



# A Weak Overdamped Limit Theorem for Langevin Processes

Mathias Rousset, Yushun Xu, Pierre-André Zitt

## ► To cite this version:

Mathias Rousset, Yushun Xu, Pierre-André Zitt. A Weak Overdamped Limit Theorem for Langevin Processes. 2017. hal-01596954v1

**HAL Id: hal-01596954**

**<https://hal.science/hal-01596954v1>**

Preprint submitted on 28 Sep 2017 (v1), last revised 8 Mar 2019 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A WEAK OVERDAMPED LIMIT THEOREM FOR LANGEVIN PROCESSES

MATHIAS ROUSSET, YUSHUN XU, AND PIERRE-ANDRÉ ZITT

ABSTRACT. In this paper, we prove convergence in distribution of Langevin processes in the overdamped asymptotics. The proof relies on the classical perturbed test function (or corrector) method, which is used both to show tightness in path space, and to identify the extracted limit with a martingale problem. The result holds assuming the continuity of the gradient of the potential energy, and a mild control of the initial kinetic energy.

## 1. INTRODUCTION

This paper focuses on the overdamped asymptotics of Langevin dynamics. The Langevin Stochastic Differential Equation (SDE) describes the dynamics of a classical mechanical system perturbed by a stochastic thermostat. The system state at time  $t \geq 0$  is encoded by its position  $Q_t$  and its momentum  $P_t$ . More formally, the equation reads:

$$\begin{cases} dQ_t &= P_t dt, \\ dP_t &= -\nabla V(Q_t) dt - P_t dt + \sqrt{2\beta^{-1}} dW_t, \end{cases}$$

where in the above,  $Q_t$  takes values in the  $d$ -dimensional torus  $\mathbb{T}^d$ ,  $P_t$  takes values in  $\times \mathbb{R}^d$ , the function  $V : \mathbb{T}^d \rightarrow \mathbb{R}$  is the particles' potential energy,  $\beta > 0$  the inverse temperature, and  $t \mapsto W_t \in \mathbb{R}^d$  is a standard  $d$ -dimensional Brownian motion. The term  $\sqrt{2\beta^{-1}} dW_t$  is a fluctuation term bringing energy into the system, while this energy is dissipated through the friction term  $-P_t dt$ ; the sum of these two terms forming the so-called thermostat part. The remaining terms are simply Newton's equation of motion. For more details on this equation, we refer to [LRS10, Section 2.2].

The case we consider here is the so-called overdamped asymptotics, where the time scale of the large damping due to friction is much smaller than the time scale of the Hamiltonian dynamics, so that the momentum becomes a fast variable compared to the slow position variable. We introduce a parameter  $\varepsilon$  for the ratio of the time scales, and consider

$$(1.1) \quad \begin{cases} dQ_t^\varepsilon &= \frac{1}{\varepsilon} P_t^\varepsilon dt, \\ dP_t^\varepsilon &= -\frac{1}{\varepsilon} \nabla V_\varepsilon(Q_t^\varepsilon) dt - \frac{1}{\varepsilon^2} P_t^\varepsilon dt + \frac{1}{\varepsilon} \sqrt{2\beta^{-1}} dW_t. \end{cases}$$

Note that we allow the potential  $V_\varepsilon \in C^1(\mathbb{T}^d)$  to depend on  $\varepsilon$  and will only suppose that it converges to a limit  $V$ ; see below for a precise statement. The Markov generator  $L_\varepsilon$  associated with (1.1) is given by

$$(1.2) \quad L_\varepsilon f(q, p) := \frac{1}{\varepsilon^2} \left( \frac{1}{\beta} \Delta_p f - p \cdot \nabla_p f \right) + \frac{1}{\varepsilon} (p \cdot \nabla_q f - \nabla_q V_\varepsilon \cdot \nabla_p f),$$

---

(M. Rousset) INRIA RENNES - BRETAGNE ATLANTIQUE, & IRMAR UNIVERSITÉ RENNES 1, FRANCE.

(Y. Xu and P. A. Zitt) UNIVERSITÉ PARIS-EST, LABORATOIRE D'ANALYSE ET DE MATHÉMATIQUES APPLIQUÉES (UMR 8050), UPEM, UPEC, CNRS, F-77454, MARNE-LA-VALLÉE, FRANCE

*E-mail addresses:* mathias.rousset@inria.fr, Yushun.Xu@u-pem.fr, Pierre-Andre.Zitt@u-pem.fr.

*Date:* September 9, 2017.

1991 *Mathematics Subject Classification.* 35P25, 35Q55.

*Key words and phrases.* Langevin dynamics, overdamped asymptotics, perturbed test function.

This work is partially supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement number 614492.

where  $f$  denotes any smooth test function of the variables  $(q, p) \in \mathbb{T}^d \times \mathbb{R}^d$ .

Overdamped processes are stochastic dynamics on the system position  $(Q_t)_{t \geq 0}$  only. The overdamped Langevin SDE is given by:

$$(1.3) \quad dQ_t = -\nabla V(Q_t)dt + \sqrt{2\beta^{-1}}dB_t,$$

where  $V : \mathbb{T}^d \rightarrow \mathbb{R}^d$  is a potential energy, limit of  $V_\varepsilon$  when  $\varepsilon \rightarrow 0$  in some appropriate sense, and  $t \mapsto B_t \in \mathbb{R}^d$  is a standard d-dimensional Wiener process. The Markov generator  $L$  associated with (1.3) acts on smooth test functions  $f$  of the variable  $q$  as follows:

$$Lf(q) := -\nabla_q V \cdot \nabla_q f + \frac{1}{\beta} \Delta_q f.$$

Our main result is the proof of the convergence in distribution of the Langevin position process  $(Q_t^\varepsilon)_{t \geq 0}$  towards its overdamped counterpart  $(Q_t)_{t \geq 0}$ , assuming the uniform convergence of the gradient potential as well as a control of moments of the initial kinetic energy.

**Theorem 1.1** (Overdamped limit of the Langevin dynamics). *For any  $\varepsilon > 0$ , suppose that  $(Q_t^\varepsilon, P_t^\varepsilon)_{t \geq 0} \in \mathbb{T}^d \times \mathbb{R}^d$  is a weak solution to the SDE (1.1). Assume that the following conditions hold:*

- (1)  $V_\varepsilon$  is  $C^1(\mathbb{T}^d)$ , and converges to  $V$  in the sense that  $\|\nabla V_\varepsilon - \nabla V\|_\infty \xrightarrow{\varepsilon \rightarrow 0} 0$ ,
- (2) The following moment bound holds true:

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \mathbb{E}(|P_0^\varepsilon|^3) = 0$$

- (3) The initial position distribution is converging to some limit:  $\text{Law}(Q_0^\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} \text{Law}(Q_0)$ .

Then, when  $\varepsilon \rightarrow 0$ , the process  $(Q_t^\varepsilon)_{t \geq 0} \in C(\mathbb{R}_+ \rightarrow \mathbb{T}^d)$  converges in distribution to the unique weak solution of the overdamped SDE (1.3).

*Remark 1.2.* In Theorem 1.1, the space of trajectories  $C(\mathbb{R}_+ \mapsto \mathbb{T}^d)$  is endowed with uniform convergence on compact sets; making it Polish (metrizable for a separable and complete metric).

The literature on diffusion approximations is very rich; we refer for instance to Stuart-Pavliotis in [PS08] for a recent pedagogical overview of related issues. Historically, a possible chain of seminal references is given by Stratonovich in [Str63], Khas'minskii in [Kha66], Papanicolaou-Varadhan in [PV73], as well as Papanicolaou-Kohler in [PK74]; complemented with the more modern viewpoint of Ethier-Kurtz in [EK86], Chapter 12 "Random evolutions".

In the present case, the momentum variable is averaged out with the diffusion approximation, so that the problem may be labeled as "diffusion approximation with averaging". Broadly speaking, the problem can be approached using strong or weak convergence techniques. For an example of the strong convergence approach, the results in [SSMD82] rely on estimating the dynamics of  $Q_t^\varepsilon$  and its limit using a Gronwall argument; this approach requires the Lipschitz continuity of  $\nabla V_\varepsilon$  uniformly in  $\varepsilon$ . On the other hand, weak convergence results rely on the so-called "perturbed" test function or "corrector" approach, that have been developed since Papanicolaou-Stroock-Varadhan in [PSV77]. The case of the overdamped limit (1.1) is not directly covered by these results. Indeed, the correctors are not bounded in the present case, due to the fact that the state space of the momentum variable is not compact.

In a series of papers [PV01, PV03, PV05], Pardoux-Veretennikov extend the classical diffusion approximation with averaging to the non-compact state space case. In the latter setting however, the slow variable has a dynamics independent of the fast one, which is not the case in the Langevin case (1.1).

We now give a physically motivated example that satisfies our assumptions but was not covered by previous works.

**Example 1.3.** *Let*

$$V_\varepsilon(q) = V(q) + \alpha_\varepsilon \chi(k_\varepsilon q),$$

where  $\chi \in C^\infty(\mathbb{T}^d)$ , and the scaling coefficients  $k_\varepsilon \in \mathbb{N}$  and  $\alpha_\varepsilon \in \mathbb{R}$  satisfy

$$k_\varepsilon \rightarrow \infty, \quad \alpha_\varepsilon k_\varepsilon \rightarrow 0.$$

*Physically, the potential  $\alpha_\varepsilon \chi(k_\varepsilon q)$  may model the interaction between a particle with unit energy and a periodic crystal of small period  $k_\varepsilon^{-1}$ , and small energy range of order  $\alpha_\varepsilon$ . When  $k_\varepsilon \rightarrow +\infty$  but  $\alpha_\varepsilon k_\varepsilon = 1$  and  $\varepsilon$  is kept constant, the effective action of the periodic crystal on the particle can not be neglected, especially for grazing velocities co-linear to the principal directions of the crystal. Indeed, in the latter case, on times of order 1, the crystal exerts on the particle a total force also of order 1, making it deviating from its trajectory.*

*Our result shows that the physically necessary condition  $\alpha_\varepsilon k_\varepsilon \rightarrow 0$  is in fact sufficient for neglecting the crystal effect in the overdamped regime. Note that if  $\alpha_\varepsilon k_\varepsilon^2 \rightarrow +\infty$ , when  $\varepsilon \rightarrow 0$ , then*

$$\|\nabla V_\varepsilon - \nabla V\|_\infty \xrightarrow{\varepsilon \rightarrow 0} 0,$$

*but still*

$$\|\nabla^2 V_\varepsilon\|_\infty \sim \alpha_\varepsilon k_\varepsilon^2 \|\nabla^2 \chi\|_\infty \xrightarrow{\varepsilon \rightarrow 0} +\infty,$$

