

LIMIT BEHAVIOR OF THE INVARIANT MEASURE FOR LANGEVIN DYNAMICS

BY

GERARDO BARRERA (HELSINKI)

Abstract. We consider the Langevin dynamics on \mathbb{R}^d with an overdamped vector field and driven by multiplicative Brownian noise of small amplitude $\sqrt{\epsilon}$, $\epsilon > 0$. Under suitable assumptions on the vector field and the diffusion coefficient, it is well-known that it has a unique invariant probability measure μ^ϵ . We prove that as ϵ tends to zero, the probability measure $\epsilon^{d/2} \mu^\epsilon(\sqrt{\epsilon} dx)$ converges in the p -Wasserstein distance for $p \in [1, 2]$ to a Gaussian measure with zero-mean vector and non-degenerate covariance matrix which solves a Lyapunov matrix equation. Moreover, the error term is estimated. We emphasize that generically no explicit formula for μ^ϵ can be found.

2020 Mathematics Subject Classification: Primary 60H10; Secondary 34D10, 37M25, 60F05, 49Q22.

Key words and phrases: coupling, Gaussian distribution, invariant distribution, Langevin dynamics, Ornstein–Uhlenbeck process, perturbations of dynamical systems, Wasserstein distance.

1. INTRODUCTION

1.1. The overdamped Langevin dynamics. Random dynamical systems arise in the modeling of a (realistic) physical system subject to noise perturbations from its surrounding environments or from intrinsic uncertainties associated with the system. The Langevin dynamics was introduced by P. Langevin in 1908 in his celebrated article *Sur la théorie du mouvement brownien*, C. R. Acad. Sci. Paris 146, 530–533. It is perhaps one of the most popular models in molecular systems. For details about the history of the Langevin equation, see [31]. For a phenomenological treatment, we recommend the monograph [7].

In the last decades, there have been many applications of Markov chain Monte Carlo methods to complex systems in computer science and statistical physics. Since sampling high-dimensional distributions is typically a difficult task, the use of stochastic equations for sampling has become important in many applications

such as artificial intelligence and Bayesian algorithms. Stochastic algorithms based on Langevin equations have been proposed to simulate and improve the rate of convergence to limiting distributions. For further details we refer to [8, 11, 13, 14, 19, 24, 35] and the references therein.

Differential equations subject to small noise perturbations are one of the classical directions of modern mathematical physics. Let $\epsilon \in (0, 1]$ be a parameter that measures the perturbation strength and let $(B_t)_{t \geq 0}$ be a standard Brownian motion on \mathbb{R}^d . For any (deterministic) $x \in \mathbb{R}^d$ we consider the unique strong solution $(X_t^\epsilon(x))_{t \geq 0}$ of the following stochastic differential equation (SDE for short) on \mathbb{R}^d :

$$(1.1) \quad \begin{cases} dX_t^\epsilon(x) = -F(X_t^\epsilon(x)) dt + \sqrt{\epsilon} \sigma(X_t^\epsilon(x)) dB_t & \text{for any } t \geq 0, \\ X_0^\epsilon(x) = x, \end{cases}$$

where the vector field $F \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ and the diffusion coefficient $\sigma \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^{d \times d})$ satisfy the following assumptions. We assume that 0_d is a fixed point for F , i.e., $F(0_d) = 0_d$, and F satisfies the following hypotheses:

Bakry–Émery condition: There exists a positive constant δ such that

$$(A) \quad \langle F(x_1) - F(x_2), x_1 - x_2 \rangle \geq \delta \|x_1 - x_2\|^2 \quad \text{for any } x_1, x_2 \in \mathbb{R}^d,$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product on \mathbb{R}^d and $\|\cdot\|$ denotes the standard Euclidean norm on \mathbb{R}^d .

Exponential growth condition: There exist positive constants c_0 and c_1 satisfying

$$(B) \quad \|D^2 F(x)\| \leq c_0 e^{c_1 \|x\|^2} \quad \text{for any } x \in \mathbb{R}^d,$$

where $D^2 F$ denotes the second order derivative of F .

For the diffusion coefficient σ we assume the following standard hypotheses:

Lipschitz continuity: There exists a positive constant ℓ such that

$$(C) \quad \|\sigma(x) - \sigma(x_0)\|_F \leq \ell \|x - x_0\| \quad \text{for all } x, x_0 \in \mathbb{R}^d,$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Ellipticity: There is a positive constant κ such that

$$(D) \quad \langle \sigma(x_0) \sigma^*(x_0) x, x \rangle \geq \kappa \|x\|^2 \quad \text{for all } x, x_0 \in \mathbb{R}^d,$$

where $*$ denotes the transpose operator.

Hypotheses (A) and (C) imply the monotone condition (3.14) given in [26, Theorem 3.5, p. 58], and hence the existence and uniqueness of the unique strong solution of (1.1). Along this article, $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space where (1.1) is defined and we denote by \mathbb{E} the expectation with respect to \mathbb{P} .

1.2. Invariant distribution. Existence of invariant measures for stochastic processes is an important feature in probability theory and mathematical physics; and typically they are not easy to describe explicitly. By using [22, Theorem 3.3.4, p. 91], it is not hard to verify that Hypotheses **(A)**, **(C)** and **(D)** yield the existence and uniqueness of an invariant probability measure μ^ϵ (absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d) for the stochastic dynamics (1.1). If in addition $F(x) = \nabla V(x) + b(x)$ and $\sigma(x) = I_d$ for any $x \in \mathbb{R}^d$, where I_d is the $d \times d$ identity matrix, $V : \mathbb{R}^d \rightarrow \mathbb{R}$ is a scalar function and $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a vector field which satisfies the divergence-free condition

$$(1.2) \quad \sum_{j=1}^d \frac{\partial}{\partial x_j} (b(x) \exp(-(2/\epsilon)V(x))) = 0 \quad \text{for any } x = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

one can verify that $\exp(-(2/\epsilon)V(x)) dx$ is a stationary measure for the random dynamics (1.1). However, it might not be a probability measure. Under some appropriate assumptions on V for $\|x\| \gg 1$, the unique invariant probability measure μ^ϵ of (1.1) is of the Gibbs type

$$(1.3) \quad \mu^\epsilon(dx) = \frac{\exp(-(2/\epsilon)V(x))}{Z^\epsilon} dx,$$

where Z^ϵ is the so-called partition function (normalizing constant). See for instance [34, Chapter 2, pp. 21–23]. Using the Laplace method (saddle-point method), the asymptotics as $\epsilon \rightarrow 0^+$ for the density of μ^ϵ can be found; see for instance [2, 18].

If we drop the free-divergence condition (1.2) and replace it by the transversality condition $\langle \nabla V(x), b(x) \rangle = 0$ for all $x \in \mathbb{R}^d$, in [33] for additive noise, a beautiful expansion in ϵ for the density of μ^ϵ is shown. However, this expansion requires smoothness of the so-called Freidlin–Wentzell quasipotential. The latter is a nontrivial mathematical problem since it is expressed by a variational principle. Using calculus of variations, [10] shows various results about the smoothness of the quasipotential under the assumptions of smoothness, boundedness and ellipticity of the coefficients of (1.1). In [9, Section 5] it is proved that the asymptotic expansion given in [33] remains valid in any open set in which the quasipotential is C^2 . For additive noise, and bounded and dissipative vector field F , in [27], by applying Watanabe’s theory and Malliavin calculus, an asymptotic expansion of μ^ϵ has been proved. Later, in [28] it is shown that μ^ϵ can be expanded in Wentzell–Kramers–Brillouin (W.K.B.) type, as $\epsilon \rightarrow 0^+$, in the set in which the quasipotential is C^∞ and each coefficient which appears in the expansion is C^∞ . More recently, in [6], using control-theoretic methods, it is proved that $\mu^\epsilon(dx) \approx \exp(-V_*(x)/\epsilon)$, $\epsilon \ll 1$, where V_* is characterized as the optimal cost of a deterministic control problem. Nevertheless, the control problem is not easy to solve explicitly.

In (1.1) we consider multiplicative noise, and no transversality condition on the vector field F is assumed. Moreover, we do not need that the Gibbs measure (1.3) remains stationary, and no smoothness of μ^ϵ or the Freidlin–Wentzell quasipoten-

tial is needed. We remark that *generically* it is not possible to compute an explicit formula for μ^ϵ .

