilon(s)}^s \|\psi(r)\|^2 dr\right)^{m/2} ds\right]$ in the following way:

$$\begin{aligned}
\mathcal{A}_1 &\leqslant \left(\mathbb{E}\int_0^T\int_{\varrho_\varepsilon(s)}^s \|\psi(r)\|^2 dr ds\right)^{m/2}\left(\mathbb{E}\int_0^T\left(1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|\right)^{2m\lambda/(2-m)}\right)^{1-m/2} \\
&\leqslant \left(\mathbb{E}\int_0^T\left(1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|\right)^{2m\lambda/(2-m)}\right)^{1-m/2}\left(\Delta\int_0^T \|\psi(r)\|^2 ds\right)^{m/2} \\
&\leqslant \tilde{C}^4 M^{m/2}\Delta^{m/2}\left(\mathbb{E}\int_0^T\left(1+\|\bar{Z}_\varepsilon^\psi(\varrho_\varepsilon(s))\|\right)^{2(\alpha-\lambda)}\right)^{1-m/2}.
\end{aligned}$$

In the above, the first inequality used Holder's inequality (with $p = 2/m$), the second used (4.1) and the third used the fact $m \leqslant 2(1-\lambda/\alpha)$ implies that $m\lambda/(2-m) \leqslant \alpha - \lambda$. Now (4.2) implies that $\mathcal{A}_2 \doteq \int_0^T \mathbb{E}\left[\|\bar{Z}^\psi_\varepsilon(s) - \bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|^m\right] ds$ can be estimated as

$$\mathcal{A}_2 \leqslant \tilde{C}^5 \left[\left(\frac{\Delta}{\varepsilon}\right)^{m/2} \mathbb{E}\int_0^T \left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{m\bar{\alpha}\vee m\lambda} ds + \left(\frac{\delta(\varepsilon)}{\varepsilon}\right)^m \Delta^{m/2}\mathbb{E}(\mathcal{I})\right]$$

$$\leqslant \tilde{C}^6 \left[\left(\frac{\Delta}{\varepsilon}\right)^{m/2} \mathbb{E}\int_0^T \left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{m\bar{\alpha}\vee m\lambda} ds + \left(\frac{\delta(\varepsilon)\Delta}{\varepsilon}\right)^m \left(\mathbb{E}\int_0^T \left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{2(\alpha-\lambda)}\right)^{1-m/2}\right].$$

The assertion (iii) now follows as (a) $\max\{\delta(\varepsilon)\Delta/\varepsilon, (\Delta/\varepsilon)^{1/2}\} \leqslant (\Delta/\varepsilon)^{1/2}$, (b) $m \leqslant 2(1-\lambda/\alpha)$ (together with $\bar{\alpha}\vee\lambda \leqslant \alpha$) implies that $m(\bar{\alpha}\vee\lambda) \leqslant 2(\alpha-\lambda)$, and (c) $1 - m/2 \leqslant 1$. (iv) now follows using the same splitting used above to obtain (ii). □

The above lemma leads to some observations that will be useful later.

Corollary 4.3. *Assume the setup of Lemma 4.1, and let* $\mathfrak{g} \geqslant 0,\ \mathfrak{g}_1 \geqslant 0,\ \mathfrak{g}_2 \geqslant 0,\ \theta \geqslant 0$. *Then the following hold*

(i) $\displaystyle \frac{1}{C^3}\,\mathbb{E}\left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{\mathfrak{g}} \leqslant \mathbb{E}\left(1 + \|\bar{Z}^\psi_\varepsilon(s)\|\right)^{\mathfrak{g}} \leqslant\ C^3\mathbb{E}\left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{\mathfrak{g}}.$

(ii) $\displaystyle \int_0^T \mathbb{E}\left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{\mathfrak{g}} 1_{\{\|\bar{Z}^\psi_\varepsilon(s)\|\leqslant\bar{B}\}} ds \leqslant C^3\left(1 + \varsigma^{\mathfrak{g}}(\varepsilon)\int_0^T \|\bar{Z}^\psi_\varepsilon(s)\|^{\mathfrak{g}} ds\right).$

(iii) $$\mathbb{E}\int_0^T \left(1 + \|\bar{Z}^\psi_\varepsilon(s)\|\right)^{\mathfrak{g}_1}\left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{\mathfrak{g}_2} \|\bar{Z}^\psi_\varepsilon(s) - \bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|^\theta ds$$
$$\leqslant \varsigma^\theta(\varepsilon) C^3\mathbb{E}\int_0^T \left(1 + \|\bar{Z}^\psi_\varepsilon(s)\|\right)^{\mathfrak{g}_1+\mathfrak{g}_2+\theta(\bar{\alpha}\vee\lambda)} ds.$$

Proof. We will give the main steps for proof of (iii). The proofs of the other assertions are simpler. Using $\|\bar{Z}^\psi_\varepsilon(s)\| \leqslant \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\| + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s)) - \bar{Z}^\psi_\varepsilon(s)\|$, we have that the integrand in the left side of (iii) is less than

$$\tilde{C}^7\left[\left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{\mathfrak{g}_2+\mathfrak{g}_1} \|\bar{Z}^\psi_\varepsilon(s) - \bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|^\theta ds + \left(1 + \|\bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|\right)^{\mathfrak{g}_2} \|\bar{Z}^\psi_\varepsilon(s) - \bar{Z}^\psi_\varepsilon(\varrho_\varepsilon(s))\|^{\mathfrak{g}_1+\theta}\right].$$

The assertion now follows quite easily by using (a) Lemma 4.1-(i), (b) the fact that $\mathbb{E}(\cdot) = \mathbb{E}(\mathbb{E}(\cdot|\mathcal{F}_{\varrho_\varepsilon(s)})$, and (c) (i) of this corollary. □

The following are the corresponding results for the original process Z^ε, which will be required in proving the CLT result, and whose proof follows analogously (and in a simpler way).

Lemma 4.4. *Let* Z^ε *be as in* (2.3)*, and assume that Condition 2.1 holds. Let* $\Delta(\varepsilon)$ *be such that* $\Delta(\varepsilon)/\varepsilon \to 0$ *as* $\varepsilon \to 0$*. Then there exist constants* C^4*, and* ε_0 *such that for all* $\varepsilon \leqslant \varepsilon_0$*,*

(i) $\mathbb{E}\left[\|Z^\varepsilon(s) - Z^\varepsilon(\varrho_\varepsilon(s))\|^m \middle| \mathcal{F}_{\varrho_\varepsilon(s)}\right] \leqslant C^4(\Delta/\varepsilon)^{m/2}\mathbb{E}\left(1 + \|Z^\varepsilon(\varrho_\varepsilon(s))\|\right)^{m\bar{\alpha}}.$

(ii) $\mathbb{E}\left[\|Z^\varepsilon(s) - Z^\varepsilon(\varrho_\varepsilon(s))\|^m\right] \leqslant C^4(\Delta/\varepsilon)^{m/2}\mathbb{E}\left(1 + \|Z^\varepsilon(s))\|\right)^{m\bar{\alpha}}.$

5 Tightness results

In the proofs of the following and the subsequent results, we will adopt the notational convention mentioned before, where we will drop the superscript ψ and use $\bar{Z}_\varepsilon$ instead of $\bar{Z}^\psi_\varepsilon$.

Lemma 5.1. *Suppose that $\bar{Z}_\varepsilon^\psi$ satisfies* (2.9) *and that* $(\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\sqrt{\varepsilon} \to 0$*. Assume that Condition 2.1 (with $\alpha > \lambda$) holds. Then for all $M > 0$, there exists an $\varepsilon_0 > 0$, such that*

$$\sup_{\varepsilon \in (0,\varepsilon_0]} \sup_{\psi \in \mathcal{P}_2^M} \mathbb{E}\left[\int_0^T \|\bar{Z}_\varepsilon^\psi(t)\|^{2(\alpha-\lambda)} dt\right] < \infty.$$

Proof. The main idea is to use Itô's lemma to the function $x \to \|x\|^{(1+\alpha-2\lambda)/2}$ and then obtain estimates on different expectations. However, if $\alpha < 2\lambda + 2$, some technical issues arise (because of singularity of the map $x \to \|x\|^{\alpha-2\lambda-2}$ at origin) for obtaining bounds on certain terms. One way to avoid them is to use a $C^\infty([0,\infty),[0,\infty))$ - function ϑ defined by

$$\vartheta(x) = \begin{cases} x^{(1+\alpha-2\lambda)/2}, & x > 1 \\ 0, & 0 < x < 0.9. \end{cases}$$

By Itô's lemma,

$$\begin{aligned}
\vartheta(\|\bar{Z}_\varepsilon(t)\|^2) =& \vartheta(\|x_0\|^2) + \bar{\mathcal{M}}_\varepsilon(t) + \frac{2}{\varepsilon}\int_0^t \vartheta'(\|\bar{Z}_\varepsilon(s)\|^2)\langle \bar{Z}_\varepsilon(s), b(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\rangle ds \\
&+ \frac{1}{\varepsilon}\int_0^t \vartheta'(\|\bar{Z}_\varepsilon(s)\|^2)\|\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\|^2 \, ds \\
&+ \frac{2\delta(\varepsilon)}{\varepsilon}\int_0^t \vartheta'(\|\bar{Z}_\varepsilon(s)\|^2)\langle \bar{Z}_\varepsilon(s), \sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\psi(s)\rangle ds \\
&+ \frac{1}{\varepsilon}\int_0^t \vartheta''(\|\bar{Z}_\varepsilon(s)\|^2)\|\bar{Z}_\varepsilon(s)\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\|^2 ds. \qquad (5.1)
\end{aligned}$$

Let $\bar{B} = B \vee 1$ (B was introduced in Condition 2.1-(i)). Splitting each term according to $\{\|\bar{Z}_\varepsilon(s)\| \leqslant \bar{B}\}$ and $\{\|\bar{Z}_\varepsilon(s)\| > \bar{B}\}$, and using Condition 2.1-(i) we get

$$\begin{aligned}
\vartheta(\|\bar{Z}_\varepsilon(t)\|^2) \leqslant& \vartheta(\|x_0\|^2) + \bar{\mathcal{M}}_\varepsilon(t) + \frac{\mathcal{C}^{1,B}(T)}{\varepsilon}\int_0^T \Big[(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda} + (1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\bar{\alpha}} \\
&+ (1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\lambda}\|\psi(s)\|\Big] 1_{\{\|\bar{Z}_\varepsilon(s)\| \leqslant \bar{B}\}} ds - \frac{\gamma(1+\alpha-2\lambda)}{\varepsilon}\int_0^t \|\bar{Z}_\varepsilon(s)\|^{2(\alpha-\lambda)} \\
&+ \frac{\mathcal{C}^2}{\varepsilon}\left[\int_0^t (1+\|\bar{Z}_\varepsilon(s)\|)^{\alpha-2\lambda-1}(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda} 1_{\{\|\bar{Z}_\varepsilon(s)\| > \bar{B}\}} ds \right.\\
&\left. + \delta(\varepsilon)\int_0^t \|\bar{Z}_\varepsilon(s)\|^{\alpha-2\lambda}(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\lambda}\|\psi(s)\| ds + A_1(t)\right], \qquad (5.2)
\end{aligned}$$

where $\bar{\mathcal{M}}_\varepsilon(t) = \frac{2}{\sqrt{\varepsilon}}\int_0^t \vartheta'(\|\bar{Z}_\varepsilon(s)\|^2)\bar{Z}_\varepsilon(s)^T \sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))dW(s)$ is a martingale, and

$$A_1(t) = \int_0^t \|\bar{Z}_\varepsilon(s)\|^{\alpha-2\lambda-1}|\langle \bar{Z}_\varepsilon(s), b(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)) - b(\bar{Z}_\varepsilon(s))\rangle| ds.$$

By Cauchy-Schwarz inequality and Corollary 4.3-(ii), $\bar{\mathcal{A}} \doteq \mathbb{E}\int_0^T (1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^\lambda \|\psi(s)\| 1_{\{\|\bar{Z}_\varepsilon(s)\| \leqslant \bar{B}\}} ds$ can be estimated as

$$\begin{aligned}
\bar{\mathcal{A}} \leqslant& \left(\mathbb{E}\int_0^T (1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda} 1_{\{\|\bar{Z}_\varepsilon(s)\| \leqslant \bar{B}\}} ds\right)^{1/2} \left(\mathbb{E}\int_0^T \psi(s)\|^2 ds\right)^{1/2} \\
\leqslant& \mathcal{C}^3(T)\left(1+\varsigma^{2\lambda}(\varepsilon)\mathbb{E}\int_0^T \|\bar{Z}_\varepsilon(s)\|^{2\lambda}\right)^{1/2} \leqslant \mathcal{C}^4(T)\left(1+\varsigma^{2\lambda}(\varepsilon)\mathbb{E}\int_0^T \|\bar{Z}_\varepsilon(s)\|^{2(\alpha-\lambda)}\right).
\end{aligned}$$

The last inequality is because $\lambda \leqslant \alpha/2$ implies $2\lambda \leqslant 2(\alpha-\lambda)$ (and the fact $\sqrt{x} \leqslant 1+x$). Similarly,

$$\mathbb{E}\int_0^T (1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda} 1_{\{\|\bar{Z}_\varepsilon(s)\|\leqslant \bar{B}\}} ds \leqslant \mathcal{C}^5(T)\left(1+\varsigma^{2\lambda}(\varepsilon)\mathbb{E}\int_0^T \|\bar{Z}_\varepsilon(s)\|^{2(\alpha-\lambda)}\right)$$

$$\mathbb{E}\int_0^T (1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\bar{\alpha}} 1_{\{\|\bar{Z}_\varepsilon(s)\|\leqslant \bar{B}\}} ds \leqslant \mathcal{C}^6(T)\left(1+\varsigma^{2\lambda}(\varepsilon)\mathbb{E}\int_0^T \|\bar{Z}_\varepsilon(s)\|^{2(\alpha-\lambda)}\right).$$

By (a) Hölder continuity of b, (b) Corollary 4.3-(iii), and (c) the fact $\alpha-2\lambda+\nu_b(\bar{\alpha}\vee\lambda) \leqslant 2(\alpha-\lambda)$,

$$\begin{aligned}\mathbb{E}(A_1(T)) &\leqslant L_b\mathbb{E}\int_0^T \|\bar{Z}_\varepsilon(s)\|^{\alpha-2\lambda}\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))-\bar{Z}_\varepsilon(s)\|^{\nu_b} 1_{\{\|\bar{Z}_\varepsilon(s)\|>\bar{B}\}} ds \\ &\leqslant \mathcal{C}^7\varsigma^{\nu_b}(\varepsilon)\int_0^T \mathbb{E}\left[1+\|\bar{Z}_\varepsilon(s)\|\right]^{\alpha-2\lambda+\nu_b(\bar{\alpha}\vee\lambda))} ds \;\leqslant\; \mathcal{C}^8\varsigma^{\nu_b}(\varepsilon)\int_0^T \mathbb{E}\left[1+\|\bar{Z}_\varepsilon(s)\|\right]^{2(\alpha-\lambda)} ds.\end{aligned}$$

