The Cutoff Phenomenon in Wasserstein Distance for Nonlinear Stable Langevin Systems with Small Lévy Noise

G. Barrera¹ · M. A. Högele² · J. C. Pardo³

Received: 13 September 2021 / Revised: 13 January 2022 / Accepted: 24 January 2022
© The Author(s) 2022

Abstract

This article establishes the cutoff phenomenon in the Wasserstein distance for systems of nonlinear ordinary differential equations with a dissipative stable fixed point subject to small additive Markovian noise. This result generalizes the results shown in Barrera, Högele, Pardo (EJP2021) in a more restrictive setting of Blumenthal–Gettoor index $\alpha > 3/2$ to the formulation in Wasserstein distance, which allows to cover the case of general Lévy processes with some given moment. The main proof techniques are based on the close control of the errors in a version of the Hartman–Grobman theorem and the adaptation of the linear theory established in Barrera, Högele, Pardo (JSP2021). In particular, they rely on the precise asymptotics of the nonlinear flow and the nonstandard shift linearity property of the Wasserstein distance, which is established by the authors in (JSP2021). Main examples are the nonlinear Fermi–Pasta–Ulam–Tsingou gradient flow and dissipative nonlinear oscillators subject to small (and possibly degenerate) Brownian or arbitrary α -stable noise.

Keywords Cutoff phenomenon · Exponential ergodicity · Lévy processes · Nonlinear Langevin dynamics · Nonstandard properties of the Wasserstein distance

Mathematics Subject Classification 60H10 · 37A25 · 60G51 · 15A16

✉ G. Barrera
gerardo.barreravargas@helsinki.fi

M. A. Högele
ma.hoegele@uniandes.edu.co

J. C. Pardo
jcpardo@ciimat.mx

¹ Department of Mathematical and Statistical Sciences, University of Helsinki, Exactum in Kumpula Campus, PL 68, Pietari Kalmin katu 5, 00560 Helsinki, Finland

² Departamento de Matemáticas, Facultad de Ciencias, Universidad de los Andes, Bogotá, Colombia

³ CIMAT, Jalisco S/N, Valenciana, CP 36240 Guanajuato, México

1 Introduction

In this paper, we study the asymptotics of the ergodic behavior of the following stochastic differential equation (SDE)

$$dX_t^\varepsilon(x) = -b(X_t^\varepsilon(x))dt + \varepsilon dL_t, \quad X_0^\varepsilon(x) = x \in \mathbb{R}^d \quad (1.1)$$

for small noise intensity $\varepsilon > 0$, where the vector field $b \in C^2(\mathbb{R}^d, \mathbb{R}^d)$ satisfies $b(0) = 0$ and the following dissipative condition.

Hypothesis 1 (*Dissipativity*) There exists a constant $\delta > 0$ such that

$$\langle b(x) - b(y), x - y \rangle \geq \delta |x - y|^2 \quad \text{for all } x, y \in \mathbb{R}^d. \quad (1.2)$$

The noise process $L = (L_t)_{t \geq 0}$ in (1.1) is a Lévy process with values in $\mathbb{R}^d$ on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is well-known that the law of L is characterized by the triplet (a, Σ, ν) , where $a \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ is a non-negative definite matrix and $\nu : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, \infty]$ is a locally finite Borel measure satisfying

$$\nu(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} (1 \wedge |z|^2) \nu(dz) < \infty.$$

For $\nu = 0$ the process L is a multidimensional Brownian motion with drift, while for $a = 0$ and $\Sigma = 0$ we have a multidimensional pure jump process such as compound Poisson processes or α -stable processes, in particular, the Cauchy process for $\alpha = 1$. We refer to [1, 14, 16, 18, 22] for further details on Lévy processes. Under Hypothesis 1, it is known that the SDE (1.1) has a pathwise unique strong solution, see for instance Theorem 1.1 in [10], here denoted by $X^\varepsilon(x) := (X_t^\varepsilon(x))_{t \geq 0}$. Moreover, $X^\varepsilon(x)$ is a Markov process and, in particular, it satisfies the Feller property see Proposition 2.1 in [21].

In order to present the main results of this paper, we formally introduce the Wasserstein distance of order p_* . We assume some finite moment for L_t and hence $X_t^\varepsilon(x)$ for all $t \geq 0$.

Hypothesis 2 (*Finite p_* -th moment*) There exists $p_* > 0$ such that

$$\int_{|z| > 1} |z|^{p_*} \nu(dz) < \infty.$$

This article shows the cutoff phenomenon for the family of processes $(X^\varepsilon(x))_{\varepsilon > 0}$ with respective invariant measures $(\mu^\varepsilon)_{\varepsilon > 0}$ under the Wasserstein distance $\mathcal{W}_{p_*}$ of order $p_* > 0$. For $p_* > 1$ we characterize the following cutoff profile asymptotics

$$\mathcal{W}_{p_*}(\text{Law}(X_{t_\varepsilon+r}^\varepsilon(x)), \mu^\varepsilon) = \varepsilon \cdot C e^{-q r} + o(\varepsilon) \quad \text{for } \varepsilon \rightarrow 0, \quad (1.3)$$

where $t_\varepsilon = \frac{1}{q} |\ln(\varepsilon)| + \frac{\ell-1}{q} \ln(|\ln(\varepsilon)|)$ for some explicit positive constants q, ℓ, C that depend on x in terms of an ω -limit set of the rotational part for the Hartman–Grobman linearization of $X^0(x)$.

For such processes $(X^\varepsilon(x))_{\varepsilon > 0}$ where (1.3) fails, we establish the following weaker window cutoff asymptotics

$$\begin{aligned} \lim_{r \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_{p_*}(\text{Law}(X_{t_\varepsilon+r}^\varepsilon(x)), \mu^\varepsilon)}{\varepsilon} &= 0 \quad \text{and} \\ \lim_{r \rightarrow -\infty} \liminf_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_{p_*}(\text{Law}(X_{t_\varepsilon+r}^\varepsilon(x)), \mu^\varepsilon)}{\varepsilon} &= \infty. \end{aligned}$$

Our results generalize the results in [2] to the nonlinear vector field and [3, 5] and [6] to the Wasserstein distance which cover second order equations with degenerate noise. For a detailed introduction on the subject we refer to the aforementioned articles, in particular, see Table 1.1 in [3]. There is a particular advantage of studying this problem under the Wasserstein distance rather than in the total variation. While the Wasserstein distance only requires the existence of moments of $X^\varepsilon(x)$ of a given order, the total variation distance needs existence of its density in addition to its regularity. The latter brings further requirements for the Lévy process L which can be quite restrictive, see [3] for further details. Furthermore the Wasserstein case, at least in case of $X^\varepsilon(x)$ moments of order $p > 1$, the cutoff phenomenon of $(X^\varepsilon(x))_{\varepsilon>0}$ is completely determined by an explicit function (see Theorem 2 below), here called as cutoff profile. On the contrary, in the total variation case the profile function can be very involved and even hard to simulate in examples.

In [4], the cutoff phenomenon with respect to the total variation distance covering SDEs of the type (1.1) in the one dimensional case, L being a standard Brownian motion and with general drift coefficient b (satisfying Hypothesis 1) is studied. Since scalar systems are gradient systems, there is always a cutoff profile which can be given explicitly in terms of the Gauss error function. The follow-up work [5] covers the multidimensional case, where the picture is considerably richer, due to the presence of strong and complicated rotational patterns. The authors characterize sharply the existence of a cutoff profile in terms of the omega limit sets appearing in the long-term behavior of the matrix exponential function $e^{-Q\tau}x$ in Lemma B.2 in [5], which plays an analogous role in this article. The paper [6] is the first attempt to study the cutoff phenomenon for such models with jumps. More precisely, [6] covers the cutoff phenomenon with respect to the total variation distance of the generalized Ornstein-Uhlenbeck processes. The previous process satisfies an SDE of the form (1.1) with L being a Lévy process and $b(x) = Qx$, where Q is a square real matrix whose eigenvalues have positive real parts. The proof methods are based on concise Fourier inversion techniques. Due to the aforementioned regularity inherited by the total variation, the results in [6] are given under the hypothesis of continuous densities of the marginals, which to date is mathematically not characterized in simple terms. The cutoff profile function in [6] is given in terms of the Lévy-Ornstein-Uhlenbeck limiting measure for $\varepsilon = 1$ and measured in the total variation distance. Such profile functions are theoretically highly insightful, but almost impossible to calculate and simulate in examples. The characterization of the existence of a cutoff-profile remains analogously to [5] in abstract terms of the behavior of the mentioned profile function on a suitably defined omega limit set. The Wasserstein case is treated in [2] where, contrary to the total variation case, it is noted that the profile function takes an explicit and simple shape. Finally, [3] treats the cutoff phenomenon with respect to the total variation distance for (1.1) with b satisfying Hypothesis 1 and driven by a Lévy process in the rather restrictive class of strongly locally layered stable processes (see Definition 1.4 in [3]).

In this article we combine a nonlinear version of the Wasserstein estimates of [2], with the Freidlin-Wentzell first order approximation of (1.1) in the spirit of [3] and the fine properties of the Wasserstein distance given in Lemma 2.1, in particular, the non-standard shift linearity of Lemma 2.1.d).

The manuscript is organized in four parts. After the exposition of the setting and the presentation of the main results in Sect. 2, we illustrate our findings for the nonlinear Fermi–Pasta–Ulam–Tsingou gradient system and a class of nonlinear oscillators in Sect. 3. The main steps of the proof of the cutoff phenomenon are given in Sect. 4 while the auxiliary technical such as exponential ergodicity in Wasserstein distance, the coupling between the original nonlinear system and the Freidlin-Wentzell linearization results are given in the “Appendix”.

2 Setting and Main Results

2.1 Fine Properties of the Wasserstein Distance

For any two probability distributions μ_1 and μ_2 on $\mathbb{R}^d$ with finite p_* -th moment for some $p_* > 0$, we define the Wasserstein p_* -distance between them as follows

$$\mathcal{W}_{p_*}(\mu_1, \mu_2) = \inf_{\Pi} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |u - v|^{p_*} \Pi(du, dv) \right)^{1 \wedge (1/p_*)},$$

where the infimum is taken over all couplings (joint distributions on $\mathbb{R}^d \times \mathbb{R}^d$) Π with marginals μ_1 and μ_2 . We refer to [12, 20] and references therein for more details. For convenience of notation we do not distinguish a random variable U and its law $\mathbb{P}_U$ as an argument of $\mathcal{W}_{p_*}$. That is, for random variables U_1, U_2 and probability measure μ we write $\mathcal{W}_{p_*}(U_1, U_2)$ instead of $\mathcal{W}_{p_*}(\mathbb{P}_{U_1}, \mathbb{P}_{U_2})$, $\mathcal{W}_{p_*}(U_1, \mu)$ instead of $\mathcal{W}_{p_*}(\mathbb{P}_{U_1}, \mu)$ etc. The next result establishes properties of the Wasserstein distance which turn out to be important for our arguments.

Lemma 2.1 (Properties of $\mathcal{W}_{p_*}$) *For $p_* > 0$, $u_1, u_2 \in \mathbb{R}^d$, $c \in \mathbb{R}$ and U_1 and U_2 being random vectors in $\mathbb{R}^d$ with finite p_* -th moment we have the following:*

- (a) *The Wasserstein distance $\mathcal{W}_{p_*}$ is a metric.*
- (b) *Translation invariance: $\mathcal{W}_{p_*}(u_1 + U_1, u_2 + U_2) = \mathcal{W}_{p_*}(u_1 - u_2 + U_1, U_2)$.*
- (c) *Homogeneity:*

$$\mathcal{W}_{p_*}(c \cdot U_1, c \cdot U_2) = \begin{cases} |c| \mathcal{W}_{p_*}(U_1, U_2) & \text{for } p_* \in [1, \infty), \\ |c|^{p_*} \mathcal{W}_{p_*}(U_1, U_2) & \text{for } p_* \in (0, 1). \end{cases}$$

- (d) *Shift linearity: For $p_* \geq 1$ it follows*

$$\mathcal{W}_{p_*}(u_1 + U_1, U_1) = |u_1|. \quad (2.1)$$

For $p_* \in (0, 1)$ we have

$$\max\{|u_1|^{p_*} - 2\mathbb{E}[|U_1|^{p_*}], 0\} \leq \mathcal{W}_{p_*}(u_1 + U_1, U_1) \leq |u_1|^{p_*}. \quad (2.2)$$

- (e) *Domination: For any given coupling $\tilde{\Pi}$ between U_1 and U_2 it follows*

$$\mathcal{W}_{p_*}(U_1, U_2) \leq \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |v_1 - v_2|^{p_*} \tilde{\Pi}(dv_1, dv_2) \right)^{1 \wedge (1/p_*)}.$$

- (f) *Characterization: Let $(U_n)_{n \in \mathbb{N}}$ be a sequence of random vectors with finite p_* -th moments and U a random vector with finite p_* -th moment. Then the following statements are equivalent:*

- (1) $\mathcal{W}_{p_*}(U_n, U) \rightarrow 0$ as $n \rightarrow \infty$.
- (2) $U_n \xrightarrow{d} U$ as $n \rightarrow \infty$ and $\mathbb{E}[|U_n|^{p_*}] \rightarrow \mathbb{E}[|U|^{p_*}]$ as $n \rightarrow \infty$.