1.3. Informal result. Our goal is to prove that the probability $\epsilon^{d/2}\mu^\epsilon(\sqrt{\epsilon} dx)$ has a Gaussian shape in the small noise limit. To be more precise, under Hypotheses (A)–(D), the probability measure

$$(1.4) \quad \epsilon^{d/2}\mu^\epsilon(\sqrt{\epsilon} dx)$$

converges in the p -Wasserstein distance ($p \in [1, 2]$) to a Gaussian \mathcal{N} distribution with zero-mean vector and covariance matrix given by the unique solution \mathbb{X} of the Lyapunov matrix equation

$$(1.5) \quad DF(0_d)\mathbb{X} + \mathbb{X}(DF(0_d))^* = \sigma(0_d)(\sigma(0_d))^*.$$

Generically, it is hard to find an explicit formula for the solution of (1.5). Nevertheless, it can be estimated via numerical algorithms; see for instance [4, 32] and the references therein. More precisely, we show an asymptotic expansion of μ^ϵ (in the Wasserstein distance):

$$(1.6) \quad \frac{\mathcal{J}^\epsilon}{\sqrt{\epsilon}} = \mathcal{N} + \mathcal{O}(\sqrt{\epsilon}) \quad \text{as } \epsilon \rightarrow 0^+,$$

where \mathcal{J}^ϵ denotes a random variable with law μ^ϵ .

We anticipate that the proof of (1.6) does not rely on explicit computations of the distribution μ^ϵ . It is based on the linearization of the nonlinear dynamics around the stationary point 0_d . It is not hard to see that the resulting linear process has the target Gaussian as invariant distribution. It is then necessary to control the difference between this linear process and the nonlinear dynamics. This is done using the so-called synchronous coupling techniques with the help of Hypotheses (A)–(C). The proof of (1.6) is *purely dynamical* and it does not require techniques like Malliavin calculus, large deviation theory for SDEs as in [15], smoothness of the quasipotential, smoothness of the density μ^ϵ , analysis of the infinitesimal generator or the W.K.B. expansion.

Quantitative bounds on the rate of convergence of Markov processes to their limiting distribution are an important and widely studied topic, particularly in the context of Markov chains; see for instance [12, 16, 25] and the references therein. We *quantify* in the Wasserstein distance the implicit error term given in (1.6). We point out that the critical regime analyzed in [1, Section 5.1] implies for additive noise the total variation convergence of (1.4) to a Gaussian distribution. However, it seems hard to obtain bounds for the total variation error term, even under our assumptions on F and σ .

1.4. Wasserstein distance. Let \mathcal{P} be the set of probability measures in the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ -algebra of \mathbb{R}^d . For $p \geq 1$ we define

$$\mathcal{P}_p := \left\{ \mu \in \mathcal{P} : \int_{\mathbb{R}^d} \|x\|^p \mu(dx) < \infty \right\},$$

the space of probability measures with finite p -moment. For any $\mu, \nu \in \mathcal{P}$ we say that a probability measure π_* in the measurable space $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ is a *coupling* between μ and ν if the marginals of π_* are μ and ν , that is, for any $B \in \mathcal{B}(\mathbb{R}^d)$ we have $\pi_*(B \times \mathbb{R}^d) = \mu(B)$ and $\pi_*(\mathbb{R}^d \times B) = \nu(B)$. Let $\Pi(\mu, \nu)$ be the set of all couplings between μ and ν . For any $\mu, \nu \in \mathcal{P}_p$, the *Wasserstein distance* of order p between μ and ν is defined by

$$\mathcal{W}_p(\mu, \nu) := \inf \left\{ \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \pi_*(dx, dy) \right)^{1/p} : \pi_* \in \Pi(\mu, \nu) \right\}.$$

Let X and Y be two random vectors on \mathbb{R}^d defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with finite p -moment. The Wasserstein distance of order p between X and Y is defined by $\mathfrak{W}_p(X, Y) := \mathcal{W}_p(\mathbb{P}_X, \mathbb{P}_Y)$, where \mathbb{P}_X and \mathbb{P}_Y are the push-forward probability measures $\mathbb{P}_X(B) := \mathbb{P}(X \in B)$ and $\mathbb{P}_Y(B) := \mathbb{P}(Y \in B)$ for any $B \in \mathcal{B}(\mathbb{R}^d)$. For simplicity, we write $\mathcal{W}_p(X, Y)$ in place of $\mathfrak{W}_p(X, Y)$. A remarkable property that we use along this article is the scaling property

$$(1.7) \quad \mathcal{W}_p(cX, cY) = |c| \mathcal{W}_p(X, Y) \quad \text{for any } c \in \mathbb{R}.$$

The Wasserstein distance metrizes the weak convergence in the space of probabilities with finite p -moment. It is a fundamental concept in optimal transport theory, probability theory and partial differential equations. The Wasserstein distance is a natural way to compare the laws of two random variables X and Y (even for degenerate cases), where one variable is derived from the other by a small perturbation. For further details and properties of the Wasserstein distance, we refer to the monographs [29] and [34].

1.5. Results. We denote by $\mathcal{N}(v, \Xi)$ the Gaussian distribution in \mathbb{R}^d with vector mean v and positive definite covariance matrix Ξ . Let I_d be the $d \times d$ identity matrix. Given a matrix $A \in \mathbb{R}^{d \times d}$, denote by A^* the transpose matrix of A and by $\text{Tr}(A)$ the trace of A .

The main result of this paper is the following.

THEOREM 1.1 (Gaussian \mathcal{W}_2 -approximation of the invariant measure μ^ϵ). *Assume Hypotheses (A)–(D) are valid. Let \mathcal{J}^ϵ be a random vector on \mathbb{R}^d with distribution μ^ϵ . Then there exists a positive constant $K := K(\delta, \ell, d, c_0, \sigma(0_d))$ such that for any $\epsilon \in (0, \epsilon_*)$ with*

$$\epsilon_* = \min \left\{ \frac{\delta}{8c_1 \|\sigma(0_d)(\sigma(0_d))^*\|_F \cdot d^2}, \frac{\delta}{2\ell^2} \right\}$$

we have

$$(1.8) \quad \mathcal{W}_2 \left(\frac{\mathcal{J}^\epsilon}{\sqrt{\epsilon}} \mathcal{N} \right) \leq K \sqrt{\epsilon},$$

where \mathcal{N} denotes the Gaussian distribution on \mathbb{R}^d with zero-mean vector and covariance matrix Σ which is the unique solution of the Lyapunov matrix equation

$$(1.9) \quad DF(0_d)\Sigma + \Sigma(DF(0_d))^* = \sigma(0_d)(\sigma(0_d))^*.$$

Using the coupling approach, rates of convergence of the time evolution to equilibrium in the Wasserstein distance for Langevin processes are given in [13] for the underdamped dynamics and in [14] for the overdamped dynamics. In [5], linking functional inequalities with dissipation to ensure a spectral gap, it is shown that the solution of the Fokker–Planck equation converges in the Wasserstein distance of order 2 to its equilibrium as the time evolution goes by. However, the authors in [5, 13, 14] do not study small random perturbations of dynamical systems, and hence no asymptotic analysis for the invariant measure is needed there.

The proof of Theorem 1.1 does not rely on explicit computations of μ^ϵ nor on an explicit formula for the Wasserstein distance of order 2 between Gaussian distributions. The Itô formula with the help of (A), (B), (C) and Grönwall’s inequality (Lemma 3.1) implies that the p -moments are bounded recursively as a function of moments of order $p - 2$. Consequently, by an analogous (but more involved) reasoning one can see that the proof of Theorem 1.1 can be adapted for the L^p -Wasserstein distance for any $p \geq 1$.

REMARK 1.1 (Total variation convergence for additive noise). We stress that (1.8) does not imply directly any convergence of the corresponding densities. In other words, the approximation of densities

$$(1.10) \quad \mu^\epsilon(dx) \approx \epsilon^{-d/2} \mathcal{N}(dx/\sqrt{\epsilon}) \quad \text{for } \epsilon \ll 1$$

cannot be deduced in a straightforward way from (1.8). For additive noise, that is, $\sigma(x) = I_d$ for all $x \in \mathbb{R}^d$, using [21, Theorem 5.1, p. 30] (implicitly the celebrated Cameron–Martin–Girsanov Theorem) it can be shown that (1.10) is valid; see [3, Proposition 3.7, p. 1190] for further details. However, no rate of convergence is given there. Multiplicative noise is implicitly discussed in [9, p. 123].

REMARK 1.2 (Formula for the constant K). The constant K on the right-hand side of (1.8) can be taken as

$$K = \frac{96c_0 d^2 \|\sigma(0_d)(\sigma(0_d))^*\|_F}{\delta^2} + \frac{2\ell C_0^{1/2}}{\delta} \quad \text{with } C_0 = 2 \operatorname{Tr}((\sigma(0_d))^* \sigma(0_d)).$$

We emphasize that the error term $K\sqrt{\epsilon}$ may not be optimal.