We now estimate $A_2(t) \doteq \int_0^t \|\bar{Z}_\varepsilon(s)\|^{\alpha-2\lambda}(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\lambda}\|\psi(s)\|ds$. By Caucy-Schwarz inequality and Corollary 4.3-(ii)

$$\begin{aligned}\mathbb{E}[A_2(T)] &\leqslant \mathcal{C}^9\left(\mathbb{E}\int_0^T \|\bar{Z}_\varepsilon(s)\|^{2(\alpha-2\lambda)}(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda}\right)^{1/2}\left(\mathbb{E}\int_0^T \|\psi(s)\|^2\, ds\right)^{1/2} \\ &\leqslant \mathcal{C}^{10}\left(\mathbb{E}\int_0^T (1+\|\bar{Z}_\varepsilon(s)\|)^{2(\alpha-\lambda)}ds\right)^{1/2}.\end{aligned}$$

Again, by Corollary 4.3-(ii)

$$\mathbb{E}\left[\int_0^T (1+\|\bar{Z}_\varepsilon(s)\|)^{\alpha-2\lambda-1}(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda} 1_{\{\|\bar{Z}_\varepsilon(s)\|>\bar{B}\}} ds\right] \leqslant \mathcal{C}^7\mathbb{E}\left[\int_0^T (1+\|\bar{Z}_\varepsilon(s)\|)^{\alpha-1}ds\right].$$

Also notice that for any $\theta>0$

$$\mathbb{E}\left(\|\bar{Z}_\varepsilon((s))\|^{\alpha-1} 1_{\{\|\bar{Z}_\varepsilon(s)\|>\bar{B}\}}\right) \leqslant \frac{1}{\bar{B}^{1-\lambda}}\mathbb{E}\left(\|\bar{Z}_\varepsilon((s))\|^{\alpha-\lambda} 1_{\{\|\bar{Z}_\varepsilon(s)\|>\bar{B}\}}\right) \leqslant \frac{1}{\bar{B}^{1-\lambda}}\left(\theta\mathbb{E}\left(\|\bar{Z}_\varepsilon((s))\|^{2(\alpha-\lambda)}\right)+\theta^{-1}\right).$$

Now, multiplying both sides of (5.2) by ε, it follows that for some constant $\hat{\mathcal{C}}_1(T)$,

$$\begin{aligned}\int_0^T \mathbb{E}\left(\|\bar{Z}_\varepsilon(s)\|^{2(\alpha-\lambda)}\right) ds \leqslant \mathcal{C}^{11}(T)\Bigg[1 &+ h(\varepsilon)\int_0^T \mathbb{E}\left(\|\bar{Z}_\varepsilon((s))\|^{2(\alpha-\lambda)}\right) ds \\ &+ \int_0^T \mathbb{E}\left(\theta\mathbb{E}\left(\|\bar{Z}_\varepsilon((s))\|^{2(\alpha-\lambda)}\right)+\theta^{-1}\right) ds\Bigg],\end{aligned}$$

where $h(\varepsilon) = \varsigma^{2\lambda\wedge\nu_b}(\varepsilon)+\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Choose $\varepsilon_0>0$ such that $\mathcal{C}^{11}(T)h(\varepsilon) \leqslant \frac{1}{4}$ for all $\varepsilon \leqslant \varepsilon_0$, and $\theta>0$ be such that $\theta\hat{\mathcal{C}}_1(T) \leqslant 1/4$. It is now immediate that

$$\frac{1}{2}\int_0^t \mathbb{E}\left(\|\bar{Z}_\varepsilon(s)\|^{2(\alpha-\lambda)}\right) ds \leqslant \mathcal{C}^{11}(T)(1+T/\theta).$$

which proves the assertion. □

Remark 5.2. Under the assumption of Lemma 5.1, it follows from Corollary 4.3-(i) that

$$\sup_{\varepsilon\in(0,\varepsilon_0]}\sup_{\psi\in\mathcal{P}_2^M}\mathbb{E}\left[\int_0^T \|\bar{Z}^{\psi}_\varepsilon(\varrho_\varepsilon(t))\|^{2(\alpha-\lambda)}dt\right] < \infty.$$

Corollary 5.3. *Under the assumptions of Lemma 5.1,*

$$\sup_{0<\varepsilon<\varepsilon_0}\sup_{\psi\in\mathcal{P}_2^M}\varepsilon\mathbb{E}\left[\sup_{r\leqslant t}\|\bar{Z}^{\psi}_\varepsilon(r)\|^{1+\alpha-2\lambda}\right] < \infty.$$

Proof. Multiplying (5.2) by ε and following exactly the same estimates used in Lemma 5.1, we readily see for some constant $\mathcal{C}^{12}(T)$

$$\mathbb{E}\left(\varepsilon\sup_{r\leqslant t}\|\bar{Z}_\varepsilon(r)\|^{1+\alpha-2\lambda}\right) \leqslant \varepsilon\|\vartheta(\|x_0\|^2)\| + \varepsilon\mathbb{E}\left(\sup_{r\leqslant t}\|\bar{\mathcal{M}}_\varepsilon(r)\|\right) + \mathcal{C}^{12}(T)\mathbb{E}\int_0^T (1+\|\bar{Z}_\varepsilon(s)\|)^{2(\alpha-\lambda)}ds$$

By Lemma 5.1, the last quantity is bounded by a constant $\mathcal{C}^{13}(T)$. The assertion follows as it is easy to show by Burkholder's inequality that $\varepsilon\mathbb{E}\left[\sup_{r\leqslant t}|\bar{\mathcal{M}}_\varepsilon(t)|\right]\leqslant \mathcal{C}^{14}(T)$ for some constant $\mathcal{C}^{14}(T)$. $\square$

Lastly, as mentioned before, for technical reasons, we also need to consider a partition $\{\tilde{t}_k\}$ which is coarser than $\{t_k\}$ and have bounds for integral moments of $Z^\varepsilon\circ\eta_\varepsilon$ and $\bar{Z}^\psi_\varepsilon\circ\eta_\varepsilon$ (i.e., for the original and the controlled processes). Here $\eta_\varepsilon(s)$ is the step function corresponding to the partition $\{\tilde{t}_k\}$.

Corollary 5.4. *Let $\{\tilde{t}_k\}$ be a partition of $[0,T]$ such that $\tilde{\Delta}=\tilde{t}_k-\tilde{t}_{k-1}\leqslant\varepsilon$. Then, under the assumptions in Lemma 5.1, for all $M>0$, there exists an $\varepsilon_0>0$ such that*

$$\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{\psi\in\mathcal{P}_2^M}\mathbb{E}\left[\int_0^T\|\bar{Z}^\psi_\varepsilon(\eta_\varepsilon(t))\|^{\mathfrak{p}}dt\right]<\infty,$$

where $\mathfrak{p}=(1+\alpha-2\lambda)\wedge 2(\alpha-\lambda)$ and $\eta_\varepsilon(t)=\tilde{t}_k$ for $\tilde{t}_k\leqslant t<\tilde{t}_{k+1}$.

Proof. The proof essentially reuses the techniques of Lemma 5.1, and we only point out the central ideas. Let $\hat{\vartheta}$ be a $C^\infty([0,\infty),[0,\infty))$ function such that $\vartheta(x)=x^{\mathfrak{p}/2}$ on $(1,\infty)$ and $=0$ on $(0,0.9)$, where $\mathfrak{p}=(1+\alpha-2\lambda)\wedge(\alpha-\lambda)$. Write

$$\begin{aligned}\int_0^t\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|^{\mathfrak{p}}ds &= \int_0^t\Big(\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|^{\mathfrak{p}}1_{\{\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|\leqslant 1\}}+\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|^{\mathfrak{p}}1_{\{\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|\geqslant 1\}}\Big)ds\\ &\leqslant \int_0^t\Big(1+\|\hat{\vartheta}\|_{\infty,1}+\hat{\vartheta}\left(\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|^2\right)-\hat{\vartheta}\left(\|\bar{Z}_\varepsilon(s)\|^2\right)+\|\bar{Z}_\varepsilon(s)\|^{\mathfrak{p}}1_{\{\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|>1\}}\Big)ds,\end{aligned}\tag{5.3}$$

and $\|\hat{\vartheta}\|_{\infty,r}$ denotes the maximum of $\hat{\vartheta}$ on $[0,r]$.

By Itô's Lemma (5.1) (with $\hat{\vartheta}$ in place of ϑ), after splitting each term according to $\{\|\bar{Z}_\varepsilon(s)\|\leqslant\bar{B}\}$ and $\{\|\bar{Z}_\varepsilon(s)\|>\bar{B}\}$ (where $\bar{B}=B\vee 1$ with B as in Condition 2.1-(iii)), we deduce that

$$\begin{aligned}Q &\doteq \hat{\vartheta}(\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|^2)-\hat{\vartheta}(\|\bar{Z}_\varepsilon(s)\|^2)\\ &\leqslant \hat{\mathcal{M}}_\varepsilon(\eta_\varepsilon(s))-\hat{\mathcal{M}}_\varepsilon(s)+\frac{\hat{\mathcal{C}}^1(T)}{\varepsilon}\int_{\eta_\varepsilon(s)}^s\Big((1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda}\\ &\quad+(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\bar{\alpha}}+(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\lambda}\|\psi(s)\|\Big)1_{\{\|\bar{Z}_\varepsilon(s)\|\leqslant\bar{B}\}}ds\\ &\quad+\frac{\hat{\mathcal{C}}^2}{\varepsilon}\int_{\eta_\varepsilon(s)}^s\big(\|\bar{Z}_\varepsilon(s)\|^{\mathfrak{p}-1}(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\bar{\alpha}}+\|\bar{Z}_\varepsilon(s)\|^{\mathfrak{p}-2}(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda}\\ &\quad+\delta(\varepsilon)\|\bar{Z}_\varepsilon(s)\|^{\mathfrak{p}-1}(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{\lambda}\|\psi(s)\|ds\Big)1_{\{\|\bar{Z}_\varepsilon(s)\|>\bar{B}\}}ds,\end{aligned}\tag{5.4}$$

where $\hat{\mathcal{M}}_\varepsilon$ is as in the proof of Lemma 5.1 with ϑ replaced by $\hat{\vartheta}$. Integrating the above display from 0 to T, we get by (a) (4.1) together with the assumption $\tilde{\Delta}\leqslant\varepsilon$ and (b) Corollary 4.3 along with $\mathfrak{p}\leqslant(1+\alpha-2\exp)\wedge 2(\alpha-\lambda)$

$$\int_0^T\mathbb{E}\left(\hat{\vartheta}(\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|^2)-\hat{\vartheta}(\|\bar{Z}_\varepsilon(s)\|^2)\right)ds\leqslant\hat{\mathcal{C}}^3(T)\mathbb{E}\left[\int_0^T(1+\|\bar{Z}_\varepsilon(s)\|)^{2(\alpha-\lambda)}ds\right]\leqslant\hat{\mathcal{C}}^4(T).$$

The last inequality in the above display is due to Lemma 5.1. The assertion now follows from (5.3). $\square$

We now state similar results for the original process Z^ε needed to prove the CLT result. Their proofs use almost exactly same estimation techniques and are actually much simpler due to the absence of the control terms. As such, they will be mostly stated without proofs. The following lemma is the analogue of Lemma 5.1, and the important point to note here is that unlike the controlled version, it holds for any exponent $q>0$.

Lemma 5.5. *Suppose that Z^ε satisfies* (2.3) *and that $\Delta(\varepsilon)/\varepsilon \to 0$. Assume that Condition 2.1 holds. Then for all $q > 0$, there exists a constant ε_0 such that*

$$\sup_{0<\varepsilon\leqslant\varepsilon_0} \mathbb{E}\left[\int_0^T \|Z^\varepsilon(t)\|^q dt\right] < \infty.$$

Proof. Let $p \geqslant 2$. Then from Itô's lemma,

$$\begin{aligned}\|Z^\varepsilon(t)\|^p =& \|x_0\|^p + \bar{\mathcal{M}}_\varepsilon(t) + \frac{p}{\varepsilon}\int_0^t \|Z^\varepsilon(s)\|^{p-2}\langle Z^\varepsilon(s), b(\bar{Z}^\varepsilon(\varrho_\varepsilon(s)))\rangle ds \\ &+ \frac{p}{2\varepsilon}\int_0^t \|Z^\varepsilon(s)\|^{p-2}\|\sigma(Z^\varepsilon(\varrho_\varepsilon(s)))\|^2\, ds + \frac{p(p-2)}{2\varepsilon}\int_0^t \|Z^\varepsilon(s)\|^{p-4}\|Z^\varepsilon(s)\sigma(Z^\varepsilon(\varrho_\varepsilon(s)))\|^2 ds,\end{aligned}$$

where $\mathcal{M}^\varepsilon(t) = \frac{p}{\sqrt{\varepsilon}}\int_0^t \|\bar{Z}_\varepsilon(s)\|^{p-2}\bar{Z}_\varepsilon(s)^T\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))dW(s)$ is a martingale. Splitting the third term according as $\|Z^\varepsilon(s)\| > B$ or not, we have for some constant $\hat{\mathcal{C}}^{01}(T)$

$$\begin{aligned}\|Z^\varepsilon(t)\|^p \leqslant \|x_0\|^p + \mathcal{M}^\varepsilon(t) + \frac{p}{\varepsilon}A^\varepsilon(t) + \frac{\hat{\mathcal{C}}^{01}(T)}{\varepsilon}\left[1 + \int_0^t \|Z^\varepsilon(s)\|^{p+\lambda-2}ds\right] \\ - \frac{p}{\varepsilon}\int_0^t \|Z^\varepsilon(s)\|^{p+\alpha-1}1_{\{\|Z^\varepsilon(s)\|>B\}}ds,\end{aligned} \tag{5.5}$$

where the term $A^\varepsilon(t) \doteq \int_0^t \|Z_\varepsilon(s)\|^{p-2}|\langle Z_\varepsilon(s), b(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)) - b(Z_\varepsilon(s))\rangle|ds$ can be easily estimated by using the Holder continuity of b and Lemma 4.1 as

$$\mathbb{E}|A^\varepsilon(t)| \leqslant \hat{\mathcal{C}}^{02}(T)\left(\frac{\Delta}{\varepsilon}\right)^{\nu_b/2}\mathbb{E}\int_0^t (1 + \|Z^\varepsilon(s)\|)^{p+\alpha-1}\, ds.$$

Now observing that $p + \lambda - 1 \leqslant p + \alpha - 1$, similar steps used in the proof of the Theorem 5.1, show that, choosing $\Delta(\varepsilon)/\varepsilon$ sufficiently small and rearranging terms in (5.5), there exists an $\varepsilon_0 > 0$ such that $\sup_{0<\varepsilon\leqslant\varepsilon_0}\mathbb{E}\left[\int_0^T \|Z^\varepsilon(t)\|^{p+\alpha-1}dt\right] < \infty$. □

The following corollary to Lemma 5.5 follows in the same way as Corollary 5.3.