For $p_* \in (0, 1)$ equality (2.1) is false in general, see Remark 2.4 in [2]. The proof of the previous lemma is given in Lemma 2.2 in [2].

The following result yields the existence of a unique invariant distribution for (1.1) under Hypotheses 1 and 2. Moreover, under the Wasserstein distance, the strong solution of (1.1) is exponentially ergodic.

Proposition 1 (Existence of a unique invariant distribution) *Under Hypothesis 1 for $p_* > 0$ and Hypothesis 2 there exists a unique invariant probability measure μ^ε such that*

$$\mathcal{W}_{p_*}(X_t^\varepsilon(x), \mu^\varepsilon) \leq e^{-(1 \wedge p_*)\delta t} \left(|x|^{1 \wedge p_*} + \int_{\mathbb{R}^d} |y|^{1 \wedge p_*} \mu^\varepsilon(dy) \right). \quad (2.3)$$

The proof is given in “Appendix 1”.

2.2 Hartman–Grobman Asymptotics

The zeroth-order approximation of a smooth dynamical systems on a finite time horizon $[0, T]$ subject to small perturbations is given by the deterministic system, that is, $(X_t^0(x))_{t \in [0, T]}$. Our main results treat small asymptotics close to the stable state 0 which translates to meaningful time scales $t_\varepsilon \rightarrow \infty$, as $\varepsilon \rightarrow 0$, in Theorems 1 and 2. Before we state our main result, we first provide the long-time asymptotics of $X_t^0(x)$ in terms of the spectral decomposition of the solution $t \mapsto e^{-Db(0)t} x^*$ of the respective linear system for some x^* in a small neighbourhood of the origin.

Lemma 2.2 (Asymptotic Hartman–Grobman) *Assume Hypothesis 1. Then for any $x \in \mathbb{R}^d \setminus \{0\}$ there exist:*

- (i) *positive constants q^x, τ^x, ℓ^x, m^x with $\ell^x, m^x \in \{1, \dots, d\}$,*
- (ii) *angular velocities $\theta_1^x, \dots, \theta_{m^x}^x \in \mathbb{R}$, where all $\theta_k^x \neq 0$ come in pairs $(\theta_{j_*}^x, \theta_{j_*+1}^x) = (\theta_{j_*}^x, -\theta_{j_*}^x)$,*
- (iii) *linearly independent vectors $v_1^x, \dots, v_{m^x}^x$ in $\mathbb{C}^d$ which are complex conjugate $(v_{j_*}^x, v_{j_*+1}^x) = (v_{j_*}^x, \bar{v}_{j_*}^x)$ whenever $(\theta_{j_*}^x, \theta_{j_*+1}^x) = (\theta_{j_*}^x, -\theta_{j_*}^x)$,*

such that

$$\lim_{t \rightarrow \infty} \left| \frac{e^{q^x t}}{t^{\ell^x - 1}} X_{t+\tau^x}^0(x) - \sum_{k=1}^{m^x} e^{i\theta_k^x t} v_k^x \right| = 0. \quad (2.4)$$

Moreover,

$$0 < \liminf_{t \rightarrow \infty} \left| \sum_{k=1}^{m^x} e^{i t \theta_k^x} v_k^x \right| \leq \limsup_{t \rightarrow \infty} \left| \sum_{k=1}^{m^x} e^{i t \theta_k^x} v_k^x \right| \leq \sum_{k=1}^{m^x} |v_k^x|. \quad (2.5)$$

The formal proof of the previous lemma is given in Lemma B.2 in Appendix B of [5].

Remark 2.3 (1) Convention: Note that $\theta_k^x = 0$ is true for at most one index $k \in \{1, \dots, m^x\}$.

If such an index shows up in $\theta_1^x, \dots, \theta_{m^x}^x$ we adopt the convention that $\theta_1^x = 0$ and $v_1^x \in \mathbb{R}^d$, and hence $m^x = 2n + 1$ for some $n \in \mathbb{N}_0$. Otherwise, $m^x = 2n$ for some $n \in \mathbb{N}_0$ and we eliminate θ_1^x and count the angular velocities as follows $\theta_2^x, \dots, \theta_{2n+1}^x$.

(2) Note that the linearly independent complex vectors $v_1^x, \dots, v_{m^x}^x$ in $\mathbb{C}^d$ not only depend on x but also crucially on the dissipation time τ^x of the deterministic system to a Hartman–Grobman domain of conjugacy U . We stress that τ^x is not unique since $X_{t+\tau^x}^0(x) \in U$ for all $t \geq 0$.

(3) A word about the parameters ℓ^x, q^x and m^x in Lemma 2.2. By the Hartman–Grobman theorem there are open sets $0 \in U, V \subset \mathbb{R}^d$ and a homeomorphism $H : U \rightarrow V$ with $H(0) = 0$ satisfying for all $u \in U$ and $t \geq 0$

$$H(X_t^0(u)) = e^{-Db(0)t} H(u). \quad (2.6)$$

In fact, by Hypothesis 1 we have that H is a $\mathcal{C}^1$ -diffeomorphism, see the original paper [8] or Theorem (Hartman), Sec. 2.8, p.127, [13]. In [8] it is shown that H can be chosen to be

$$H(x) = x + o(|x|)_{|x| \rightarrow 0}.$$

Let $\tilde{u} = X_{\tau^x}^0(x) \in U$. With the help of a linear coordinate change W we obtain the Jordan normal form $Db(0) = W^{-1}J(Db(0))W$ and (using the linearity of the semigroup)

$$H(X_{t+\tau^x}^0(x)) = W^{-1}e^{-J(Db(0))t}(WH(\tilde{u})).$$

We denote $\tilde{w} = WH(\tilde{u})$. Now, the parameters ℓ^x , q^x and m^x are given as follows. Consider the sequence of generalized eigenspaces H_j of $J(Db(0))$ such that

$$\mathbb{R}^d = H_1 \oplus \cdots \oplus H_{k_*}.$$

By construction, $\tilde{w} \in G(\tilde{w}) := \text{span}(\{H_k \mid \text{where } 1 \leq k \leq k_* : \text{proj}(\tilde{w}, H_k) \neq 0\})$. Note that $G(\tilde{w})$ is unique. We consider the restriction

$$\tilde{J}(\tilde{w}) := J(Db(0))|_{G(\tilde{w})}.$$

Now, q^x is the smallest real part of the spectrum of $\tilde{J}(\tilde{w})$, ℓ^x is the dimension of the largest Jordan block of $\tilde{J}(\tilde{w})$ which has the real part q^x and m^x is the number of Jordan blocks associated to q^x and ℓ^x . Note that in case of a non real eigenvalue with real part q^x and Jordan block size ℓ^x , we have $m^x \geq 2$. For an extensive numerical example for a linear chain of oscillators we refer to Sect. 4.3.2 in [2].

2.3 Main Results

Our first main result establishes $\infty/0$ collapse of the Wasserstein distance between the law of the current state $X_t^\varepsilon(x)$ and the dynamical equilibrium μ^ε along the critical time scale t_ε^x given in (2.7) under mild conditions.

Theorem 1 (Window cutoff) *Let b satisfy Hypothesis 1 and v satisfy Hypothesis 2 for some $p_* > 0$. Fix $x \in \mathbb{R}^d \setminus \{0\}$ and consider the notation in the asymptotic Hartman–Grobman representation $q^x > 0$, $\ell^x, m^x \in \{1, \dots, d\}$, $\theta_1^x, \dots, \theta_{m^x}^x \in [0, 2\pi)$, $v_1^x, \dots, v_{m^x}^x \in \mathbb{C}^d$ and $\tau^x > 0$ of Lemma 2.2.*

Then the family of processes $(X^\varepsilon(x))_{\varepsilon>0}$ exhibits a window cutoff phenomenon on the time scale

$$t_\varepsilon^x = \frac{1}{q^x} |\ln(\varepsilon)| + \frac{\ell^x - 1}{q^x} \ln(|\ln(\varepsilon)|) \quad (2.7)$$

and for all asymptotically constant window sizes w_ε , that is, $w_\varepsilon \rightarrow w > 0$ as $\varepsilon \rightarrow 0$, in the following sense. For all $0 < p < p^$ we have*

$$\begin{aligned} \lim_{r \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} &= 0 \quad \text{and} \\ \lim_{r \rightarrow -\infty} \liminf_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} &= \infty. \end{aligned} \quad (2.8)$$

The second main result provides two characterizations for the proper limits ($\varepsilon \rightarrow 0$) of the expressions in (2.8) for any fixed $r \in \mathbb{R}$. That is to say, we characterize under which

conditions the asymptotics (1.3) is satisfied. In addition, it yields the precise shape of the limit which turn out to be a simple exponential function for $p \in [1, p_*)$.

Theorem 2 (Dynamical profile cutoff characterization for $p_* > 0$) *Let the assumptions (and the notation) of Theorem 1 be valid for some $p_* > 0$. Consider the unique strong solution $(\mathcal{O}_t)_{t \geq 0}$ of the linear system*

$$d\mathcal{O}_t = -Db(0)\mathcal{O}_t + dL_t, \quad (2.9)$$

where $\mathcal{O}_\infty$ is the unique invariant probability distribution of (2.9).

(1) *Then for any $0 < p < p_*$ the following statements are equivalent.*

(i) *For any $\lambda > 0$, the function $\omega(x) \ni u \mapsto \mathcal{W}_p(\lambda u + \mathcal{O}_\infty, \mathcal{O}_\infty)$ is constant, where*

$$\omega(x) := \left\{ \text{accumulation points of } \sum_{k=1}^m e^{it\theta_k^x} v_k^x \text{ as } t \rightarrow \infty \right\}.$$

(ii) *The family of processes $(X^\varepsilon(x))_{\varepsilon > 0}$ exhibits a profile cutoff for any $0 < p < p_*$ as follows*

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} = \mathcal{P}_p^x(r) \quad \text{for any } r \in \mathbb{R},$$

where

$$\mathcal{P}_p^x(r) := \mathcal{W}_p(\kappa^x(r) \cdot v + \mathcal{O}_\infty, \mathcal{O}_\infty) \quad \text{for any } v \in \omega(x) \quad (2.10)$$

and

$$\kappa^x(r) = \frac{e^{-q^x r \cdot w}}{e^{q^x r^x} (q^x)^{\ell^x - 1}}.$$

(2) *For $p_* > 1$ and $p \in [1, p_*)$ the profile has the shape*

$$\mathcal{P}_p^x(r) = \kappa^x(r) \cdot |v| \quad \text{for all } v \in \omega(x)$$

if and only if $\omega(x)$ is contained in a sphere in $\mathbb{R}^d$ with respect to the Euclidean norm.

(3) *We recall the convention of Remark 2.3. Let $p_* > 1$ and $p \in [1, p_*)$. If the angles $\theta_2^x, \dots, \theta_{2n}^x$ satisfy the following non-resonance condition*

$$h_1 \theta_2 + \dots + h_n \theta_{2n} \in 2\pi \cdot \mathbb{Z} \quad \text{for all } (h_1, \dots, h_n) \in \mathbb{Z}^n \setminus \{0\}, \quad (2.11)$$

then the statements (i) and (ii) in item (1) are equivalent to the following normal growth condition of the asymptotic Hartman–Grobman linearization: The family of limiting vectors $(v_1^x, \operatorname{Re} v_2^x, \operatorname{Im} v_2^x, \dots, \operatorname{Re} v_{2n}^x, \operatorname{Im} v_{2n}^x)$ is orthogonal in $\mathbb{R}^d$ and satisfies

$$|\operatorname{Re} v_{2k}^x| = |\operatorname{Im} v_{2k}^x| \quad \text{for all } k = 1, \dots, n.$$

Remark 2.4 We stress that $\mathcal{O}_\infty = \lim_{t \rightarrow \infty} \mathcal{O}_t$ in $\mathcal{W}_{p_*}$ and due to Hypothesis 1 (in combination with Hypothesis 2) the distribution of $\mathcal{O}_\infty$ does not depend on any deterministic initial condition of (2.9).

Due to its relevance as physical observables, we formulate the corresponding window cutoff result for the respective moments.

Corollary 2.5 (Moments cutoff) *Let the assumptions (and the notation) of Theorem 1 be valid for some $p_* > 0$. Then for any $0 < p < p_*$ it follows*

$$\lim_{r \rightarrow \infty} \liminf_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[|X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x)|^p]}{\varepsilon^p} = \lim_{r \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[|X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x)|^p]}{\varepsilon^p} = \mathbb{E}[|\mathcal{O}_\infty|^p],$$

$$\lim_{r \rightarrow -\infty} \liminf_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[|X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x)|^p]}{\varepsilon^p} = \lim_{r \rightarrow -\infty} \limsup_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[|X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x)|^p]}{\varepsilon^p} = \infty.$$

3 Examples

In this section we present two examples which illustrate the applicability of Theorem 1 and Theorem 2 to nonlinear dynamics with degenerate noise.