*preventing  $\nabla V_\varepsilon$  from being Lipschitz uniformly in  $\varepsilon$ ; and hence forbidding results based on strong convergence.*

In order to prove Theorem 1.1, we will establish a more general weak convergence result. We consider a sequence (indexed by a small parameter  $\varepsilon > 0$ ) of Markov processes of the form  $t \mapsto (Q_t^\varepsilon, P_t^\varepsilon) \in \mathbb{T}^d \times \mathbb{R}^d$  taking value in the Skorokhod path space  $\mathbb{D}_{\mathbb{T}^d \times \mathbb{R}^d}$ . Our general convergence result, namely Theorem 3.5, gives general conditions under which  $(Q_t^\varepsilon)_{t \geq 0}$  converges in distribution to the unique solution of a particular martingale problem. The proof follows the usual pattern: first we prove tightness for the family of distributions of  $(Q_t^\varepsilon)$ , and then characterize the limit through martingale problems. For both steps, we use the perturbed test function method. The key sufficient criteria yielding the results of both steps is given in Assumption 3.4, which states that to any smooth  $f : \mathbb{T}^d \rightarrow \mathbb{R}$ , we can associate a perturbed test function  $f_\varepsilon : \mathbb{T}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  such that for all  $T > 0$ ,

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{t \leq T} |f(Q_t^\varepsilon) - f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)| \right) = 0 \text{ and } \lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \int_0^T |Lf(Q_t^\varepsilon) - L_\varepsilon f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)| dt \right) = 0.$$

*Remark 1.4* (On the choice of the state space). Theorem 3.5 can be useful for càd-làg processes, which explains the fact that we work in Skorokhod space. We have chosen to work in  $\mathbb{T}^d \times \mathbb{R}^d$  for notational simplicity, but Theorem 3.5 could be extended to more general product spaces of the type  $E \times F$ , where  $E$  and  $F$  are Polish spaces. If  $E$  is compact, the extension is straightforward. If  $E$  is locally compact, then one can work with  $E \cup \{\infty\}$ , the one point compactification of  $E$  at infinity (see [EK86, Chapter 4]). If  $E$  is not locally compact, then one needs to use Theorem 9.1 in [EK86, Chapter 3] instead of Theorem 2.12 below which is a corollary of the former. In the latter case: (i) the *a priori* compact containment condition (9.1) of Theorem 9.1 in [EK86, Chapter 3] has to be proven; and (ii) one has to show the tightness of  $(\text{Law}(f(Q_t^\varepsilon))_{t \geq 0})_{\varepsilon \geq 0}$  for all  $f$  in a space of functions dense in  $C_b(E)$  for the topology of uniform convergence on compacts. Such extensions to infinite dimensional spaces are left for future work.

The paper is organized as follows. Section 2 starts with some notation and preliminaries. In Section 3, we state and prove the general convergence result Theorem 3.5. This

general method is then applied in Section 4 to the overdamped Langevin limit, proving Theorem 1.1.

## 2. NOTATION AND PRELIMINARIES

In what follows, we introduce notation and recall some known results.

**2.1. General notation.** Let  $(E, d)$  be a Polish space, that is, a topological space which is metric, complete and separable. Denote  $C(E)$  the Banach space of all continuous functions and  $C_b(E)$  the Banach space of all bounded continuous functions. We denote by  $\mathcal{P}(E)$  the space of probability measures on the Borel  $\sigma$ -field  $\mathcal{B}(E)$ . The notation  $\mathcal{F}_t^X$  means the natural filtration of càd-làg processes  $(X_t)_{t \geq 0}$ , that is  $\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$ . For any  $(s, t) \in \mathbb{R} \times \mathbb{R}$ , we denote by  $s \wedge t$  the minimum of  $s$  and  $t$ .

**2.2. The Skorokhod space.** A càd-làg function (from the French "continu à droite, limité à gauche", also called RCLL for "right continuous with left limits") is a function defined on  $\mathbb{R}_+$  that is everywhere right-continuous and has left limits everywhere. The collection of càd-làg functions on a given domain is known as the Skorokhod space. We denote  $\mathbb{D}_E$  the space of càd-làg functions with values in a Polish space  $E$ . We recall that this path space  $\mathbb{D}_E$  may be equipped with the Skorokhod topology (see Section 5 of [EK86, Chapter 3]): a family of trajectories  $(q_s^\varepsilon)_{s \geq 0}$  indexed by  $\varepsilon$  converges to a limit trajectory  $(q_s^0)_{s \geq 0}$  if there exists a sequence  $(\lambda_\varepsilon)_{\varepsilon \geq 0}$  in the space of strictly increasing continuous bijections of  $[0, \infty[$ , such that for each  $T > 0$ :  $\lim_{\varepsilon \rightarrow 0} \sup_{t \leq T} |\lambda_\varepsilon(t) - t| = 0$  and  $\lim_{\varepsilon \rightarrow 0} \sup_{t \leq T} d(q_t^\varepsilon, q_{\lambda_\varepsilon(t)}^0) = 0$ . The following result will be useful in the proof of Theorem 3.5.

**Lemma 2.1.** *Integration with respect to time is continuous with respect to the Skorokhod topology: if  $(q_t^\varepsilon)_{t \geq 0}$  converges to  $(q_t^0)_{t \geq 0}$  in  $\mathbb{D}_E$ , and  $\psi : E \rightarrow \mathbb{R}$  is bounded and continuous, then for each  $T > 0$ ,*

$$\int_0^T \psi(q_t^\varepsilon) dt \xrightarrow{\varepsilon \rightarrow 0} \int_0^T \psi(q_t^0) dt.$$

*Proof.* Let us denote by  $J_T := \{t \in [0, T], q_{t-}^0 \neq q_t^0\}$  the countable set of jump times in  $[0, T]$  of  $q^0$ . By definition of convergence in the Skorokhod space,

$$\lim_{\varepsilon \rightarrow 0} q_s^\varepsilon = q_s^0 \quad \forall s \in [0, T] \setminus J_T.$$

Since  $J_T$  has Lebesgue measure 0 and  $\psi$  is continuous and bounded, dominated convergence yields the result.  $\square$

**2.3. Martingale problems.** Let us first recall some basics on martingales and stochastic calculus. Let  $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t \geq 0})$  a filtered probability space. A càd-làg real-valued process  $(X_t)_{t \geq 0}$  is said to be adapted if  $X_t$  is  $\mathcal{F}_t$ -measurable for all  $t \geq 0$ , and is called a  $(\mathcal{F}_t)_{t \geq 0}$ -martingale if  $\mathbb{E}(|X_t| | \mathcal{F}_s) < +\infty$  and  $\mathbb{E}(X_t | \mathcal{F}_s) = X_s$  for any  $0 \leq s \leq t$ .

We will often need the technical tool of localization by stopping times, to deal with the unboundedness of the momentum variable. We follow here the presentation of [EK86, Chapter 4].

**Definition 2.2** (Local martingale). *A càd-làg real-valued process  $(X_t)_{t \geq 0}$  defined on  $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t \geq 0})$  is called a local martingale with respect to  $(\mathcal{F}_t)_{t \geq 0}$  if there exists a non-decreasing sequence  $(\tau_n)_{n \in \mathbb{N}}$  of  $(\mathcal{F}_t)_{t \geq 0}$ -stopping times such that  $\tau_n \rightarrow \infty$   $\mathbf{P}$ -almost surely, and for every  $n \in \mathbb{N}$ ,  $(X_{t \wedge \tau_n})_{t \geq 0}$  is an  $(\mathcal{F}_t)_{t \geq 0}$ -martingale.*

Let us now state precisely what it means for a process to solve a martingale problem.

**Definition 2.3** (Martingale problem). *Let  $E$  be a Polish space. Let  $L$  be a linear operator mapping a given space  $\mathcal{D} \subset C_b(E)$  into bounded measurable functions. Let  $\mu$  be a probability distribution on  $E$ . A càd-làg process  $(X_t)_{t \geq 0}$  with values in  $E$  solves the martingale problem for the generator  $L$  on the space  $\mathcal{D}$  with initial measure  $\mu$  — in short,  $X$  solves  $\mathbf{MP}(L, \mathcal{D}, \mu)$  — if  $\text{Law}(X_0) = \mu$  and if, for any  $\varphi \in \mathcal{D}$ ,*

$$(2.1) \quad t \mapsto M_t(\varphi) := \varphi(X_t) - \varphi(X_0) - \int_0^t L\varphi(X_s) ds$$

*is a martingale with respect to the natural filtration  $(\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t))_{t \geq 0}$ .*

*Moreover, the martingale problem  $\mathbf{MP}(L, \mathcal{D}, \mu)$  is said to be well-posed if:*

- *There exists a probability space and a càd-làg process defined on it that solves the martingale problem (existence);*
- *whenever two processes solve  $\mathbf{MP}(L, \mathcal{D}, \mu)$ , then they have the same distribution on  $\mathbb{D}_E$  (uniqueness).*

**2.4. Weak solutions of SDEs.** Let  $b : \mathbb{R}^d \mapsto \mathbb{R}^d$  and  $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d \times n}$  be locally bounded. Consider a stochastic differential equation in  $\mathbb{R}^d$  of the form:

$$(2.2) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

with an initial condition  $\text{Law}(X_0) = \mu_0$ . Let  $L$  be the formal generator

$$(2.3) \quad L := \sum_{i=1}^d b_i \partial_i + \frac{1}{2} \sum_{i,j=1}^d a_{ij} \partial_i \partial_j,$$

where  $a = \sigma \sigma^T$ .

**Definition 2.4** (Weak solution of the SDE). *A continuous process  $(X_t)_{t \geq 0}$  is a weak solution of (2.2) if there exists a filtered probability space  $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t \geq 0})$  such that:*

- *$t \mapsto W_t$  is a  $(\mathcal{F}_t)_{t \geq 0}$ -Brownian motion, that is, an  $(\mathcal{F}_t)_{t \geq 0}$ -adapted process such that  $\text{Law}(W_{t+h} - W_t | \mathcal{F}_t) = \mathcal{N}(0, h)$ .*
- *$X$  is a continuous,  $(\mathcal{F}_t)_{t \geq 0}$ -adapted process and satisfies the stochastic integral equation*

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dW_s \quad \text{a.s.}$$

We now quote two results from [EK86] concerning existence and uniqueness of solutions to SDEs and martingale problems. The first is an existence result, and can be found in [EK86, Section 5.3] (Corollary 3.4 and Theorem 3.10).