Corollary 5.6. *Under the assumptions of Lemma 5.5, for any $q > 0$, there exists a constant ε_0 such that* $\sup_{0<\varepsilon<\varepsilon_0}\varepsilon\mathbb{E}\left[\sup_{r\leqslant t}\|Z^\varepsilon(r)\|^q\right] < \infty$.

Remark 5.7. Similar to Corollary 5.4, we have for the following result for the original process Z^ε corresponding to the coarser partition $\{\tilde{t}_k\}$: for any $q > 0$, there exist a constant $\hat{\mathcal{C}}^{03}(T)$ and ε_0 such that $\sup_{0<\varepsilon\leqslant\varepsilon_0}\sup_{\psi\in\mathcal{P}_2^M}\mathbb{E}\left[\int_0^T \|Z^\varepsilon(\eta_\varepsilon(t))\|^q dt\right] \leqslant \hat{\mathcal{C}}^{03}(T)$.

Proposition 5.8. *Let $\{\psi_\varepsilon\}$ be such that $\int_0^T \|\psi_\varepsilon(s)\|^2 ds \leqslant M$ for some constant $M > 0$. Let $\bar{Z}_\varepsilon \equiv \bar{Z}_\varepsilon^{\psi_\varepsilon}$ satisfy* (2.9) *with ψ replaced by ψ_ε, and define the occupation measure $\bar{R}_\varepsilon$ on $\mathbb{B}_T$ by*

$$\bar{R}_\varepsilon([0,t]\times A\times B) = \int_0^t 1_{\{\bar{Z}_\varepsilon(s)\in A\}}1_{\{\psi_\varepsilon(s)\in B\}}ds. \tag{5.6}$$

Assume that

(i) the step size $\Delta(\varepsilon)$ is such that $(\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\sqrt{\varepsilon} \to 0$, as $\varepsilon \to 0$ ($\nu = \nu_b \wedge \nu_\sigma$);

(ii) $f : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^n$ satisfies Assumption 2.3, with $\mathfrak{r}(\Delta) = \sqrt{\Delta}$;

(iii) Condition 2.1 (with $\alpha > \lambda$) and Assumption 2.8 hold.

Then $(\bar{R}_\varepsilon, \bar{\Upsilon}_\varepsilon(f))$ is tight in $\mathcal{M}_1(\mathbb{B}_T)\times C([0,T]:\mathbb{R}^d)$, and any limit point (R,ξ) satisfies (3.1) - (3.3), *where $\bar{\Upsilon}_\varepsilon(f)$ was defined before* (2.9). *Moreover, the same assertion is true for $(\bar{R}_\varepsilon, \Xi_\varepsilon^R(f)/\delta(\varepsilon))$ if Assumption 2.5-(A) or Assumption 2.5-(B) (with $p_0' \leqslant \alpha - 2\lambda$) holds.*

Proof. We start by establishing the tightness of $\bar{R}_\varepsilon$ and toward this end, we need to show that for every $\eta > 0$, there exists a constant C_η such that

$$\sup_\varepsilon \mathbb{E}\bar{R}_\varepsilon \{\mathbf{y} : \|x\| + \|z\| > C_\eta\} \leqslant \eta \tag{5.7}$$

where recall that $\mathbf{y}$ denotes a typical tuple (s, x, z) in $\mathbb{B}_T$. Note that for all $0 < \varepsilon < 1$,

$$\int_0^T \|\psi^\varepsilon(s)\|^2 ds = \int_{\mathbb{B}_T} \|z\|^2 \bar{R}_\varepsilon(d\mathbf{y}) \leqslant M, \tag{5.8}$$

and by Lemma 5.1,

$$\sup_\varepsilon \mathbb{E} \int_{\mathbb{B}_T} \|x\|^{2(\alpha-\lambda)} \bar{R}_\varepsilon(d\mathbf{y}) = \sup_\varepsilon \mathbb{E} \int_0^T \|\bar{Z}_\varepsilon(s)\|^{2(\alpha-\lambda)} ds < \infty. \tag{5.9}$$

(5.7) now follows after an application of Markov inequality.

Let $\{\tilde{t}_k\}_{k=0}^N$ be a partition of $[0, T]$ such that $\tilde{\Delta} = \tilde{t}_k - \tilde{t}_{k-1} = \varepsilon$. Applying Itô-Krylov lemma [19] to each component u_l, we have for $r \in [t_k, t_{k+1}]$,

$$\begin{aligned} u_l(\tilde{t}_k, \bar{Z}_\varepsilon(r)) = u_l(\tilde{t}_k, \bar{Z}_\varepsilon(\tilde{t}_k)) &+ \frac{1}{\varepsilon}\left(\int_{\tilde{t}_k}^r \nabla^T u_l(\tilde{t}_k, \bar{Z}_\varepsilon(s)) b(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))ds \right. \\ &\left. + \frac{1}{2}\int_{\tilde{t}_k}^r tr\left(D^2 u_l(\tilde{t}_k, \bar{Z}_\varepsilon(s)) a(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\right) ds\right) \\ &+ \frac{\delta(\varepsilon)}{\varepsilon}\int_{\tilde{t}_k}^r \nabla^T u_l(\tilde{t}_k, \bar{Z}_\varepsilon(s)) \sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\psi_\varepsilon(s) ds \\ &+ \frac{1}{\varepsilon^{1/2}}\int_{\tilde{t}_k}^r \nabla^T u_l(\tilde{t}_k, \bar{Z}_\varepsilon(s)) \sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))dW(s). \end{aligned}$$

Let $k_0 = \max\{k : \tilde{t}_k < t\}$ and without loss of generality assume that $\tilde{t}_{k_0+1} = t$. Summing over k, we can write $\tilde{\mathcal{A}}_0 \doteq \sum_{k=0}^{k_0} \left(u_l(\tilde{t}_k, \bar{Z}_\varepsilon(\tilde{t}_{k+1})) - u_l(\tilde{t}_k, \bar{Z}_\varepsilon(\tilde{t}_k))\right)$ as

$$\begin{aligned} \tilde{\mathcal{A}}_0 = \frac{1}{\varepsilon}\int_0^t \mathcal{L}u(\eta_\varepsilon(s), \cdot)(\bar{Z}_\varepsilon(s))ds + \delta(\varepsilon)\mathcal{E}_0^\varepsilon(t)/\varepsilon + \frac{\delta(\varepsilon)}{\varepsilon}\int_0^t \nabla^T u_l(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\psi_\varepsilon(s)ds \\ + \frac{1}{\varepsilon^{1/2}}\int_0^t \nabla^T u_l(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))dW(s), \end{aligned} \tag{5.10}$$

where, as before, $\eta_\varepsilon(s) = \tilde{t}_k$ if $\tilde{t}_k < s \leqslant \tilde{t}_{k+1}$, and

$$\begin{aligned} \mathcal{E}_{0,l}^\varepsilon(t) = \frac{1}{\delta(\varepsilon)}\left(\int_0^t \left[\nabla^T u_l(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s)) b(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))ds + \frac{1}{2}tr\left(D^2 u_l(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s)) a(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\right)\right] ds \right. \\ \left. - \int_0^t \mathcal{L}u_l(\eta_\varepsilon(s), \cdot)(\bar{Z}_\varepsilon(s))ds\right). \end{aligned}$$

Therefore from (5.10), $\mathcal{U}_t^\varepsilon \doteq \frac{\varepsilon}{\delta(\varepsilon)}\left(u(t, \bar{Z}_\varepsilon(t)) - u(0, x_0)\right)$ can be written as

$$\begin{aligned} \mathcal{U}_t^\varepsilon =& \frac{\varepsilon}{\delta(\varepsilon)}\sum_{k=0}^{k_0}\left(u(\tilde{t}_{k+1}, \bar{Z}_\varepsilon(\tilde{t}_{k+1})) - u(\tilde{t}_k, \bar{Z}_\varepsilon(\tilde{t}_{k+1}))\right) + \frac{\varepsilon}{\delta(\varepsilon)}\sum_{k=0}^{k_0}\left(u(\tilde{t}_k, \bar{Z}_\varepsilon(\tilde{t}_{k+1})) - u(\tilde{t}_k, \bar{Z}_\varepsilon(\tilde{t}_k))\right) \\ =& \mathcal{E}_0^\varepsilon(t) + \mathcal{E}_1^\varepsilon(t) - \frac{1}{\delta(\varepsilon)}\int_0^t f(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))ds + \int_0^t Du(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(s))\psi_\varepsilon(s)ds \\ &+ \frac{\sqrt{\varepsilon}}{\delta(\varepsilon)}\int_0^t Du(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))dW(s) \\ =& -\bar{\Upsilon}_\varepsilon(f)(s) + \int_{\mathbb{B}_t} Du(s, x)\sigma(x)z\bar{R}_\varepsilon(d\mathbf{y}) + \frac{\sqrt{\varepsilon}}{\delta(\varepsilon)}\int_0^t Du(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))dW(s) \\ &+ \mathcal{E}_0^\varepsilon(t) + \mathcal{E}_1^\varepsilon(t) + \mathcal{E}_2^\varepsilon(t) + \mathcal{E}_3^\varepsilon(t), \end{aligned} \tag{5.11}$$

where the quantities $\mathcal{E}_i^\varepsilon(t)$ are defined below:

$$\mathcal{E}_1^\varepsilon(t) \doteq \frac{\varepsilon}{\delta(\varepsilon)} \sum_k \left(u(\tilde{t}_{k+1}, \bar{Z}_\varepsilon(\tilde{t}_{k+1})) - u(\tilde{t}_k, \bar{Z}_\varepsilon(\tilde{t}_{k+1}))\right);$$

$$\mathcal{E}_2^\varepsilon(t) \doteq \frac{1}{\delta(\varepsilon)} \left(\int_0^t f(s, \bar{Z}_\varepsilon(s))ds - \int_0^t f(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\right) ds;$$

$$\mathcal{E}_3^\varepsilon(t) \doteq \int_0^t Du(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\psi_\varepsilon(s)ds - \int_0^t Du(s, \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(s))\psi_\varepsilon(s)ds,$$

and $\mathcal{E}_0^\varepsilon$ is of course given by $\mathcal{E}_0^\varepsilon = (\mathcal{E}_{0,1}^\varepsilon, \dots, \mathcal{E}_{0,n}^\varepsilon)^T$.

Recalling that $\tilde{\Delta} = \varepsilon$, it follows from Assumption 2.3-(ii), the fact that $q_0 \leqslant 2(\alpha - \lambda)$ (Assumption 2.8-(iii)), and Lemma 5.1 that

$$\mathbb{E}\left(\sup_{s \leqslant t} \|\mathcal{E}_2^\varepsilon(s)\|\right) \leqslant \bar{\mathfrak{C}}^1(T) \frac{\sqrt{\varepsilon}}{\delta(\varepsilon)} \left(\int_0^t (1 + \mathbb{E}(\|\bar{Z}_\varepsilon(s)\|^{q_0}) ds\right) \to 0, \tag{5.12}$$

as $\varepsilon \to 0$. Next, by Assumption 2.3-(iii) and Hölder continuity of σ, we have for some constant $\tilde{\mathfrak{C}}^1(T)$

$$\|\mathcal{E}_3^\varepsilon(s)\| \leqslant \bar{\mathfrak{C}}^2(T) \left[\sqrt{\varepsilon} \int_0^t (1 + \|\bar{Z}_\varepsilon(s)\|)^{q_2+\lambda} \|\psi_\varepsilon(s)\| ds + \int_0^t (1 + \|\bar{Z}_\varepsilon(s)\|)^{p_2} \|\bar{Z}_\varepsilon(s) - \bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\nu_\sigma} \|\psi_\varepsilon(s)\| ds\right]$$
$$\equiv \mathcal{E}_{3,1}^\varepsilon(t) + \mathcal{E}_{3,2}^\varepsilon(t).$$

By (a) the assumption that $q_2 \leqslant \alpha - 2\lambda$ (which implies $q_2 + \lambda \leqslant \alpha - \lambda$) (Assumption 2.8-(iii)), (b) Lemma 5.1, and (c) Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}\left(\sup_{s \leqslant t} |\mathcal{E}_{3,1}^\varepsilon(s)|\right) &\leqslant \bar{\mathfrak{C}}^3(T)\sqrt{\varepsilon} \left(\int_0^t \mathbb{E}(1 + \|\bar{Z}_\varepsilon(s)\|)^{2(q_2+\lambda)} ds \; \mathbb{E}\int_0^t \|\psi_\varepsilon(s)\|^2 ds\right)^{1/2} \\
&\leqslant \bar{\mathfrak{C}}^3(T) M^{1/2} \sup_\varepsilon \left(\int_0^t \mathbb{E}\left(1 + \|\bar{Z}_\varepsilon(s)\|\right)^{2(q_2+\lambda)} ds\right)^{1/2} \sqrt{\varepsilon} \; \to \; 0,
\end{aligned}$$

as $\varepsilon \to 0$. Also, since $\nu_\sigma \leqslant 1 - \lambda/\alpha$

$$\begin{aligned}
\mathbb{E}\left(\sup_{s \leqslant t} |\mathcal{E}_{3,2}^\varepsilon(s)|\right) &\leqslant \bar{\mathfrak{C}}^4(T)\mathbb{E}\left(\int_0^t (1 + \|\bar{Z}_\varepsilon(s)\|)^{p_2} \|\bar{Z}_\varepsilon(s) - \bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\nu_\sigma} \|\psi_\varepsilon(s)\| ds\right) \\
&\leqslant \bar{\mathfrak{C}}^4(T)\mathbb{E}\left(\sup_{s \leqslant T}(1 + \|\bar{Z}_\varepsilon(s)\|)^{p_2} \int_0^t \|\bar{Z}_\varepsilon(s) - \bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\nu_\sigma} \|\psi_\varepsilon(s)\| ds\right) \\
&\leqslant \bar{\mathfrak{C}}^4(T) L_\sigma \mathbb{E}\left[\sup_{s \leqslant T}(1 + \|\bar{Z}_\varepsilon(s)\|)^{p_2} \left(\int_0^t \|\bar{Z}_\varepsilon(s) - \bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{2\nu_\sigma}\right)^{1/2} \left(\int_0^t \|\psi_\varepsilon(s)\|^2 ds\right)^{1/2}\right] \\
&\leqslant \bar{\mathfrak{C}}^4(T) L_\sigma M^{1/2} \mathbb{E}\left[\sup_{s \leqslant T}(1 + \|\bar{Z}_\varepsilon(s)\|)^{p_2} \left(\int_0^t \|\bar{Z}_\varepsilon(s) - \bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{2\nu_\sigma} ds\right)^{1/2}\right] \\
&\leqslant \bar{\mathfrak{C}}^4(T) L_\sigma M^{1/2} \left[\mathbb{E}\left(\sup_{s \leqslant T}(1 + \|\bar{Z}_\varepsilon(s)\|)^{2p_2}\right)\right]^{1/2} \left[\int_0^t \mathbb{E}\|\bar{Z}_\varepsilon(s) - \bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{2\nu_\sigma} ds\right]^{1/2} \\
&\leqslant \bar{\mathfrak{C}}^5(T) \frac{(\Delta(\varepsilon)/\varepsilon)^{\nu_\sigma/2}}{\sqrt{\varepsilon}} \left[\varepsilon \mathbb{E}\left(\sup_{s \leqslant T}(1 + \|\bar{Z}_\varepsilon(s)\|)^{2p_2}\right)\right]^{1/2} \\
&\quad \times \left[\int_0^t (1 + \mathbb{E}\|\bar{Z}_\varepsilon(s)\|^{2(\alpha-\lambda)}) ds\right]^{1/2} \leqslant \; \bar{\mathfrak{C}}^6(T) \frac{(\Delta(\varepsilon)/\varepsilon)^{\nu_\sigma/2}}{\sqrt{\varepsilon}} \overset{\varepsilon \to 0}{\to} 0,
\end{aligned}$$

by the choice of discretization step $\Delta(\varepsilon)$ (see (i) in the hypotheses of the proposition). The last inequality above used the (a) Corollary 5.3 along with the assumption that $p_2 \leqslant (1 + \alpha - 2\lambda)/2$ (Assumption 2.8-(ii)), and (b) Lemma 5.1.