Example 3.1 (The Fermi–Pasta–Ulam–Tsingou potential) We consider the nonlinear Langevin gradient system

$$dX_t^\varepsilon = -\nabla \mathcal{U}(X_t^\varepsilon)dt + \varepsilon dL_t \quad (3.1)$$

for the strongly convex quartic Fermi–Pasta–Ulam–Tsingou potential $\mathcal{U}(x) = \frac{1}{2}|x|^2 + \frac{1}{4}|x|^4$, $x \in \mathbb{R}^d$ subject to degenerate noise dL_t . For any Lévy process L satisfying Hypothesis 2 for some $p_* > 0$ the system (3.1) exhibits a profile cutoff due to Theorem 2 where the cutoff time is given by $t_\varepsilon^x = |\ln(\varepsilon)|$. For $p_* > 1$ and any $p \in [1, p_*)$ the profile function in $\mathcal{W}_p$ is always of the following exponential shape

$$\mathcal{P}_p^x(r) = e^{-wr - \tau^x} \left| \sum_{k=1}^m v_k^x \right|, \quad (3.2)$$

where $\tau^x := \min\{t \geq 0 : |X_t^0(x)| \leq R_0/2\}$ and R_0 being an small radius inside of which Hartman–Grobman conjugation is valid. Note that τ^x can be replaced by any upper bound of τ^x such as for instance $(1/\delta) \ln(2|x|/R_0)$ given by Hypothesis 1.

In particular, the profile cutoff (3.2) is valid for $L = L^\alpha$ being an (possibly degenerate) α -stable process with index $\alpha \in (1, 2]$. Note that for the limiting case of a possibly degenerate Cauchy process ($\alpha = 1$) and in fact of any L^α with index $\alpha \in (0, 1)$, Theorem 2 also yields a profile cutoff. However, the profile function remains not explicit. This is due to the absence of a finite first moment and the lack of the shift linearity (2.2). In other words, the profile function is given in (2.10) for $p \in (0, \alpha)$ and up to our knowledge unknown how to simplify further. Note that the case of $\alpha \in (0, 3/2]$ is new and is not covered in [3].

Example 3.2 (Nonlinear non-gradient with degenerate noise) For $F, \mathcal{H} \in \mathcal{C}^2(\mathbb{R}^2, \mathbb{R})$ we consider the following perturbed simple harmonic oscillator with unit angular frequency given in Section 4 of [19] subject to a small noise perturbation

$$d \begin{pmatrix} X_t^{\varepsilon,1} \\ X_t^{\varepsilon,2} \end{pmatrix} = - \begin{pmatrix} X_t^{\varepsilon,2} F(X_t^{\varepsilon,1}, X_t^{\varepsilon,2}) - \partial_1 \mathcal{H}(X_t^{\varepsilon,1}, X_t^{\varepsilon,2}) \\ -X_t^{\varepsilon,1} F(X_t^{\varepsilon,1}, X_t^{\varepsilon,2}) - \partial_2 \mathcal{H}(X_t^{\varepsilon,1}, X_t^{\varepsilon,2}) \end{pmatrix} dt + \varepsilon d \begin{pmatrix} 0 \\ \mathcal{L}_t \end{pmatrix},$$

where $\mathcal{L} = (\mathcal{L}_t)_{t \geq 0}$ is a one dimensional Lévy process with finite p_* -th moments. The Jacobian matrix $Jb(v_1, v_2)$ at (v_1, v_2) of the respective vector field $b : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is given by

$$\begin{pmatrix} v_2 \partial_1 F(v_1, v_2) - \partial_{11} \mathcal{H}(v_1, v_2) & F(v_1, v_2) + v_2 \partial_2 F(v_1, v_2) - \partial_{12} \mathcal{H}(v_1, v_2) \\ -F(v_1, v_2) - v_1 \partial_1 F(v_1, v_2) - \partial_{12} \mathcal{H}(v_1, v_2) & -v_1 \partial_2 F(v_1, v_2) - \partial_{22} \mathcal{H}(v_1, v_2) \end{pmatrix}.$$

It is enough to prove the existence of a positive constant δ such that for any $u_1, u_2, v_1, v_2 \in \mathbb{R}$ it follows

$$\begin{aligned}(u_1, u_2)Jb(v_1, v_2)(u_1, u_2)^* &= (v_2\partial_1 F(v_1, v_2) - \partial_{11}\mathcal{H}(v_1, v_2))u_1^2 \\ &\quad + (-v_1\partial_2 F(v_1, v_2) - \partial_{22}\mathcal{H}(v_1, v_2))u_2^2 \\ &\quad + (v_2\partial_2 F(v_1, v_2) - v_1\partial_1 F(v_1, v_2) - 2\partial_{12}\mathcal{H}(v_1, v_2))u_1u_2 \\ &\geq \delta(u_1^2 + u_2^2).\end{aligned}\quad (3.3)$$

For instance, for a nonlinear perturbation of a linear oscillator, that is, $F(v_1, v_2) = \eta$ for some $\eta > 0$, the preceding condition reads

$$-\left(\partial_{11}\mathcal{H}(v_1, v_2)u_1^2 + \partial_{22}\mathcal{H}(v_1, v_2)u_2^2 + 2\partial_{12}\mathcal{H}(v_1, v_2)u_1u_2\right) \geq \delta(u_1^2 + u_2^2).$$

For $\mathcal{L}$ satisfying Hypothesis 2 with p_* , and $F, \mathcal{H}$ fulfilling (3.3) Theorem 1 implies window cutoff for any initial condition $(X_0^{\varepsilon,1}, X_0^{\varepsilon,2}) = x \in \mathbb{R}^2 \setminus \{0\}$ and any $p \in (0, p_*)$. The cutoff time is given by

$$t_\varepsilon^x = \frac{1}{q^x} |\ln(\varepsilon)| + \frac{\ell^x - 1}{q^x} \ln(|\ln(\varepsilon)|).$$

Note that this result is new even in the Brownian case since the results of [3] and [5] are stated for the total variation distance which requires regularity on the transition probabilities given in the setting of non-degenerate noise. In our case, the Wasserstein distance circumvents this difficulty by the continuity of $\mathcal{W}_p(x + X, X)$ for any $X \in L^p$ as $|x| \rightarrow 0$ and $|x| \rightarrow \infty$, while for total variation distance it requires absolutely continuity on the distribution of X . We refer to [3], Lemma 1.17 in Subsection 1.3.5, for an example where the continuity of the total variation distance under shifts is not valid.

In the sequel, we characterize the existence of a profile cutoff under (3.3) in terms of the linearization at the stable state $(0, 0)$. Let $a := -\partial_{11}^2 \mathcal{H}(0, 0)$, $b := -\partial_{22}^2 \mathcal{H}(0, 0)$, $c := -\partial_{12}^2 \mathcal{H}(0, 0)$ and $\eta_0 := -F(0, 0)$. Then

$$Jb(0, 0) = \begin{pmatrix} a & -\eta_0 + c \\ \eta_0 + c & b \end{pmatrix}.$$

Note that $\eta_0 = c$ implies that the eigenvalues of $Jb(0, 0)$ are the numbers a and b which are positive and hence by Theorem 2 profile cutoff is valid. In the sequel we assume $\eta_0 \neq c$. Then the eigenvalues of $Jb(0, 0)$ are given by

$$\lambda_\pm := \frac{(a+b) \pm \sqrt{\Delta}}{2}, \quad \Delta := (a-b)^2 + 4(c^2 - \eta_0^2),$$

with corresponding eigenvectors

$$v_\pm := \left(1, -\frac{a-b \mp \sqrt{\Delta}}{2(-\eta_0 + c)}\right).$$

In addition,

$$\operatorname{Re}(v_\pm) = \begin{cases} \left(1, -\frac{a-b \mp \sqrt{\Delta}}{2(-\eta_0 + c)}\right) & \text{if } \Delta \geq 0, \\ \left(1, -\frac{a-b}{2(-\eta_0 + c)}\right) & \text{if } \Delta < 0, \end{cases} \quad \text{and} \quad \operatorname{Im}(v_\pm) = \begin{cases} (0, 0) & \text{if } \Delta \geq 0, \\ \pm \left(0, \frac{\sqrt{|\Delta|}}{2(-\eta_0 + c)}\right) & \text{if } \Delta < 0. \end{cases}$$

For $\Delta \geq 0$ Theorem 2 yields a profile cutoff phenomenon. For $\Delta < 0$ Theorem 1 implies the weaker window cutoff phenomenon, however, by part (3) of Theorem 2 the stronger profile cutoff for $p_* > 1$ and $p \in [1, p_*)$ is valid if and only if

$$|\operatorname{Re}(v_+)|^2 = |\operatorname{Im}(v_+)|^2 \text{ and } \langle \operatorname{Re}(v_+), \operatorname{Im}(v_+) \rangle = 0$$

which is equivalent to special case $a = b$ and $c = 0$. In other words, $e^{-Jb(0,0)t} = e^{-at} R(\theta t)$, where $R(\theta t)$ is an orthogonal 2×2 matrix with angle θt .

Remark 3.3 (A word about the linear dynamics) In [2] the authors study (1.1) for the linear vector field $b(x) = Qx$ for any Hurwitz stable matrix $-Q$, that is, $\operatorname{Re}(\lambda) < 0$ for any eigenvalue λ of $-Q$. Under these assumptions, the results of Theorem 1 and Theorem 2 are obtained.

It is not hard to see that Hypothesis 1 implies $\operatorname{Re}(\lambda) \leq -\delta$ for any eigenvalue λ of $-Q$ and hence Hurwitz stability. However, the dissipativity condition (1.2) which is assumed in order to control the nonlinear vector field, is strictly stronger than Hurwitz stability. For instance, the vector field $b : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given by $b(x) = Qx$ with

$$-Q = \begin{pmatrix} 0 & -1 \\ \lambda & \lambda \end{pmatrix} \text{ with } \lambda \in (0, 1/2)$$

has eigenvalues with real part $-\lambda/2 < 0$, but it does not satisfy Hypothesis 1. Note that the dissipativity condition (1.2) is not even satisfied locally in a neighborhood of the origin.

4 Proofs of the Main Results

4.1 The First Order Approximation

We define the Freidlin-Wentzell first order approximation given by

$$Y_t^\varepsilon(x) = X_t^0(x) + \varepsilon \mathcal{Y}_t^x, \quad t \geq 0, \quad (4.1)$$

where $(\mathcal{Y}_t^x)_{t \geq 0}$ is the unique strong solution of the linear inhomogeneous SDE

$$\begin{cases} d\mathcal{Y}_t^x = -Db(X_t^0(x))\mathcal{Y}_t^x dt + dL_t & \text{for any } t \geq 0, \\ \mathcal{Y}_0^x = 0. \end{cases} \quad (4.2)$$

In [3], Lemma C.4 in Section C.4 it is shown that $Y_t^\varepsilon(x)$ converges in total variation distance to a unique limiting distribution μ_*^ε as $t \rightarrow \infty$. Moreover, it is shown there that $\mu_*^\varepsilon \stackrel{d}{=} \varepsilon \mathcal{O}_\infty$, where $\mathcal{O}_\infty$ is the unique invariant probability distribution of the homogeneous Ornstein-Uhlenbeck dynamics

$$d\mathcal{O}_t = -Db(0)\mathcal{O}_t + dL_t. \quad (4.3)$$

In the sequel we reduce the nonlinear ergodic convergence of $X_t^\varepsilon(x)$ to the ergodic convergence of the Freidlin-Wentzell linearization $Y_t^\varepsilon(x)$ in (4.4) up to error terms. For any $0 < p \leq p_*$, by the triangle inequality it follows that

$$\mathcal{W}_p(X_t^\varepsilon(x), \mu^\varepsilon) \leq \mathcal{W}_p(X_t^\varepsilon(x), Y_t^\varepsilon(x)) + \mathcal{W}_p(Y_t^\varepsilon(x), \mu_*^\varepsilon) + \mathcal{W}_p(\mu_*^\varepsilon, \mu^\varepsilon)$$

for any $t \geq 0, x \in \mathbb{R}^d$. Analogously we estimate

$$\mathcal{W}_p(Y_t^\varepsilon(x), \mu_*^\varepsilon) \leq \mathcal{W}_p(Y_t^\varepsilon(x), X_t^\varepsilon(x)) + \mathcal{W}_p(X_t^\varepsilon(x), \mu^\varepsilon) + \mathcal{W}_p(\mu^\varepsilon, \mu_*^\varepsilon).$$

Combining the preceding inequalities we obtain the linear approximation

$$|\mathcal{W}_p(X_t^\varepsilon(x), \mu^\varepsilon) - \mathcal{W}_p(Y_t^\varepsilon(x), \mu_*^\varepsilon)| \leq \mathcal{W}_p(X_t^\varepsilon(x), Y_t^\varepsilon(x)) + \mathcal{W}_p(\mu^\varepsilon, \mu_*^\varepsilon) \quad (4.4)$$

for any $t \geq 0$, $x \in \mathbb{R}^d$. In Proposition 2 given in “Appendix B.2” we show that for any $t_\varepsilon = O(|\ln(\varepsilon)|)$ and $0 < p < p_*$ the following limit holds

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon}^\varepsilon(x), Y_{t_\varepsilon}^\varepsilon(x))}{\varepsilon^{1 \wedge p}} = 0. \quad (4.5)$$

Moreover, in Lemma B.2 we show that for $0 < p < p_*$

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(\mu_*^\varepsilon, \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} = 0. \quad (4.6)$$