**Theorem 2.5.** *Assume that  $b, \sigma$  are continuous. If there exists a constant  $K$  such that for any  $t \geq 0, x \in \mathbb{R}^d$ :*

$$(2.4) \quad |\sigma|^2 \leq K(1 + |x|^2);$$

$$(2.5) \quad x \cdot b(x) \leq K(1 + |x|^2),$$

*then there exists a weak solution of the stochastic differential equation (2.2) corresponding to  $(\sigma, b, \mu)$ , which is also solution of the martingale problem  $\mathbf{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$ ,  $C_c^\infty(\mathbb{R}^d)$  being the set of smooth functions with compact support.*

**Remark 2.6.** For the Langevin equation (1.1)) we first remark that the latter can be set in  $\mathbb{R}^d \times \mathbb{R}^d$  using the  $\mathbb{Z}^d$ -periodic extension of  $V_\varepsilon$ . Then  $b(q, p) = \left(\frac{1}{\varepsilon}p, -\frac{1}{\varepsilon}\nabla V_\varepsilon(q) - \frac{1}{\varepsilon^2}p\right)$  and  $\sigma = (0, \frac{1}{\varepsilon}\sqrt{2\beta^{-1}}\text{Id}_{\mathbb{R}^d})$  are continuous since  $V_\varepsilon \in C^1(\mathbb{R}^d)$ . Moreover,  $|\sigma|^2 = \sigma \sigma^T = (0, \frac{2}{\beta \varepsilon^2}\text{Id}_{\mathbb{R}^d})$ , and on the other hand

$$(q, p) \cdot b(q, p) = \frac{1}{\varepsilon}pq - \frac{1}{\varepsilon}p \nabla V_\varepsilon(q) - \frac{1}{\varepsilon^2}p^2 \leq \frac{1}{2\varepsilon}(1 + \|\nabla V_\varepsilon\|_\infty)(1 + |p|^2 + |q|^2),$$

which implies the existence of weak solution of (1.1) in  $\mathbb{R}^d$ . One then obtains existence of a weak solution in  $\mathbb{T}^d$  of the original (1.1) using the canonical continuous mapping  $\mathbb{R}^d \rightarrow \mathbb{T}^d := \mathbb{R}^d / \mathbb{Z}^d$ .

The next result follows from [EK86] (Theorem 1.7 in Section 8.1) and [SV07] (Theorem 10.2.2 and the discussion following their Corollary 10.1.2).

**Theorem 2.7.** *Assume that the bounds (2.4) and (2.5) hold. Suppose that  $a := \sigma \sigma^\top$  is continuous and uniformly elliptic:*

$$\exists C_a > 0, \forall \xi \in \mathbb{R}^d, \forall x \in \mathbb{R}^d, \quad \xi^\top a(x) \xi \geq C_a |\xi|^2.$$

*Then for any initial condition  $\mu$ , there is a unique weak solution of the stochastic differential equation (2.2). This solution is also the unique solution of the martingale problem  $\mathbf{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$ .*

*Remark 2.8.* For the overdamped Langevin equation (1.3), we remark again that the latter can be set in  $\mathbb{R}^d$  using the  $\mathbb{Z}^d$ -periodic extension of  $V_\varepsilon$ . One then obtains well-posedness of the martingale problem  $\mathbf{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$  in  $\mathbb{R}^d$  since  $\nabla V$  is bounded and continuous by assumption. This solution obviously solves  $\mathbf{MP}(L, C_c^\infty(\mathbb{T}^d), \mu)$  in  $\mathbb{T}^d$ . The fact that uniqueness of  $\mathbf{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$  implies uniqueness of  $\mathbf{MP}(L, C_c^\infty(\mathbb{T}^d), \mu)$  is technically less obvious. It can be treated using the localization technique of Theorem A.1 stated in appendix. More precisely, using the notation of Theorem A.1, one can define the covering of  $\mathbb{R}^d$  by the open sets

$$U_k := \{(x_1, \dots, x_d) \in \mathbb{R}^d \mid |x_i - k_i/8| \leq 1/4 \ \forall i = 1 \dots d\}$$

where  $k \in \mathbb{Z}^d$  and then remark that by partition of unity for smooth functions, any  $\varphi \in C_c^\infty(\mathbb{R}^d)$  can be written as a finite sum of smooth functions with compact support in each given  $U_k$ ,  $k \in \mathbb{Z}^d$ .

**2.5. Convergence in distribution.** As we said before, we are interested here in proving convergence in distribution for processes. Let us briefly recall several key results that will be used later.

For completeness, we start by recalling the very classical Prohorov theorem, characterizing relative compactness by tightness (see for example Section 2 in [EK86, Chapter 3]).

**Theorem 2.9** (Prohorov theorem). *Let  $(\mu_\varepsilon)_\varepsilon$  be a family of probability measures on a Polish space  $E$ . Then the following are equivalent:*

- (1)  $(\mu_\varepsilon)_\varepsilon$  is relatively compact for the topology of convergence in distribution.
- (2)  $(\mu_\varepsilon)_\varepsilon$  is tight, that is to say, for any  $\delta > 0$ , there is a compact set  $K_\delta$  such that

$$\inf_\varepsilon \mu_\varepsilon(K_\delta) \geq 1 - \delta.$$

Over the years several relative compactness criteria in Skorokhod space have been developed. We will use the following one [EK86, Theorem 8.6, Chapter 4].

**Theorem 2.10** (Kurtz-Aldous tightness criterion). *Consider a family of stochastic processes  $((X_t^\varepsilon)_{t \geq 0})_\varepsilon$  in  $\mathbb{D}_\mathbb{R}$ . Assume that  $(\text{Law}(X_0^\varepsilon))_\varepsilon$  is tight.  $\forall \delta \in (0, 1)$  and  $T > 0$ , there exists a family of nonnegative random variable  $\Gamma_{\varepsilon, \delta}$ , such that:  $\forall 0 \leq t \leq t+h \leq t+\delta \leq T$*

$$(2.6) \quad \mathbb{E}(|X_{t+h}^\varepsilon - X_t^\varepsilon|^2 | \mathcal{F}_t^{X^\varepsilon}) \leq \mathbb{E}(\Gamma_{\varepsilon, \delta} | \mathcal{F}_t^{X^\varepsilon});$$

with

$$(2.7) \quad \lim_{\delta \rightarrow 0} \sup_\varepsilon \mathbb{E}(\Gamma_{\varepsilon, \delta}) = 0.$$

*Then the family of distributions  $(\text{Law}((X_t^\varepsilon)_{t \geq 0}))_\varepsilon$  is tight.*

*Remark 2.11* (On using sequences). If  $\varepsilon > 0$  is a real number and that instead of (2.7), one considers the condition  $\lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0^+} \mathbb{E}(\Gamma_{\varepsilon, \delta}) = 0$ , then the conclusion becomes the following:  $(\text{Law}((X_t^{\varepsilon_n})_{t \geq 0}))_{\varepsilon_n}$  is tight for any  $(\varepsilon_n)_{n \geq 1}$ -sequence such that  $\varepsilon_n > 0$  and  $\lim_{n \rightarrow +\infty} \varepsilon_n = 0$ . This version will be the one used in the present paper.

If the processes, say  $(Q_t^\varepsilon)_{t \geq 0}$ , is defined in a general state space  $E$ , it is natural to consider the image processes  $(f(Q_t^\varepsilon))_{t \geq 0}$  for various observables, or test functions,  $f$ . The following result enables us to recover the tightness for the original process from the tightness of the observed processes (Corollary 9.3 Chapter 3 in [EK86]).

**Theorem 2.12** (Tightness from observables). *Let  $E$  be a compact Polish space and  $((Q_t^\varepsilon)_{t \geq 0})_{\varepsilon > 0}$  be a family of stochastic processes in  $\mathbb{D}_E$ . Assume that there is an algebra of test functions  $\mathcal{D} \subset C_b(E)$ , dense for the uniform convergence, such that for any  $f \in \mathcal{D}$ ,  $((f(Q_t^\varepsilon))_{t \geq 0})_{\varepsilon > 0}$  is tight in  $\mathbb{D}_{\mathbb{R}}$ . Then  $(\text{Law}(Q_t^\varepsilon)_{t \geq 0})_{\varepsilon > 0}$  is tight in  $\mathbb{D}_E$ .*

*Remark 2.13.* Again, the above theorem will be used for families indexed by sequences  $(\varepsilon_n)_{n \geq 1}$  such that  $\varepsilon_n > 0$  and  $\lim_{n \rightarrow +\infty} \varepsilon_n = 0$ .

Finally, the following two lemmas will be useful when we considering martingale problems. The first one states that the distribution of jumps of càd-làg processes have atoms in a countable set (see Lemma 7.7 Chapter 3 in [EK86]).

**Lemma 2.14.** *Let  $(X_t)_{t \geq 0}$  be a random process in the Skorokhod path space  $\mathbb{D}_E$ . The set of instants where no jump occurs almost surely:*

$$\mathcal{C}_{\text{Law}(X)} := \{t \in \mathbb{R}^+ | \mathbb{P}(X_{t-} = X_t) = 1\},$$

*has countable complement in  $\mathbb{R}^+$ . In particular, it is a dense set.*

The second one is a very useful way to check whether a process is a martingale or not (see page 174 in Ethier-Kurtz[EK86]).

**Lemma 2.15** (Martingale equivalent condition). *Let  $(M_t)_{t \geq 0}$  and  $(X_t)_{t \geq 0}$  be two càd-làg processes and let  $\mathcal{C}$  be an arbitrary dense subset of  $\mathbb{R}_+$ . Then  $(M_t)_{t \geq 0}$  is  $\mathcal{F}_t^X$ -martingale if and only if*

$$\mathbb{E}[(M_{t_{k+1}} - M_{t_k})\varphi_k(X_{t_k})\dots\varphi_1(X_{t_1})] = 0,$$

*for any time ladder  $t_1 \leq \dots \leq t_{k+1} \in \mathcal{C} \subset \mathbb{R}_+$ ,  $k \geq 1$ , and  $\varphi_1, \dots, \varphi_k \in C_b(E)$ .*

### 3. A GENERAL PERTURBED TEST FUNCTION METHOD

In this section, we consider a sequence of stochastic processes, indexed by a small parameter  $\varepsilon > 0$ , of the form

$$t \mapsto (Q_t^\varepsilon, P_t^\varepsilon) \in \mathbb{T}^d \times \mathbb{R}^d,$$

taking value in the Skorokhod path space  $\mathbb{D}_{\mathbb{T}^d \times \mathbb{R}^d}$  associated with the (Polish) product state space  $\mathbb{T}^d \times \mathbb{R}^d$ . Our goal is to describe a general framework to prove the convergence of the (slow) variables  $Q$  towards a well-identified dynamics. We use standard tightness arguments and characterization through martingale problems, emphasizing the technical role of perturbed test functions.