We now consider $\mathcal{E}_1^\varepsilon$. Note that because of Assumption 2.3-(iii)-(c)

$$|\mathcal{E}_1^\varepsilon(t)| \leqslant \frac{\varepsilon}{\delta(\varepsilon)} \sum_{k=0}^{k_0} \mathcal{C}_1(T)\left(1+\|\bar{Z}_\varepsilon(\tilde{t}_{k+1})\|^{q_1}\right)\tilde{\Delta}^{1/2} = \frac{\sqrt{\varepsilon}}{\delta(\varepsilon)} \sum_{k=1}^{k_0+1} \mathcal{C}_1(T)\left(1+\|\bar{Z}_\varepsilon(\tilde{t}_{k})\|^{q_1}\right)\varepsilon$$
$$\leqslant \frac{\sqrt{\varepsilon}}{\delta(\varepsilon)}\mathcal{C}_1(T)\int_0^T\left(1+\|\bar{Z}_\varepsilon(\eta_\varepsilon(s))\|^{q_1}\right)ds,$$

and because of Assumption 2.8-(iii), it follows by Corollary 5.4 that $\mathbb{E}\left(\sup_{s\leqslant T}\|\mathcal{E}_1^\varepsilon(s)\|\right) \overset{\varepsilon\to 0}{\to} 0$.

To estimate $\mathcal{E}_0^\varepsilon$, note that for each l, by Assumption 2.3-(iii)-(b) & (e),

$$\sup_{t\leqslant T}\|\mathcal{E}_{0,l}^\varepsilon(t)\| \leqslant \frac{1}{\delta(\varepsilon)}\left(\int_0^T \|\nabla^T u_l(\eta_\varepsilon(s),\bar{Z}_\varepsilon(s))\|\|b(\bar{Z}_\varepsilon(s))-b(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\|ds\right.$$
$$\left.+\int_0^T \|D^2 u_l(\eta_\varepsilon(s),\bar{Z}_\varepsilon(s))\|\|a(\bar{Z}_\varepsilon(s))-a(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\|ds\right)$$
$$=\frac{1}{\delta(\varepsilon)}\left(\mathcal{E}_{0,l}^{1,\varepsilon}(t)+\mathcal{E}_{0,l}^{2,\varepsilon}(t)\right)$$

for some constant $\bar{C}^1(T)$. Notice that by Corollary 4.3-(iv)

$$\mathbb{E}\left(\sup_{t\leqslant T}\|\mathcal{E}_{0,l}^{1,\varepsilon}(t)\|\right) \leqslant \mathcal{C}_1(T)L_b\mathbb{E}\int_0^T(1+\|\bar{Z}_\varepsilon(s)\|)^{p_2}\|\bar{Z}_\varepsilon(s)-\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\nu_b}ds,$$
$$\leqslant \bar{\mathcal{C}}^7(T)\varsigma^{\nu_b}(\varepsilon)\mathbb{E}\int_0^T(1+\|\bar{Z}_\varepsilon(s)\|)^{p_2+\nu_b(\bar{\alpha}\vee\lambda)}ds \leqslant \bar{\mathcal{C}}^8(T)\varsigma^{\nu_b}(\varepsilon).$$

The last step used (a) the fact that $p_2\leqslant\alpha-2\lambda$ (c.f. Assumption 2.8-(ii)) implies $p_2+\nu_b(\bar{\alpha}\vee\lambda)\leqslant 2(\alpha-\lambda)$ and (b) Lemma 5.1. Again by Corollary 4.3-(iv)

$$\mathbb{E}\left(\sup_{t\leqslant T}\|\mathcal{E}_{0,l}^{2,\varepsilon}(t)\|\right) \leqslant \mathcal{C}_1(T)\mathbb{E}\int_0^T(1+\|\bar{Z}_\varepsilon(s)\|^{p_3})\Big(\|\sigma(\bar{Z}_\varepsilon(s))\|\|\bar{Z}_\varepsilon(s)-\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\nu_\sigma}$$
$$+\|\sigma(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\|\|\bar{Z}_\varepsilon(s)-\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\nu_\sigma}ds\Big),$$
$$\leqslant \bar{\mathcal{C}}^9(T)\left(\mathbb{E}\int_0^T(1+\|\bar{Z}_\varepsilon(s)\|^{p_3+\lambda})\|\|\bar{Z}_\varepsilon(s)-\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\nu_\sigma}\ ds\right.$$
$$\left.+\int_0^T(1+\|\bar{Z}_\varepsilon(s)\|^{p_3})(1+\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\lambda})\|\bar{Z}_\varepsilon(s)-\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{\nu_\sigma}ds\right),$$
$$\leqslant \bar{\mathcal{C}}^{10}(T)\varsigma^{\nu_\sigma}(\varepsilon)\mathbb{E}\int_0^T(1+\|\bar{Z}_\varepsilon(s)\|)^{p_3+\lambda+\nu_\sigma(\bar{\alpha}\vee\lambda)}ds \leqslant \bar{\mathcal{C}}^{11}(T)\varsigma^{\nu_\sigma}(\varepsilon).$$

The last step used (a) the fact that $p_3\leqslant\alpha-2\lambda$ (c.f. Assumption 2.8-(ii)) and $\nu_\sigma\leqslant 1-\lambda/\alpha$ implies $p_3+\lambda+\nu_\sigma(\bar{\alpha}\vee\lambda)\leqslant 2(\alpha-\lambda)$ and (b) Lemma 5.1. It follows from Remark 4.2 that

$$\mathbb{E}\left(\sup_{t\leqslant T}\|\mathcal{E}_{0,l}^\varepsilon(t)\|\right) \leqslant \bar{\mathcal{C}}^{12}(T)\varsigma^{\nu}(\varepsilon)/\delta(\varepsilon) \overset{\varepsilon\to 0}{\to} 0, \quad \nu=\nu_b\wedge\nu_\sigma.$$

We next show that as $\varepsilon\to 0$

$$\frac{\varepsilon}{\delta(\varepsilon)}\mathbb{E}\left[\sup_{t\leqslant T}\|u(t,\bar{Z}_\varepsilon(t))-u(0,x_0)\|\right]\to 0. \tag{5.13}$$

Since $p_1\leqslant(1+\alpha-2\lambda)/2$ (Assumption 2.8-(i)),

$$\frac{\varepsilon}{\delta(\varepsilon)}\mathbb{E}\left(\sup_{s\leqslant t}(1+\|\bar{Z}_\varepsilon(s)\|)^{p_1}\right) \leqslant \bar{\mathcal{C}}^{13}\frac{\varepsilon}{\delta(\varepsilon)}\mathbb{E}\left(\sup_{s\leqslant t}(1+\|\bar{Z}_\varepsilon(s)\|)^{(1+\alpha-2\lambda)/2}\right)$$
$$=\bar{\mathcal{C}}^{13}\frac{\sqrt{\varepsilon}}{\delta(\varepsilon)}\left[\varepsilon\mathbb{E}\left(\sup_{s\leqslant t}(1+\|\bar{Z}_\varepsilon(s)\|)^{1+\alpha-2\lambda}\right)\right]^{1/2}\to 0,$$

as $\varepsilon\to 0$ by Corollary 5.3, and (5.13) follows because of Assumption 2.3-(iii)-(a).

For each $l = 1, 2, \ldots, n$, the martingale terms, $\tilde{\mathcal{M}}^l_\varepsilon(t) \doteq \frac{\sqrt{\varepsilon}}{\delta(\varepsilon)} \int_0^r \nabla^T u_l(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(s))dW(s)$, can be estimated as below by (a) Burkholder-Davis-Gundy inequality, (b) Corollary 4.3, (c) Lemma 5.1 along with the assumption $p_2 < \alpha - 2\lambda$:

$$
\begin{aligned}
\mathbb{E}\left[\sup_{r\leqslant t} |\tilde{\mathcal{M}}^l_\varepsilon(r)|^2\right] &\leqslant \frac{\varepsilon}{\delta^2(\varepsilon)}\mathbb{E}\int_0^t \|\nabla^T u_l(\eta_\varepsilon(s), \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(s))\|^2 ds \\
&\leqslant \mathcal{C}_1^2(T)\bar{\mathcal{B}}^2 \frac{\varepsilon}{\delta^2(\varepsilon)}\mathbb{E}\int_0^t (1 + \|\bar{Z}_\varepsilon(s)\|)^{2p_2}(1 + \|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|)^{2\lambda} ds \\
&\leqslant \bar{\mathcal{C}}^{14}(T)\frac{\varepsilon}{\delta^2(\varepsilon)}\mathbb{E}\int_0^t (1 + \|\bar{Z}_\varepsilon(s)\|)^{2(p_2+\lambda)}\, ds \leqslant\ \bar{\mathcal{C}}^{15}(T)\frac{\varepsilon}{\delta^2(\varepsilon)} \overset{\varepsilon\to 0}{\to} 0.
\end{aligned}
$$

It now follows from (5.11) that for tightness $\bar{\Upsilon}_\varepsilon(f)$ we only need to show tightness of $\bar{\Lambda}_\varepsilon$, where

$$
\bar{\Lambda}_\varepsilon(t) = \int_{\mathbb{B}_t} Du(s,x)\sigma(x)z\bar{R}_\varepsilon(d\mathbf{y}) = \int_0^t Du(s, \bar{Z}_\varepsilon(s))\sigma(\bar{Z}_\varepsilon(s))\psi_\varepsilon(s)ds.
$$

Toward this end, notice that by Assumption 2.3-(iii)-(b) for any $K > 0$,

$$
\begin{aligned}
\|\bar{\Lambda}_\varepsilon(t+h) - \bar{\Lambda}_\varepsilon(t)\| &\leqslant \mathcal{C}_1(T)\bar{\mathcal{B}}\int_t^{t+h} (1 + \|\bar{Z}_\varepsilon(s)\|)^{p_2+\lambda}\|\psi_\varepsilon(s)\|ds \\
&\leqslant \bar{\mathcal{C}}^{16}(T)\left[(1+K)^{p_2+\lambda}\int_t^{t+h}\|\psi_\varepsilon(s)\|ds + \int_t^{t+h}(1 + \|\bar{Z}_\varepsilon(s)\|)^{p_2+\lambda}1_{\{\|\bar{Z}_\varepsilon(s)\|>K\}}\|\psi_\varepsilon(s)\|ds\right] \\
&\leqslant \bar{\mathcal{C}}^{16}(T)\left[(1+K)^{p_2+\lambda}M^{1/2}h^{1/2} + \frac{1}{(1+K)^{\alpha-2\lambda-p_2}}\int_t^{t+h}(1 + \|\bar{Z}_\varepsilon(s)\|)^{\alpha-\lambda}\|\psi_\varepsilon(s)\|ds\right] \\
&\leqslant \bar{\mathcal{C}}^{16}(T)\left[(1+K)^{p_2}M^{1/2}h^{1/2} + \frac{1}{(1+K)^{\alpha-2\lambda-p_2}}\left(\int_0^T(1 + \|\bar{Z}_\varepsilon(s)\|)^{2(\alpha-\lambda)}ds + M\right)\right],
\end{aligned}
$$

where $\bar{\mathcal{C}}^{16}(T)$ is a constant independent of K. Taking $(1 + K) = h^{-1/4p_2}$, and using Lemma 5.1 we have that for some constant $\bar{\mathcal{C}}^{17}(T)$,

$$
\mathbb{E}\left[\sup_{0\leqslant t\leqslant t+h\leqslant T}\|\bar{\Lambda}_\varepsilon(t+h) - \bar{\Lambda}_\varepsilon(t)\|\right] \leqslant \bar{\mathcal{C}}^{17}(T)\left(h^{1/4} + h^{(\alpha-2\lambda-p_2)/4p_2}\right).
$$

Recalling that $p_2 < \alpha - 2\lambda$, tightness of $\bar{\Lambda}_\varepsilon$ is now immediate. Here, of course, we assumed $p_2 > 0$. The argument for $p_2 = 0$ (that is, when Du is bounded) is much simpler.

Let (R, ξ) be a limit point of $\{(\bar{R}_\varepsilon, \bar{\Upsilon}_\varepsilon(f))\}$ and by Skorohod representation theorem assume without loss of generality that $(\bar{R}_\varepsilon, \bar{\Upsilon}_\varepsilon(f)) \to (R, \xi)$ a.s in $\mathcal{M}_1(\mathbb{B}_T) \times C([0, T] : \mathbb{R}^d)$ as $\varepsilon \to 0$, at least, along some subsequence. Note that (3.1) follows from (5.8), continuity of $z \to \|z\|$ and (generalized) Fatou's lemma. (see [10, Theorem 1.1]).