4.2 Derivation of the Cutoff Phenomenon

In the sequel, we analyze the asymptotic behavior of $\mathcal{W}_p(Y_t^\varepsilon(x), \mu_*^\varepsilon) \cdot \varepsilon^{-(1 \wedge p)}$ from which we recognize the cutoff of the Freidlin-Wentzell linearization $Y_t^\varepsilon(x)$. By the triangle inequality, translation invariance, homogeneity and shift linearity given in Lemma 2.1 we obtain for $0 < p \leq p_*$

$$\begin{aligned} \mathcal{W}_p(Y_t^\varepsilon(x), \mu_*^\varepsilon) &= \mathcal{W}_p(X_t^0(x) + \varepsilon \mathcal{Y}_t^x, \varepsilon \mathcal{O}_\infty) \\ &\leq \mathcal{W}_p(X_t^0(x) + \varepsilon \mathcal{Y}_t^x, X_t^0(x) + \varepsilon \mathcal{O}_\infty) + \mathcal{W}_p(X_t^0(x) + \varepsilon \mathcal{O}_\infty, \varepsilon \mathcal{O}_\infty) \\ &= \varepsilon^{1 \wedge p} \cdot \mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{O}_\infty) + \varepsilon^{1 \wedge p} \cdot \mathcal{W}_p(\varepsilon^{-1} \cdot X_t^0(x) + \mathcal{O}_\infty, \mathcal{O}_\infty). \end{aligned}$$

Analogously we deduce

$$\mathcal{W}_p(Y_t^\varepsilon(x), \mu_*^\varepsilon) \geq \varepsilon^{1 \wedge p} \cdot \mathcal{W}_p(\varepsilon^{-1} \cdot X_t^0(x) + \mathcal{O}_\infty, \mathcal{O}_\infty) - \varepsilon^{1 \wedge p} \cdot \mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{O}_\infty).$$

Consequently,

$$\left| \frac{\mathcal{W}_p(Y_t^\varepsilon(x), \mu_*^\varepsilon)}{\varepsilon^{1 \wedge p}} - \mathcal{W}_p(\varepsilon^{-1} \cdot X_t^0(x) + \mathcal{O}_\infty, \mathcal{O}_\infty) \right| \leq \mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{O}_\infty). \quad (4.7)$$

The right-hand side of (4.7) does not depend of ε and by Lemma B.3 it tends to 0 as $t \rightarrow \infty$. It is therefore enough to study the precise longterm behavior of $\mathcal{W}_p(\varepsilon^{-1} \cdot X_t^0(x) + \mathcal{O}_\infty, \mathcal{O}_\infty)$ in order to derive the cutoff phenomenon.

4.3 Proof of Theorem 1

For any $0 < p < p_*$, t_ε^x and w_ε being given in statement and $r \in \mathbb{R}$, (4.4), (4.5), (4.6), (4.7) yield

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} &= \limsup_{\varepsilon \rightarrow 0} \mathcal{W}_p\left(\frac{X_t^0(x)}{\varepsilon} + \mathcal{O}_\infty, \mathcal{O}_\infty\right), \\ \liminf_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} &= \liminf_{\varepsilon \rightarrow 0} \mathcal{W}_p\left(\frac{X_t^0(x)}{\varepsilon} + \mathcal{O}_\infty, \mathcal{O}_\infty\right). \end{aligned}$$

For short, we define

$$\mathfrak{T}_\varepsilon^x = t_\varepsilon^x + r \cdot w_\varepsilon - \tau^x \quad \text{and} \quad \Lambda^x(\varepsilon) := \frac{(\mathfrak{T}_\varepsilon^x)^{\ell-1}}{\varepsilon e^{q^x \mathfrak{T}_\varepsilon^x}} \sum_{k=1}^m e^{i \mathfrak{T}_\varepsilon^x \theta_k^x} v_k^x. \quad (4.8)$$

Claim A.

$$\limsup_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} = \limsup_{\varepsilon \rightarrow 0} \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty)$$

and

$$\liminf_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} = \liminf_{\varepsilon \rightarrow 0} \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty).$$

for any $0 < p < p_*$. In particular, the limit

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} \text{ exists iff } \lim_{\varepsilon \rightarrow 0} \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty) \text{ exists.} \quad (4.9)$$

Proof of Claim A. In the sequel we study the asymptotics of the drift term $X_t^0(x) \cdot \varepsilon^{-1}$. A straightforward calculation shows

$$\lim_{\varepsilon \rightarrow 0} \frac{(\mathfrak{T}_\varepsilon^x)^{\ell-1} e^{-q^x \mathfrak{T}_\varepsilon^x}}{\varepsilon} = e^{-q^x \tau} (q^x)^{1-\ell} e^{-q^x r \cdot w}. \quad (4.10)$$

The preceding limit implies with the help of the spectral decomposition (2.4) given in Lemma 2.2 and the triangle inequality that

$$\begin{aligned} \mathcal{W}_p\left(\frac{X_{t_\varepsilon^x + r \cdot w_\varepsilon}^0(x)}{\varepsilon} + \mathcal{O}_\infty, \mathcal{O}_\infty\right) &\leq \mathcal{W}_p\left(\left(\frac{X_{\tau + \mathfrak{T}_\varepsilon^x}^0}{\varepsilon} - \Lambda^x(\varepsilon)\right) + \mathcal{O}_\infty, \mathcal{O}_\infty\right) \\ &\quad + \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty). \end{aligned} \quad (4.11)$$

We set

$$R_\varepsilon^x := \mathcal{W}_p\left(\left(\frac{X_{\tau + \mathfrak{T}_\varepsilon^x}^0}{\varepsilon} - \Lambda^x(\varepsilon)\right) + \mathcal{O}_\infty, \mathcal{O}_\infty\right).$$

Analogous reasoning yields

$$\mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty) \leq \mathcal{W}_p\left(\frac{X_{t_\varepsilon^x + r \cdot w_\varepsilon}^0}{\varepsilon} + \mathcal{O}_\infty, \mathcal{O}_\infty\right) + R_\varepsilon^x.$$

In the sequel it remains to show that $R_\varepsilon^x \rightarrow 0$ as $\varepsilon \rightarrow 0$. By the continuity of $z \rightarrow \mathcal{W}_p(z + \mathcal{O}_\infty, \mathcal{O}_\infty)$ at $z = 0$ it is enough to prove

$$\left| \frac{X_{\tau + \mathfrak{T}_\varepsilon^x}^0}{\varepsilon} - \Lambda^x(\varepsilon) \right| \rightarrow 0, \quad \varepsilon \rightarrow 0,$$

which is valid due to the limit (2.4) and (4.10). This finishes the proof of Claim A. $\square$

In the sequel, we prove the window cutoff asymptotics in (2.8). Note that $\Lambda^x(\varepsilon)$ is uniformly bounded on $\varepsilon \in (0, 1]$. For any accumulation point U (as $\varepsilon \rightarrow 0$) of $(\mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty))_{\varepsilon \in (0, 1]}$ there exists a sequence $(\varepsilon_k)_{k \in \mathbb{N}}$, $\varepsilon_k \rightarrow 0$ as $k \rightarrow \infty$, such that

$$U = \lim_{k \rightarrow \infty} \mathcal{W}_p(\Lambda^x(\varepsilon_k) + \mathcal{O}_\infty, \mathcal{O}_\infty).$$

The Bolzano–Weierstrass theorem for the sequence $(\Lambda(\varepsilon_k))_{k \in \mathbb{N}}$, the limit (4.10) and the continuity of $\mathcal{W}_p$ yield

$$U = \mathcal{W}_p(e^{-q^x \tau^x} (q^x)^{1-\ell^x} e^{-q^x w r} u + \mathcal{O}_\infty, \mathcal{O}_\infty) \quad \text{for some } u \in \omega(x). \quad (4.12)$$

In particular,

$$\limsup_{\varepsilon \rightarrow 0} \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty) = \mathcal{W}_p(e^{-q^x \tau^x} (q^x)^{1-\ell^x} e^{-q^x w r} \hat{u} + \mathcal{O}_\infty, \mathcal{O}_\infty),$$

$$\liminf_{\varepsilon \rightarrow 0} \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty) = \mathcal{W}_p(e^{-q^x \tau^x} (q^x)^{1-\ell^x} e^{-q^x w r} \check{u} + \mathcal{O}_\infty, \mathcal{O}_\infty),$$

where $\hat{u}, \check{u} \in \omega(x)$ and $\check{u} \neq 0$ by (2.5). Hence item (d) in Lemma 2.1 implies

$$\begin{aligned} \lim_{r \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty) &= 0 \quad \text{and} \\ \lim_{r \rightarrow -\infty} \liminf_{\varepsilon \rightarrow 0} \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty) &= \infty. \end{aligned}$$

This finishes the proof of Theorem 1. $\square$

4.4 Proof of Theorem 2

We keep the notation (4.8) of the proof of Theorem 1. By (4.9) it is enough to prove that the limit

$$\lim_{\varepsilon \rightarrow 0} \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty) \quad \text{exists.} \quad (4.13)$$

We recall the definition of $\Lambda^x(\varepsilon)$ (4.8) and the limit (4.10). By (4.12) we have

$$\begin{aligned} &\left\{ \text{accumulation points of } \mathcal{W}_p(\Lambda^x(\varepsilon) + \mathcal{O}_\infty, \mathcal{O}_\infty) \text{ as } \varepsilon \rightarrow 0 \right\} \\ &= \left\{ \mathcal{W}_p((e^{-q^x \tau^x} (q^x)^{1-\ell^x} e^{-q^x w r} u + \mathcal{O}_\infty, \mathcal{O}_\infty) : u \in \omega(x)) \right\}. \end{aligned} \quad (4.14)$$

For $p \geq 1$, the shift linearity given in item d) of Lemma 2.1 implies

$$\mathcal{W}_p(e^{-q^x \tau^x} (q^x)^{1-\ell^x} e^{-q^x w r} u + \mathcal{O}_\infty, \mathcal{O}_\infty) = e^{-q^x \tau^x} (q^x)^{1-\ell^x} e^{-q^x w r} |u|. \quad (4.15)$$

Combining (4.14) and (4.15) we infer

$$\begin{aligned} &\left\{ \mathcal{W}_p((e^{-q^x \tau^x} (q^x)^{1-\ell^x} e^{-q^x w r} u + \mathcal{O}_\infty, \mathcal{O}_\infty) : u \in \omega(x)) \right\} \\ &= \left\{ e^{-q^x \tau^x} (q^x)^{1-\ell^x} e^{-q^x w r} |u| : u \in \omega(x) \right\}. \end{aligned} \quad (4.16)$$

Hence (4.14) and (4.16) imply that the limit (4.13) exists if and only if the right-hand side of (4.16) has exactly one element. This is equivalent to $\omega(x)$ being contained in a sphere in $\mathbb{R}^d$ with respect to the Euclidean distance. For $p \in (0, 1)$ the shift linearity is not valid and we are stuck after (4.14). Consequently, (4.14) holds true and the limit (4.13) exists if and only if for all $\lambda > 0$ the function

$$\omega(x) \ni u \mapsto \mathcal{W}_p(\lambda u + \mathcal{O}_\infty, \mathcal{O}_\infty) \quad \text{is constant.}$$

This finishes the proof of Theorem 2. $\square$

Acknowledgements The authors would like to thank the anonymous referee for her/his valuable comments which has led to significant improvement of the manuscript. The authors would like to thank Carlos Gustavo Tamm de Araújo Moreira (Gugu) at IMPA for clarifying comments on the Hartman–Grobman theorem.

Author Contributions All authors have contributed equally to the paper.

Funding Open Access funding provided by University of Helsinki including Helsinki University Central Hospital. The research of GB has been supported by the Academy of Finland, via the Matter and Materials

Profi4 University Profiling Action, an Academy Project (Project No. 339228) and the Finnish Centre of Excellence in Randomness and Structures (Project No. 346306). GB also would like to express his gratitude to University of Helsinki for all the facilities used along the realization of this work. The research of MAH has been supported by the proyecto de la Convocatoria 2020–2021: “Stochastic dynamics of systems perturbed with small Markovian noise with applications in biophysics, climatology and statistics” of the Facultad de Ciencias at Universidad de los Andes.

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A. Existence of the Invariant Measure

A.1 Invariant distribution μ^ε

In the sequel we show the existence of a unique invariant distribution μ^ε of the solution of (1.1) for any $\varepsilon > 0$. We stress that beyond the existence of moments (Hypothesis 2), this does not include any regularity such as absolute continuity whatsoever in our setting. For instance, our setting covers nonlinear oscillators with degenerate noise in Example 3.2.