REMARK 1.3 (Existence, uniqueness and integral representation for the covariance matrix Σ). By (A) and (D), [23, Theorem 1, p. 443] implies that (1.9) has a unique solution. Moreover, [23, Theorem 3, p. 414] yields the integral representation for the solution:

$$\Sigma = \int_0^\infty e^{-DF(0_d)s} \sigma(0_d)(\sigma(0_d))^* e^{-(DF(0_d))^*s} ds.$$

As a consequence of Theorem 1.1 we have the following corollaries.

COROLLARY 1.1 (\mathcal{W}_p convergence for $p \in [1, 2]$). *Assume that Hypotheses (A)–(D) are valid. Let \mathcal{J}^ϵ , \mathcal{N} , K and ϵ_* be as in Theorem 1.1. For any $p \in [1, 2]$,*

$$\mathcal{W}_p\left(\frac{\mathcal{J}^\epsilon}{\sqrt{\epsilon}} \mathcal{N}\right) \leq K\sqrt{\epsilon} \quad \text{for all } \epsilon \in (0, \epsilon_*).$$

Proof. This follows immediately by the Hölder inequality and (1.8). ■

COROLLARY 1.2 (Concentration). *Assume Hypotheses (A)–(D) are valid. Let \mathcal{J}^ϵ be as in Theorem 1.1. For any $p \in [1, 2]$ and $\beta < 1/2$,*

$$\lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon^\beta} \mathcal{W}_p(\mathcal{J}^\epsilon, 0_d) = 0.$$

Proof. This follows by the triangle inequality for \mathcal{W}_p , property (1.7) and Theorem 1.1. ■

Concentration of the equilibrium measure has been of considerable interest to physicists. Theorem 1 in [6] implies that $\mathcal{J}^\epsilon \rightarrow 0_d$ as $\epsilon \rightarrow 0^+$ in distribution sense. However, it does not say anything about the rate of convergence, like Corollary 1.2 does. Results about quantitative concentration of stationary measures on attractors and repellers for multiplicative noise are given in [17, 20].

The rest of the paper is organized as follows. Section 2 outlines the proof of Theorem 1.1. Section 3 is devoted to the proofs of the results skipped in Section 2. Finally, in the Appendix we provide polynomials and exponential moment estimates for the Ornstein–Uhlenbeck process that we use in Section 3.

2. OUTLINE OF THE PROOF

2.1. Linear diffusion approximation. Due to the dissipativity condition (A), the nonlinear random dynamics $(X_t^\epsilon(x))_{t \geq 0}$ is pushed back to the origin with high probability. In a neighbourhood of the origin, it is reasonable that an Ornstein–Uhlenbeck process helps us to understand $(X_t^\epsilon(x))_{t \geq 0}$ for large times. Let $(Y_t(x))_{t \geq 0}$ be the unique strong solution of the linear SDE

$$(2.1) \quad \begin{cases} dY_t^\epsilon(x) = -DF(0_d)Y_t^\epsilon(x) dt + \sqrt{\epsilon} \sigma(0_d) dB_t & \text{for any } t \geq 0, \\ Y_0^\epsilon(x) = x, \end{cases}$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion on \mathbb{R}^d and $DF(0_d)$ denotes the Jacobian matrix at 0_d . Variation of parameters yields

$$(2.2) \quad \begin{aligned} Y_t^\epsilon(x) &= e^{-DF(0_d)t} x \\ &+ \sqrt{\epsilon} e^{-DF(0_d)t} \int_0^t e^{DF(0_d)s} \sigma(0_d) dB_s \quad \text{for any } t \geq 0. \end{aligned}$$

This implies that for any $t > 0$, $Y_t^\epsilon(x)$ has Gaussian distribution with vector mean $m_t(x) := e^{-DF(0_d)t}x$ and covariance matrix $\Sigma_t^\epsilon := \epsilon\Sigma_t$ for any $t \geq 0$, where $(\Sigma_t)_{t \geq 0}$ solves the matrix differential equation

$$(2.3) \quad \begin{cases} \frac{d}{dt}\Sigma_t = -DF(0_d)\Sigma_t - \Sigma_t(DF(0_d))^* + \sigma(0_d)(\sigma(0_d))^* & \text{for any } t \geq 0, \\ \Sigma_0 = 0_{d \times d}, \end{cases}$$

where $0_{d \times d}$ is the $d \times d$ zero matrix. We refer to [30, Section 3.7] for further details. By (A) one can easily see that the eigenvalues of $DF(0_d)$ are contained in the set $\{z \in \mathbb{C} : \Re(z) \geq \delta\}$. As a consequence,

$$\|m_t(x)\| \leq e^{-\delta t}\|x\| \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

If we assume in addition that $\sigma(0_d)$ is invertible, Lemma 4.2 in Appendix implies

$$\|\Sigma_t - \Sigma\|_F \leq \|\Sigma\|_F^2 e^{-2\delta t} \rightarrow 0 \quad \text{as } t \rightarrow \infty,$$

where Σ is the unique solution of the matrix Lyapunov equation (1.9). Therefore, the limiting distribution of $Y_t^\epsilon(x)$ is a Gaussian law with zero-mean vector and positive definite covariance matrix $\epsilon\Sigma$. Moreover, [30, Proposition 3.5] implies that $\mathcal{N}(0_d, \epsilon\Sigma)$ is the unique invariant probability measure for the dynamics given by (2.1).

2.2. Disintegration. For short we write \mathcal{N} in place of $\mathcal{N}(0_d, \Sigma)$. Recall that \mathcal{J}^ϵ denotes a random vector on \mathbb{R}^d with distribution μ^ϵ . Let $t \geq 0$ and $x_0 \in \mathbb{R}^d$. The triangle inequality yields

$$(2.4) \quad \begin{aligned} \mathcal{W}_2(\mathcal{J}^\epsilon, \sqrt{\epsilon}\mathcal{N}) &\leq \mathcal{W}_2(\mathcal{J}^\epsilon, X_t^\epsilon(x_0)) + \mathcal{W}_2(X_t^\epsilon(x_0), Y_t^\epsilon(x_0)) \\ &\quad + \mathcal{W}_2(Y_t^\epsilon(x_0), \sqrt{\epsilon}\mathcal{N}). \end{aligned}$$

Since μ^ϵ is invariant for the dynamics (1.1), for any $t \geq 0$, $X_t^\epsilon(\mathcal{J}^\epsilon)$ has distribution μ^ϵ . By disintegration, the first term of the right-hand side of (2.4) can be estimated as follows:

$$(2.5) \quad \mathcal{W}_2(\mathcal{J}^\epsilon, X_t^\epsilon(x_0)) \leq \int_{\mathbb{R}^d} \mathcal{W}_2(X_t^\epsilon(x), X_t^\epsilon(x_0)) \mu^\epsilon(dx).$$

Analogously,

$$(2.6) \quad \mathcal{W}_2(\sqrt{\epsilon}\mathcal{N}, Y_t^\epsilon(x_0)) \leq \int_{\mathbb{R}^d} \mathcal{W}_2(Y_t^\epsilon(x), Y_t^\epsilon(x_0)) \mathcal{N}(0_d, \epsilon\Sigma)(dx),$$

where $\mathcal{N}(0_d, \epsilon\Sigma)(dx)$ denotes the density of $\sqrt{\epsilon}\mathcal{N}$. Combining (2.4)–(2.6) we obtain

$$\begin{aligned} \mathcal{W}_2(\mathcal{J}^\epsilon, \sqrt{\epsilon}\mathcal{N}) &\leq \int_{\mathbb{R}^d} \mathcal{W}_2(X_t^\epsilon(x), X_t^\epsilon(x_0)) \mu^\epsilon(dx) + \mathcal{W}_2(X_t^\epsilon(x_0), Y_t^\epsilon(x_0)) \\ &\quad + \int_{\mathbb{R}^d} \mathcal{W}_2(Y_t^\epsilon(x), Y_t^\epsilon(x_0)) \mathcal{N}(0_d, \epsilon\Sigma)(dx) \end{aligned}$$

for any $t \geq 0$ and $x_0 \in \mathbb{R}^d$. In particular, for any $t \geq 0$ we have

$$(2.7) \quad \mathcal{W}_2(\mathcal{J}^\epsilon, \sqrt{\epsilon} \mathcal{N}) \leq \int_{\mathbb{R}^d} \mathcal{W}_2(X_t^\epsilon(x), X_t^\epsilon(0_d)) \mu^\epsilon(dx) + \mathcal{W}_2(X_t^\epsilon(0_d), Y_t^\epsilon(0_d)) \\ + \int_{\mathbb{R}^d} \mathcal{W}_2(Y_t^\epsilon(x), Y_t^\epsilon(0_d)) \mathcal{N}(0_d, \epsilon \Sigma)(dx).$$

In what follows, we provide tools for estimating the right-hand side of (2.7). The following lemma allows us to couple two solutions of (1.1) with different initial conditions.