Next notice that $\|Du(s, x)\sigma(x)z\| = o(\|x\|^{\alpha-\lambda}\|z\|)$ by Assumption 2.3-(iii)-(b), and the fact that $p_2 < \alpha - 2\lambda$ (Assumption 2.8-(ii)). Hence (5.8), (5.9), and an application of Lemma B.4 imply that as $\varepsilon \to 0$,

$$
\int_{\mathbb{B}_t} Du(s,x)\sigma(x)z\bar{R}_\varepsilon(d\mathbf{y}) \to \int_{\mathbb{B}_t} Du(s,x)\sigma(x)zR(d\mathbf{y}).
$$

Thus from (5.11) and the above calculations it follows that (3.2) holds, that is,

$$
\xi(t) = \int_{\mathbb{B}_t} Du(s,x)\sigma(x)zR(d\mathbf{y}).
$$

Finally, for (3.3), let $g \in C_b^2(\mathbb{R}^d, \mathbb{R})$. Then a simpler version of (5.10) with u replaced by g and much easier calculations reveal that

$$
\int_{\mathbb{B}_t} \mathcal{L}g(x)R(d\mathbf{y}) = 0, \quad 0 \leqslant t \leqslant T.
$$

For the result on $\bar{\Xi}^R_\varepsilon(f)$, notice we only need to show that $\tilde{\mathcal{E}}^\varepsilon_2$, defined by

$$\tilde{\mathcal{E}}^\varepsilon_2(t) \doteq \frac{1}{\delta(\varepsilon)}\left(\int_0^t f(\eta_\varepsilon(s),\bar{Z}_\varepsilon(s))ds - \int_0^t f(\varrho_\varepsilon(s),\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))ds\right) \stackrel{\varepsilon\to 0}{\longrightarrow} 0.$$

We work only under Assumption 2.5-B. The steps under Assumption 2.5-A are simpler. Writing

$$\begin{aligned}\tilde{\mathcal{E}}^\varepsilon_2(t) = & \frac{1}{\delta(\varepsilon)}\int_0^t \big(f(\eta_\varepsilon(s),\bar{Z}_\varepsilon(s)) - f(\varrho_\varepsilon(s),\bar{Z}_\varepsilon(s))\big)\,ds - \frac{1}{\delta(\varepsilon)}\int_0^t \big(f(\varrho_\varepsilon(s),\bar{Z}_\varepsilon(s)) - f(\varrho_\varepsilon(s),\bar{Z}_\varepsilon(\varrho_\varepsilon(s)))\big)\,ds \\ \doteq & \tilde{\mathcal{E}}^\varepsilon_{2,1}(t) + \tilde{\mathcal{E}}^\varepsilon_{2,2}(t),\end{aligned}$$

it is immediate that (c.f. (5.12)) $\mathbb{E}\left[\sup_{t\leqslant T}|\tilde{\mathcal{E}}^\varepsilon_{2,1}(t)|\right] \to 0$, as $\varepsilon\to 0$.

Next, for each $l=1,\dots,n$, by the mean value theorem,

$$\tilde{\mathcal{E}}^\varepsilon_{2,2,l}(t) = \frac{1}{\delta(\varepsilon)}\int_0^t \nabla f_l\left(\eta_\varepsilon(\varrho_\varepsilon(s)),\theta_l(s)\bar{Z}_\varepsilon(\varrho_\varepsilon(s)) + (1-\theta_l(s))\bar{Z}_\varepsilon(s)\right)\left(\bar{Z}_\varepsilon(\varrho_\varepsilon(s)) - \bar{Z}_\varepsilon(s)\right)ds$$

for some $\theta_l(s)\in(0,1)$. Thus by Assumption 2.5-B, and Corollary 4.3,

$$\begin{aligned}\sup_{s\leqslant T}|\tilde{\mathcal{E}}^\varepsilon_{2,2,l}(t)| \leqslant & \frac{\bar{\mathfrak{C}}^{18}(T)}{\delta(\varepsilon)}\int_0^t \left(\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s))\|^{p_0'} + \|\bar{Z}_\varepsilon(s)\|^{p_0'}\right)\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s)) - \bar{Z}_\varepsilon(s)\|ds \\ \leqslant & \frac{\bar{\mathfrak{C}}^{19}(T)}{\delta(\varepsilon)}\int_0^t \left(\|\bar{Z}_\varepsilon(s)\|^{p_0'}\|\bar{Z}_\varepsilon(\varrho_\varepsilon(s)) - \bar{Z}_\varepsilon(s)\| + \|\bar{Z}_\varepsilon(\varrho_\varepsilon(s)) - \bar{Z}_\varepsilon(s)\|^{p_0'+1}\right)ds \\ \leqslant & \bar{\mathfrak{C}}^{20}(T)\frac{\varsigma(\varepsilon)}{\delta(\varepsilon)}\mathbb{E}\int_0^T \left(1+\|\bar{Z}_\varepsilon(s)\|\right)^{p_0'+\bar{\alpha}\vee\lambda}ds \leqslant \bar{\mathfrak{C}}^{21}(T)\frac{\varsigma(\varepsilon)}{\delta(\varepsilon)} \stackrel{\varepsilon\to 0}{\to} 0.\end{aligned}$$

The last step used Lemma 5.1 along with the fact that $p_0' + \bar{\alpha}\vee\lambda \leqslant 2(\alpha-\lambda)$, and Remark 4.2. □

Remark 5.9. Analogous to Lemma 5.5, we have (by much simpler methods) that

$$\sup_{0<\varepsilon\leqslant\varepsilon_0}\mathbb{E}\left[\int_0^T \|X^\varepsilon(t)\|^q dt\right] \equiv \sup_{0<\varepsilon\leqslant\varepsilon_0}\mathbb{E}\left[\int_{\mathbb{R}^d\times[0,T]} \|x\|^q \Gamma_\varepsilon(dx\times ds)\right] =< \infty$$

under (i), (iii) and (iv) of Condition 2.1. Now again by much simpler calculations than that used in proof of Proposition 5.8, any limit point Γ satisfies

$$\int_{\mathbb{R}^d\times[0,t]} \mathcal{L}g(x)\Gamma(dx\times ds) = 0, \quad 0\leqslant t\leqslant T.$$

Writing $\Gamma(dx\times ds) = \Gamma_{2|1}(dx|s)ds$, it follows from the uniqueness of the invariant measure π (c.f. beginning of Section 2.1) that $\Gamma(dx\times ds) = \pi(dx)ds$. Hence, it follows from continuity of $x\to\|x\|$, (generalized) Fatou's lemma (see [10, Theorem 1.1]) , that for any $q\geqslant 0$, $\int_{\mathbb{R}^d}\|x\|^q\pi(dx) < \infty$.

5.1 Proof of Theorem 2.7

Notice because of Lemma 5.5, Ξ^ε is tight and as in the above remark

$$\Xi^\varepsilon \Rightarrow \Xi, \quad \text{where } \Xi(dx\times ds) = \pi(dx)ds. \tag{5.14}$$

Now, again using the coarser partition $\{\tilde{t}_k\}$ with $\tilde{\Delta}(\varepsilon) = t_k - t_{k-1} = \varepsilon$, similar to (5.11), for the original process Z^ε, we can write $\mathcal{U}_0 \doteq \sqrt{\varepsilon}\,(u(t,Z^\varepsilon(t)) - u(0,x_0))$ as

$$\begin{aligned}\mathcal{U}_0 = & \sqrt{\varepsilon}\sum_{k=0}^{k_0}\left(u(\tilde{t}_{k+1},Z^\varepsilon(\tilde{t}_{k+1})) - u(\tilde{t}_k,Z^\varepsilon(\tilde{t}_{k+1}))\right) + \sqrt{\varepsilon}\sum_{k=0}^{k_0}\left(u(\tilde{t}_k,Z^\varepsilon(\tilde{t}_{k+1})) - u(\tilde{t}_k,Z^\varepsilon(\tilde{t}_k))\right) \\ = & \hat{\mathcal{E}}^\varepsilon_0(t) + \hat{\mathcal{E}}^\varepsilon_1(t) - \frac{1}{\sqrt{\varepsilon}}\int_0^t f(\eta_\varepsilon(s),Z^\varepsilon(s))ds + \int_0^t Du(\eta_\varepsilon(s),Z^\varepsilon(s))\sigma(Z^\varepsilon(\varrho_\varepsilon(s)))dW(s) \\ = & -\varepsilon^{-1/2}\Xi_\varepsilon(f)(t) + \hat{\mathfrak{M}}^\varepsilon(t) + \hat{\mathcal{E}}^\varepsilon_0(t) + \hat{\mathcal{E}}^\varepsilon_1(t) + \hat{\mathcal{E}}^\varepsilon_2(t) + \hat{\mathcal{E}}^\varepsilon_3(t),\end{aligned} \tag{5.15}$$

where

$$\hat{\mathcal{E}}^{\varepsilon}_{l,0}(t) = \frac{1}{\sqrt{\varepsilon}}\left(\int_0^t \left[\nabla^T u_l(\eta_\varepsilon(s), Z^\varepsilon(s))b(Z^\varepsilon(\varrho_\varepsilon(s)))ds + \frac{1}{2}tr\left(D^2u_l(\eta_\varepsilon(s), Z^\varepsilon(s))a(Z^\varepsilon(\varrho_\varepsilon(s)))\right)\right]ds\right.$$
$$\left. - \int_0^t \mathcal{L}u_l(\eta_\varepsilon(s),\cdot)(Z^\varepsilon(s))ds\right)$$
$$\hat{\mathcal{E}}^{\varepsilon}_1(t) \doteq \sqrt{\varepsilon}\sum_k \left(u(\tilde{t}_{k+1}, Z^\varepsilon(\tilde{t}_{k+1})) - u(\tilde{t}_k, Z^\varepsilon(\tilde{t}_{k+1}))\right);$$
$$\hat{\mathcal{E}}^{\varepsilon}_2(t) \doteq \frac{1}{\sqrt{\varepsilon}}\left(\int_0^t f(s, Z^\varepsilon(s))ds - \int_0^t f(\eta_\varepsilon(s), Z^\varepsilon(s))\right)ds;$$
$$\hat{\mathcal{E}}^{\varepsilon}_3(t) \doteq \int_0^t Du(\eta_\varepsilon(s), Z^\varepsilon(s)\sigma(Z^\varepsilon(\varrho_\varepsilon(s))) - Du(s, Z^\varepsilon(s))\sigma(Z^\varepsilon(s))dW(s),$$
$$\hat{\mathcal{M}}^\varepsilon(t) \doteq \int_0^t Du(s, Z^\varepsilon(s))\sigma(Z^\varepsilon(s))dW(s).$$

Notice that by Remark 5.7,

$$\mathbb{E}[\sup_{t\leqslant T}\|\hat{\mathcal{E}}^{\varepsilon}_1(t)\|] \leqslant \sqrt{\varepsilon}\mathcal{C}_1(T)\sum_k \mathbb{E}(1+\|Z^\varepsilon(\tilde{t}_{k+1})\|^{q_1})\mathfrak{r}(\varepsilon) \leqslant \frac{\mathfrak{r}(\varepsilon)}{\sqrt{\varepsilon}}\mathcal{C}_1(T)\int_0^T \mathbb{E}(1+\|Z^\varepsilon(\eta_\varepsilon(s))\|^{q_1})ds \xrightarrow{\varepsilon\to 0} 0.$$

Also, since $\mathfrak{r}(\varepsilon)\to 0$, by Lemma 4.4 and Lemma 5.5

$$\mathbb{E}[\sup_{t\leqslant T}\|\hat{\mathcal{E}}^{\varepsilon}_3(t)\|] \leqslant \mathcal{C}_1(T)\bar{\mathcal{B}}\mathfrak{r}(\varepsilon)\int_0^T \mathbb{E}(1+\|Z^\varepsilon(s)\|^{q_2+\lambda})ds$$
$$+ L_b\mathcal{C}_1(T)\int_0^T \mathbb{E}(1+\|Z^\varepsilon(s)\|^{p_2})\|Z^\varepsilon(s) - Z^\varepsilon(\varrho_\varepsilon(s)))\|^{\nu_\sigma}ds \xrightarrow{\varepsilon\to 0} 0.$$

Similarly, it easily follows that $\mathbb{E}[\sup_{s\leqslant T}\|\hat{\mathcal{E}}^{\varepsilon}_2(s)\|]\to 0$, and by similar techniques used in the proof of Proposition 5.8, $(\Delta(\varepsilon)/\varepsilon)^{\nu/2}/\sqrt{\varepsilon}\to 0$ implies that $\mathbb{E}[\sup_{s\leqslant T}\|\hat{\mathcal{E}}^{\varepsilon}_0(s)\|]\to 0$ as $\varepsilon\to 0$.

Moreover, since Corollary 5.6 holds for any q, using Assumption 2.3-(iii)-(a), it could be seen that $\sqrt{\varepsilon}\mathbb{E}\left[\sup_{t\leqslant T}|u(t,Z^\varepsilon(t))|\right]\to 0$, as $\varepsilon\to 0$ (c.f. the proof of (5.13)). For the martingale term we look at its quadratic variation. By (5.14) and Lemma B.4, it follows that as $\varepsilon\to 0$,

$$[\hat{\mathcal{M}}^\varepsilon]_t = \int_0^t Du(s,Z^\varepsilon(s))a(Z^\varepsilon(s))(Du(s,Z^\varepsilon(s)))^T ds = \int_{\mathbb{R}^d\times[0,t]} Du(s,x)a(x)(Du(s,x))^T\Xi^\varepsilon(dx\times ds)$$
$$\to \int_{\mathbb{R}^d\times[0,t]} Du(s,x)a(x)(Du(s,x))^T\pi(dx)ds = \int_0^t M_f(s)ds.$$

The result now follows from the martingale central limit theorem [9, Chapter 7].

Finally, just as in the last part of the proof of Proposition 5.8, and using the same techniques,

$$\varepsilon^{-1/2}\mathbb{E}\left[\sup_{t\leqslant T}\left|\int_0^t \left(f(\eta_\varepsilon(s),Z^\varepsilon(s)) - f(\varrho_\varepsilon(s),Z^\varepsilon(\varrho_\varepsilon(s)))\right)ds\right|\right]\to 0,$$

as $\varepsilon\to 0$, and the assertion for $\Xi^R_\varepsilon(f)$ follows. □

5.2 LDP / Laplace principle upper bound - Theorem 2.9

The objective of this section is to prove the Laplace principle upper bound, that is, to show that

$$\limsup_{\varepsilon\to 0}\beta(\varepsilon)\ln\mathbb{E}\left[\exp\Big(-F(\Upsilon_\varepsilon(f))/\beta(\varepsilon)\Big)\right] \leqslant -\inf_{\xi\in C([0,T],\mathbb{R}^d)}[I_f(\xi)+F(\xi)]. \tag{5.16}$$

Note that (2.8) implies that for every $\varepsilon>0$, there exists a sequence of $\{\psi^\varepsilon\}$ such that

$$-\beta(\varepsilon)\ln\mathbb{E}\left[\exp\Big(-F(\Upsilon_\varepsilon(f))/\beta(\varepsilon)\Big)\right] \geqslant \frac{1}{2}\mathbb{E}\left[\int_0^T\|\psi^\varepsilon(s)\|^2ds + F(\bar{\Upsilon}_\varepsilon(f))\right] - \varepsilon, \tag{5.17}$$

Let $\bar{R}_\varepsilon$ be as in Proposition 5.8. Since F is bounded, by a standard localization argument [3, 6], one can assume without loss of generality that $\sup_{0<\varepsilon<1} \int_0^T \|\psi^\varepsilon(s)\|^2 ds \leqslant M$ for some constant $M > 0$. By Proposition 5.8, $(\bar{R}_\varepsilon, \bar{\Upsilon}_\varepsilon(f))$ is tight and any limit point (R, ξ) satisfies (3.1) - (3.3). Hence $(R, \xi) \in \mathcal{R}_\xi$, where $\mathcal{R}_\xi$ was introduced in Section 3. Assume, without loss of generality, that $(\bar{R}_\varepsilon, \bar{\Upsilon}_\varepsilon(f)) \Rightarrow (R, \xi)$ along the full sequence. It follows from (5.17) and (generalized) Fatou's lemma that

$$\begin{aligned}\liminf_{\varepsilon\to 0} -\beta(\varepsilon) \ln \mathbb{E}\left[\exp\Big(-F(\Upsilon_\varepsilon(f))/\beta(\varepsilon)\Big)\right] &\geqslant \frac{1}{2}\liminf_{\varepsilon\to 0} \mathbb{E}\left[\int_0^T \|\psi^\varepsilon(s)\|^2 ds + F(\bar{\Upsilon}_\varepsilon(f))\right] \\ &\geqslant \mathbb{E}\left[\frac{1}{2}\int_{\mathbb{B}_T} \|z\|^2 R(d\mathbf{y}) + F(\xi)\right] \geqslant I(\xi) + F(\xi),\end{aligned}$$

which proves (5.16). Here we used the equivalent form of the rate function given in Lemma 3.2.