**3.1. Notation and Assumptions.** For each  $\varepsilon$ , we consider a càd-làg process  $t \mapsto (Q_t^\varepsilon, P_t^\varepsilon) \in \mathbb{T}^d \times \mathbb{R}^d$ . The natural filtration of the full process and the process  $(Q_t^\varepsilon)_{t \geq 0}$  are denoted respectively by  $\mathcal{F}_t^{Q^\varepsilon, P^\varepsilon} := \sigma((Q_s^\varepsilon, P_s^\varepsilon), 0 \leq s \leq t)$ , and  $\mathcal{F}_t^{Q^\varepsilon} := \sigma(Q_s^\varepsilon, 0 \leq s \leq t)$ . We now state the key assumptions that will imply convergence in distribution of the process  $(Q_t^\varepsilon)_{t \geq 0}$  towards the solution of a martingale problem.



**Assumption 3.1** (Generator of the process  $(Q_t^\varepsilon, P_t^\varepsilon)$ ). *There exists a linear operator  $L_\varepsilon$  acting on  $C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$  which is the extended Markov generator of  $(Q_t^\varepsilon, P_t^\varepsilon)_{t \geq 0}$  in the sense that, for all  $f \in C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$ ,  $L_\varepsilon f$  is locally bounded and*

$$t \mapsto M_t^\varepsilon(f) := f(Q_t^\varepsilon, P_t^\varepsilon) - f(Q_0^\varepsilon, P_0^\varepsilon) - \int_0^t L_\varepsilon f(Q_s^\varepsilon, P_s^\varepsilon) ds$$

*is a  $(\mathcal{F}_t^{Q^\varepsilon, P^\varepsilon})_{t \geq 0}$ -local martingale.*

**Assumption 3.2** (The limit process). *There exists a linear operator  $L$  mapping  $C^\infty(\mathbb{T}^d)$  to  $C(\mathbb{T}^d)$  such that the martingale problem  $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$  is well-posed for any initial condition  $\mu$ .*

**Assumption 3.3** (Initial condition). *The initial condition  $(\text{Law}(Q_0^\varepsilon))_{\varepsilon > 0}$  converge to a limit  $\mu_0$ , when  $\varepsilon \rightarrow 0$ .*

**Assumption 3.4** (Existence of perturbed test functions). *For all  $f \in C^\infty(\mathbb{T}^d)$ , there exists a perturbed test function  $f_\varepsilon \in C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$ , such that for all  $T$ , the rest terms*

$$R_{1,t}^\varepsilon(f) := |f(Q_t^\varepsilon) - f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)| \quad \text{and} \quad R_{2,t}^\varepsilon(f) := |Lf(Q_t^\varepsilon) - L_\varepsilon f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)|$$

*satisfy the following bounds:*

$$(3.1) \quad \lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{0 \leq t \leq T} R_{1,t}^\varepsilon(f) \right) = 0,$$

$$(3.2) \quad \lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \int_0^T R_{2,t}^\varepsilon(f) dt \right) = 0.$$

**3.2. The general convergence theorem.** We are now in position to state our main abstract result.

**Theorem 3.5.** *Under the Assumptions 3.1, 3.2, 3.3, and 3.4, the family  $(\text{Law}(Q_t^\varepsilon)_{t \geq 0})_{\varepsilon > 0}$  converges when  $\varepsilon \rightarrow 0$  to the unique solution of martingale problem  $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ .*

The proof follows the classical pattern, in two steps: we first prove that the processes  $Q_t^\varepsilon$  are relatively compact in  $\mathbb{D}_{\mathbb{T}^d}$ ; then we show that any possible limit must solve the martingale problem  $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ .

**3.2.1. Step one: The proof of tightness.** We want to prove that for each sequence  $(\varepsilon_n)_{n \geq 1}$  satisfying  $\lim_n \varepsilon_n = 0$ ,  $(\text{Law}(Q_t^{\varepsilon_n}))_{n \geq 1}$  is tight. By Theorem 2.12, it is enough to prove the tightness of  $(\text{Law}(f(Q_t^{\varepsilon_n})))_{n \geq 1}$  for all  $f \in C^\infty(\mathbb{T}^d)$ . The latter fact will follow from Theorem 2.10, if we are able to construct, for any function  $f \in C^\infty(\mathbb{T}^d)$  and any  $\varepsilon, \delta > 0$  and any  $T > 0$ , a random variable  $\Gamma_{\varepsilon, \delta}(f)$  such that for all  $0 \leq t \leq t+h \leq t+\delta \leq T$ , one has

$$(3.3) \quad \mathbb{E} \left[ (f(Q_{t+h}^\varepsilon) - f(Q_t^\varepsilon))^2 \middle| \mathcal{F}_t^{Q^\varepsilon} \right] \leq \mathbb{E} \left[ \Gamma_{\varepsilon, \delta}(f) \middle| \mathcal{F}_t^{Q^\varepsilon} \right],$$

$$(3.4) \quad \text{where} \quad \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \geq 0} \mathbb{E} [\Gamma_{\varepsilon, \delta}(f)] = 0.$$

We claim that the following variant:

**Lemma 3.6.** *For any  $g \in C^\infty(\mathbb{T}^d)$ , and any  $\delta, \varepsilon, T > 0$ , there exists a random variable  $\Gamma'_{\varepsilon, \delta}(g)$  such that for all  $0 \leq t \leq t+h \leq t+\delta \leq T$ ,*

$$(3.5) \quad \left| \mathbb{E} \left[ g(Q_{t+h}^\varepsilon) - g(Q_t^\varepsilon) \middle| \mathcal{F}_t^{Q^\varepsilon} \right] \right| \leq \mathbb{E} \left[ \Gamma'_{\varepsilon, \delta}(g) \middle| \mathcal{F}_t^{Q^\varepsilon} \right],$$

$$(3.6) \quad \text{where} \quad \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \geq 0} \mathbb{E} [\Gamma'_{\varepsilon, \delta}(g)] = 0.$$

is a sufficient condition. Indeed, the required estimates (3.3), (3.4) will follow easily from the basic decomposition

$$(f(Q_t^\varepsilon) - f(Q_{t+h}^\varepsilon))^2 = (f(Q_{t+h}^\varepsilon))^2 - (f(Q_t^\varepsilon))^2 - 2f(Q_t^\varepsilon)(f(Q_{t+h}^\varepsilon) - f(Q_t^\varepsilon)).$$

since we get

$$(3.7) \quad \mathbb{E} \left[ (f(Q_{t+h}^\varepsilon) - f(Q_t^\varepsilon))^2 \middle| \mathcal{F}_t^{Q^\varepsilon} \right] \leq \mathbb{E} \left[ \Gamma'_{\varepsilon, \delta}(f^2) \middle| \mathcal{F}_t^{Q^\varepsilon} \right] + 2\|f\|_\infty \mathbb{E} \left[ \Gamma'_{\varepsilon, \delta}(f) \middle| \mathcal{F}_t^{Q^\varepsilon} \right],$$

and it is enough to let  $\Gamma_{\varepsilon, \delta}(f) = \Gamma'_{\varepsilon, \delta}(f^2) + 2\|f\|_\infty \Gamma'_{\varepsilon, \delta}(f)$  to conclude.

Let us now prove the Lemma 3.6. Let  $g$  be an arbitrary smooth function, and let  $g_\varepsilon$  be the perturbed test function given by Assumption 3.4. An elementary rewriting leads to

$$(3.8) \quad \begin{aligned} g(Q_{t+h}^\varepsilon) - g(Q_t^\varepsilon) &= (g(Q_{t+h}^\varepsilon) - g_\varepsilon(Q_{t+h}^\varepsilon, P_{t+h}^\varepsilon)) - (g(Q_t^\varepsilon) - g_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)) \\ &\quad - \int_t^{t+h} (Lg(Q_s^\varepsilon) - L_\varepsilon g_\varepsilon(Q_s^\varepsilon, P_s^\varepsilon)) ds + \int_t^{t+h} Lg(Q_s^\varepsilon) ds \\ &\quad - M_t^\varepsilon(g_\varepsilon) + M_{t+h}^\varepsilon(g_\varepsilon), \end{aligned}$$

where  $(M_t^\varepsilon(g_\varepsilon))_{t \geq 0}$  is a local  $\mathcal{F}^{Q^\varepsilon, P^\varepsilon}$ -martingale by Assumption 3.1. Let  $\tau_n$  be an associated localizing sequence of stopping times. Applying (3.8) at times  $t \wedge \tau_n$  and  $(t+h) \wedge \tau_n$ , we get

$$\begin{aligned} &g(Q_{(t+h) \wedge \tau_n}^\varepsilon) - g(Q_{t \wedge \tau_n}^\varepsilon) \\ &= g(Q_{(t+h) \wedge \tau_n}^\varepsilon) - g_\varepsilon(Q_{(t+h) \wedge \tau_n}^\varepsilon, P_{(t+h) \wedge \tau_n}^\varepsilon) - (g(Q_{t \wedge \tau_n}^\varepsilon) - g_\varepsilon(Q_{t \wedge \tau_n}^\varepsilon, P_{t \wedge \tau_n}^\varepsilon)) \\ &\quad - \int_t^{t+h} (Lg(Q_s^\varepsilon) - L_\varepsilon g_\varepsilon(Q_s^\varepsilon, P_s^\varepsilon)) \mathbf{1}_{s \leq \tau_n} ds + \int_t^{t+h} Lg(Q_s^\varepsilon) \mathbf{1}_{s \leq \tau_n} ds \\ &\quad - M_{t \wedge \tau_n}^\varepsilon(g_\varepsilon) + M_{(t+h) \wedge \tau_n}^\varepsilon(g_\varepsilon). \end{aligned}$$