LEMMA 2.1 (Synchronous coupling I). *Assume Hypotheses (A) and (C) are valid. Let $x, x_0 \in \mathbb{R}^d$. Then*

$$\mathcal{W}_2(X_t^\epsilon(x), X_t^\epsilon(x_0)) \leq e^{-(\delta/2)t} \|x - x_0\| \quad \text{for all } t \geq 0, \epsilon \in (0, \delta/\ell^2],$$

where $\delta > 0$ is the dissipativity constant that appears in (A) and ℓ is the Lipschitz constant that appears in (C). In particular,

$$\mathcal{W}_2(X_t^\epsilon(x), X_t^\epsilon(0_d)) \leq e^{-(\delta/2)t} \|x\| \quad \text{for all } t \geq 0, \epsilon \in (0, \delta/\ell^2].$$

The following lemma provides second moment estimates for the marginals of the process (1.1) and also for its invariant probability measure μ^ϵ .

LEMMA 2.2 (Second moment estimates). *Assume Hypotheses (A) and (C) are valid. For any $x \in \mathbb{R}^d$ we have*

$$\mathbb{E}[\|X_t^\epsilon(x)\|^2] \leq \|x\|^2 e^{-\delta t} + \frac{\epsilon C_0}{\delta} \quad \text{for all } t \geq 0, \epsilon \in (0, \delta/(2\ell^2)],$$

where $\delta > 0$ is the dissipativity constant that appears in (A), ℓ is the Lipschitz constant that appears in (C), and $C_0 = 2 \operatorname{Tr}((\sigma(0_d))^* \sigma(0_d))$. In addition,

$$(2.8) \quad \int_{\mathbb{R}^d} \|x\|^2 \mu^\epsilon(dx) \leq \frac{\epsilon C_0}{\delta} \quad \text{for all } \epsilon \in (0, \delta/(2\ell^2)].$$

The next lemma is crucial in our argument. Due to the contracting nature of the dynamics, the random dynamics around zero, $(X_t^\epsilon(0_d))_{t \geq 0}$, can be approximated by its linearization $(Y_t^\epsilon(0_d))_{t \geq 0}$.

LEMMA 2.3 (Synchronous coupling II). *Assume Hypotheses (A)–(D) are valid. Then there exists a positive constant $C := C(\delta, \ell, d, c_0, \sigma(0_d))$ such that for any $\epsilon \in (0, \epsilon_*)$ with $\epsilon_* := \min \left\{ \frac{\delta}{8c_1 \|\sigma(0_d)(\sigma(0_d))^*\|_{\mathbb{F} \cdot d^2}}, \frac{\delta}{2\ell^2} \right\}$, and for all $t \geq 0$,*

$$\mathcal{W}_2(X_t^\epsilon(0_d), Y_t^\epsilon(0_d)) \leq C\epsilon,$$

where $\delta > 0$ is the dissipativity constant that appears in (A), c_0, c_1 are the positive constants that appear in (B) and ℓ is the Lipschitz constant that appears in (C).

We point out that the constant C can be taken as

$$C = \frac{48c_0d^2\|\sigma(0_d)(\sigma(0_d))^*\|_F}{\delta^2} + \frac{\ell C_0^{1/2}}{\delta};$$

this can be deduced from (3.17). Recall that Σ is the solution of (2.3). Since for any $t \geq 0$, $Y_t^\epsilon(\mathcal{N}(0_d, \epsilon\Sigma_t))$ has distribution $\mathcal{N}(0_d, \epsilon\Sigma)$, a reasoning analogous to those used in the proofs of Lemmas 2.1 and 2.2 yields the following lemma.

LEMMA 2.4 (Synchronous coupling III). *Assume Hypotheses (A) and (C) are valid. For any $x \in \mathbb{R}^d$,*

$$\mathcal{W}_2(Y_t^\epsilon(x), Y_t^\epsilon(0_d)) \leq e^{-(\delta/2)t}\|x\| \quad \text{for all } t \geq 0, \epsilon \in (0, \delta/\ell^2],$$

where $\delta > 0$ is the dissipativity constant that appears in (A) and ℓ is the Lipschitz constant that appears in (C). Assume in addition that $\sigma(0_d)$ is invertible. Then

$$(2.9) \quad \int_{\mathbb{R}^d} \|x\|^2 \mathcal{N}(0_d, \epsilon\Sigma)(dx) \leq \epsilon d \|\Sigma^{1/2}\|_F^2.$$

For simplicity we assume that $\sigma(0_d)$ is invertible in Lemma 2.4. Actually, this is not needed to obtain an estimate like (2.9). Nevertheless, it is necessary when defining the so-called generalized Gaussian distribution with degenerate covariance matrix and hence the notion of Moore–Penrose pseudoinverse is required. The assumption that $\sigma(0_d)$ is invertible can be removed and (2.8) in Lemma 2.3 remains valid with μ^ϵ replaced by the law of $\mathcal{N}(0_d, \epsilon\Sigma)$.

We stress that Theorem 1.1 is just a consequence of what we have already stated up to now.

Proof of Theorem 1.1. By (2.7) and Lemmas 2.1–2.4 we have

$$(2.10) \quad \mathcal{W}_2(\mathcal{J}^\epsilon, \sqrt{\epsilon}\mathcal{N}) \leq \sqrt{\frac{\epsilon C_0}{\delta}} e^{-(\delta/2)t} + C\epsilon + \sqrt{\epsilon d \|\Sigma^{1/2}\|_F^2} e^{-(\delta/2)t}$$

for any $t \geq 0$ and $\epsilon \in (0, \epsilon_*]$. Due to (1.7), (2.10) implies

$$\mathcal{W}_2\left(\frac{\mathcal{J}^\epsilon}{\sqrt{\epsilon}}, \mathcal{N}\right) \leq \sqrt{\frac{C_0}{\delta}} e^{-(\delta/2)t} + C\sqrt{\epsilon} + \sqrt{d \|\Sigma^{1/2}\|_F^2} e^{-(\delta/2)t}$$

for any $t \geq 0$ and $\epsilon \in (0, \epsilon_*]$. The cunning choice

$$t_\epsilon = \max \left\{ \frac{1}{\delta} \ln \left(\frac{4C_0}{\delta C^2 \epsilon} \right), \frac{1}{\delta} \ln \left(\frac{4d \|\Sigma^{1/2}\|_F^2}{C^2 \epsilon} \right) \right\}$$

yields

$$\mathcal{W}_2(\mathcal{J}^\epsilon/\sqrt{\epsilon}\mathcal{N}) \leq 2C\sqrt{\epsilon},$$

which concludes the proof. ■

3. PROOFS

In this section, we give the proofs of Lemmas 2.1–2.3. Along their proofs we use, several times, the celebrated Grönwall inequality; we recall it for completeness.

LEMMA 3.1 (Grönwall’s inequality). *Let $T > 0$ be fixed, $\mathfrak{g} : [0, T] \rightarrow \mathbb{R}$ be a C^1 -function and $\mathfrak{h} : [0, T] \rightarrow \mathbb{R}$ be a C^0 -function. Assume that*

$$\frac{d}{dt}\mathfrak{g}(t) \leq -a\mathfrak{g}(t) + \mathfrak{h}(t) \quad \text{for any } t \in [0, T],$$

where $a \in \mathbb{R}$, and the derivatives at 0 and T are understood as the right and left derivatives, respectively. Then

$$\mathfrak{g}(t) \leq e^{-at}\mathfrak{g}(0) + e^{-at} \int_0^t e^{as}\mathfrak{h}(s) ds \quad \text{for any } t \in [0, T].$$

3.1. Synchronous coupling I. For $x, x_0 \in \mathbb{R}^d$, let $(X_t^\epsilon(x))_{t \geq 0}$ and $(X_t^\epsilon(x_0))_{t \geq 0}$ be the solutions of (1.1) with initial conditions x and x_0 , respectively. In the following, we consider the so-called synchronous coupling, i.e., both processes $(X_t^\epsilon(x))_{t \geq 0}$ and $(X_t^\epsilon(x_0))_{t \geq 0}$ have the same driving noise $(B_t)_{t \geq 0}$.