We recall the standing assumptions Hypothesis 1 with $\delta > 0$ and Hypothesis 2 with $p_* > 0$. For the existence of the invariant probability measure μ^ε it is enough to verify the following condition by [7], p. 388. For some $x \in \mathbb{R}^d$, the limit

$$\lim_{R \rightarrow \infty} \liminf_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{P}(|X_t^\varepsilon(x)| > R) dt = 0. \quad (\text{A.1})$$

Hypotheses 1 and 2 imply inequality (D.3) p. 71 in [3]. That is to say, for $\gamma \in (0, 1 \wedge p_*)$ there exist positive constants C_1, C_2, C_3 such that for all $x \in \mathbb{R}^d$, $\varepsilon > 0$, $t \geq 0$, $A = \varepsilon \Pi$, $c = \varepsilon$

$$\mathbb{E}[|X_t^\varepsilon(x)|^\gamma] \leq e^{-\delta \gamma t} |x|^\gamma + C_3, \quad (\text{A.2})$$

where $C_3 = c^\gamma + \frac{1}{\gamma \delta} (\gamma \delta c^\gamma + C_1 \|A\|^\gamma + C_2 c^{\gamma-2} \|A\|^2) = \varepsilon^\gamma \cdot (2 + \frac{1}{\gamma \delta} (C_1 \|\Pi\|^\gamma + C_2 \|\Pi\|^2))$. Inequality (A.2) implies (A.1) with the help of the Markov inequality.

For the uniqueness, it enough to verify the following condition given in Theorem 11.4.3 in [9]. For any given positive numbers η, δ and R , there exists a positive constant S such that

$$\frac{1}{T} \int_0^T \mathbb{P}(|X_t^\varepsilon(x) - X_t^\varepsilon(y)| \geq \delta) dt < \eta \quad \text{for all } |x|, |y| \leq R \quad \text{and } T > S. \quad (\text{A.3})$$

Hypotheses 1, 2 and the additivity of the noise imply (D.5) p. 71 in [3]. In other words, for any $\gamma \in (0, 1 \wedge p_*)$, $x, y \in \mathbb{R}^d$, $t \geq 0$, $\varepsilon > 0$, $c = \varepsilon$ we have

$$\mathbb{E}[|X_t^\varepsilon(x) - X_t^\varepsilon(y)|^\gamma] \leq |x - y|^\gamma e^{-\delta\gamma t} + 2\varepsilon^\gamma.$$

The preceding inequality implies (A.3) with the help of the Markov inequality.

A.2 Convergence to μ^ε in $\mathcal{W}_{p_*}$ for $p_* > 0$

Due to Hypothesis 1 and the additive of the noise the natural coupling yields

$$|X_t^\varepsilon(x) - X_t^\varepsilon(y)| \leq |x - y|e^{-\delta t} \quad \text{for all } x, y \in \mathbb{R}^d, t \geq 0. \quad (\text{A.4})$$

Since μ^ε is an invariant measure and X^ε is a Feller process, disintegration and (A.4) imply

$$\begin{aligned} \mathcal{W}_{p_*}(X_t^\varepsilon(x), \mu^\varepsilon) &\leq \int_{\mathbb{R}^d} \mathcal{W}_{p_*}(X_t^\varepsilon(x), X_t^\varepsilon(y)) \mu^\varepsilon(dy) \leq e^{-(1 \wedge p_*)\delta t} \int_{\mathbb{R}^d} |x - y|^{1 \wedge p_*} \mu^\varepsilon(dy) \\ &\leq e^{-(1 \wedge p_*)\delta t} |x|^{1 \wedge p_*} + e^{-(1 \wedge p_*)\delta t} \int_{\mathbb{R}^d} |y|^{1 \wedge p_*} \mu^\varepsilon(dy). \end{aligned} \quad (\text{A.5})$$

The preceding right-hand side tends to zero as $t \rightarrow \infty$ provided that $\int_{\mathbb{R}^d} |y|^{1 \wedge p_*} \mu^\varepsilon(dy) < \infty$ which is shown in (2.84) p. 48 in [3].

Appendix B. L^p Estimates for $p \in (0, p_*)$

We recall the Lévy-Khinchin formula of L with characteristic triple (a, Σ, ν)

$$\ln(\mathbb{E}[e^{i\langle u, L_t \rangle}]) = t \left(i\langle a, u \rangle - \frac{1}{2} \langle u, \Sigma u \rangle + \int_{\mathbb{R}^d} \left(e^{i\langle u, z \rangle} - 1 - i\langle u, z \rangle \mathbf{1}_{(0,1)}(|z|) \right) \nu(dz) \right)$$

and the pathwise Lévy-Itô representation

$$L_t = at + \Sigma^{1/2} B_t + \int_0^t \int_{|z| \leq 1} z \tilde{N}(dsdz) + \int_0^t \int_{|z| > 1} z N(dsdz), \quad (\text{B.1})$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion in $\mathbb{R}^d$, N is a Poisson random measure on $[0, \infty) \times \mathbb{R}^d$ with intensity measure $dt \otimes \nu(dz)$ and $\tilde{N}$ is the compensated counterpart of N . See [16] for further details on Lévy processes.

We recall the standing assumptions Hypothesis 1 with $\delta > 0$ and Hypothesis 2 with $p_* > 0$.

B.1 Localization

We start with the probability estimate of the event

$$\mathcal{D}_t^x = \left\{ \sup_{0 \leq s \leq t} |\mathcal{Y}_s^x| > \vartheta \right\}, \quad \vartheta > 0.$$

where $\mathcal{Y}^x$ is given in (4.2). Note that $\mathcal{Z}_t(0) = \mathcal{Y}_t^0$ satisfies

$$d\mathcal{Z}_t(x) = -Db(0)\mathcal{Z}_t(x)dt + dL_t, \quad \mathcal{Z}_0(x) = x \quad (\text{B.2})$$

for $x = 0$.

Lemma B.1 For any $\gamma \in (0, p_* \wedge 1]$ there is a positive constant C such that for any $\vartheta \geq 1$, $x \in \mathbb{R}^d$ and $t \geq 0$ we have

$$\mathbb{P}(\mathcal{D}_t^x) \leq C t \vartheta^{-\gamma}. \quad (\text{B.3})$$

Proof By Theorem 1 in [17] we have

$$\sup_{0 \leq s \leq t} |\mathcal{Y}_s^x| \leq 6\sqrt{[\mathcal{Y}^x(0)]_s} + 2 \int_0^t H_{s-} \, dL_s, \quad \text{where} \quad H_{s-} = \frac{\mathcal{Y}_{s-}^x}{\sqrt{\sup_{s \leq t} (|\mathcal{Y}_{s-}^x|^2 + [\mathcal{Y}^x]_{s-})}}.$$

In particular, it follows

$$\begin{aligned} [\mathcal{Y}^x]_t &= [L]_t = \int_0^t \int_{|z| \leq 1} |z|^2 N(\mathrm{d}s \mathrm{d}z) \quad \text{such that} \\ \int_0^t H_{s-} \, d\mathcal{Y}_s^x &= \int_0^t \langle H_s, -Db(0)\mathcal{Y}_s^x \rangle \mathrm{d}s + \int_0^t \int_{|z| \leq 1} \langle H_{s-}, z \rangle \tilde{N}(\mathrm{d}s \mathrm{d}z) \\ &\quad + \int_0^t \int_{|z| > 1} \langle H_{s-}, z \rangle N(\mathrm{d}s \mathrm{d}z). \end{aligned}$$

By Hypothesis 1 we obtain $\int_0^t \langle H_{s-}, -Db(X_s^0(x))\mathcal{Y}_s^x \rangle \mathrm{d}s \leq 0$ a.s. Hence

$$\begin{aligned} &\mathbb{P}\left(\sup_{0 \leq s \leq t} |\mathcal{Y}_s^x| > \vartheta\right) \\ &\leq \mathbb{P}\left(6\left(\int_0^t \int_{|z| \leq 1} |z|^2 N(\mathrm{d}s \mathrm{d}z)\right)^{\frac{1}{2}} + 2 \int_0^t \int_{|z| \leq 1} \langle H_{s-}, z \rangle \tilde{N}(\mathrm{d}s \mathrm{d}z) \right. \\ &\quad \left. + 2 \int_0^t \int_{|z| > 1} \langle H_{s-}, z \rangle N(\mathrm{d}s \mathrm{d}z) > \vartheta\right) \\ &\leq \mathbb{P}\left(\int_0^t \int_{|z| \leq 1} |z|^2 N(\mathrm{d}s \mathrm{d}z) > \frac{\vartheta^2}{18^2}\right) + \mathbb{P}\left(\int_0^t \int_{|z| \leq 1} \langle H_{s-}, z \rangle \tilde{N}(\mathrm{d}s \mathrm{d}z) > \frac{2\vartheta}{3}\right) \\ &\quad + \mathbb{P}\left(\int_0^t \int_{|z| > 1} \langle H_{s-}, z \rangle N(\mathrm{d}s \mathrm{d}z) > \frac{2\vartheta}{3}\right). \end{aligned}$$

We continue term by term. By the Chebyshev inequality we obtain

$$\mathbb{P}\left(\int_0^t \int_{|z| \leq 1} |z|^2 N(\mathrm{d}s \mathrm{d}z) > \frac{\vartheta^2}{18^2}\right) \leq \frac{18^2 t}{\vartheta^2} \int_{|z| \leq 1} |z|^2 \nu(\mathrm{d}z) =: C_1 \frac{t}{\vartheta^2}$$

and

$$\begin{aligned} \mathbb{P}\left(\int_0^t \int_{|z| \leq 1} \langle H_{s-}, z \rangle \tilde{N}(\mathrm{d}s \mathrm{d}z) > \frac{2\vartheta}{3}\right) &\leq \left(\frac{3}{2}\right)^2 \frac{1}{\vartheta^2} \mathbb{E}\left[\left(\int_0^t \int_{|z| \leq 1} \langle H_{s-}, z \rangle \tilde{N}(\mathrm{d}s \mathrm{d}z)\right)^2\right] \\ &= \left(\frac{3}{2}\right)^2 \frac{1}{\vartheta^2} \mathbb{E}\left[\int_0^t \int_{|z| \leq 1} \langle H_{s-}, z \rangle^2 \nu(\mathrm{d}z) \mathrm{d}s\right] \\ &\leq \left(\frac{3}{2}\right)^2 \frac{t}{\vartheta^2} \int_{|z| \leq 1} |z|^2 \nu(\mathrm{d}z) =: C_2 \frac{t}{\vartheta^2}. \end{aligned}$$

Finally, for $\gamma \in (0, p_* \wedge 1]$ we have

$$\begin{aligned} \mathbb{P}\left(\int_0^t \int_{|z|>1} \langle H_{s-}, z \rangle N(ds dz) > \frac{2\vartheta}{3}\right) &\leq \mathbb{P}\left(\int_0^t \int_{|z|>1} |z| N(ds dz) > \frac{2\vartheta}{3}\right) \\ &\leq \left(\frac{3}{2}\right)^\gamma \frac{1}{\vartheta^\gamma} \mathbb{E}\left[\left(\int_0^t \int_{|z|>1} |z| N(ds dz)\right)^\gamma\right] \\ &\leq \left(\frac{3}{2}\right)^\gamma \frac{1}{\vartheta^\gamma} \mathbb{E}\left[\int_0^t \int_{|z|>1} |z|^\gamma N(ds dz)\right] \\ &= \left(\frac{3}{2}\right)^\gamma \frac{t}{\vartheta^\gamma} \int_{|z|>1} |z|^\gamma \nu(dz) =: C_3 \frac{t}{\vartheta^\gamma}, \end{aligned}$$

where we have used the subadditivity of the power γ in the sense of Subsection 1.1.2, see formula (1.6) in [15]. This finishes the proof of the statement. $\square$

B.2 First Order Approximation

We start with some technical preliminaries. In order to overcome that $u \mapsto |u|^p$ for $p \in (0, 2)$ is not twice continuously differentiable which turns out to be necessary for applying Itô's formula we use the following C^2 norm approximation $|x|_c := \sqrt{|x|^2 + c^2}$, $c > 0$, with the limiting case $|x|_0 = |x|$. It is well-behaved in the following sense. For any $c > 0$ we have

$$c \leq |x|_c \leq |x| + c, \quad \nabla |x|_c := \frac{x}{|x|_c} \quad \text{and} \quad 0 \leq \frac{|x|}{|x|_c} < 1.$$

Furthermore, it is straightforward to verify for $G(x) = |x|_c^p$ the following calculations

$$\nabla G(x) = p|x|_c^{p-1} \frac{x}{|x|_c} = p|x|_c^{p-2} x \quad \text{and} \quad |\nabla G(x)| \leq p|x|_c^{p-1}.$$

The L_1 -matrix norm $\|\cdot\|_1$ of the respective Hessian $H_G(x)$, $x \in \mathbb{R}^d$, can be estimated as follows

$$\|H_G(x)\|_1 \leq pd|x|_c^{p-2} + pd(2-p)|x|_c^{p-2} = C(p, d)|x|_c^{p-2}.$$

For details of the estimates, we refer to p. 69 in [3]. Since $p \in (0, 2)$ and $c \leq |x|_c$, we obtain

$$\sup_{x \in \mathbb{R}^d} |H_G(x)|_1 \leq C(p, d) c^{p-2}. \quad (\text{B.4})$$

Proposition 2 *We keep the notation of Theorem 1. Then for any $x \in \mathbb{R}^d$, $r \in \mathbb{R}$ and $p \in (0, p_*)$ it follows*

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x), Y_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(x))}{\varepsilon^{1 \wedge p}} = 0. \quad (\text{B.5})$$