Taking the conditional expectation with respect to  $\mathcal{F}_t^{Q^\varepsilon}$ , the martingale terms cancel out, and we get:

$$\begin{aligned} &\left| \mathbb{E} \left[ g(Q_{(t+h) \wedge \tau_n}^\varepsilon) - g(Q_{t \wedge \tau_n}^\varepsilon) \middle| \mathcal{F}_t^{Q^\varepsilon} \right] \right| \\ &\leq \left| \mathbb{E} \left[ g(Q_{(t+h) \wedge \tau_n}^\varepsilon) - g_\varepsilon(Q_{(t+h) \wedge \tau_n}^\varepsilon, P_{(t+h) \wedge \tau_n}^\varepsilon) \middle| \mathcal{F}_t^{Q^\varepsilon} \right] \right| \\ &\quad + \left| \mathbb{E} \left[ g(Q_{t \wedge \tau_n}^\varepsilon) - g_\varepsilon(Q_{t \wedge \tau_n}^\varepsilon, P_{t \wedge \tau_n}^\varepsilon) \middle| \mathcal{F}_t^{Q^\varepsilon} \right] \right| \\ &\quad + \int_t^{t+h} \left| \mathbb{E} \left[ Lg(Q_s^\varepsilon) - L_\varepsilon g_\varepsilon(Q_s^\varepsilon, P_s^\varepsilon) \middle| \mathcal{F}_t^{Q^\varepsilon} \right] \right| ds + h \sup_{q \in \mathbb{T}^d} |Lg(q)| \\ &\leq \mathbb{E} \left[ R_{1, (t+h) \wedge \tau_n}^\varepsilon + R_{1, t \wedge \tau_n}^\varepsilon \middle| \mathcal{F}_t^{Q^\varepsilon} \right] + \int_t^{t+h} \mathbb{E} \left[ R_{2, s}^\varepsilon \middle| \mathcal{F}_t^{Q^\varepsilon} \right] ds + \delta \sup_{q \in \mathbb{T}^d} |Lg(q)| \\ &\leq 2\mathbb{E} \left[ \sup_{s \in [0, T]} R_{1, s}^\varepsilon \middle| \mathcal{F}_t^{Q^\varepsilon} \right] + \int_0^T \mathbb{E} \left[ R_{2, s}^\varepsilon \middle| \mathcal{F}_t^{Q^\varepsilon} \right] ds + \delta \sup_{q \in \mathbb{T}^d} |Lg(q)|. \end{aligned}$$

The right hand side does not depend on  $n$  any longer. On the left hand side, we apply dominated convergence for  $n \rightarrow \infty$  to get

$$\left| \mathbb{E} \left[ g(Q_{t+h}^\varepsilon) - g(Q_t^\varepsilon) \middle| \mathcal{F}_t^{Q^\varepsilon} \right] \right| \leq \mathbb{E} \left[ \Gamma'_{\varepsilon, \delta}(g) \middle| \mathcal{F}_t^{Q^\varepsilon} \right]$$

for  $\Gamma'_{\varepsilon, \delta}(g) = 2 \sup_{[0, T]} R_{1, t}^\varepsilon + \int_0^T R_{2, s}^\varepsilon ds + \delta \|Lg\|_\infty$ . The controls on the rest terms given by Assumption 3.4, and the continuity of  $Lg$  (Assumption 3.2) ensure that

$$\lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \Gamma'_{\varepsilon, \delta}(g) = 0,$$

and the proof of tightness is concluded.

**3.2.2. Step two: identification of the limit.** In this step, we suppose that a sequence  $Q_t^n = Q_t^{\varepsilon_n}$  converges in distribution to a limit  $Q_t^0$ , and we prove that necessarily,  $Q^0$  solves the martingale problem for the generator  $L$ .

Let  $f \in C^\infty(\mathbb{T}^d)$ , we have to check that

$$(3.9) \quad M_t(Q_t^0) := f(Q_t^0) - f(Q_0^0) - \int_0^t Lf(Q_s^0)ds$$

is a martingale with respect to  $\mathcal{F}_t^{Q^0} = \sigma(Q_s^0, 0 \leq s \leq t)$ . Consider a time sequence  $0 \leq t_1 \leq \dots \leq t_p \leq t_{p+1}$  for  $p \geq 1$ , taken in the continuity set  $\mathcal{C}_{\text{Law}}(Q)$  given by Lemma 2.14. Recall that  $\mathcal{C}_{\text{Law}}(Q)$  is dense in  $\mathbb{R}$ . Let  $\varphi_1, \dots, \varphi_p \in C_b(\mathbb{T}^d)$  be  $p$  test functions. By Lemma 2.15, it is enough to prove that

$$I_0 := \mathbb{E} \left[ \left( f(Q_{t_{p+1}}^0) - f(Q_{t_p}^0) - \int_{t_p}^{t_{p+1}} Lf(Q_s^0)ds \right) \varphi_1(Q_{t_1}^0) \cdots \varphi_p(Q_{t_p}^0) \right] = 0.$$

Let  $I_\varepsilon$  be the corresponding quantity for  $\varepsilon > 0$ , that is,

$$I_\varepsilon := \mathbb{E} \left[ \left( f(Q_{t_{p+1}}^\varepsilon) - f(Q_{t_p}^\varepsilon) - \int_{t_p}^{t_{p+1}} Lf(Q_s^\varepsilon)ds \right) \varphi_1(Q_{t_1}^\varepsilon) \cdots \varphi_p(Q_{t_p}^\varepsilon) \right].$$

Let us first show that  $I_\varepsilon$  converges to 0. We first condition on  $\mathcal{F}_{t_p}^{Q^\varepsilon}$  to get:

$$\begin{aligned} |I_\varepsilon| &\leq \mathbb{E} \left[ \mathbb{E} \left[ \left| f(Q_{t_{p+1}}^\varepsilon) - f(Q_{t_p}^\varepsilon) - \int_{t_p}^{t_{p+1}} Lf(Q_s^\varepsilon)ds \right| \middle| \mathcal{F}_{t_p}^{Q^\varepsilon} \right] |\varphi_1(Q_{t_1}^\varepsilon)| \cdots |\varphi_p(Q_{t_p}^\varepsilon)| \right] \\ &\leq \mathbb{E} \left[ \mathbb{E} \left[ \left| f(Q_{t_{p+1}}^\varepsilon) - f(Q_{t_p}^\varepsilon) - \int_{t_p}^{t_{p+1}} Lf(Q_s^\varepsilon)ds \right| \middle| \mathcal{F}_{t_p}^{Q^\varepsilon} \right] \|\varphi_1\|_\infty \cdots \|\varphi_p\|_\infty \right]. \end{aligned}$$

Using again the perturbed test function  $f_\varepsilon$  and the decomposition (3.8), we get by the same localization argument as in Step 1 that

$$|I_\varepsilon| \leq \mathbb{E} \left[ R_{1,t_{p+1}}^\varepsilon(f) + R_{1,t_p}^\varepsilon(f) + \int_{t_p}^{t_{p+1}} R_{2,s}^\varepsilon(f) \right] \|\varphi_1\|_\infty \cdots \|\varphi_p\|_\infty.$$

The estimates on the rest term from Assumption 3.4 then imply that  $I_\varepsilon \rightarrow 0$ .

Let us now prove that  $I_\varepsilon$  converges to  $I_0$ . Let  $\Phi : \mathbb{D}_{\mathbb{T}^d} \rightarrow \mathbb{R}$  be the functional

$$\Phi : (q_t)_{t \geq 0} \mapsto \left( f(q_{t_{p+1}}) - f(q_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(q_s)ds \right) \varphi_1(q_{t_1}) \cdots \varphi_p(q_{t_p})$$

so that  $I_\varepsilon = \mathbb{E}[\Phi((Q_t^\varepsilon)_{t \geq 0})]$  and  $I_0 = \mathbb{E}[\Phi((Q_t^0)_{t \geq 0})]$ . Let us first check that, if  $q^0 \in \mathbb{D}_{\mathbb{T}^d}$  satisfies  $q_{t_k}^0 = q_{t_k}^0$  for each  $1 \leq k \leq p+1$ , then the functional  $\Phi$  is continuous at the trajectory  $q^0$ . Indeed, since  $Lf$  is continuous and bounded by Assumption 3.2, Lemma 2.1 shows that the map  $(q_t)_{t \geq 0} \mapsto \int_{t_p}^{t_{p+1}} Lf(q_s)ds$  is continuous with respect to Skorokhod topology; moreover, by assumption,  $q^0$  is continuous at the time  $t_k$  for each  $1 \leq k \leq p+1$ , so the map  $(q_t)_{t \geq 0} \mapsto \varphi_k(q_{t_k})$  is continuous at  $q^0 \in \mathbb{D}_{\mathbb{T}^d}$ .

Let now  $(\varepsilon_n)_{n \geq 1}$  be any sequence such that  $\varepsilon_n \rightarrow 0$  and  $(Q_t^{\varepsilon_n})_{t \geq 0}$  converges in distribution to  $(Q_t^0)_{t \geq 0}$ . The Skorokhod representation theorem (Theorem 1.8 in [EK86, Chapter 3]) ensures that one can construct a probability space where the distribution of  $(Q_t^{\varepsilon_n})_{t \geq 0}$  for each  $n$  is unchanged but for which  $\lim_{n \rightarrow +\infty} Q^{\varepsilon_n} = Q^0$  almost surely in  $\mathbb{D}_{\mathbb{T}^d}$ . Since  $t_k \in \mathcal{C}_{\text{Law}}(Q^0)$  for each  $k = 1 \dots p+1$ ,  $\Phi$  is almost surely continuous at  $Q^0$  and we can apply the dominated convergence theorem to obtain  $\lim_{n \rightarrow +\infty} I_{\varepsilon_n} = I_0$ . Since the choice of the vanishing sequence  $(\varepsilon_n)_{n \geq 1}$  is arbitrary, we conclude that  $\lim_{\varepsilon \rightarrow 0} I_\varepsilon = I_0$ . The limit process thus solves the martingale problem  $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ .

**3.2.3. Conclusion.** For each sequence  $(\varepsilon_n)_{n \geq 1}$  satisfying  $\lim_n \varepsilon_n = 0$ , we have proven that  $(\text{Law}(Q_t^{\varepsilon_n}))_{n \geq 1}$  is tight and that any converging subsequence is solution to the martingale problem  $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ . By uniqueness of the latter according to Assumption 3.2, this identifies the limit, showing that  $(\text{Law}(Q_t^{\varepsilon_n}))_{n \geq 1}$  converges to the solution of  $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ . Since the sequence  $(\varepsilon_n)_{n \geq 1}$  is arbitrary and convergence in distribution is metrizable,  $(\text{Law}(Q_t^\varepsilon))_{\varepsilon > 0}$  also converges to the solution of  $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ , proving Theorem 3.5.

#### 4. OVERDAMPED LIMIT OF THE LANGEVIN DYNAMICS

In the section, we will use the perturbed test function method presented in last section to prove Theorem 1.1. We will first state the key estimates on  $(|P_t^\varepsilon|)_{t \geq 0}$ . These estimates are then used to check the assumptions of our general Theorem 3.5 in the specific case of Langevin processes. In a last section we will detail the proof of the key estimates.

**4.1. Some moments estimates for Langevin processes.** We start by giving a few facts about the solution to the Langevin SDE (1.1). We first check that the operator  $L_\varepsilon$  acting on  $C^\infty(\mathbb{T}^d, \mathbb{R}^d)$  by

$$L_\varepsilon f(q, p) := \frac{1}{\varepsilon^2} \left( \frac{1}{\beta} \Delta_p f - p \cdot \nabla_p f \right) + \frac{1}{\varepsilon} (p \cdot \nabla_q f - \nabla_q V_\varepsilon \cdot \nabla_p f)$$

is the generator the process, in the sense that Assumption 3.1 holds.