Proof of Lemma 2.1. By the Itô formula we have

$$\begin{aligned} d\|X_t^\epsilon(x) - X_t^\epsilon(x_0)\|^2 &= -2\langle X_t^\epsilon(x) - X_t^\epsilon(x_0), F(X_t^\epsilon(x)) - F(X_t^\epsilon(x_0)) \rangle dt \\ &\quad + \epsilon \operatorname{Tr}[(\sigma(X_t^\epsilon(x)) - \sigma(X_t^\epsilon(x_0)))^*(\sigma(X_t^\epsilon(x)) - \sigma(X_t^\epsilon(x_0)))] dt \\ &\quad + 2\sqrt{\epsilon} \langle X_t^\epsilon(x) - X_t^\epsilon(x_0), (\sigma(X_t^\epsilon(x)) - \sigma(X_t^\epsilon(x_0))) dB_t \rangle. \end{aligned}$$

By (C) we have

$$(3.1) \quad \begin{aligned} \operatorname{Tr}[(\sigma(X_t^\epsilon(x)) - \sigma(X_t^\epsilon(x_0)))^*(\sigma(X_t^\epsilon(x)) - \sigma(X_t^\epsilon(x_0)))] \\ \leq \ell^2 \|X_t^\epsilon(x) - X_t^\epsilon(x_0)\|^2. \end{aligned}$$

A localization argument with the help of (A) and (3.1) implies

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[\|X_t^\epsilon(x) - X_t^\epsilon(x_0)\|^2] &\leq -2\delta\mathbb{E}[\|X_t^\epsilon(x) - X_t^\epsilon(x_0)\|^2] \\ &\quad + \epsilon\ell^2\mathbb{E}[\|X_t^\epsilon(x) - X_t^\epsilon(x_0)\|^2] \\ &\leq -(2\delta - \epsilon\ell^2)\mathbb{E}[\|X_t^\epsilon(x) - X_t^\epsilon(x_0)\|^2] \end{aligned}$$

for all $t \geq 0$. Since $\mathbb{E}[\|X_0^\epsilon(x) - X_0^\epsilon(x_0)\|^2] = \|x - x_0\|^2$, Lemma 3.1 yields

$$\mathbb{E}[\|X_t^\epsilon(x) - X_t^\epsilon(x_0)\|^2] \leq e^{-(2\delta - \epsilon\ell^2)t} \|x - x_0\|^2 \quad \text{for any } t \geq 0.$$

Therefore, for any $\epsilon \in (0, \delta/\ell^2]$ we have

$$\mathcal{W}_2(X_t^\epsilon(x), X_t^\epsilon(x_0)) \leq e^{-(\delta/2)t} \|x - x_0\| \quad \text{for any } x, x_0 \in \mathbb{R}^d, t \geq 0. \quad \blacksquare$$

3.2. Second moment estimates. For any $x \in \mathbb{R}^d$, let $(X_t^\epsilon(x))_{t \geq 0}$ be the solution of (1.1) with initial condition x .

Proof of Lemma 2.2. We wish to estimate $\mathbb{E}[\|X_t^\epsilon(x)\|^2]$. The Itô formula and (A) yield

$$\begin{aligned} d\|X_t^\epsilon(x)\|^2 &= -2\langle X_t^\epsilon(x), F(X_t^\epsilon(x)) \rangle dt + \epsilon \operatorname{Tr}[(\sigma(X_t^\epsilon(x)))^* \sigma(X_t^\epsilon(x))] + M_t^\epsilon(x) \\ &\leq -2\delta\|X_t^\epsilon(x)\|^2 dt + \epsilon \operatorname{Tr}[(\sigma(X_t^\epsilon(x)))^* \sigma(X_t^\epsilon(x))] dt + M_t^\epsilon(x), \end{aligned}$$

where $M_t^\epsilon(x) := \langle 2\sqrt{\epsilon} X_t^\epsilon(x), dB_t \rangle$ for every $t \geq 0$. Since

$$\begin{aligned} \operatorname{Tr}[(\sigma(X_t^\epsilon(x)))^* \sigma(X_t^\epsilon(x))] &\leq 2\operatorname{Tr}[(\sigma(X_t^\epsilon(x)) - \sigma(0_d))^* (\sigma(X_t^\epsilon(x)) - \sigma(0_d))] \\ &\quad + 2\operatorname{Tr}((\sigma(0_d))^* \sigma(0_d)), \end{aligned}$$

Hypothesis (C) implies

$$(3.2) \quad \operatorname{Tr}[(\sigma(X_t^\epsilon(x)))^* \sigma(X_t^\epsilon(x))] \leq 2\ell^2\|X_t^\epsilon(x)\|^2 + C_0,$$

where $C_0 := 2\operatorname{Tr}((\sigma(0_d))^* \sigma(0_d))$. A localization argument with the help of (A) and (3.2) implies

$$\frac{d}{dt} \mathbb{E}[\|X_t^\epsilon(x)\|^2] \leq -(2\delta - 2\epsilon\ell^2) \mathbb{E}[\|X_t^\epsilon(x)\|^2] + \epsilon C_0 \quad \text{for any } t \geq 0.$$

Since $\mathbb{E}[\|X_0^\epsilon(x)\|^2] = \|x\|^2$, for any $\epsilon \in (0, \delta/(2\ell^2)]$ Lemma 3.1 yields

$$(3.3) \quad \mathbb{E}[\|X_t^\epsilon(x)\|^2] \leq e^{-\delta t} \|x\|^2 + \frac{\epsilon C_0}{\delta} (1 - e^{-\delta t}) \leq e^{-\delta t} \|x\|^2 + \frac{\epsilon C_0}{\delta}$$

for any $t \geq 0$ and $x \in \mathbb{R}^d$. Following the reasoning in [3, p. 39], it is not hard to see that (3.3) implies

$$\int_{\mathbb{R}^d} \|x\|^2 \mu^\epsilon(dx) \leq \frac{\epsilon C_0}{\delta} \quad \text{for all } \epsilon \in (0, \delta/(2\ell^2)]. \quad \blacksquare$$

3.3. Synchronous coupling II. We consider the solution of (1.1) with initial condition $x = 0_d$, $(X_t^\epsilon(0_d))_{t \geq 0}$. Let $(Y_t^\epsilon(0_d))_{t \geq 0}$ be as in (2.1). In this section, we use the synchronous coupling between $X_t^\epsilon(0_d)$ and $Y_t^\epsilon(0_d)$, i.e., both processes $(X_t^\epsilon(0_d))_{t \geq 0}$ and $(Y_t^\epsilon(0_d))_{t \geq 0}$ have the same driving noise $(B_t)_{t \geq 0}$.

Proof of Lemma 2.3. We wish to estimate $\mathbb{E}[\|X_t^\epsilon(0_d) - Y_t^\epsilon(0_d)\|^2]$. Note that $X_0^\epsilon(0_d) = Y_0^\epsilon(0_d) = 0_d$. Let $\Delta_t^\epsilon(0_d) := X_t^\epsilon(0_d) - Y_t^\epsilon(0_d)$, $t \geq 0$. Then

$$\begin{aligned} d\Delta_t^\epsilon(0_d) &= -[F(X_t^\epsilon(0_d)) - F(Y_t^\epsilon(0_d))] dt + [DF(0_d)Y_t^\epsilon(0_d) - F(Y_t^\epsilon(0_d))] dt \\ &\quad + \sqrt{\epsilon}(\sigma(X_t^\epsilon(0_d)) - \sigma(0_d)) dB_t. \end{aligned}$$

Hence, the Itô formula reads

$$\begin{aligned} d\|\Delta_t^\epsilon(0_d)\|^2 &= -2\langle \Delta_t^\epsilon(0_d), F(X_t^\epsilon(0_d)) - F(Y_t^\epsilon(0_d)) \rangle dt \\ &\quad + 2\langle \Delta_t^\epsilon(0_d), DF(0_d)Y_t^\epsilon(0_d) - F(Y_t^\epsilon(0_d)) \rangle dt \\ &\quad + \epsilon \operatorname{Tr}[(\sigma(X_t^\epsilon(0_d)) - \sigma(0_d))^*(\sigma(X_t^\epsilon(0_d)) - \sigma(0_d))] dt \\ &\quad + 2\sqrt{\epsilon} \langle \Delta_t^\epsilon(0_d), (\sigma(X_t^\epsilon(0_d)) - \sigma(0_d)) dB_t \rangle. \end{aligned}$$

By (C) we have

$$(3.4) \quad \operatorname{Tr}[(\sigma(X_t^\epsilon(0_d)) - \sigma(0_d))^*(\sigma(X_t^\epsilon(0_d)) - \sigma(0_d))] \leq \ell^2 \|X_t^\epsilon(0_d)\|^2.$$