The proof for the Laplace principle upper bound for $\Xi^R_\varepsilon(f)/\delta(\varepsilon)$ follows by the exact same steps. □

6 LDP / Laplace principle lower bound - Theorem 2.9

The goal of this section is to prove the Laplace principle lower bound, which is equivalent to proving the LDP lower bound. Specifically, we will show that

$$\liminf_{\varepsilon\to 0} \beta(\varepsilon) \ln \mathbb{E}\left[\exp\Big(-F(\Upsilon_\varepsilon(f))/\beta(\varepsilon)\Big)\right] \geqslant -\inf_{\xi\in C([0,T],\mathbb{R}^d)} [I_f(\xi) + F(\xi)] \tag{6.18}$$

for a bounded Lipschitz continuous function $F : C([0,T], \mathbb{R}^n) \to \mathbb{R}$, Fix $\kappa > 0$. Let ξ be such that

$$I_f(\xi) + F(\xi) \leqslant \inf_{\xi\in C([0,T],\mathbb{R}^d)} [I_f(\xi) + F(\xi)] + \kappa/2.$$

Recall that by Theorem 3.2, $I_f = \bar{I}_f$. Choose $\phi \in \mathcal{A}_\xi$ such that

$$\frac{1}{2}\int_{\mathbb{R}^d\times[0,T]} \|\phi(x,s)\|^2 \pi(dx)ds + F(\xi) \leqslant I_f(\xi) + F(\xi) + \kappa/2 \leqslant \inf_{\xi\in C([0,T],\mathbb{R}^d)} [I_f(\xi) + F(\xi)] + \kappa.$$

Using the denseness of $C_c^\infty([0,T], \mathbb{R}^d)$ in $L^2(\pi \times \lambda_T)$, find $\phi_\kappa \in C_c^\infty([0,T], \mathbb{R}^d)$ such that $\|\phi_\kappa - \phi\|_2 \leqslant \kappa$. Define ξ^κ by

$$\xi^\kappa(t) = \int_{\mathbb{R}^d\times[0,t]} Du(x,s)\sigma(x)\phi_\kappa(x,s)\pi(dx)ds. \tag{6.19}$$

Notice that by Assumption 2.3-(iii)-(b), there exists a constant $\tilde{\mathcal{C}}_1(T)$ such that

$$\begin{aligned}|\xi(t) - \xi^\kappa(t)| &\leqslant \int_{\mathbb{R}^d\times[0,t]} \|Du(x,s)\|_{op}\|\sigma(x)\|_{op}\|\phi(x,s) - \phi_\kappa(x,s)\|\pi(dx)ds \\ &\leqslant \tilde{\mathcal{C}}_1(T)\int_{\mathbb{R}^d\times[0,t]} (1+\|x\|)^{p_2+\lambda}\|\phi(x,s) - \phi_\kappa(x,s)\|\pi(dx)ds \\ &\leqslant \tilde{\mathcal{C}}_1(T)\left(\int_{\mathbb{R}^d\times[0,t]} (1+\|x\|)^{2(p_2+\lambda)}\pi(dx)ds \int_{\mathbb{R}^d\times[0,t]} \|\phi(x,s) - \phi_\kappa(x,s)\|^2\pi(dx)ds\right)^{1/2} \\ &\leqslant \Theta_2\tilde{\mathcal{C}}_1(T)T^{1/2}\kappa,\end{aligned} \tag{6.20}$$

where $\Theta_2 \equiv \int_{\mathbb{R}^d} (1+\|x\|)^{2(p_2+\lambda)}\pi(dx) < \infty$ (by Remark 2.2).

Let $\bar{Z}^\kappa_\varepsilon$ be the solution to the following SDE

$$\bar{Z}^\kappa_\varepsilon(t) = x_0 + \frac{1}{\varepsilon}\int_0^t b(\bar{Z}^\kappa_\varepsilon(\varrho_\varepsilon(s)))ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t \sigma(\bar{Z}^\kappa_\varepsilon(\varrho_\varepsilon(s)))dW(s) + \frac{\delta(\varepsilon)}{\varepsilon}\int_0^t \sigma(\bar{Z}^\kappa_\varepsilon(\varrho_\varepsilon(s)))\phi_\kappa(\bar{Z}^\kappa_\varepsilon(s), s)ds. \tag{6.21}$$

Since $\phi_\kappa \in C_c^\infty([0,T], \mathbb{R}^d)$ and hence Lipschitz, it readily follows that there exists a unique solution to

(6.21). Define $\psi_\varepsilon(s) = \phi_\kappa(\bar Z^\kappa_\varepsilon(s), s)$ and by the variational representation we have

$$-\beta(\varepsilon)\ln\mathbb{E}\left[\exp\Big(-F(\Upsilon_\varepsilon(f))/\beta(\varepsilon)\Big)\right] \leqslant \mathbb{E}\left[\frac{1}{2}\int_0^T \|\psi^\varepsilon(s)\|^2 ds + F(\bar\Upsilon^\kappa_\varepsilon(f))\right], \tag{6.22}$$

where $\bar\Upsilon^\kappa_\varepsilon(f) = \frac{1}{\delta(\varepsilon)}\int_0^t f(s, \bar Z^\kappa_\varepsilon(s))ds$. Let $\bar\Xi^\kappa_\varepsilon$, defined by

$$\bar\Xi^\kappa_\varepsilon(A\times[0,t]) = \int_0^t 1_{\{\bar Z^\kappa_\varepsilon(s)\in A\}}ds,$$

denote the occupation measure of $\bar Z^\kappa_\varepsilon(s)$ on $\mathbb{R}^d\times[0,T]$. We now study the limit of $\bar\Upsilon^\kappa_\varepsilon(f)$. Since

$$\sup_\varepsilon \mathbb{E}\left[\int_{\mathbb{R}^d\times[0,T]}\|x\|^{2\alpha}\bar\Xi^\kappa_\varepsilon(dx\times ds)\right] = \sup_\varepsilon\mathbb{E}\left[\int_0^T \|\bar Z^\kappa_\varepsilon(s)\|^{2\alpha}ds\right] < \infty,$$

$\bar\Xi^\kappa_\varepsilon$ is tight in $\mathcal{M}_1(\mathbb{R}^d\times[0,T])$. Let Ξ^κ be a limit point of $\bar\Xi^\kappa_\varepsilon$ and assume without loss of generality that $\bar\Xi^\kappa_\varepsilon \to \Xi^\kappa$ as $\varepsilon\to 0$. Now observe that from (5.11) using ϕ_k in place of ϕ

$$\begin{aligned}\frac{\varepsilon}{\delta(\varepsilon)}\big(u(t,\bar X^\kappa_\varepsilon(t)) - u(0,x_0)\big) = & -\bar\Upsilon^\kappa_\varepsilon(f)(s) + \int_{\mathbb{R}^d\times[0,t]} Du(x,s)\sigma(x)\phi_\kappa(x,s)\bar\Xi^\kappa_\varepsilon(dx\times ds)\\ & + \frac{\sqrt{\varepsilon}}{\delta(\varepsilon)}\int_0^t Du(\eta_\varepsilon(s), \bar Z^\kappa_\varepsilon(s))\sigma(\bar Z^\kappa_\varepsilon(\varrho_\varepsilon(s)))dW(s)\\ & + \mathcal{E}^\varepsilon_0(t) + \mathcal{E}^\varepsilon_1(t) + \mathcal{E}^\varepsilon_2(t) + \mathcal{E}^\varepsilon_3(t),\end{aligned}\tag{6.23}$$

where $\mathcal{E}^\varepsilon_j$ are defined analogously. Thus invoking the same calculations in the proof of Proposition 5.8, $\mathbb{E}(\sup_{s\leqslant t}\|\mathcal{E}^\varepsilon_j(s)\|)\to 0,\ j=0,\dots,3$ and

$$\mathbb{E}\left[\sup_{r\leqslant t}|\frac{\sqrt{\varepsilon}}{\delta(\varepsilon)}\int_0^r \nabla^T u_l(\eta_\varepsilon(s), \bar Z^\kappa_\varepsilon(s))\sigma(\bar Z^\kappa_\varepsilon(\varrho_\varepsilon(s)))dW(s)|^2\right]\to 0,$$

as $\varepsilon\to 0$. Next by Lemma B.4

$$\int_{\mathbb{R}^d\times[0,t]} Du(x,s)\sigma(x)\phi_\kappa(x,s)\bar\Xi^\kappa_\varepsilon(dx\times ds) \overset{\varepsilon\to 0}{\longrightarrow} \int_{\mathbb{R}^d\times[0,t]} Du(x,s)\sigma(x)\phi_\kappa(x,s)\Xi^\kappa(dx\times ds).$$

Consequently, it follows that $\bar\Upsilon^\kappa_\varepsilon(f) \to \int_{\mathbb{R}^d\times[0,\cdot]} Du(x,s)\sigma(x)\phi_\kappa(x,s)\Xi^\kappa(dx\times ds)$. Now just as in the proof of Proposition 5.8, much easier calculation shows that for any $g\in C^2_b(\mathbb{R}^d)$, $\int_{\mathbb{R}^d\times[0,t]}\mathcal{L}g(x)\Xi^\kappa(dx\times ds) = 0$ for all $t\in[0,T]$. Writing $\Xi^\kappa(dx\times ds) = \gamma^\kappa_s(dx)ds$, we have by the uniqueness of the invariant distribution of π, $\Xi^\kappa(dx\times ds) = \pi(dx)ds$. Thus $\bar\Upsilon^\kappa_\varepsilon(f)\to\xi^\kappa$, where ξ^κ is defined by (6.19).

Next we observe that since $(x,s)\to\phi_\kappa(x,s)$ is continuous and bounded, and $\Xi^\kappa_\varepsilon \Rightarrow \Xi^\kappa$, where $\Xi^\kappa(dx\times ds) = \pi(dx)ds$,

$$\int_0^T \phi_\kappa(\bar Z^\kappa_\varepsilon(s), s)ds = \int_{\mathbb{R}^d\times[0,T]}\phi_\kappa(x,s)\Xi^\kappa_\varepsilon(dx\times ds) \to \int_{\mathbb{R}^d\times[0,T]}\phi_\kappa(x,s)\pi(dx)ds.$$

Now taking limits in (6.22), we have that $\mathcal{A} \doteq -\beta(\varepsilon)\ln\mathbb{E}\left[\exp\Big(-F(\Upsilon_\varepsilon(f))/\beta(\varepsilon)\Big)\right]$ satisfies

$$\begin{aligned}\mathcal{A} &\leqslant \limsup_{\varepsilon\to 0}\mathbb{E}\left[\frac{1}{2}\int_0^T \|\phi_\kappa(\bar Z^\kappa_\varepsilon(s), s)\|^2 ds + F(\bar\Upsilon^\kappa_\varepsilon(f))\right] = \frac{1}{2}\int_0^T\|\phi_\kappa(x,s)\|^2\pi(dx)ds + F(\xi^\kappa)\\ &\leqslant \frac{1}{2}\int_0^T \|\phi(x,s)\|^2\pi(dx)ds + F(\xi) + \|\phi-\phi_\kappa\|_2^2 + L^F_{lip}\|\xi-\xi_\kappa\|_T\\ &\leqslant \frac{1}{2}\int_0^T\|\phi(x,s)\|^2\pi(dx)ds + F(\xi) + \kappa^2 + L^F_{lip}\Theta_2\tilde{\mathfrak{C}}_1(T)T^{1/2}\kappa\\ &\leqslant \inf_{\xi\in C([0,T],\mathbb{R}^d)}[I_f(\xi)+F(\xi)] + \kappa + \kappa^2 + L^F_{lip}\Theta_2\tilde{\mathfrak{C}}_1(T)T^{1/2}\kappa.\end{aligned}$$

Here L^F_{lip} is the Lipschitz constant of F, and the fourth step used (6.20). Sending $\kappa\to 0$, we have (6.18). Again, the proof for the Laplace principle lower bound for $\Xi^R_\varepsilon(f)/\delta(\varepsilon)$ follows similarly.

Appendix

A Poisson equation

The goal of this section is to characterize (at least partially) the set of functions $f = (f_1, f_2, \dots, f_n) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^n$ satisfying Assumption 2.3. As discussed in Remark 2.4, this requires us to study the solution of the Poisson equation for each $t > 0$:

$$\mathcal{L}u(t,\cdot)(x) = -f(t,x), \tag{A.1}$$

The equation above is component-wise, and so without loss of generality, *we will simply assume* $n = 1$. A solution u, when it exists, is given by

$$u(t,x) = \int_0^\infty P_s f(t,\cdot)(x) ds = \int_0^\infty \int_{\mathbb{R}^d} f(t,y) P_s(x,dy) ds, \tag{A.2}$$

Notice since t is fixed, it is just playing the role of a parameter here. This section summarizes the work of Pardoux and Veretennikov [29] on existence and regularity of the solution u under the assumption $a = \sigma\sigma^T$ is uniformly elliptic and bounded:

- there exist strictly positive constants c_1 and c_2 such that for all $x, y \in \mathbb{R}^d$

$$0 < c_1 \leqslant y^T \left(\sigma(x)\sigma^T(x)\right) y/\|y\|^2 \leqslant c_2. \tag{A.3}$$

From [29, Theorems 1, 2], under Condition 2.1-(i) and (A.3), if f satisfies Assumption 2.3-(i), then for each $t > 0$ (which is just playing the role of a parameter here), (A.1) admits a unique solution $u(t,\cdot)$ in the class of functions belonging to $W^{2,p}_{loc}$ for any $p > 1$. $u(t,\cdot)$. Note that by choosing $p > d$ and using Sobolev embedding theorem [11, Section 7.7], it follows that for each $t > 0$, $Du(t,\cdot)$ is continuous. Moreover, if we assume that the coefficients b and a are C^1, f is (weakly) differentiable and $\sup_{t \leqslant T} \|Df(t,x)\| \leqslant \mathcal{C}(T)(1 + \|x\|)^{p_0'}$ for some $p_0' \in \mathbb{R}$, then by [11, Theorem 9.19], it follows that $f \in W^{3,p}_{loc}$ for all $p > 1$. As before, choosing $p > d$ and using Sobolev embedding theorem, it now follows that $D^2 u(t,\cdot)$ is continuous.