**Proposition 4.1.** *If  $(Q_t^\varepsilon, P_t^\varepsilon)_{t \geq 0}$  is a weak solution of the Langevin SDE (1.1), then for any smooth function  $f : \mathbb{T}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ , the process*

$$t \mapsto M_t^\varepsilon(f) = f(Q_t^\varepsilon, P_t^\varepsilon) - f(Q_0^\varepsilon, P_0^\varepsilon) - \int_0^t L_\varepsilon f(Q_s^\varepsilon, P_s^\varepsilon) ds,$$

*is a  $(\mathcal{F}_t^{Q^\varepsilon, P^\varepsilon})_{t \geq 0}$ -local martingale.*

*Proof.* This is a very classical result. By Itô calculus we write

$$df_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon) = L_\varepsilon f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon) dt + \frac{1}{\varepsilon} \sqrt{2\beta^{-1}} \nabla_p f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon) dW_t.$$

Defining the sequence of  $(\mathcal{F}_t^{Q^\varepsilon, P^\varepsilon})_{t \geq 0}$ -stopping time

$$(4.1) \quad \tau_n = \inf\{t \geq 0, |P_t^\varepsilon| \geq n\},$$

which converge almost surely to infinity, we obtain that

$$M_t^{\varepsilon, n}(f_\varepsilon) := \frac{1}{\varepsilon} \sqrt{2\beta^{-1}} \int_0^t \nabla_p f_\varepsilon(Q_s^\varepsilon, P_s^\varepsilon) 1_{s \leq \tau_n} dW_s$$

is a  $(\mathcal{F}_t^{Q^\varepsilon, P^\varepsilon})_{t \geq 0}$ -martingale for any  $n \geq 0$ , which is the definition of a local martingale.  $\square$

We now state several bounds on the momentum variable  $P_t^\varepsilon$ , which are the key technical estimates needed later to control the rest terms appearing in the perturbed test function method. For any continuous  $V : \mathbb{T}^d \rightarrow \mathbb{R}$  we denote by  $\text{osc}(V)$  the oscillation defined by

$$\text{osc}(V) = \max V - \min V.$$

**Lemma 4.2** (Propagation of moments). *For any  $\gamma \geq 1$ , any  $M > 0$  and any  $\beta > 0$ , there is a numerical constant  $C(\gamma, M, \beta)$  such that for any  $\varepsilon > 0$ , if  $\text{osc}(V_\varepsilon) \leq M$ , then*

$$(4.2) \quad \sup_{t \geq 0} \mathbb{E} \left[ |P_t^\varepsilon|^{2\gamma} \right] \leq C(\gamma, M, \beta) \left( \mathbb{E} \left[ |P_0^\varepsilon|^{2\gamma} \right] + 1 \right).$$

**Lemma 4.3** (Moment of suprema). *For any  $M > 0$ , any  $\beta > 0$  and any  $T > 0$ , there is a numerical constant  $C(M, \beta, T)$  such that for any  $\varepsilon \in (0, 1)$ , if  $\text{osc}(V_\varepsilon) \leq M$ , then*

$$(4.3) \quad \mathbb{E} \left[ \sup_{0 \leq t \leq T} |P_t^\varepsilon|^2 \right] \leq \mathbb{E} [|P_0^\varepsilon|^2] + \frac{1}{\varepsilon} C(M, \beta, T) \left( \mathbb{E} [|P_0^\varepsilon|^2] + 1 \right)^{1/2}.$$

In particular, if  $\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \mathbb{E} [|P_0^\varepsilon|^2] = 0$ , then

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \mathbb{E} \left[ \sup_{0 \leq t \leq T} |P_t^\varepsilon|^2 \right] = 0.$$

The proofs of these estimates use classical techniques of stochastic calculus and are postponed to Section 4.3.

**4.2. The perturbed test functions in the Langevin case.** In this section we apply the general method described in Section 3 to the specific Langevin case, in order to prove Theorem 1.1.

We will use the following standard notation for multidimensional derivatives:

$$\nabla^k f(p_1, \dots, p_k) := \sum_{i_1, \dots, i_k=1}^d \partial_{i_1} \dots \partial_{i_k} f \times p_1^{i_1} \times \dots \times p_k^{i_k}$$

where in the above  $p_1, \dots, p_k \in \mathbb{R}^d$ . Note that as usual  $\Delta f = \text{Tr}(\nabla^2 f)$ .

We first construct explicitly, for any  $f \in C^\infty(\mathbb{T}^d)$ , a perturbed test function  $f_\varepsilon \in C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$ . Let us look for  $f_\varepsilon$  in the following form (see [PSV77])

$$(4.4) \quad f_\varepsilon(q, p) = f(q) + \varepsilon g_1(q, p) + \varepsilon^2 g_2(q, p).$$

Applying the generator  $L_\varepsilon$ , using the fact that  $f$  does not depend on  $p$ , and grouping terms with respect to powers of  $\varepsilon$ , we get

$$\begin{aligned} L_\varepsilon f_\varepsilon(q, p) &= \frac{1}{\varepsilon} p \cdot \nabla_q [f(q) + \varepsilon g_1(q, p) + \varepsilon^2 g_2(q, p)] - \frac{1}{\varepsilon} \nabla_q V(q) \cdot \nabla_p [\varepsilon g_1(q, p) + \varepsilon^2 g_2(q, p)] \\ &\quad - \frac{1}{\varepsilon^2} p \cdot \nabla_p [\varepsilon g_1(q, p) + \varepsilon^2 g_2(q, p)] + \frac{1}{\varepsilon^2 \beta} \Delta_p [\varepsilon g_1(q, p) + \varepsilon^2 g_2(q, p)] \\ &= \frac{1}{\varepsilon} \left( p \cdot \nabla_q f - p \cdot \nabla_p g_1 + \frac{1}{\beta} \Delta_p g_1 \right) \\ &\quad + \left( p \cdot \nabla_q g_1 - \nabla_q V \cdot \nabla_p g_1 - p \cdot \nabla_p g_2 + \frac{1}{\beta} \Delta_p g_2 \right) \\ (4.5) \quad &+ \varepsilon (p \cdot \nabla_q g_2 - \nabla_p g_2 \cdot \nabla_q V_\varepsilon). \end{aligned}$$

In order for  $L_\varepsilon f_\varepsilon$  to converge to  $Lf$ , the  $\varepsilon^{-1}$ -order terms should vanish, and the  $\varepsilon^0$ -order terms should converge at least formally to  $L(f)$ . As a consequence  $g_1$  and  $g_2$  should solve the following equations:

$$(4.6) \quad 0 = p \cdot \nabla_q f - p \cdot \nabla_p g_1 + \frac{1}{\beta} \Delta_p g_1,$$

$$(4.7) \quad Lf(q) = p \cdot \nabla_q g_1 - \nabla_q V \cdot \nabla_p g_1 - p \cdot \nabla_p g_2 + \frac{1}{\beta} \Delta_p g_2.$$

The function  $g_1(q, p) = p \cdot \nabla_q f(q)$  clearly solves (4.6). With this choice, (4.7) becomes

$$Lf(q) = \nabla_q^2 f(p, p) - \nabla_q V \cdot \nabla_q f - p \cdot \nabla_p g_2 + \frac{1}{\beta} \Delta_p g_2.$$

Since  $Lf(q) = \frac{1}{\beta} \Delta_q f - \nabla_q V \cdot \nabla_q f$ , it is easy to check that  $g_2(q, p) = \frac{1}{2} \nabla_q^2 f(p, p)$  solves the equation.

Therefore, in view of Eq. (4.4), we defined the perturbed test function by :

$$(4.8) \quad f_\varepsilon(q, p) = f(q) + \varepsilon p \cdot \nabla_q f + \frac{1}{2} \varepsilon^2 \nabla_q^2 f(p, p).$$

With this choice, we get using previous calculations and the last line of (4.5)

$$(4.9) \quad \begin{aligned} L_\varepsilon f_\varepsilon(q, p) - Lf(q) &= (\nabla_q V - \nabla_q V_\varepsilon) \cdot \nabla_q f + \varepsilon (p \cdot \nabla_q g_2 - \nabla_p g_2 \cdot \nabla_q V_\varepsilon) \\ &= (\nabla_q V - \nabla_q V_\varepsilon) \cdot \nabla_q f + \frac{1}{2} \varepsilon \left( \nabla_q^3 f(q)(p, p, p) - \nabla_q^2 f(p, \nabla_q V_\varepsilon) \right). \end{aligned}$$

We now need to show that Assumption 3.4 holds for this choice of a perturbed test function, that is, we want to show that the differences  $f_\varepsilon - f$  and  $L_\varepsilon f_\varepsilon - Lf$  are small in the following appropriate sense. Recalling the notation

$$R_{1,t}^\varepsilon(f) = |f(Q_t^\varepsilon) - f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)|, \quad R_{2,t}^\varepsilon(f) = |Lf(Q_t^\varepsilon) - L_\varepsilon f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)|,$$

we need to prove that

$$(4.10) \quad \lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{0 \leq t \leq T} R_{1,t}^\varepsilon(f) \right) = 0,$$

$$(4.11) \quad \lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \int_0^T R_{2,t}^\varepsilon(f) dt \right) = 0.$$

Since  $f \in C^\infty(\mathbb{T}^d)$ , there exists a  $C_f = \max(\|\nabla f\|_\infty, \|\nabla^2 f\|_\infty)$  such that for all  $(q, p)$  and all  $\delta \in (0, 1/2)$

$$\begin{aligned} |f_\varepsilon(q, p) - f(q)| &= \varepsilon |p \cdot \nabla_q f(q)| + \frac{1}{2} \varepsilon^2 \left| \nabla_q^2 f(q) \cdot (p, p) \right| \\ &\leq C_f (\varepsilon |p| + \varepsilon^2 |p|^2) \\ &\leq \delta C_f + \frac{1}{\delta} C_f \varepsilon^2 |p|^2, \end{aligned}$$

where we have used that for any  $\delta > 0$ ,  $\varepsilon |p| \leq \frac{1}{2} \delta + \frac{1}{2} \varepsilon^2 |p|^2 / \delta$ . Therefore

$$\mathbb{E} \left[ \sup_{t \in [0, T]} R_{1,t}^\varepsilon(f) \right] \leq \delta C_f + \frac{1}{\delta} C_f \varepsilon^2 \mathbb{E} \left[ \sup_{t \in [0, T]} |P_t^\varepsilon|^2 \right].$$

By assumption,  $\lim_{\varepsilon \rightarrow 0} \varepsilon \mathbb{E} [|P_0^\varepsilon|^3] = 0$ , so  $\varepsilon^2 \mathbb{E} [|P_0^\varepsilon|^2] \leq \varepsilon^{4/3} (\varepsilon \mathbb{E} [|P_0^\varepsilon|^3])^{2/3}$  also goes to zero by Jensen's inequality. By the key Lemma 4.3 this entails that the last term in the previous display disappears in the limit and we get

$$\limsup_{\varepsilon \rightarrow 0} \mathbb{E} \left[ \sup_{t \in [0, T]} R_{1,t}^\varepsilon(f) \right] \leq \delta C_f,$$

which proves (4.10) since  $\delta$  is arbitrary.