Proof By the domination property of the Wasserstein distance in Lemma 2.1 it is enough to show the preceding limit in the respective L^p space. By (4.1) we have

$$dY_t^\varepsilon(x) = (-Db(X_t^0(x))Y_t^\varepsilon(x) + Db(X_t^0(x))X_t^0(x) - b(X_t^0(x)))dt + \varepsilon dL_t.$$

Let $\Delta_t^\varepsilon := X_t^\varepsilon(x) - Y_t^\varepsilon(x)$, $t \geq 0$. Then

$$d\Delta_t^\varepsilon = -(b(X_t^\varepsilon(x)) - b(Y_t^\varepsilon(x)))dt - (b(Y_t^\varepsilon(x)) - b(X_t^0(x)) - Db(X_t^0(x))\varepsilon Y_t^x)dt,$$

where $(\mathcal{Y}_t^x)_{t \geq 0}$ is given in (4.2). An elementary estimate of the p_* -th power of a sum yields for all $t \geq 0$

$$\begin{aligned} \mathbb{E}[|\Delta_t^\varepsilon|^{p_*}] &= \mathbb{E}[|(X_t^\varepsilon(x) - X_t^0(x)) + \varepsilon \mathcal{Y}_t^x|^{p_*}] \\ &\leq C_{p_*} (\mathbb{E}[|X_t^\varepsilon(x) - X_t^0(x)|^{p_*}] + \varepsilon^{p_*} \mathbb{E}[|\mathcal{Y}_t^x|^{p_*}]), \end{aligned} \quad (\text{B.6})$$

where C_{p_*} is a positive constant. Since $(\mathcal{Y}_t^x)_{t \geq 0}$ satisfies a dissipative linear equation, it exhibits the same integrability as L , which is straightforward to verify. There are a positive constant $\tilde{C}_{p_*}$ and a function $S_{p_*}(t)$ of at most polynomial order such that

$$\mathbb{E}[|\mathcal{Y}_t^x|^{p_*}] \leq \tilde{C}_{p_*} \mathbb{E}[|L_t|^{p_*}] \leq \tilde{C}_{p_*} S_{p_*}(t) \quad \text{for all } t \geq 0. \quad (\text{B.7})$$

For the first term of the right-hand side of (B.6), Lemmas B.4 and B.5 yield the following estimate. For any $\eta \in (0, p_*)$ there is a map $R_\eta : [0, \infty) \rightarrow [0, \infty)$ which increases with polynomial order as t tends to infinity, such that

$$\mathbb{E}[|X_t^\varepsilon(x) - X_t^0(x)|^{p_*}] \leq \varepsilon^{p_* - \eta} R_\eta(t) \quad \text{for any } t \geq 0. \quad (\text{B.8})$$

We start with the case $p_* > 1$ and $p \in (1, p_*)$. The Hölder inequality implies

$$\mathbb{E}[|\Delta_t^\varepsilon|^{p-1}] \leq (\mathbb{E}[|\Delta_t^\varepsilon|^{p_*}])^{\frac{p-1}{p_*}} \leq \varepsilon^{\frac{p_* - \eta}{p_*}(p-1)} \tilde{R}_\eta(t) = \varepsilon^{p-1-\eta'} \tilde{R}_\eta(t) \quad \text{for any } t \geq 0, \quad (\text{B.9})$$

where $\tilde{R}_\eta$ is a function of at most polynomial order as t tends to infinity and $\eta' = \eta(p-1)p_*^{-1}$. For η small enough we fix $\eta' \in (0, 1/4)$. Since $p_* > 1$, we may choose $p \in (1, p_*)$ and $\theta \in (0, 1/4)$. We split

$$\mathbb{E}[|\Delta_t^\varepsilon|^p] = \mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] + \mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}((\mathcal{A}_t^\varepsilon)^c)], \quad (\text{B.10})$$

where

$$\mathcal{A}_t^\varepsilon := \left\{ \sup_{0 \leq s \leq t} |\varepsilon \mathcal{Y}_s^x| \leq \varepsilon^{1-\theta} \right\}. \quad (\text{B.11})$$

First we prove that

$$(\mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}(\mathcal{A}_t^\varepsilon)])^{\frac{1}{p}} \leq \left(pC(|x|) \int_0^t \tilde{R}_\eta(s) ds \right)^{\frac{1}{p}} \varepsilon^{1 + \frac{1-\eta'-2\theta}{p}},$$

where $C(|x|) = \max_{|u| \leq |x|+1} |D^2 b(u)|$. The choice of η' and θ yields $1 - \eta' - 2\theta > 1/4$.

For notational convenience, we use the differential formalism, however, we stress that all differential inequalities are understood in the integral sense. Since $p > 1$, the chain rule, Hypothesis 1 and Cauchy-Schwarz inequality imply

$$\begin{aligned} d|\Delta_t^\varepsilon|^p &= -p|\Delta_t^\varepsilon|^{p-2} \langle \Delta_t^\varepsilon, b(X_t^\varepsilon(x)) - b(Y_t^\varepsilon(x)) \rangle dt \\ &\quad - p|\Delta_t^\varepsilon|^{p-2} \langle \Delta_t^\varepsilon, b(Y_t^\varepsilon(x)) - b(X_t^0(x)) - Db(X_t^0(x))\varepsilon \mathcal{Y}_t^x \rangle dt \\ &\leq -\delta p|\Delta_t^\varepsilon|^p + p|\Delta_t^\varepsilon|^{p-1} |b(Y_t^\varepsilon(x)) - b(X_t^0(x)) - Db(X_t^0(x))\varepsilon \mathcal{Y}_t^x| dt. \end{aligned}$$

On the event $\mathcal{A}_t^\varepsilon$, Taylor's theorem applied to b implies

$$d|\Delta_t^\varepsilon|^p \leq -\delta p|\Delta_t^\varepsilon|^p dt + pC(|x|)|\Delta_t^\varepsilon|^{p-1} \varepsilon^{2-2\theta} dt.$$

Taking expectation, the integral monotonicity, Fubini's theorem and (B.9) yield

$$d\mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}(\mathcal{A}_t^\varepsilon)]$$

$$\begin{aligned}
 &\leq -\delta p \mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] dt + pC(|x|) \mathbb{E}[|\Delta_t^\varepsilon|^{p-1} \mathbf{1}(\mathcal{A}_t^\varepsilon)] \varepsilon^{2-2\theta} dt \\
 &\leq pC(|x|) \mathbb{E}[|\Delta_t^\varepsilon|^{p-1}] \varepsilon^{2-2\theta} dt \\
 &\leq pC(|x|) \tilde{R}_\eta(t) \varepsilon^{p+1-\eta'-2\theta} dt.
 \end{aligned}$$

Bearing in mind $|\Delta_0^\varepsilon|^p = 0$, we have

$$\mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] \leq pC(|x|) \varepsilon^{p+1-\eta'-2\theta} \int_0^t \tilde{R}_\eta(s) ds.$$

Therefore

$$\left(\mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] \right)^{\frac{1}{p}} \leq \left(pC(|x|) \int_0^t \tilde{R}_\eta(s) ds \right)^{\frac{1}{p}} \varepsilon^{1+\frac{1-\eta'-2\theta}{p}}. \quad (\text{B.12})$$

We continue with the estimate on the complement of $\mathcal{A}_t^\varepsilon$. We show

$$\mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}((\mathcal{A}_t^\varepsilon)^c)] \leq \varepsilon^{p-\eta'} \mathcal{R}(t) \cdot \mathbb{P}((\mathcal{A}_t^\varepsilon)^c)^{\frac{p_*-p}{p_*}},$$

where $\mathcal{R}(t)$ is a function of at most polynomial order. Indeed, by Hölder's inequality and the inequalities (B.6), (B.7) and (B.8) we have

$$\begin{aligned}
 \mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}((\mathcal{A}_t^\varepsilon)^c)] &\leq \mathbb{E}[|\Delta_t^\varepsilon|^{p_*}]^{\frac{p}{p_*}} \cdot \mathbb{P}((\mathcal{A}_t^\varepsilon)^c)^{\frac{p_*-p}{p_*}} \\
 &\leq \left(C_{p_*} \varepsilon^{p_*-\eta} R_\eta(t) + C_{p_*} \varepsilon^{p_*} \tilde{C}_{p_*} S_{p_*}(t) \right)^{\frac{p}{p_*}} \cdot \mathbb{P}((\mathcal{A}_t^\varepsilon)^c)^{\frac{p_*-p}{p_*}} \\
 &\leq \left((C_{p_*} \varepsilon^{p_*-\eta} R_\eta(t))^{\frac{p}{p_*}} + (C_{p_*} \varepsilon^{p_*} \tilde{C}_{p_*} S_{p_*}(t))^{\frac{p}{p_*}} \right) \cdot \mathbb{P}((\mathcal{A}_t^\varepsilon)^c)^{\frac{p_*-p}{p_*}} \\
 &= \left((C_{p_*} R_\eta(t))^{\frac{p}{p_*}} \varepsilon^{(p_*-\eta)\frac{p}{p_*}} + (C_{p_*} \tilde{C}_{p_*} S_{p_*}(t))^{\frac{p}{p_*}} \varepsilon^p \right) \cdot \mathbb{P}((\mathcal{A}_t^\varepsilon)^c)^{\frac{p_*-p}{p_*}} \\
 &\leq \varepsilon^{p-\eta} \frac{p}{p_*} \mathcal{R}(t) \cdot \mathbb{P}((\mathcal{A}_t^\varepsilon)^c)^{\frac{p_*-p}{p_*}},
 \end{aligned}$$

where $\mathcal{R}(t) := \max\{(C_{p_*} R_\eta(t))^{\frac{p}{p_*}}, (C_{p_*} \tilde{C}_{p_*} S_{p_*}(t))^{\frac{p}{p_*}}\}$. As a consequence,

$$\left(\mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}((\mathcal{A}_t^\varepsilon)^c)] \right)^{1/p} \leq \varepsilon^{1-\frac{\eta}{p_*}} (\mathcal{R}(t))^{\frac{1}{p}} \cdot \mathbb{P}((\mathcal{A}_t^\varepsilon)^c)^{\frac{p_*-p}{p_* p}}. \quad (\text{B.13})$$

Combining estimates (B.12), (B.13) in decomposition (B.10) we obtain a positive constant $C := C(p_*, p, \delta, |x|, |D^2 F|)$ such that for any $t \geq 0$

$$\begin{aligned}
 \mathcal{W}_p(X_t^\varepsilon(x), Y_t^\varepsilon(x)) &\leq \left(\mathbb{E}[|\Delta_t^\varepsilon|^p] \right)^{\frac{1}{p}} \\
 &\leq \left(pC(|x|) \int_0^t \tilde{R}_\eta(s) ds \right)^{\frac{1}{p}} \varepsilon^{1+\frac{1-\eta'-2\theta}{p}} + \varepsilon^{1-\frac{\eta}{p_*}} (\mathcal{R}(t))^{\frac{1}{p}} \cdot \mathbb{P}((\mathcal{A}_t^\varepsilon)^c)^{\frac{p_*-p}{p_* p}}. \quad (\text{B.14})
 \end{aligned}$$

By Lemma B.1 there exists a positive constant C such that for all $\gamma \in (0, 1)$ for the choice $\vartheta = \varepsilon^{-\theta/\gamma}$ and any $t \geq 0$ it follows

$$\mathbb{P}((\mathcal{A}_t^\varepsilon)^c) \leq Ct \varepsilon^\theta. \quad (\text{B.15})$$

We further restrict θ such that additionally $0 < \theta < \min\{\frac{2\eta p}{p_*-p}, \frac{1}{4}\}$. Hence, with the help of inequality (B.14) and (B.15) we have

$$\mathcal{W}_p(X_t^\varepsilon(x), Y_t^\varepsilon(x)) \leq \mathcal{R}_1(t) \varepsilon^{1+\frac{1}{4p}} + \mathcal{R}_2(t) \varepsilon^{1+\frac{\eta}{p_*}},$$

where $\mathcal{R}_1$ and $\mathcal{R}_2$ are functions of at most polynomial order. Consequently we obtain the desired limit

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x+r \cdot w_\varepsilon}^\varepsilon(x), Y_{t_\varepsilon^x+r \cdot w_\varepsilon}^\varepsilon(x))}{\varepsilon} = 0.$$

We continue with the case $p_* > 0$ and $p \in (0, 1 \wedge p_*]$. Let $\theta \in (0, 1/4)$ and recall the event $\mathcal{A}_t^\varepsilon$ in (B.11). For $p \in (0, 1 \wedge p_*]$ we split

$$\mathbb{E}[|\Delta_t^\varepsilon|^p] = \mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] + \mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}((\mathcal{A}_t^\varepsilon)^c)] =: J_1 + J_2.$$