A localization argument with the help of (A), the Cauchy–Schwarz inequality and (3.4) implies

$$(3.5) \quad \begin{aligned} \frac{d}{dt} \mathbb{E}[\|\Delta_t^\epsilon(0_d)\|^2] &\leq -2\delta \mathbb{E}[\|\Delta_t^\epsilon(0_d)\|^2] \\ &\quad + 2\mathbb{E}[\|\Delta_t^\epsilon(0_d)\| \cdot \|F(Y_t^\epsilon(0_d)) - DF(0_d)Y_t^\epsilon(0_d)\|] + \epsilon \ell^2 \mathbb{E}[\|X_t^\epsilon(0_d)\|^2]. \end{aligned}$$

The differential inequality (3.5) and the Young inequality (for $p = 2$) yield

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[\|\Delta_t^\epsilon(0_d)\|^2] &\leq -\delta \mathbb{E}[\|\Delta_t^\epsilon(0_d)\|^2] \\ &\quad + \frac{1}{\delta} \mathbb{E}[\|F(Y_t^\epsilon(0_d)) - DF(0_d)Y_t^\epsilon(0_d)\|^2] + \epsilon \ell^2 \mathbb{E}[\|X_t^\epsilon(0_d)\|^2]. \end{aligned}$$

By Lemma 2.2 we have

$$\mathbb{E}[\|X_t^\epsilon(0_d)\|^2] \leq \frac{\epsilon C_0}{\delta} \quad \text{for all } t \geq 0, \epsilon \in (0, \delta/(2\ell^2)],$$

where $C_0 = 2 \operatorname{Tr}((\sigma(0_d))^* \sigma(0_d))$. Since $\Delta_t^\epsilon(0_d) = 0$, Lemma 3.1 implies

$$(3.6) \quad \begin{aligned} \mathbb{E}[\|\Delta_t^\epsilon(0_d)\|^2] &\leq \frac{1}{\delta} e^{-\delta t} \int_0^t e^{\delta s} \mathbb{E}[\|F(Y_s^\epsilon(0_d)) - DF(0_d)Y_s^\epsilon(0_d)\|^2] ds + \frac{\epsilon^2 \ell^2 C_0}{\delta^2} \\ &\leq \frac{1}{\delta^2} \sup_{0 \leq s \leq t} \mathbb{E}[\|F(Y_s^\epsilon(0_d)) - DF(0_d)Y_s^\epsilon(0_d)\|^2] + \frac{\epsilon^2 \ell^2 C_0}{\delta^2} \end{aligned}$$

for all $t \geq 0$ and $\epsilon \in (0, \delta/(2\ell^2)]$. Next, we estimate

$$\sup_{0 \leq s \leq t} \mathbb{E}[\|F(Y_s^\epsilon(0_d)) - DF(0_d)Y_s^\epsilon(0_d)\|^2].$$

Let $s \in [0, t]$. Recall that $F \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d)$. Since $F(0_d) = 0_d$, the mean value theorem yields

$$F(Y_s^\epsilon(0_d)) - F(0_d) = \int_0^1 DF(\theta_1 Y_s^\epsilon(0_d)) d\theta_1 \cdot Y_s^\epsilon(0_d),$$

where DF denotes the derivative of F . Since $F(0_d) = 0_d$, we have

$$(3.7) \quad F(Y_s^\epsilon(0_d)) - DF(0_d)Y_s^\epsilon(0_d) = \int_0^1 [DF(\theta_1 Y_s^\epsilon(0_d)) - DF(0_d)] d\theta_1 \cdot Y_s^\epsilon(0_d).$$

Applying the mean value theorem to (3.7) we deduce

$$(3.8) \quad \|F(Y_s^\epsilon(0_d)) - DF(0_d)Y_s^\epsilon(0_d)\| \leq C_s^\epsilon \|Y_s^\epsilon(0_d)\|^2,$$

where

$$C_s^\epsilon := \int_0^1 \int_0^1 \|D^2 F(\theta_1 \theta_2 Y_s^\epsilon(0_d))\| d\theta_1 d\theta_2$$

and $D^2 F$ denotes the second order derivative of F . Note that

$$(3.9) \quad Y_t^\epsilon(0_d) = \sqrt{\epsilon} Y_t \quad \text{for any } t \geq 0,$$

where $(Y_t)_{t \geq 0}$ is the unique strong solution of

$$(3.10) \quad \begin{cases} dY_t = -DF(0_d)Y_t dt + \sigma(0_d) dB_t & \text{for any } t \geq 0, \\ Y_0 = 0_d. \end{cases}$$

By (3.9) and **(B)** we have

$$\|D^2 F(\theta_1 \theta_2 Y_s^\epsilon(0_d))\| = \|D^2 F(\theta_1 \theta_2 \sqrt{\epsilon} Y_s)\| \leq c_0 e^{c_1 \theta_1^2 \theta_2^2 \epsilon \|Y_s\|^2}.$$

Since $\theta_1, \theta_2 \in [0, 1]$, we obtain

$$(3.11) \quad \|D^2 F(\theta_1 \theta_2 Y_s^\epsilon(0_d))\| \leq c_0 e^{c_1 \epsilon \|Y_s\|^2}.$$

Inequality (3.11) with the help of (3.8) and (3.9) yields

$$(3.12) \quad \|F(Y_s^\epsilon(0_d)) - DF(0_d)Y_s^\epsilon(0_d)\|^2 \leq c_0^2 e^{2c_1 \epsilon \|Y_s\|^2} \epsilon^2 \|Y_s\|^4$$

for any $s \geq 0$, where $(Y_t)_{t \geq 0}$ is the solution of (3.10). By Lemma 4.1(i) in Appendix it follows that

$$(3.13) \quad \mathbb{E}[\|Y_s\|^8] \leq 24C_*^4 \quad \text{for any } s \geq 0,$$

where

$$(3.14) \quad C_* = \frac{\|\sigma(0_d)(\sigma(0_d))^*\|_{\mathbb{F}} \cdot d^2}{\delta}$$

and $\|\cdot\|_F$ denotes the Frobenius norm. Due to **(D)**, we note that $C_* > 0$. Moreover, by Lemma 4.1(ii), $\epsilon \in (0, \frac{1}{4c_1C_*})$ we have

$$(3.15) \quad \mathbb{E}[e^{4c_1\epsilon\|Y_s\|^2}] \leq \frac{1}{1 - 4\epsilon c_1 C_*} \quad \text{for any } s \geq 0.$$

Estimate (3.12) with the help of the Cauchy–Schwarz inequality, (3.13) and (3.15) implies

$$\begin{aligned} \mathbb{E}[\|F(Y_s^\epsilon(0_d)) - DF(0_d)Y_s^\epsilon(0_d)\|^2] &\leq \epsilon^2 c_0^2 (\mathbb{E}[e^{4c_1\epsilon\|Y_s\|^2}] \mathbb{E}[\|Y_s\|^8])^{1/2} \\ &\leq \tilde{C}(\delta, d, c_0) \epsilon^2 \left(\frac{1}{1 - 4\epsilon c_1 C_*}\right)^{1/2} \end{aligned}$$

for any $s \geq 0, \epsilon \in (0, \frac{1}{4c_1C_*})$, where $\tilde{C}(\delta, d, c_0) = \sqrt{24} c_0^2 C_*^2$ is a positive constant. Consequently, for $\epsilon \in (0, \frac{1}{4c_1C_*})$ we obtain

$$(3.16) \quad \sup_{0 \leq s \leq t} \mathbb{E}[\|F(Y_s^\epsilon(0_d)) - DF(0_d)Y_s^\epsilon(0_d)\|^2] \leq \tilde{C}(\delta, d, c_0) \epsilon^2 \left(\frac{1}{1 - 4\epsilon c_1 C_*}\right)^{1/2}.$$

Note that if $\epsilon \in (0, \frac{1}{8c_1C_*})$, then $(1 - 4\epsilon c_1 C_*) \geq 1/2$. Let $\epsilon_* := \min\{\frac{1}{8c_1C_*}, \frac{\delta}{2\ell^2}\}$. By (3.6) and (3.16) we have, for all $\epsilon \in (0, \epsilon_*]$ and all $t \geq 0$,

$$\mathbb{E}[\|X_t^\epsilon(0_d) - Y_t^\epsilon(0_d)\|^2] \leq \frac{\sqrt{2}}{\delta^2} \tilde{C}(\delta, d, c_0) \epsilon^2 + \frac{\epsilon^2 \ell^2 C_0}{\delta^2}.$$