Remark A.1. As mentioned in the introduction, we do note that the proof of the estimate on the growth rate of Du, [29, Theorem 2, eq. (21)], requires the drift b to be globally bounded. This is not explicitly mentioned in [29], where b is said to be locally bounded (although in the statement of Theorem 1 of [29], it did mention once that the constant depends on $\sup_{i,x} |b_i(x)|$). To see why this is indeed the case, first observe that the proof uses the result on interior L^p-estimates of solutions of the elliptic equation from Gilbarg and Trudinger [11, Theorem 9.1]. However, the constant in this result depends on the bounds of the coefficients, b and a, in the domain of interest, Ω. The coefficient a is assumed to be bounded, but the drift term b in most examples will not be. More specifically, since the domain $\Omega = B(x,1)$ in the part (e) of proof of [11, Theorem 9.1], the constant C in [11, Eq. (9.4)], and hence the constant C' in the first display of [29, Page 1070] will actually depend on x. For example, for Ornstein-Uhlenbeck SDE, where $b(x) \sim -x$, it is not hard to see following the chain of arguments leading to [11, Eq. (9.4)] that this particular $C \sim x^2$. This affects the growth rate of the gradient of the solution u in [29, Theorem 2, eq. (21)].

The statement as stated in [29, Theorem 2, eq. (21)] might still be true for more general b, but unfortunately, we cannot find a way to adapt the proof given by Pardoux and Veretennikov or find an alternate proof – except in one-dimension. For one-dimensional SDEs, the original statement of [29] (at least, a very similar one) is indeed true, and we were able to find an alternate way to prove it. For multi-dimensional SDEs, a closer inspection of the proof of [11, Theorem 9.1], shows that a modified statement with a different growth rate of Du holds (c.f. Proposition A.2 below). We claim no originality of its proof, and we just kept track of certain constants in the original proof of [29] to arrive at the correct exponents for u and its derivatives.

Proposition A.2. *Suppose that Condition 2.1 and Assumption 2.3-(i) hold, and that $a = \sigma\sigma^T$ satisfies* (A.3). *Then $u \in C^1(\mathbb{R}^d, \mathbb{R})$, and (iii)-(a) and (iii)-(b) of Assumption 2.3 hold, with the following relations between the exponents:*

$$p_1 = (p_0 - \alpha + 1)^+, \quad p_2 = \max\{p_1 + 2\bar{\alpha}, p_0\}.$$

If in addition, Assumption 2.3-(ii) holds, then (iii)-(c) and (iii)-(d) of Assumption 2.3 also hold with

$$q_1 = (q_0 - \alpha + 1)^+, \quad q_2 = \max\{q_1 + 2\bar{\alpha}), q_0\}.$$

Here p_0 and q_0 are as in Assumption 2.3.

Furthermore, assume that b and a are in $C^1(\mathbb{R}^d)$, $\|D^2 a\|_\infty < \infty$, $\|Db\| \leqslant \mathcal{B}(1+\|x\|^{\bar{\alpha}})$, f is (weakly) differentiable and $\sup_{t \leqslant T} \|Df(t,x)\| \leqslant \mathcal{C}(T)(1+\|x\|)^{p_0'}$ for some $p_0' \in \mathbb{R}$ and some constant $\mathcal{C}(T) > 0$. Then $u \in C^2(\mathbb{R}^d, \mathbb{R})$, and Assumption 2.3-(iii)-(e) also holds with

$$p_3 = \max\{p_0 + 2\bar{\alpha}, p_1 + 4\bar{\alpha}\}.$$

Proof. The fact that $u \in C^1(\mathbb{R}^d, \mathbb{R}^n)$ (or $C^2(\mathbb{R}^d, \mathbb{R}^n)$, under additional hypotheses) follows from the discussion above Remark A.1. Assumption 2.3-(iii)-(a) follows from [29, Theorem 2]. Assumption 2.3-(iii)-(b) and Assumption 2.3-(iii)-(e) now follow from Lemma B.1, Remark B.2 and Lemma B.3, applied to f and u.

In fact, a closer observation of the proof of [29, Theorem 2] reveals the following more detailed assertion: if $\|g_\kappa(x)\| \leqslant \kappa(1+\|x\|^{p_0})$ for some parameter κ, and v_κ given by (A.2) (with f_l replaced by g_κ) is the solution to the Poisson equation $\mathcal{L}v_\kappa = -g_\kappa$, then

$$\|v_\kappa(x)\| \leqslant \tilde{\mathfrak{m}}^0 \kappa(1+\|x\|^{p_1}), \tag{A.4}$$

with $p_1 = (p_0 - \alpha + 1)^+$, and where the constant $\tilde{\mathfrak{m}}^0$ does not depend on κ. Next, Remark B.2 shows that for some constant $\tilde{\mathfrak{m}}^1$ not depending on the parameter κ,

$$\|\nabla v_\kappa(y)\| \leqslant \tilde{\mathfrak{m}}^1 \kappa(1+\|y\|)^{p_2}. \tag{A.5}$$

with $p_2 = \max\{p_1 2\bar{\alpha}, p_0\}$.

In other words, the parameter κ appears in the bound of the solution, u_κ and its gradient, in the same linear way it appears in the bound of the input function, g_κ. This key observation is the reason behind the validity of of (c) and (d) of Assumption 2.3-(iii).

To see this notice that for a fixed t and Δ

$$\bar{u}^{t,\Delta}(x) \doteq u(t+\Delta, x) - u(t,x) = \int_0^\infty P_s \bar{f}^{t,\Delta}(x) ds$$

is the solution to the equation $\mathcal{L}v = -\bar{f}^{t,\Delta}$, where $\bar{f}^{t,\Delta}(x) \doteq f(t+\Delta, x) - f(t,x)$ satisfies $\|\bar{f}^{t,\Delta}(x)\| \leqslant \mathcal{C}(T)\mathfrak{r}(\Delta)(1+\|x\|)^{q_0}$ (by Assumption 2.3-(ii)). It follows from (A.4) and (A.5) with q_1 and q_2 as in the statement of the proposition (and with $\mathfrak{r}(\Delta)$ playing the role of κ) that

$$|u(t+\Delta, x) - u(t,x)| \equiv |\bar{u}^{t,\Delta}(x)| \leqslant \bar{\mathcal{C}}(T)\mathfrak{r}(\Delta)(1+\|x\|)^{q_1}, \quad \text{and}$$
$$|\nabla u(t+\Delta, x) - \nabla u(t,x)| \equiv |\nabla \bar{u}^{t,\Delta}(x)| \leqslant \bar{\mathcal{C}}(T)\mathfrak{r}(\Delta)(1+\|x\|)^{q_2}$$

Hence (c) and (d) of Assumption 2.3-(iii) hold. □

Although the above theorem is nice and might be the only tool available to check Assumption 2.3-(iii) and Assumption 2.8 for many stochastic models, it is not optimal. Consider an one dimensional model, where we have $xb(x) = -|x|^{1+\alpha}$. Then clearly, $|b(x)| \sim |x|^\alpha$. Then if Proposition A.2 is used to determine the exponents of u, Du, then it follows from Assumption 2.8 that f has to be chosen from the class for which $p_0 < -1$, that is, $|f(x)| \sim 1/(1+|x|)$. This restricts the applicability of the theorem to a smaller class of functions than desired.

However, for one-dimensional SDEs, Proposition A.2 could actually be vastly improved, and tighter bounds on growth rate of u and u' can be obtained. This result is presented in Proposition A.4. This makes our MDP results applicable to a wide class of stochastic models, and to functions f having polynomial-like growth – without doing any extra work for checking regularity of Poisson equation.

Regularity of Poisson equation for one dimensional SDE

When $d = 1$, the invariant distribution of X is given by

$$\pi(z) = \frac{\mathcal{B}}{a(z)} \exp\left(2\int_0^z \frac{b(y)}{a(y)}dy\right),$$

where $\mathcal{B}$ is the normalizing constant, and by a slight abuse of notation, we used $\pi(\cdot)$ to denote the density of the invariant distribution π . In this case the solution of the Poisson equation, $u(t,\cdot)$, have the following explicit representation:

$$u_f(t,x) \equiv u(t,x) = -\int_{-\infty}^x \frac{2}{a(z)\pi(z)} \int_{-\infty}^z f(t,y)\pi(y)dy\ dz. \tag{A.6}$$

Since t is just a parameter in (A.6), for convenience, we will drop t from the following result.

Assumption A.3. *There exist exponents p, $\theta(> -1)$ and constants $\mathfrak{c}_0, \mathfrak{c}_1$ and $\mathfrak{b}$ such that*

$$(i)\ |f(x)| = O(|x|^{p_0}), |b(x)| = O(|x|^{\alpha}), \quad (ii)\ |f(x)/b(x)| = O(|x|^{p_0-\alpha}),$$
$$(iii)\ \mathfrak{c}_0|x|^{\theta} \leqslant |b(x)/a(x)| \leqslant \mathfrak{c}_1|x|^{\theta}, \quad \textit{for } |x| \geqslant \mathfrak{b}.$$

Proposition A.4. *Suppose that Condition 2.1-(i) and Assumption A.3 hold. Then, u_f defined by* (A.6)*, is a solution to the Poisson equation, and*

$$(i)\ |u_f(x)| = O(|x|^{p_0-\alpha+1}), \quad \textit{for } p_0-\alpha \neq -1, \quad |u(x)| = O(|\ln x|), \quad \textit{for } p_0-\alpha = -1,$$
$$(ii)\ |u_f'(x)| = O(|x|^{p_0-\alpha}), \quad (iii)\ |u_f''(x)| = O(|x|^{p_0-\alpha+\theta})$$

Proof. Direct computation shows that u_f defined by (A.6), is a solution to the Poisson equation. Notice that

$$(\pi(z)a(z))' = 2b(z)\pi(z). \tag{A.7}$$

Also, it is clear from (a) Assumption A.3-(iii), (b) the expression of invariant distribution π, and (c) the fact that $\theta + 1 > 0$, that for any m

$$x^m a(x)\pi(x) \to 0, \qquad \text{as } |x| \to \infty. \tag{A.8}$$

Notice that since f is centered, that is $\pi(f) = 0$,

$$u_f'(x) = -\frac{2}{a(x)\pi(x)} \int_{-\infty}^x f(y)\pi(y)dy = \frac{2}{a(x)\pi(x)} \int_x^{\infty} f(y)\pi(y)dy \tag{A.9}$$

Since for $|x| > B$ (B was introduced Condition 2.1-(i)) , $xb(x) < 0$, we have that $b(x) < 0$ for all $x > B$ and $b(x) > 0$ for $x < -B$ For our purposes, the second equality in (A.9) needs to be used when $x > B$, and the first needs to be used when $x < -B$.

We first consider the case when $x > B$. Observe by Assumption A.3-(ii) and the fact that for $x > B$, $|b(x)| = -b(x)$, we have for some constant $\mathfrak{c}_2$

$$|u_f'(x)| \leqslant \frac{2}{a(x)\pi(x)} \int_x^{\infty} \left|\frac{f(y)}{b(y)}\right| |b(y)|\pi(y)dy \leqslant -\frac{2\mathfrak{c}_2}{a(x)\pi(x)} \int_x^{\infty} y^{p_0-\alpha}b(y)\pi(y)dy.$$

If $p_0 \leqslant \alpha$, then by (A.7) and (A.8), it follows that $|u_f'(x)| = O(|x|^{p_0-\alpha})$. If $p_0 > \alpha$, then we use (A.7) and integration by parts to get,

$$\begin{aligned}
|u_f'(x)| \leqslant\ & -\frac{\mathfrak{c}_2}{a(x)\pi(x)} \left[y^{p_0-\alpha}a(y)\pi(y)\Big|_x^{\infty} - \int_x^{\infty} y^{p_0-\alpha-1}a(y)\pi(y)dy \right] \\
&= \mathfrak{c}_2 x^{p_0-\alpha} + \frac{\mathfrak{c}_2}{a(x)\pi(x)} \int_x^{\infty} y^{p_0-\alpha-1}a(y)\pi(y)dy \\
&= \mathfrak{c}_2 x^{p_0-\alpha} + \frac{\mathfrak{c}_2}{a(x)\pi(x)} \int_x^{\infty} y^{p_0-\alpha-1}\frac{a(y)}{|b(y)|}|b(y)|\pi(y)dy \\
&\leqslant \mathfrak{c}_2 x^{p_0-\alpha} - \frac{\mathfrak{c}_2/\mathfrak{c}_0}{a(x)\pi(x)} \int_x^{\infty} y^{p_0-\alpha-1-\theta}b(y)\pi(y)dy.
\end{aligned}$$

If $p-\alpha-\theta \leqslant 1$, then it follows that

$$|u'_f(x)| \leqslant \mathfrak{c}_2 x^{p_0-\alpha} - \frac{\mathfrak{c}_2 x^{p_0-\alpha-1-\theta}/\mathfrak{c}_0}{a(x)\pi(x)} \int_x^\infty b(y)\pi(y)dy \; = \; 2\mathfrak{c}_2 x^{p_0-\alpha} + \mathfrak{c}_2 x^{p-\alpha-1-\theta}/2\mathfrak{c}_0 \; = \; O(|x|^{p_0-\alpha})$$

where we have used (A.7) and (A.8). If $p_0-\alpha-\theta > 1$, then let $k>1$ be the smallest integer such that $p_0-\alpha-\theta \leqslant k$. Now we repeat the integration by parts technique k times to prove the assertion.

If $x < -B$ then we use the first equality in (A.9) and the same techniques to prove the assertion. To prove the bound on u''_f simply observe that $|a(x)u''_f(x)| \leqslant \; |b(x)u'_f(x)| + |f(x)|$. Now the assertion follows from (ii) and (iii) of Assumption A.3. □

Example A.5. For a *mean-reverting* Ornstein-Uhlenbeck process ($b(x) = \mu - x$, and $\sigma(x) = \sigma$), the invariant distribution is of course the Normal$(\mu, \sigma^2/2\kappa)$. Here for f satisfying $|f(x)| \leqslant \mathcal{C}(1+\|x\|)^{p_0}$, Proposition A.4 gives the exponents of Assumption 2.3-(iii) : $p_1 = p_0$, $p_2 = p_0 - 1$. Note that p_3 is not needed as the diffusion coefficient is constant σ (see Remark 2.10). Thus Assumption 2.8 (and hence Theorem 2.9) holds for functions f with $p_0 \leqslant 1$. For Cox-Ingersoll-Ross (CIR) model ($b(x) = \kappa(\mu - x)$ and $\sigma(x) = s\sqrt{x}$), the invariant distribution is given by *Gamma*$(2\mu\kappa/\sigma^2, 2\kappa/\sigma^2)$. Here Proposition A.4 implies that Theorem 2.9 holds for all functions f with growth exponent $p_0 \leqslant 1/2$.