We start with the term J_1 . Since $|\cdot|^p$ is not differentiable, we apply the chain rule for the smooth approximation $|x|_c^p = (\sqrt{|x|^2 + c^2})^p$. Hypothesis 1 then yields

$$\begin{aligned} d|\Delta_t^\varepsilon|_c^p &= -p|\Delta_t^\varepsilon|_c^{p-2} \langle \Delta_t^\varepsilon, b(X_t^\varepsilon(x)) - b(Y_t^\varepsilon(x)) \rangle dt \\ &\quad + p|\Delta_t^\varepsilon|_c^{p-2} \langle \Delta_t^\varepsilon, b(Y_t^\varepsilon(x)) - b(X_t^0(x)) - Db(X_t^0(x))\varepsilon \mathcal{Y}_t^x \rangle dt \\ &\leq -p\delta|\Delta_t^\varepsilon|_c^{p-2} |\Delta_t^\varepsilon|^2 dt + p|\Delta_t^\varepsilon|_c^{p-1} |b(Y_t^\varepsilon(x)) - b(X_t^0(x)) - Db(X_t^0(x))\varepsilon \mathcal{Y}_t^x| dt \\ &\leq -p\delta|\Delta_t^\varepsilon|_c^p dt + p\delta c^p dt + pc^{p-1} |b(Y_t^\varepsilon(x)) - b(X_t^0(x)) - Db(X_t^0(x))\varepsilon \mathcal{Y}_t^x| dt. \end{aligned}$$

Due to $|X_t^0(x)| \leq e^{-\delta t}|x|$ for all $t \geq 0$ and $x \in \mathbb{R}^d$, Taylor's expansion for b on the event $\mathcal{A}_t^\varepsilon$ implies

$$d|\Delta_t^\varepsilon|_c^p \leq -p\delta|\Delta_t^\varepsilon|_c^p dt + p\delta c^p dt + pc^{p-1} C(|x|)\varepsilon^{2(1-\theta)},$$

where $C(|x|) = \max_{|u| \leq |x|+1} |D^2 b(u)|$. Hence

$$d\mathbb{E}[|\Delta_t^\varepsilon|_c^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] \leq -p\delta\mathbb{E}[|\Delta_t^\varepsilon|_c^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] dt + p\delta c^p dt + pc^{p-1} C(|x|)\varepsilon^{2(1-\theta)} dt.$$

The integral version of the Grönwall inequality with negative linearity given in Lemma 1 in [11] implies for all $t \geq 0$

$$\mathbb{E}[|\Delta_t^\varepsilon|_c^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] \leq \mathbb{E}[|\Delta_t^\varepsilon|_c^p \mathbf{1}(\mathcal{A}_t^\varepsilon)] \leq c^p + \frac{1}{\delta} c^{p-1} C(|x|)\varepsilon^{2(1-\theta)}. \quad (\text{B.16})$$

For $p \neq 1$ we have the following. Since $c > 0$ is arbitrary and $\theta \in (0, 1/4)$, the choice $c = \varepsilon^{1+\eta/p}$ with $\eta \in (0, \frac{p}{2(1-p)})$ in (B.16) yields for any $r \in \mathbb{R}$

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^p} \mathbb{E}[|\Delta_{t_\varepsilon^x+r \cdot w_\varepsilon}^\varepsilon|^p \mathbf{1}(\mathcal{A}_{t_\varepsilon^x+r \cdot w_\varepsilon}^\varepsilon)] = 0. \quad (\text{B.17})$$

The case of $p = 1$ follows by the choice $c = \varepsilon^2$ in (B.16).

We continue with the term J_2 . By the subadditivity of the power $p \leq 1$ and the Hölder inequality for the index p'/p where $p' \in (p, p_*)$ and r is such that $p/p' + 1/r = 1$ we have

$$\begin{aligned} \mathbb{E}[|\Delta_t^\varepsilon|^p \mathbf{1}((\mathcal{A}_t^\varepsilon)^c)] &\leq \mathbb{E}[|X_t^\varepsilon(x)|^p \mathbf{1}((\mathcal{A}_t^\varepsilon)^c)] + \mathbb{E}[|Y_t^\varepsilon(x)|^p \mathbf{1}((\mathcal{A}_t^\varepsilon)^c)] \\ &\leq (\mathbb{E}[|X_t^\varepsilon(x)|^{p'}])^{p/p'} (\mathbb{P}((\mathcal{A}_t^\varepsilon)^c))^{1/r} + (\mathbb{E}[|Y_t^\varepsilon(x)|^{p'}])^{p/p'} (\mathbb{P}((\mathcal{A}_t^\varepsilon)^c))^{1/r}. \end{aligned} \quad (\text{B.18})$$

By Lemma B.5 we obtain for all $t \geq 0$

$$\begin{aligned} (\mathbb{E}[|X_t^\varepsilon(x)|^{p'}])^{p/p'} &\leq (\mathbb{E}[|X_t^\varepsilon(x) - X_t^0(x)|^{p'}] + |X_t^0(x)|^{p'})^{p/p'} \\ &\leq \varepsilon^p (1 + C_{p'} \cdot t)^{p/p'} + |X_t^0(x)|^p. \end{aligned} \quad (\text{B.19})$$

Note that for all $t \geq 0$ it follows

$$(\mathbb{E}[|Y_t^\varepsilon(x)|^{p'}])^{p/p'} \leq \varepsilon^p (\mathbb{E}[|\mathcal{Y}_t^x|^{p'}])^{p/p'} + |X_t^0(x)|^p. \quad (\text{B.20})$$

Lemma A.1 in [3] yields the existence of a positive constant $C(r, |x|)$ such that

$$|X_{t_\varepsilon^x + r \cdot w_\varepsilon}^0(x)| \leq C(r, |x|)\varepsilon. \quad (\text{B.21})$$

Combining (B.18) with inequalities (B.15), (B.19), (B.20) and (B.21) gives

$$\begin{aligned} \mathbb{E}[|\Delta_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon|^{p'} \mathbf{1}((\mathcal{A}_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon)^c)] &\leq (C(t_\varepsilon^x + r \cdot w_\varepsilon)\varepsilon^\theta)^{1/r} C^p(r, |x|)\varepsilon^p \\ &\quad + (C(t_\varepsilon^x + r \cdot w_\varepsilon)\varepsilon^\theta)^{1/r} (\varepsilon^p (1 + C_{p'} \cdot (t_\varepsilon^x + r \cdot w_\varepsilon))^{p/p'}) \\ &\quad + (C(t_\varepsilon^x + r \cdot w_\varepsilon)\varepsilon^\theta)^{1/r} (\varepsilon^p (\mathbb{E}[|\mathcal{Y}_{t_\varepsilon^x + r \cdot w_\varepsilon}^x|^{p'}])^{p/p'}). \end{aligned}$$

Since $\mathbb{E}[|\mathcal{Y}_t^x|^{p'}] \leq \mathcal{R}(t)$, where $\mathcal{R}$ is a function of at most polynomial order, we have

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^{p'}} \mathbb{E}[|\Delta_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon|^{p'} \mathbf{1}((\mathcal{A}_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon)^c)] \\ \leq \limsup_{\varepsilon \rightarrow 0} \varepsilon^{\theta/r} (C(t_\varepsilon^x + r \cdot w_\varepsilon))^{1/r} (C^p(r, |x|) \\ + (1 + C_{p'} \cdot (t_\varepsilon^x + r \cdot w_\varepsilon))^{p/p'} + \mathcal{R}(t_\varepsilon^x + r \cdot w_\varepsilon)). \end{aligned}$$

The right-hand side of the preceding inequality equals zero. The preceding argument combined with (B.17) yields the desired limit (B.5). $\square$

B.3 Asymptotic First Order Approximation

Lemma B.2 For any $p \in (0, p_*)$ we have

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(\mu_*^\varepsilon, \mu^\varepsilon)}{\varepsilon^{1 \wedge p}} = 0.$$

Proof First we observe that $Y_t^\varepsilon(0) = \mathcal{Z}_t^\varepsilon(0)$ for any $t \geq 0$, $\varepsilon > 0$, where $(\mathcal{Z}_t^\varepsilon(0))_{t \geq 0}$ is given in (B.2). In abuse of notation, we write $(X_t^\varepsilon(\mu^\varepsilon))_{t \geq 0}$ (and analogously respectively $(\mathcal{Z}_t^\varepsilon(\mu_*^\varepsilon))_{t \geq 0}$) for the process starting at the random vector with distribution μ^ε independent of the noise process L . Since $X_t^\varepsilon(\mu^\varepsilon) = \mu^\varepsilon$ and $\mathcal{Z}_t^\varepsilon(\mu_*^\varepsilon) = \mu_*^\varepsilon$ for any $t \geq 0$, the triangle inequality yields

$$\begin{aligned} \mathcal{W}_p(\mu^\varepsilon, \mu_*^\varepsilon) &= \mathcal{W}_p(X_t^\varepsilon(\mu^\varepsilon), \mathcal{Z}_t^\varepsilon(\mu_*^\varepsilon)) \leq \mathcal{W}_p(X_t^\varepsilon(\mu^\varepsilon), X_t^\varepsilon(0)) \\ &\quad + \mathcal{W}_p(X_t^\varepsilon(0), \mathcal{Z}_t^\varepsilon(0)) + \mathcal{W}_p(\mathcal{Z}_t^\varepsilon(0), \mathcal{Z}_t^\varepsilon(\mu_*^\varepsilon)). \end{aligned} \quad (\text{B.22})$$

By Proposition 2 for $x = 0$, we have

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(0), \mathcal{Z}_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(0))}{\varepsilon^{1 \wedge p}} = 0. \quad (\text{B.23})$$

By disintegration, inequalities (A.4) and (2.84) in [3] imply

$$\begin{aligned} \mathcal{W}_p(X_t^\varepsilon(\mu^\varepsilon), X_t^\varepsilon(0)) &\leq \int_{\mathbb{R}^d} \mathcal{W}_p(X_t^\varepsilon(u), X_t^\varepsilon(0)) \mu^\varepsilon(du) \\ &\leq e^{-\delta(1 \wedge p)t} \int_{\mathbb{R}^d} |u|^{1 \wedge p} \mu^\varepsilon(du) \leq C e^{-\delta(1 \wedge p)t} \varepsilon^{1 \wedge p} \end{aligned}$$

for some positive constant C . As a consequence,

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(\mu^\varepsilon), X_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(0))}{\varepsilon^{1 \wedge p}} = 0. \quad (\text{B.24})$$

Analogously,

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{W}_p(\mathcal{Z}_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(\mu^\varepsilon), \mathcal{Z}_{t_\varepsilon^x + r \cdot w_\varepsilon}^\varepsilon(0))}{\varepsilon^{1 \wedge p}} = 0. \quad (\text{B.25})$$

Combining (B.22) with the estimates (B.23), (B.24) and (B.25) completes the proof. $\square$

Lemma B.3 *For any $p \in (0, p_*)$ we have*

$$\lim_{t \rightarrow \infty} \mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{O}_\infty) = 0. \quad (\text{B.26})$$

Proof Recall that $\mathcal{O}_\infty$ is the limiting and invariant distribution of the homogeneous Ornstein-Uhlenbeck process $(\mathcal{Z}(x)_t)_{t \geq 0}$ defined in (B.2). That is $\mathcal{O}_\infty \stackrel{d}{=} \mathcal{Z}_\infty$. Since $-Db(X_t^0(x))$ converges exponentially fast to $-Db(0)$, it is natural to expect that the flow of $(\mathcal{Y}_t^x)_{t \geq 0}$ behaves as the flow of $(\mathcal{Z}_t(x))_{t \geq 0}$ for large t . In [3], Lemma C.3, it is shown that $\mathcal{Y}_t^x \rightarrow \mathcal{O}_\infty$ as $t \rightarrow \infty$ in law. However, the law $\mathcal{O}_\infty$ is not invariant under the random dynamics of $(\mathcal{Y}_t^x)_{t \geq 0}$ due to the time inhomogeneity. Analogously as in (A.5) we deduce

$$\mathcal{W}_p(\mathcal{Z}_t(x), \mathcal{O}_\infty) \rightarrow 0, \quad \text{as } t \rightarrow \infty. \quad (\text{B.27})$$

We start with the proof of the statement. The triangle inequality yields

$$\mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{O}_\infty) \leq \mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{Z}_t(0)) + \mathcal{W}_p(\mathcal{Z}_t(0), \mathcal{O}_\infty), \quad (\text{B.28})$$

where the second term on the right-hand side tends to 0 as $t \rightarrow \infty$ due to (B.27). Thus it remains to prove $\mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{Z}_t(0)) \rightarrow 0$, as $t \rightarrow \infty$. Since

$$\mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{Z}_t(0)) \leq (\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p])^{1 \wedge (1/p)},$$

we derive the respective L^p estimates. By (4.2) and (B.2) we obtain

$$d(|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|) = -Db(X_t^0(x))(\mathcal{Y}_t^x - \mathcal{Z}_t(0))dt + (Db(0) - Db(X_t^0(x)))\mathcal{Z}_t(0)dt.$$

We first consider the case $p_* > 1$ and $p \in (1, p_*)$. The chain rule and Hypothesis 1 yield

$$\begin{aligned} d|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p &= -p|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^{p-2} \langle \mathcal{Y}_t^x - \mathcal{Z}_t(0), Db(X_t^0(x))(\mathcal{Y}_t^x - \mathcal{Z}_t(0)) \rangle dt \\ &\quad + p|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^{p-2} \langle \mathcal{Y}_t^x - \mathcal{Z}_t(0), (Db(0) - Db(X_t^0(x)))\mathcal{Z}_t(0) \rangle dt \\ &\leq -p\delta|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p dt + p|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^{p-1} |Db(0) \\ &\quad - Db(X_t^0(x))||\mathcal{Z}_t(0)|| dt \\ &\leq -p\delta|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p dt + p|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^{p-1} C(|x|)|X_t^0(x)||\mathcal{Z}_t(0)|| dt, \end{aligned}$$

where $C(|x|) = \max_{|u| \leq |x|+1} |D^2b(u)|$. Taking expectation, using the monotonicity of the integrals and Fubini's theorem imply

$$\begin{aligned} d\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p] &\leq -p\delta\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p]dt \\ &\quad + pC(|x|)|X_t^0(x)|\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^{p-1} \cdot |\mathcal{Z}_t(0)|]dt. \end{aligned}$$