We now turn to the proof of (4.11), that is, we want to compare  $L_\varepsilon f_\varepsilon$  and  $Lf$ . By the expression (4.9), we have for some constant  $C_f = \max(\|\nabla f\|_\infty, \|\nabla^2 f\|_\infty, \|\nabla^3 f\|_\infty)$

$$|L_\varepsilon f_\varepsilon(q, p) - Lf(q)| \leq C_f \|\nabla_q V - \nabla_q V_\varepsilon\|_\infty + C_f \varepsilon \left( |p|^3 + \|\nabla_q V_\varepsilon\|_\infty |p| \right).$$

We get rid of the product term with Young's inequality  $ab \leq a^3/3 + \frac{2}{3}b^{3/2} \leq a^3 + b^{3/2}$  and get

$$\mathbb{E} [R_{2,t}^\varepsilon] \leq C_f \|\nabla_q V - \nabla_q V_\varepsilon\|_\infty + \varepsilon C_f \mathbb{E} \left[ 2 |P_t^\varepsilon|^3 + \|\nabla V_\varepsilon\|_\infty^{3/2} \right].$$

We integrate in  $t$  to obtain

$$\int_0^T \mathbb{E} [R_{2,t}^\varepsilon] dt \leq C_f \|\nabla_q V - \nabla_q V_\varepsilon\|_\infty T + \varepsilon C_f T \left( \sup_{t \in [0,T]} \mathbb{E} [2|P_t^\varepsilon|^3] + \|\nabla V_\varepsilon\|_\infty^{3/2} \right).$$

By assumption,  $\lim_{\varepsilon \rightarrow 0} \varepsilon \mathbb{E} [|P_0^\varepsilon|^3] = 0$ , and by the uniform convergence of  $\nabla V_\varepsilon$  to  $\nabla V$  we can find a uniform bound  $M$  such that  $\text{osc}(V_\varepsilon) \leq M$  for all  $\varepsilon$ , so we may apply Lemma 4.2 with  $\gamma = 3/2$  and get

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \sup_{t \in [0,T]} \mathbb{E} [|P_t^\varepsilon|^3] = 0,$$

for any  $T \geq 0$ . Together with the convergence of  $\nabla V_\varepsilon$  to  $\nabla V$  this yields

$$\lim_{\varepsilon \rightarrow 0} \int_0^T \mathbb{E} [R_{2,t}^\varepsilon] dt = 0.$$

from which (4.11) follows.

**4.3. Proofs of the moment bounds.** We now come back to the proofs of the moment bounds (Lemmas 4.2 and 4.3). It will prove useful to work with the Hamiltonian of the system rather than directly with  $P_t^\varepsilon$ . For convenience's sake we assume without loss of generality that  $0 \leq V_\varepsilon(q) \leq \text{osc}(V_\varepsilon)$ .

**Definition 4.4** (Hamiltonian). *We denote by  $H^\varepsilon$  the Hamiltonian of the system:*

$$H^\varepsilon(q, p) = \frac{1}{2} |p|^2 + V_\varepsilon(q).$$

We will also write  $H_t^\varepsilon := H^\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)$ .

By Itô's formula,

$$\begin{aligned} dH_t^\varepsilon &= P_t^\varepsilon dP_t^\varepsilon + \nabla_q V_\varepsilon(Q_t^\varepsilon) dQ_t^\varepsilon + \frac{1}{2} \sum_{i,j=1}^d d\langle (P^\varepsilon)^i, (P^\varepsilon)^j \rangle_t \\ (4.12) \quad &= \left( -\frac{1}{\varepsilon^2} |P_t^\varepsilon|^2 + \frac{1}{\varepsilon^2 \beta} \right) dt + \frac{1}{\varepsilon} \sqrt{2\beta^{-1}} P_t^\varepsilon dW_t \end{aligned}$$

$$(4.13) \quad = \left( -\frac{2}{\varepsilon^2} H_t^\varepsilon + \frac{2}{\varepsilon^2} V_\varepsilon(Q_t^\varepsilon) + \frac{1}{\varepsilon^2 \beta} \right) dt + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} P_t^\varepsilon dW_t.$$

Again, by Itô's formula, we thus get for any smooth function  $(t, h) \mapsto \phi(t, h)$

$$(4.14) \quad d\phi(t, H_t^\varepsilon) = \partial_t \phi(t, H_t^\varepsilon) dt + \partial_h \phi(t, H_t^\varepsilon) dH_t^\varepsilon + \frac{1}{\varepsilon^2 \beta} \partial_h^2 \phi(t, H_t^\varepsilon) |P_t^\varepsilon|^2 dt.$$

*Proof of Lemma 4.2.* Let  $\gamma \geq 1$ . We apply (4.14) to  $\phi(t, x) = e^{\alpha t} h^\gamma$  and plug in (4.13) to get:

$$\begin{aligned} d(e^{\alpha t} (H_t^\varepsilon)^\gamma) &= \gamma (H_t^\varepsilon)^{\gamma-1} \left( \frac{\alpha}{\gamma} H_t^\varepsilon - \frac{2}{\varepsilon^2} H_t^\varepsilon + \frac{2}{\varepsilon^2} V_\varepsilon(Q_t^\varepsilon) + \frac{1}{\varepsilon^2 \beta} \right) e^{\alpha t} dt \\ &\quad + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} \gamma (H_t^\varepsilon)^{\gamma-1} P_t^\varepsilon e^{\alpha t} dW_t + \frac{\gamma(\gamma-1)}{\varepsilon^2 \beta} (H_t^\varepsilon)^{\gamma-2} |P_t^\varepsilon|^2 e^{\alpha t} dt. \end{aligned}$$

The choice

$$\alpha = 2\gamma/\varepsilon^2$$

cancels the higher order term in the first bracket. We integrate in time, multiply by  $e^{-\alpha t}$  and regroup the finite variation terms to get:

$$\begin{aligned} (H_t^\varepsilon)^\gamma &= (H_0^\varepsilon)^\gamma + \int_0^t \left( \gamma (H_s^\varepsilon)^{\gamma-1} \left( \frac{2}{\varepsilon^2} V_\varepsilon(Q_s^\varepsilon) + \frac{1}{\varepsilon^2 \beta} \right) + \frac{\gamma(\gamma-1)}{\varepsilon^2 \beta} (H_s^\varepsilon)^{\gamma-2} |P_s^\varepsilon|^2 \right) e^{-\alpha(t-s)} ds \\ &\quad + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} \int_0^t \gamma (H_s^\varepsilon)^{\gamma-1} P_s^\varepsilon e^{-\alpha(t-s)} dW_s. \end{aligned}$$

Since  $(1/2) |P_s^\varepsilon|^2 \leq H_s^\varepsilon \leq (1/2) |P_s^\varepsilon|^2 + \text{osc}(V_\varepsilon)$ ,

$$(4.15) \quad \begin{aligned} (H_t^\varepsilon)^\gamma &\leq (H_0^\varepsilon)^\gamma + \frac{2\gamma}{\varepsilon^2} \left( \text{osc}(V_\varepsilon) + \frac{\gamma}{\beta} \right) \int_0^t (H_s^\varepsilon)^{\gamma-1} e^{-\alpha(t-s)} ds \\ &\quad + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} \int_0^t (H_s^\varepsilon)^{\gamma-1} P_s^\varepsilon e^{-\alpha(t-s)} dW_s. \end{aligned}$$

To deal with the unboundedness of the momentum  $P$ , we define the following stopping times:

$$(4.16) \quad \tau_n := \inf\{t : |P_t| = n\}.$$

When  $s \leq \tau_n$ , we have  $|P_s^\varepsilon| \leq n$  and  $H_s^\varepsilon \leq (\text{osc}(V_\varepsilon) + \frac{n^2}{2})$ . This entails that  $t \mapsto \int_0^{t \wedge \tau_n} (H_s^\varepsilon)^{\gamma-1} P_s^\varepsilon dW_s$  is martingale. Writing (4.15) at time  $t \wedge \tau_n$  and taking expectations, the martingale part disappears; recalling that  $\alpha = 2\gamma/\varepsilon^2$  we get

$$\begin{aligned} \mathbb{E}[(H_{t \wedge \tau_n}^\varepsilon)^\gamma] &\leq \mathbb{E}[(H_0^\varepsilon)^\gamma] + \left( \text{osc}(V_\varepsilon) + \frac{\gamma}{\beta} \right) \alpha \mathbb{E} \left[ \int_0^{t \wedge \tau_n} (H_s^\varepsilon)^{\gamma-1} e^{-\alpha(t-s)} ds \right] \\ &\leq \mathbb{E}[(H_0^\varepsilon)^\gamma] + \left( \text{osc}(V_\varepsilon) + \frac{\gamma}{\beta} \right) \sup_{s \leq t} \mathbb{E}[(H_s^\varepsilon)^{\gamma-1}] ds. \end{aligned}$$

Sending  $n$  to infinity, we apply Fatou's lemma to get

$$\mathbb{E}[(H_t^\varepsilon)^\gamma] \leq \mathbb{E}[(H_0^\varepsilon)^\gamma] + \left( \text{osc}(V_\varepsilon) + \frac{\gamma}{\beta} \right) \sup_{s \leq t} \mathbb{E}[(H_s^\varepsilon)^{\gamma-1}],$$

and thus

$$(4.17) \quad \sup_{t \geq 0} \mathbb{E}[(H_t^\varepsilon)^\gamma] \leq \mathbb{E}[(H_0^\varepsilon)^\gamma] + \left( \text{osc}(V_\varepsilon) + \frac{\gamma}{\beta} \right) \sup_{t \geq 0} \mathbb{E}[(H_s^\varepsilon)^{\gamma-1}].$$