As a consequence, for any $\epsilon \in (0, \epsilon_*]$ and $t \geq 0$ we have

$$(3.17) \quad \begin{aligned} \mathcal{W}_2(X_t^\epsilon(0_d), Y_t^\epsilon(0_d)) &\leq \frac{\epsilon}{\delta} (\sqrt{2} \tilde{C}(\delta, d, c_0) + \ell^2 C_0)^{1/2} \\ &\leq \frac{\epsilon}{\delta} (48c_0 C_* + \ell C_0^{1/2}), \end{aligned}$$

where in the last inequality we use the subadditivity property of the root map. Inequality (3.17) with the help of (3.14) implies the statement. ■

4. APPENDIX

In this section, we compute the even moments and exponential moments of the Ornstein–Uhlenbeck process starting at zero. Let $(Z_t)_{t \geq 0}$ be the unique strong solution of the linear SDE

$$(4.1) \quad \begin{cases} dZ_t = -U Z_t dt + V dB_t & \text{for any } t \geq 0, \\ Z_0 = 0_d, \end{cases}$$

where $U, V \in \mathbb{R}^{d \times d}$ are given matrices. The drift matrix U satisfies the following condition: there exists a positive δ such that

$$(4.2) \quad \langle Ux, x \rangle \geq \delta \|x\|^2 \quad \text{for all } x \in \mathbb{R}^d.$$

We recall the definitions of 1-norm $\|\cdot\|_1$ and the Frobenius norm $\|\cdot\|_F$. For a given matrix $A = (a^{i,j})_{i,j=1}^d$ they are given by

$$\|A\|_1 := \sum_{i,j=1}^d |a^{i,j}| \quad \text{and} \quad \|A\|_F := \sqrt{\sum_{j=1}^d |a^{i,j}|^2}.$$

LEMMA 4.1 (Polynomial and exponential moments). *Assume that (4.2) is valid and let $(Z_t)_{t \geq 0}$ be the unique strong solution of the SDE (4.1). Then the following holds.*

(i) For each $j \in \mathbb{N}$,

$$(4.3) \quad \mathbb{E}[\|Z_t\|^{2j}] \leq C_*^j j! \quad \text{for all } t \geq 0, \quad \text{where} \quad C_* := \frac{\|VV^*\|_F \cdot d^2}{\delta}.$$

(ii) Let C_* be as in (i). For any $\lambda \in (0, 1/C_*)$ and all $t \geq 0$,

$$\mathbb{E}[e^{\lambda \|Z_t\|^2}] \leq \frac{1}{1 - \lambda C_*}.$$

Proof. (i) The proof is by induction on j . We start with the base case, $j = 1$. The Itô formula yields

$$(4.4) \quad d\|Z_t\|^2 = -2\langle Z_t, UZ_t \rangle dt + \text{Tr}[V^*V] dt + 2\langle Z_t, V dB_t \rangle.$$

A localization argument in (4.4) with the help of (4.2) implies

$$\frac{d}{dt} \mathbb{E}[\|Z_t\|^2] \leq -2\delta \mathbb{E}[\|Z_t\|^2] + \text{Tr}[V^*V].$$

Since $Z_t = 0_d$, Lemma 3.1 yields

$$(4.5) \quad \mathbb{E}[\|Z_t\|^2] \leq \frac{\text{Tr}[V^*V]}{2\delta} \quad \text{for all } t \geq 0.$$

Note that

$$(4.6) \quad \frac{\text{Tr}[V^*V]}{2\delta} \leq \frac{\|VV^*\|_1}{2\delta} \leq \frac{d\|VV^*\|_F}{2\delta} \leq \frac{d^2\|VV^*\|_F}{\delta}.$$

Combining (4.5) and (4.6) we get the base case.

We now assume that (4.3) holds for $j = n$ and prove it for $j = n + 1$. The Itô formula for the function $f(x) = \|x\|^{2(n+1)}$, $x \in \mathbb{R}^d$, reads

$$(4.7) \quad d\|Z_t\|^{2(n+1)} = -2(n+1)\|Z_t\|^{2n} \langle Z_t, U Z_t \rangle dt + (1/2) \operatorname{Tr}[V^* H(Z_t) V] dt + 2(n+1)\|Z_t\|^{2n} \langle Z_t, V dB_t \rangle,$$

where the matrix valued function $\mathbb{R}^d \ni x \mapsto H(x) := (H_{i,j}(x))_{i,j}^d \in \mathbb{R}^{d \times d}$ is given by

$$H_{i,j}(x) := \begin{cases} 4(n+1)n\|x\|^{2(n-1)}x_i^2 + 2(n+1)\|x\|^{2n} & \text{for } i = j, \\ 4(n+1)n\|x\|^{2(n-1)}x_i x_j & \text{for } i \neq j. \end{cases}$$

By definition of $\|\cdot\|_1$ it follows that

$$(4.8) \quad \begin{aligned} \|H(x)\|_1 &\leq 2d(n+1)\|x\|^{2n} + 4d(n+1)n\|x\|^{2n} \\ &= 2d(n+1)(1+2n)\|x\|^{2n} \end{aligned}$$

for all $x \in \mathbb{R}^d$. Note that $\operatorname{Tr}[V^* H(Z_t) V] = \operatorname{Tr}[H(Z_t) V V^*]$. By (4.8) we obtain

$$(4.9) \quad \begin{aligned} |\operatorname{Tr}[H(Z_t) V V^*]| &\leq d\|H(Z_t)\|_1 \|V V^*\|_F \\ &\leq 2d^2(n+1)(1+2n)\|V V^*\|_F \|Z_t\|^{2n}. \end{aligned}$$

Using a localization argument in (4.7) with the help of (4.2) and (4.9) yields

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[\|Z_t\|^{2(n+1)}] &\leq -2(n+1)\delta \mathbb{E}[\|Z_t\|^{2(n+1)}] \\ &\quad + d^2(n+1)(1+2n)\|V V^*\|_F \mathbb{E}[\|Z_t\|^{2n}]. \end{aligned}$$

By induction hypothesis we have $\mathbb{E}[\|Z_t\|^{2n}] \leq C_*^n n!$ for all $t \geq 0$. Since $Z_0 = 0_d$, Lemma 3.1 yields, for all $t \geq 0$,

$$\mathbb{E}[\|Z_t\|^{2(n+1)}] \leq \frac{d^2(n+1)(1+2n)\|V V^*\|_F C_*^n n!}{2(n+1)\delta} \leq C_*^{n+1}(n+1)!,$$

which finishes the induction step. This concludes the proof of (i).

(ii) By the Monotone Convergence Theorem we have

$$\mathbb{E}[e^{\lambda\|Z_t\|^2}] = \sum_{j=0}^{\infty} \frac{\lambda^j \mathbb{E}[\|Z_t\|^{2j}]}{j!} \quad \text{for all } \lambda \geq 0.$$

By (i) for all $\lambda \in (0, 1/C_*)$ and $t \geq 0$ it follows that

$$\mathbb{E}[e^{\lambda\|Z_t\|^2}] \leq \sum_{j=0}^{\infty} (\lambda C_*)^j = \frac{1}{1 - \lambda C_*}. \quad \blacksquare$$

LEMMA 4.2 (Covariance). *Assume that (4.2) holds and that the diffusion matrix V is invertible. Let $(Z_t)_{t \geq 0}$ be the unique strong solution of the SDE (4.1) and let $\Theta_t := \mathbb{E}[Z_t Z_t^*]$ for any $t \geq 0$. Then*

$$\|\Theta_t - \Theta\|_F \leq \|\Theta\|_F^2 e^{-2\delta t} \quad \text{for all } t \geq 0,$$

where $\Theta \in \mathbb{R}^{d \times d}$ is the unique symmetric and positive definite solution of the Lyapunov matrix equation

$$(4.10) \quad U\Theta + \Theta U^* = VV^*.$$

Proof. The proof follows the reasoning used in [3, proof of Lemma C.4]. We give it here for completeness of presentation.