B Other results

The following version of [11, Theorem 9.11] is used in proving Proposition A.2. The proof just requires tracking of constants in the proof of [11, Theorem 9.11] and is omitted.

Lemma B.1. *Let $g \in L^p_{loc}(\mathbb{R}^d, \mathbb{R})$ and $v \in W^{2,p}_{loc}(\mathbb{R}^d, \mathbb{R})$ a solution to the elliptic equation $\mathcal{L}v = g$, where the coefficient a is uniformly continuous and satisfies* (A.3)*, and b satisfies*

$$\|b(x)\| \leqslant \mathcal{B}(1+\|x\|)^{\bar{\alpha}}$$

for some constant $\mathcal{B} > 0$ and exponent $\bar{\alpha} > 0$. Then for any $R > 0$ and $0 < \theta < 1$, there exists a constant $\bar{\mathfrak{c}}^0$ depending on $\mathcal{B}, \lambda_1, \lambda_2, \theta, R, d$ and p such that

$$\|v\|_{W^{2,p}B(y,\theta R)} \leqslant \bar{\mathfrak{c}}^0(\|g\|_{L^p(B(y,R))} + (1+\|y\|^{2\bar{\alpha}})\|v\|_{L^p(B(y,R))}).$$

Remark B.2. Let κ be a parameter, and $|g_\kappa(x)| \leqslant \mathfrak{m}_0\kappa(1+\|x\|)^{p_0}$, $|v_\kappa(x)| \leqslant \mathfrak{m}_1\kappa(1+\|x\|)^{p_1}$. Then notice that for some constants $\mathfrak{m}_{01}$

$$\|g_\kappa\|_{L^p(B(y,r))} \leqslant \mathfrak{m}_{01}\kappa(1+\|y\|)^{p_0}, \quad \|v_\kappa\|_{L^p(B(y,r))} \leqslant \mathfrak{m}_{11}\kappa(1+\|y\|)^{p_1}$$

where the constants $\mathfrak{m}_{01}$ (resp. $\mathfrak{m}_{11}$) depend on r, $\mathfrak{m}_0$ (resp. $\mathfrak{m}_1$), and p_0 (resp. p_1), but not on the parameter κ. It then follows from Lemma B.1,

$$\|v_\kappa\|_{W^{2,p}B(y,\theta r)} \leqslant \bar{\mathfrak{m}}^0\kappa(1+\|y\|)^{p_2},$$

where $p_2 = \max\{p_0, p_1 + 2\bar{\alpha}\}$, and the constant $\bar{\mathfrak{m}}^0$ does not depend on κ. Next choose $p > d$. Then by Sobolev's embedding theorem, there exists a constant $\bar{\mathfrak{m}}^1 \equiv \bar{\mathfrak{m}}^1(p, d, \theta r)$ such that

$$\|\nabla v_\kappa(y)\| \leqslant \bar{\mathfrak{m}}^1\|v_\kappa\|_{W^{2,p}B(y,\theta r)} \leqslant \bar{\mathfrak{m}}^1\bar{\mathfrak{m}}^0\kappa(1+\|y\|)^{p_2}.$$

We now state the result on pointwise bounds of $\|D^2 v(\cdot)\|$.

Lemma B.3. *Assume the setup and hypothesis of Lemma B.1. Furthermore, suppose that the coefficients of $\mathcal{L}$ are in $C^1(\mathbb{R}^d)$, and that for each k, $\|a^{(k)}\|_\infty < \infty$ and $\|b^{(k)}(x)\| \leqslant \mathcal{B}(1+\|x\|^{\bar{\alpha}})$ for some constant $\mathcal{B}$ and some exponent $\bar{\alpha}'$, where*

$$a^{(k)}_{ij}(x) = \partial_k a_{ij}(x), \qquad b^{(k)}_i(x) = \partial_k b_i(x).$$

Also, as in Remark B.2, assume that $|g(x)| = O(\|x\|^{p_0}), \|\nabla g(x)\| = O(\|x\|^{p_0})$, and $|v(x)| = O(\|x\|^{p_1})$ for some exponents p_0 and p_1. Then for some constant $\bar{\mathfrak{c}}^2$

$$\|D^2 v(y)\| \leqslant \bar{\mathfrak{c}}^2(1+\|y\|)^{p_3}$$

where $p_3 = \max\{p_0 + 2\bar{\alpha}, p_1 + 4\bar{\alpha}\}$.

Proof. First notice that $v^{(k)} = \partial_k v$ satisfies $\mathcal{L}v^{(k)} = \tilde{g}_k$, where $\tilde{g}_k = g^{(k)} - b^{(k)} \cdot \nabla v - \frac{1}{2}tr(a^{(k)}D^2 v)$. It now follows from Lemma B.1 that for each $k = 1, 2, \ldots, d$,

$$\begin{aligned}\|v^{(k)}\|_{W^{2,p}B(y,\theta r)} &\leqslant \bar{\mathfrak{c}}_7 \left(\|\tilde{g}_k\|_{L^p(B(y,r))} + (1 + \|y\|^{2\bar{\alpha}})\|v^{(k)}\|_{L^p(B(y,r))}\right)\\ &\leqslant \bar{\mathfrak{c}}_8\Big(\|g^{(k)}\|\|_{L^p(B(y,r))} + (1 + \|y\|^{\bar{\alpha}})\|\nabla v\|_{L^p(B(y,r))} + \|D^2 v\|_{L^p(B(y,R))}\\ &\qquad + (1 + \|y\|^{2\bar{\alpha}})\|\nabla v\|_{L^p(B(y,r))}\Big).\end{aligned}$$

Thus,

$$\begin{aligned}\|v\|_{W^{3,p}B(y,\theta r)} &\leqslant \bar{\mathfrak{c}}_9 \ \left(\|\nabla g\|_{L^p(B(y,r))} + (1 + \|y\|^{2\bar{\alpha}})\|v\|_{W^{2,p}(B(y,r))}\right)\\ &\leqslant \bar{\mathfrak{c}}_{10} \ (\|\nabla g\|_{L^p(B(y,r))} + (1 + \|y\|^{2\bar{\alpha}})\|g\|_{L^p(B(y,r))} + (1 + \|y\|^{4\bar{\alpha}})\|v\|_{L^p(B(y,r))}).\end{aligned}$$

The desired pointwise bound now again follows from Sobolev's embedding theorem (by choosing $p > d$), and the assumption on g and ∇g. □

The following lemma is about convergence of integrals with respect to random probability measures under uniform integrability like condition. The proof is similar to [17, Proposition 3.12] in the deterministic case and is omitted.

Lemma B.4. *Let E be a separable Banach space, and $\{\mu_n\}$ a sequence of $\mathcal{P}(E)$-valued random variables such that $\mu_n(\omega) \Rightarrow \mu(\omega)$ ω-a.s as $n \to \infty$. Suppose that $h : E \to \mathbb{R}^d$ is a continuous function satisfying $\|h(x)\|/f(x) \to 0$ as $\|x\| \to \infty$, where $f : E \to [0, \infty)$ is a lower semicontinuous function such that $K_0 \doteq \sup_n \mathbb{E}\int_E f(x)\ \mu_n(dx) < \infty$. Then as $n \to \infty$, $\mathbb{E}\left\|\int_E h(x)\mu_n(dx) - \int_E h(x)\mu(dx)\right\| \to 0$.*

Lemma B.5. *Let $\mathcal{L}^2_d \equiv \mathcal{L}^2(\Omega, \mathbb{R}^d)$ denote the space of square integrable $\mathbb{R}^d$-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and H an $n \times d$ random matrix. Assume that $M = \mathbb{E}(HH^T)$ is invertible. Then for any $b \in \mathbb{R}^n$,*

$$\min\{\mathbb{E}\|Y\|^2 : Y \in \mathcal{L}^2_d,\ \mathbb{E}(HY) = b\} = b^T M^{-1} b,$$

Proof. Let Y be such that $E(HY) = b$. Then notice that

$$\begin{aligned}0 &\leqslant \mathbb{E}\|Y - H^T M^{-1} b\|^2 = \mathbb{E}\|Y\|^2 - 2\mathbb{E}\left(\langle Y, H^T M^{-1} b\rangle\right) + \mathbb{E}\|H^T M^{-1} b\|^2\\ &= \mathbb{E}\|Y\|^2 - 2\mathbb{E}\left(\langle HY, M^{-1} b\rangle\right) + \mathbb{E}(b^T M^{-1}(HH^T)M^{-1} b)\\ &= \mathbb{E}\|Y\|^2 - 2\langle\mathbb{E}(HY), M^{-1} b\rangle + b^T M^{-1}\mathbb{E}(HH^T)M^{-1} b \ = \ \mathbb{E}\|Y\|^2 - b^T M^{-1} b,\end{aligned}$$

which proves that $\mathbb{E}\|Y\|^2 \geqslant b^T M^{-1} b$. Finally, observe that equality holds for $Y = H^T M^{-1} b$. □

References

[1] David F. Anderson, Arnab Ganguly, and Thomas G. Kurtz. Error analysis of tau-leap simulation methods. *Ann. Appl. Probab.*, 21(6):2226–2262, 2011. ISSN 1050-5164.

[2] Jaya P. N. Bishwal. *Parameter estimation in stochastic differential equations*, volume 1923 of *Lecture Notes in Mathematics*. Springer, Berlin, 2008. ISBN 978-3-540-74447-4.

[3] Michelle Boué and Paul Dupuis. A variational representation for certain functionals of Brownian motion. *Ann. Probab.*, 26(4):1641–1659, 1998. ISSN 0091-1798.

[4] Amarjit Budhiraja, Paul Dupuis, and Vasileios Maroulas. Large deviations for infinite dimensional stochastic dynamical systems. *Ann. Probab.*, 36(4):1390–1420, 2008.

[5] Amarjit Budhiraja, Paul Dupuis, and Vasileios Maroulas. Variational representations for continuous time processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(3):725–747, 2011. ISSN 0246-0203.

[6] Amarjit Budhiraja, Paul Dupuis, and Arnab Ganguly. Moderate deviation principles for stochastic differential equations with jumps. *Ann. Probab.*, 44(3):1723–1775, 2016. ISSN 0091-1798.

[7] Paul Dupuis and Richard S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[8] Paul Dupuis and Dane Johnson. Moderate deviations for recursive stochastic algorithms. *Stoch. Syst.*, 5 (1):87–119, 2015. ISSN 1946-5238.

[9] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. ISBN 0-471-08186-8. Characterization and convergence.

[10] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk. Fatou's lemma for weakly converging probabilities. *Theory Probab. Appl.*, 58(4):683–689, 2014. ISSN 0040-585X.

[11] David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order.* Classics in Mathematics. Springer-Verlag, Berlin, 2001. ISBN 3-540-41160-7. Reprint of the 1998 edition.

[12] Carl Graham and Denis Talay. *Stochastic simulation and Monte Carlo methods*, volume 68 of *Stochastic Modelling and Applied Probability.* Springer, Heidelberg, 2013. ISBN 978-3-642-39362-4; 978-3-642-39363-1. Mathematical foundations of stochastic simulation.

[13] A. Guillin. Averaging principle of SDE with small diffusion: moderate deviations. *Ann. Probab.*, 31(1): 413–443, 2003. ISSN 0091-1798.

[14] A. Guillin and R. Liptser. Examples of moderate deviation principle for diffusion processes. *Discrete Contin. Dyn. Syst. Ser. B*, 6(4):803–828, 2006. ISSN 1531-3492.

[15] Arnaud Guillin. Moderate deviations of inhomogeneous functionals of Markov processes and application to averaging. *Stochastic Process. Appl.*, 92(2):287–313, 2001. ISSN 0304-4149.

[16] Desmond J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, 43(3):525–546, 2001. ISSN 0036-1445.

[17] Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

[18] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, 1992. ISBN 3-540-54062-8.

[19] N. V. Krylov. *Controlled diffusion processes*, volume 14 of *Applications of Mathematics.* Springer-Verlag, New York-Berlin, 1980. ISBN 0-387-90461-1. Translated from the Russian by A. B. Aries.

[20] Yury A. Kutoyants. *Statistical inference for ergodic diffusion processes.* Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2004. ISBN 1-85233-759-1.

[21] Damien Lamberton and Gilles Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002. ISSN 1350-7265.

[22] Damien Lamberton and Gilles Pagès. Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift. *Stoch. Dyn.*, 3(4):435–451, 2003. ISSN 0219-4937.

[23] R. Liptser and V. Spokoiny. Moderate deviations type evaluation for integral functionals of diffusion processes. *Electron. J. Probab.*, 4:no. 17, 25, 1999. ISSN 1083-6489.

[24] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002. ISSN 0304-4149.

[25] Jonathan C. Mattingly, Andrew M. Stuart, and M. V. Tretyakov. Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.*, 48(2):552–577, 2010. ISSN 0036-1429.

[26] Gilles Pagès and Fabien Panloup. Ergodic approximation of the distribution of a stationary diffusion: rate of convergence. *Ann. Appl. Probab.*, 22(3):1059–1100, 2012. ISSN 1050-5164.

[27] Gilles Pagès and Fabien Panloup. A mixed-step algorithm for the approximation of the stationary regime of a diffusion. *Stochastic Process. Appl.*, 124(1):522–565, 2014. ISSN 0304-4149.

[28] Fabien Panloup. Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. *Ann. Appl. Probab.*, 18(2):379–426, 2008. ISSN 1050-5164.

[29] E. Pardoux and A. Yu. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001. ISSN 0091-1798.

[30] È. Pardoux and A. Yu. Veretennikov. On Poisson equation and diffusion approximation. II. *Ann. Probab.*, 31(3):1166–1192, 2003. ISSN 0091-1798.

[31] Eckhard Platen. An introduction to numerical methods for stochastic differential equations. In *Acta numerica, 1999*, volume 8 of *Acta Numer.*, pages 197–246. Cambridge Univ. Press, Cambridge, 1999.

[32] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002. ISSN 1024-2953. Inhomogeneous random systems (Cergy-Pontoise, 2001).

[33] Denis Talay. Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics and Stochastic Reports*, 29(1):13–36, 1990.

[34] Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990. ISSN 0736-2994.

[35] A. Yu. Veretennikov. Estimates of the mixing rate for stochastic equations. *Teor. Veroyatnost. i Primenen.*, 32(2):299–308, 1987. ISSN 0040-361X.