We are now ready to conclude. Say that  $\gamma$  is good if there exists a  $C(\gamma, M, \beta)$  such that for all  $\varepsilon$ ,

$$\sup_t \mathbb{E}[(H_t^\varepsilon)^\gamma] \leq C(\gamma, M, \beta)(1 + \mathbb{E}[(H_0^\varepsilon)^\gamma]),$$

whenever  $\text{osc}(V_\varepsilon) \leq M$ . The bound (4.17) immediately shows that  $\gamma = 1$  is good. If  $\gamma$  is good and  $\gamma \leq \gamma' \leq \gamma + 1$ , using the elementary inequality  $x^a \leq 1 + x^b$  valid for any  $x > 0$  and any  $1 \leq a < b$ , we get

$$\begin{aligned} \sup_{t \geq 0} \mathbb{E}[(H_t^\varepsilon)^{\gamma'}] &\leq \mathbb{E}[(H_0^\varepsilon)^{\gamma'}] + \left( \text{osc}(V_\varepsilon) + \frac{\gamma'}{\beta} \right) \sup_{t \geq 0} \mathbb{E}[(H_s^\varepsilon)^{\gamma'-1}] \\ &\leq \mathbb{E}[(H_0^\varepsilon)^{\gamma'}] + \left( M + \frac{\gamma'}{\beta} \right) \left( 1 + \sup_{t \geq 0} \mathbb{E}[(H_s^\varepsilon)^\gamma] \right) \\ &\leq \mathbb{E}[(H_0^\varepsilon)^{\gamma'}] + \left( M + \frac{\gamma'}{\beta} \right) (1 + C(\gamma, M, \beta) \mathbb{E}[(H_0^\varepsilon)^\gamma]) \\ &\leq \mathbb{E}[(H_0^\varepsilon)^{\gamma'}] + \left( M + \frac{\gamma'}{\beta} \right) (1 + C(\gamma, M, \beta) (1 + \mathbb{E}[(H_0^\varepsilon)^{\gamma'}])) \end{aligned}$$

showing that  $\gamma'$  is itself good. Therefore all  $\gamma \geq 1$  are good. Using the bounds  $(1/2)p^2 \leq H^\varepsilon(q, p) \leq (1/2)p^2 + M$  it is easy to translate this into bounds on  $\mathbb{E}[|P_t^\varepsilon|^{2\gamma}]$ , concluding the proof of Lemma 4.2.  $\square$

*Proof of Lemma 4.3.* Let us fix an arbitrary  $T > 0$ , and prove (4.3), that is, prove the existence of a numerical constant  $C(\beta, M, T)$  such for any  $\varepsilon \in (0, 1)$ ,

$$(4.18) \quad \mathbb{E} \left[ \sup_{0 \leq t \leq T} |P_t^\varepsilon|^2 \right] \leq \mathbb{E}[|P_0^\varepsilon|^2] + \frac{1}{\varepsilon} C(\beta, M, T) \left( \mathbb{E}[|P_0^\varepsilon|^2] + 1 \right)^{1/2}$$



whenever  $\text{osc}(V_\varepsilon) \leq M$ . As before, since  $2H_t^\varepsilon - 2M \leq (P_t^\varepsilon)^2 \leq 2H_t^\varepsilon$ , it is enough to prove the statement with  $H_t^\varepsilon$  instead of  $|P_t^\varepsilon|^2$ .

We start by recalling (4.15) for  $\gamma = 1$  and  $\alpha = 2/\varepsilon^2$ :

$$(4.19) \quad H_t^\varepsilon \leq H_0^\varepsilon + \left( \text{osc}(V_\varepsilon) + \frac{1}{\beta} \right) + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} \int_0^t e^{-\alpha(t-s)} P_s^\varepsilon dW_s.$$

Recall that this led by a localization argument to the following bound (4.17):

$$(4.20) \quad \sup_{t \geq 0} \mathbb{E}[H_t^\varepsilon] \leq \mathbb{E}[H_0^\varepsilon] + \left( M + \frac{1}{\beta} \right).$$

In order to control the expectation of the supremum, we must control the stochastic integral. Define  $M_t = \int_0^t P_s^\varepsilon dW_s$  and integrate by parts:

$$\begin{aligned} \left| \int_0^t e^{-\alpha(t-s)} P_s^\varepsilon dW_s \right| &= \left| \int_0^t e^{-\alpha(t-s)} dM_s \right| = \left| M_t - \alpha \int_0^t e^{-\alpha(t-s)} M_s ds \right| \\ &\leq |M_t| + \sup_{s \in [0, t]} |M_s| \\ &\leq 2 \sup_{s \in [0, T]} |M_s|. \end{aligned}$$

Plugging this in (4.19) yields

$$(4.21) \quad \sup_{t \in [0, T]} H_t \leq H_0^\varepsilon + \left( \text{osc}(V_\varepsilon) + \frac{1}{\beta} \right) + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} 2 \sup_{t \in [0, T]} |M_t|.$$

By Doob's martingale maximal inequality, Itô's isometry and the bound (4.20) we get

$$\begin{aligned} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |M_t^\varepsilon|^2 \right] &\leq 4\mathbb{E} [ |M_T^\varepsilon|^2 ] = 4\mathbb{E} \left[ \left| \int_0^T P_s^\varepsilon dW_s \right|^2 \right] = 4\mathbb{E} \left[ \int_0^T (P_s^\varepsilon)^2 ds \right] \\ &\leq 8T \left( \text{osc}(V_\varepsilon) + \sup_{t \in [0, T]} \mathbb{E}[H_t^\varepsilon] \right) \\ &\leq 8T \left( 2 \text{osc}(V_\varepsilon) + \mathbb{E}[H_0^\varepsilon] + \frac{1}{\beta} \right). \end{aligned}$$

Injecting this in (4.21) and applying Cauchy-Schwarz inequality yields

$$(4.22) \quad \mathbb{E} \left[ \sup_{t \in [0, T]} H_t \right] \leq \mathbb{E}[H_0^\varepsilon] + \left( \text{osc}(V_\varepsilon) + \frac{1}{\beta} \right) + 8 \frac{\sqrt{T\beta^{-1}}}{\varepsilon} \left( 2 \text{osc}(V_\varepsilon) + \mathbb{E}[H_0^\varepsilon] + \frac{1}{\beta} \right)^{1/2},$$

concluding the proof of (4.18).  $\square$

#### APPENDIX A. STOPPED MARTINGALE PROBLEM

Let  $E$  be a Polish space. Let  $L$  be a linear operator mapping a given space  $\mathcal{D} \subset C_b(E)$  into bounded measurable functions. Let  $\mu$  be a probability distribution on  $E$ . Let  $U \subset E$  be an open set. A càd-làg process  $(X_t)_{t \geq 0}$  with values in  $E$  solves the *stopped martingale problem* for the generator  $L$  on the space  $\mathcal{D}$  with initial measure  $\mu$  and domain  $U$  — in short,  $X$  solves **sMP**( $L, D(L), \mu, U$ ) — if, denoting

$$\tau_U := \inf \{ t \geq 0 | X_t \notin U \text{ or } X_{t-} \notin U \},$$

(i)  $\text{Law}(X_0) = \mu$ ; (ii)  $X_t = X_{t \wedge \tau_U}$ ; and (iii) if for any  $\varphi \in \mathcal{D}$ ,

$$t \mapsto M_t(\varphi) := \varphi(X_t) - \varphi(X_0) - \int_0^{t \wedge \tau_U} L\varphi(X_s) ds$$

is a martingale with respect to the natural filtration  $(\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t))_{t \geq 0}$ .

Moreover, the stopped martingale problem **sMP**( $L, \mathcal{D}, \mu, U$ ) is said to be *well-posed* if:

- There exists a probability space and a càd-làg process defined on it that solves the stopped martingale problem (existence);
- whenever two processes solve  $\mathbf{sMP}(L, \mathcal{D}, \mu, U)$ , then they have the same distribution on  $\mathbb{D}_E$  (uniqueness).

The following theorem is a synthesis of the localization technique of Theorem 6.1 and 6.2 of [EK86, Chapter 4]. It gives a simple criteria ensuring equivalence of uniqueness between (i) a global martingale problem, and (ii) local stopped martingale problems.

**Theorem A.1.** *Let  $(U_k)_{k \in K}$  be a countable family of open subsets of  $E$  such that  $\bigcup_{k \in K} U_k = E$ . Assume that for any initial  $\nu$ , there exists a solution to  $\mathbf{MP}(L, \mathcal{D}, \mu)$ . Then uniqueness of  $\mathbf{MP}(L, \mathcal{D}, \mu)$  for all  $\mu$  is equivalent to uniqueness of  $\mathbf{sMP}(L, \mathcal{D}, \mu, U_k)$  for all  $\mu$  and all  $k \in K$ .*

## REFERENCES

- [EK86] S. N. Ethier and T. G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986, Characterization and convergence. MR 838085
- [Kha66] R. Z. Khas'minskii, *A limit theorem for the solutions of differential equations with random right-hand sides*, Theory of Probability & Its Applications **11** (1966), no. 3, 390–406.
- [LRS10] T. Lelièvre, M. Rousset, and G. Stoltz, *Free energy computations*, Imperial College Press, London, 2010, A mathematical perspective. MR 2681239
- [PK74] G. C. Papanicolaou and W. Kohler, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math. **27** (1974), 641–668. MR 0368142
- [PS08] G. A. Pavliotis and A. M. Stuart, *Multiscale methods*, Texts in Applied Mathematics, vol. 53, Springer, New York, 2008, Averaging and homogenization. MR 2382139
- [PSV77] G. C. Papanicolaou, D. Stroock, and S. R. S. Varadhan, *Martingale approach to some limit theorems*, ii+120 pp. Duke Univ. Math. Ser., Vol. III. MR 0461684
- [PV73] G. C. Papanicolaou and S. R. S. Varadhan, *A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations*, Comm. Pure Appl. Math. **26** (1973), 497–524. MR 0383530
- [PV01] E. Pardoux and A. Y. Veretennikov, *On the Poisson equation and diffusion approximation. I*, Ann. Probab. **29** (2001), no. 3, 1061–1085. MR 1872736
- [PV03] ———, *On Poisson equation and diffusion approximation. II*, Ann. Probab. **31** (2003), no. 3, 1166–1192. MR 1988467
- [PV05] ———, *On the Poisson equation and diffusion approximation. III*, Ann. Probab. **33** (2005), no. 3, 1111–1133. MR 2135314
- [SSMD82] J. M. Sancho, M. San Miguel, and D. Dürr, *Adiabatic elimination for systems of Brownian particles with nonconstant damping coefficients*, J. Statist. Phys. **28** (1982), no. 2, 291–305. MR 666513
- [Str63] R. L. Stratonovich, *Topics in the theory of random noise. Vol. I: General theory of random processes. Nonlinear transformations of signals and noise*, Revised English edition. Translated from the Russian by Richard A. Silverman, Gordon and Breach Science Publishers, New York-London, 1963. MR 0158437
- [SV07] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, Springer, 2007.