Hypothesis (4.2) and [23, Theorem 1, p. 443] imply that (4.10) has a unique solution. By [30, Proposition 3.5] we have

$$(4.11) \quad \begin{cases} \frac{d}{dt}\Theta_t = -U\Theta_t - \Theta_t U^* + VV^* & \text{for any } t \geq 0, \\ \Theta_0 = 0_{d \times d}, \end{cases}$$

where $0_{d \times d} \in \mathbb{R}^{d \times d}$. Let $t \geq 0$ be fixed. Write $r_t := \|\Theta_t - \Theta\|_F^2$, $\Theta_t = (\Theta_t^{i,j})_{i,j=1}^d$, $\Theta = (\Theta^{i,j})_{i,j=1}^d$, $U = (U^{i,j})_{i,j=1}^d$ and $VV^* = ((VV^*)^{i,j})_{i,j=1}^d$. By (4.10) we obtain

$$(4.12) \quad \sum_{k=1}^d (U^{i,k}\Theta^{k,j} + \Theta^{i,k}U^{j,k}) = (VV^*)^{i,j} \quad \text{for all } i, j \in \{1, \dots, d\}.$$

The differential equation (4.11) with the help of (4.12) reads

$$(4.13) \quad \begin{aligned} \frac{d}{dt}\Theta_t^{i,j} &= \sum_{k=1}^d (-U^{i,k}\Theta_t^{k,j} - \Theta_t^{i,k}U^{j,k} + (VV^*)^{i,j}) \\ &= -\sum_{k=1}^d (U^{i,k}(\Theta_t^{k,j} - \Theta^{k,j}) + (\Theta_t^{i,k} - \Theta^{i,k})U^{j,k}). \end{aligned}$$

The chain rule and (4.13) imply

$$\begin{aligned} \frac{d}{dt}r_t &= 2 \sum_{i,j=1}^d (\Theta_t^{i,j} - \Theta^{i,j}) \frac{d}{dt}(\Theta_t^{i,j} - \Theta^{i,j}) \\ &= -2 \sum_{i,j=1}^d (\Theta_t^{i,j} - \Theta^{i,j}) \sum_{k=1}^d (U^{i,k}(\Theta_t^{k,j} - \Theta^{k,j}) + (\Theta_t^{i,k} - \Theta^{i,k})U^{j,k}) \\ &= -2 \sum_{j=1}^d \sum_{i,k=1}^d (\Theta_t^{i,j} - \Theta^{i,j}) U^{i,k} (\Theta_t^{k,j} - \Theta^{k,j}) \\ &\quad - 2 \sum_{i=1}^d \sum_{j,k=1}^d (\Theta_t^{i,j} - \Theta^{i,j}) U^{j,k} (\Theta_t^{i,k} - \Theta^{i,k}), \end{aligned}$$

where in the last equality we rearrange the sums. By (4.2) we deduce the differential inequality

$$\frac{d}{dt}r_t \leq -4\delta \sum_{i,j=1}^d (\Theta_t^{i,j} - \Theta^{i,j})^2 = -4\delta r_t \quad \text{for all } t \geq 0.$$

Lemma 3.1 yields $r_t \leq r_0 e^{-4\delta t}$ for all $t \geq 0$ and consequently the statement. ■

Acknowledgments. This research has been supported by the Academy of Finland via the Matter and Materials Profi4 university profiling action. He gratefully acknowledges support from a post-doctorate grant (2020-2023) held at the Department of Mathematical and Statistical Sciences at University of Helsinki and expresses his gratitude for all the facilities used along the realization of this work. The author is grateful to the reviewers for the thorough examination of the manuscript, which has led to significant improvements.

REFERENCES

- [1] A. Arapostathis, A. Biswas, and V. Borkar, *Controlled equilibrium selection in stochastically perturbed dynamics*, Ann. Probab. 46 (2018), 2749–2799.
- [2] K. Athreya and C. Hwang, *Gibbs measures asymptotics*, Sankhyā A 72 (2018), 191–207.
- [3] G. Barrera and M. Jara, *Thermalisation for small random perturbations of dynamical systems*, Ann. Appl. Probab. 30 (2020), 1164–1208.
- [4] P. Benner, J. Li, and T. Penzl, *Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems*, Numer. Linear Algebra Appl. 15 (2008), 755–777.
- [5] F. Bolley, I. Gentil, and A. Guillin, *Convergence to equilibrium in Wasserstein distance for Fokker–Planck equations*, J. Funct. Anal. 263 (2012), 2430–2457.
- [6] A. Biswas and V. Borkar, *Small noise asymptotics for invariant densities for a class of diffusions: a control theoretic view*, J. Math. Anal. Appl. 360 (2009), 476–484.
- [7] W. Coffey and Y. Kalmykov, *The Langevin Equation: with Applications in Physics, Chemistry and Electrical Engineering*, 3rd ed., World Sci., 2012.
- [8] A. Dalalyan, *Theoretical guarantees for approximate sampling from smooth and log-concave densities*, J. Roy. Statist. Soc. Ser. B. Statist. Methodol. 79 (2017), 651–676.
- [9] M. Day, *Recent progress on the small parameter exit problem*, Stochastics 20 (1987), 121–150.
- [10] M. Day and T. Darden, *Some regularity results on the Ventcel–Freidlin quasi-potential function*, Appl. Math. Optim. 13 (1985), 259–282.
- [11] A. Duncan, N. Nüsken, and G. Pavliotis, *Using perturbed underdamped Langevin dynamics to efficiently sample from probability distributions*, J. Statist. Phys. 169 (2017), 1098–1131.
- [12] A. Durmus and É. Moulines, *Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis adjusted Langevin algorithm*, Statist. Comput. 25 (2015), 5–19.
- [13] A. Eberle, A. Guillin, and R. Zimmer, *Couplings and quantitative contraction rates for Langevin dynamics*, Ann. Probab. 47 (2019), 1982–2010.
- [14] A. Eberle, A. Guillin, and R. Zimmer, *Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes*, Trans. Amer. Math. Soc. 371 (2019), 7135–7173.
- [15] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems*, 3rd ed., Springer, 2012.
- [16] R. Gareth and J. Rosenthal, *Hitting time and convergence rate bounds for symmetric Langevin diffusions*, Methodol. Comput. Appl. Probab. 21 (2019), 921–929.

- [17] W. Huang, M. Ji, Z. Liu, and Y. Yi, *Concentration and limit behaviors of stationary measures*, Phys. D 369 (2018), 1–17.
- [18] C. Hwang, *Laplace’s method revisited: weak convergence of probability measures*, Ann. Probab. 8 (1980), 1177–1182.
- [19] C. Hwang, S. Hwang-Ma, and S. Sheu, *Accelerating Gaussian diffusions*, Ann. Appl. Probab. 3 (1993), 897–913.
- [20] M. Ji, Z. Shen, and Y. Yi, *Quantitative concentration of stationary measures*, Phys. D 399 (2019), 73–85.
- [21] Y. Kabanov, R. Liptser, and A. Shiryaev, *On the variation distance for probability measures defined on a filtered space*, Probab. Theory Related Fields 71 (1986), 19–35.
- [22] A. Kulik, *Ergodic Behavior of Markov Processes with Applications to Limit Theorems*, De Gruyter Stud. Math. 67, De Gruyter, 2018.
- [23] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed., Academic Press, 1985.
- [24] T. Lelièvre, F. Nier, and G. Pavliotis, *Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion*, J. Statist. Phys. 152 (2013), 237–274.
- [25] N. Madras and D. Sezer, *Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances*, Bernoulli 16 (2010), 882–908.
- [26] X. Mao, *Stochastic Differential Equations and Applications*, 2nd ed., Horwood, 2008.
- [27] T. Mikami, *Asymptotic expansions of the invariant density of a Markov process with a small parameter*, Ann. Inst. H. Poincaré Probab. Statist. 24 (1988), 403–424.
- [28] T. Mikami, *Asymptotic analysis of invariant density of randomly perturbed dynamical systems*, Ann. Probab. 18 (1990), 524–536.
- [29] V. Panaretos and Y. Zemel, *An Invitation to Statistics in Wasserstein Space*, Springer, 2020.
- [30] G. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*, Springer, 2014.
- [31] Y. Pomeau and J. Piasecki, *The Langevin equation*, C. R. Phys. 18 (2017), 570–582.
- [32] A. Scottedward, B. Tenison, and K. Poolla, *Numerical solution of the Lyapunov equation by approximate power iteration*, Linear Algebra Appl. 236 (1996), 205–230.
- [33] S. Sheu, *Asymptotic behavior of the invariant density of a diffusion Markov process with small diffusion*, SIAM J. Math. Anal. 17 (1986), 451–460.
- [34] C. Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer, 2009.
- [35] S. Wu, C. Hwang, and M. Chu, *Attaining the optimal Gaussian diffusion acceleration*, J. Statist. Phys. 155 (2014), 571–590.

Gerardo Barrera
Department of Mathematical and Statistical Sciences
University of Helsinki
PL 68, Pietari Kalmin katu 5
Postal Code: 00560. Helsinki, Finland
E-mail: gerardo.barreravargas@helsinki.fi

Received 10.5.2021;
accepted 14.4.2022