By Young's inequality and $|X_t^0(x)| \leq e^{-\delta t}|x|$ for any $t \geq 0$ and $x \in \mathbb{R}^d$ it follows

$$d\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p] \leq -p\delta\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p]dt$$

$$+ pC(|x|)|x|e^{-\delta t} (\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p]dt + \mathbb{E}[|\mathcal{Z}_t(0)|^p]) dt.$$

A straightforward calculation yields (for any $p > 0$) that there exist functions $P_1(t)$ and $P_2(t)$ of polynomial order (depending of $p, \delta, |x|$) such that

$$\mathbb{E}[|\mathcal{Z}_t(0)|^p] \leq P_1(t) \quad \text{and} \quad \mathbb{E}[|\mathcal{Y}_t^x|^p] \leq P_2(t) \quad \text{for any } t \geq 0. \quad (\text{B.29})$$

Therefore,

$$d \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p] \leq -p\delta \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p]dt + p2^p C(|x|)|x|e^{-\delta t} (P_1(t) + P_2(t))dt.$$

The integral version of the Grönwall inequality with negative linearity given in Lemma 1 in [11] yields

$$\begin{aligned} \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p] &\leq p2^p C(|x|)|x|e^{-p\delta t} \int_0^t e^{p\delta s} e^{-\delta s} (P_1(s) + P_2(s))ds \\ &\leq \frac{p2^p C(|x|)|x|}{\delta(p-1)} \max_{0 \leq s \leq t} \{P_1(s), P_2(s)\} e^{-\delta t}. \end{aligned}$$

Therefore,

$$\lim_{t \rightarrow \infty} \mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{Z}_t(0)) \leq \lim_{t \rightarrow \infty} \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p] = 0. \quad (\text{B.30})$$

Combining (B.27) and (B.30) in (B.28) we conclude (B.26).

We continue with the case $p \in (0, p_* \wedge 1]$. Note that the case $p_* > 1$ and $p \in (0, 1]$ is also covered in the sequel. By Lemma B.1 there exists a positive constant C such that for the choice $\gamma = p, \vartheta = e^{\frac{\delta}{2}t}$ and any $t \geq 0$ it follows

$$\mathbb{P}(\mathcal{D}_t^0) \leq Cte^{-\frac{\delta p}{2}t}, \quad \text{where we recall} \quad \mathcal{D}_t^0 = \left\{ \sup_{0 \leq s \leq t} |\mathcal{Z}_s(0)| > \vartheta \right\}. \quad (\text{B.31})$$

We split

$$\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p] = \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p \mathbf{1}((\mathcal{D}_t^0)^c)] + \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p \mathbf{1}(\mathcal{D}_t^0)] =: I_1 + I_2.$$

We start with the term I_1 . The chain rule for $|x|_c^p = (\sqrt{|x|^2 + c^2})^p$ and Hypothesis 1 yield

$$\begin{aligned} d|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^p &= -p|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^{p-2} \langle \mathcal{Y}_t^x - \mathcal{Z}_t(0), Db(X_t^0(x))(\mathcal{Y}_t^x - \mathcal{Z}_t(0)) \rangle dt \\ &\quad + p|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^{p-2} \langle \mathcal{Y}_t^x - \mathcal{Z}_t(0), (Db(0) - Db(X_t^0(x)))\mathcal{Z}_t(0) \rangle dt \\ &\leq -p\delta|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^{p-2} |\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^2 dt \\ &\quad + p|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^{p-1} C(|x|)|X_t^0(x)||\mathcal{Z}_t(0)|dt \\ &= -p\delta|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^p dt + p\delta c^p dt + pc^{p-1} C(|x|)|X_t^0(x)||\mathcal{Z}_t(0)|dt, \end{aligned}$$

where $C(|x|) = \max_{|u| \leq |x|+1} |D^2b(u)|$. On the event $(\mathcal{D}_t^0)^c$ we have

$$d|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^p \leq -p\delta|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^p dt + p\delta c^p dt + pc^{p-1} C(|x|)|x|e^{-(\delta/2)t} dt$$

due to $|X_t^0(x)| \leq e^{-\delta t}|x|$ for all $t \geq 0$ and $x \in \mathbb{R}^d$. Hence

$$\begin{aligned} d \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^p \mathbf{1}((\mathcal{D}_t^0)^c)] &\leq -p\delta \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^p \mathbf{1}((\mathcal{D}_t^0)^c)]dt \\ &\quad + p\delta c^p dt + pc^{p-1} C(|x|)|x|e^{-(\delta/2)t} dt. \end{aligned}$$

The Grönwall inequality in [11] implies

$$\begin{aligned} \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p \mathbf{1}((\mathcal{D}_t^0)^c)] &\leq \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|_c^p \mathbf{1}((\mathcal{D}_t^0)^c)] \\ &\leq c^p + pc^{p-1} C(|x|)|x|e^{-p\delta t} \int_0^t e^{p\delta s} e^{-(\delta/2)s} ds. \end{aligned}$$

Then

$$\limsup_{t \rightarrow \infty} \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p \mathbf{1}((\mathcal{D}_t^0)^c)] \leq c^p \quad \text{for all } c > 0,$$

which yields $\lim_{t \rightarrow \infty} \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p \mathbf{1}((\mathcal{D}_t^0)^c)] = 0$.

We continue with the term I_2 . By the Hölder inequality for the index p'/p where $p' = (p + p_*)/2$ and r the conjugate index of p'/p we have

$$\begin{aligned} \mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p \mathbf{1}(D_t)] &\leq \mathbb{E}[|\mathcal{Y}_t^x|^p \mathbf{1}(D_t)] + \mathbb{E}[|\mathcal{Z}_t(0)|^p \mathbf{1}(D_t)] \\ &\leq (\mathbb{E}[|\mathcal{Y}_t^x|^{p'}])^{p/p'} (\mathbb{P}(D_t))^{1/r} + (\mathbb{E}[|\mathcal{Z}_t(0)|^{p'}])^{p/p'} (\mathbb{P}(D_t))^{1/r}. \end{aligned} \quad (\text{B.32})$$

By (B.29) and (B.31) the right-hand side of (B.32) tends to zero as $t \rightarrow \infty$. As a consequence we have $\mathcal{W}_p(\mathcal{Y}_t^x, \mathcal{Z}_t(0)) \leq (\mathbb{E}[|\mathcal{Y}_t^x - \mathcal{Z}_t(0)|^p])^{1 \wedge (1/p)}$ which tends to zero as $t \rightarrow \infty$. By (B.27) and (B.28) we obtain (B.26). $\square$

B.4 Auxiliary Moment Estimates

Lemma B.4 *For any $2 \leq p < p_*$ (and $p = 2$ if $p_* = 2$) there is a function of at most polynomial order $R(t)$ as $t \rightarrow \infty$ and $\varepsilon_0 \in (0, 1]$ such that for any $t \geq 0$ and $0 < \varepsilon < \varepsilon_0$ we have*

$$\mathbb{E}[|X_t^\varepsilon(x) - X_t^0(x)|^{p_*}] \leq \varepsilon^p R(t).$$

Proof First note that for $G(u) = |u|^{p_*}$, $p_* \geq 2$ we have

$$\begin{aligned} \nabla G(u) &= p_* |u|^{p_*-2} u = p_* (|u|^2)^{\frac{p_*-2}{2}} u, \quad \text{with } \partial_i G(u) = p_* |u|^{p_*-2} u_i, \\ \sum_{ij} \partial_i \partial_j G(u) &\leq p_* |u|^{p_*-4} (d|u|^2 + \sum_{ij} \frac{(p_*-2)}{2} (u_j^2 + u_i^2)) = p_* (p_* - 1) d |u|^{p_*-4} |u|^2. \end{aligned}$$

Recall the notation (B.1) for L . The Itô formula for $\Theta_t^\varepsilon = X_t^\varepsilon(x) - X_t^0(x)$ yields

$$\begin{aligned} d|\Theta_t^\varepsilon|^{p_*} &= -p_* |\Theta_t^\varepsilon|^{p_*-2} \langle \Theta_t^\varepsilon, b(X_t^\varepsilon(x)) - b(X_t^0(x)) \rangle dt + p_* |\Theta_t^\varepsilon|^{p_*-2} \langle \Theta_t^\varepsilon, \varepsilon \Sigma^{1/2} dB_t \rangle \\ &\quad + \frac{\varepsilon^2}{2} \text{trace}(\Sigma^{1/2} \text{Hess} G(\Theta_t^\varepsilon) (\Sigma^{1/2})^*) dt \\ &\quad + \int_{\mathbb{R}^d} (|\Theta_t^\varepsilon + \varepsilon z|^{p_*} - |\Theta_t^\varepsilon|^{p_*} - p_* |\Theta_t^\varepsilon|^{p_*-2} \langle \Theta_t^\varepsilon, \varepsilon z \rangle \mathbf{1}\{|z| \leq 1\}) \nu(dz) dt \\ &\quad + \int_{\mathbb{R}^d} (|\Theta_t^\varepsilon + \varepsilon z|^{p_*} - |\Theta_t^\varepsilon|^{p_*}) \tilde{N}(dt, dz). \end{aligned}$$

Taking expectation yields

$$\begin{aligned} \mathbb{E}[|\Theta_t^\varepsilon|^{p_*}] &\leq -\delta p_* \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}] ds + \varepsilon^2 p_* (p_* - 1) d \text{trace}(\Sigma^{1/2} (\Sigma^{1/2})^*) \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*-2}] ds \end{aligned}$$

$$+ \int_0^t \int_{\mathbb{R}^d} \mathbb{E} \left[|\Theta_t^\varepsilon + \varepsilon z|^{p_*} - |\Theta_t^\varepsilon|^{p_*} - p_* |\Theta_t^\varepsilon|^{p_*-2} \langle \Theta_t^\varepsilon, \varepsilon z \rangle \right] \nu(dz) ds.$$

By the mean value theorem we have

$$\begin{aligned} & \mathbb{E} \left[|\Theta_t^\varepsilon + \varepsilon z|^{p_*} - |\Theta_t^\varepsilon|^{p_*} - p_* |\Theta_t^\varepsilon|^{p_*-2} \langle \Theta_t^\varepsilon, \varepsilon z \rangle \right] \\ & \leq \mathbb{E} \left[p_*(p_* - 1) d \int_0^1 |\Theta_t^\varepsilon + \theta \vartheta \varepsilon z|^{p_*-2} d\theta d\vartheta \right] |\varepsilon z|^2 \\ & \leq (1 \vee 2^{p_*-2}) \mathbb{E} \left[p_*(p_* - 1) d (|\Theta_t^\varepsilon|^{p_*-2} + |\varepsilon z|^{p_*-2}) \right] |\varepsilon z|^2 \\ & \leq (1 \vee 2^{p_*-2}) p_*(p_* - 1) d \mathbb{E} \left[|\Theta_t^\varepsilon|^{p_*-2} \right] (|\varepsilon z|^2 + |\varepsilon z|^{p_*}) \end{aligned}$$

and

$$\begin{aligned} & \int_{\mathbb{R}^d} \mathbb{E} \left[|\Theta_t^\varepsilon + \varepsilon z|^{p_*} - |\Theta_t^\varepsilon|^{p_*} - p |\Theta_t^\varepsilon|^{p_*-2} \langle \Theta_t^\varepsilon, \varepsilon z \rangle \right] \nu(dz) \\ & \leq C_{p_*, d} \left(\int_{\mathbb{R}^d} |z|^2 \nu(dz) + \int_{\mathbb{R}^d} |z|^{p_*} \nu(dz) \right) \varepsilon^2 \mathbb{E} \left[|\Theta_t^\varepsilon|^{p_*-2} \right]. \end{aligned}$$

Hence there is a positive constant K such that

$$\mathbb{E}[|\Theta_t^\varepsilon|^{p_*}] \leq -\delta p_* \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}] ds + \varepsilon^2 K \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*-2}] ds. \quad (\text{B.33})$$

For $p_* = 2$ we have directly $\mathbb{E}[|\Theta_t^\varepsilon|^{p_*}] \leq \varepsilon^2 K t$. For $p_* > 2$ we continue in (B.33) with Young's inequality

$$\begin{aligned} \mathbb{E}[|\Theta_t^\varepsilon|^{p_*}] & \leq -\delta p_* \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}] ds + \varepsilon^2 K \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*-2}] ds \\ & \leq -\delta p_* \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}] ds + \varepsilon^2 K \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}] ds + \varepsilon^2 K t \\ & \leq -\frac{\delta p_*}{2} \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}] ds + \varepsilon^2 K t \end{aligned}$$

for $\varepsilon < (\frac{\delta p_*}{2K})^{1/2}$. Grönwall's lemma applied to the preceding estimate yields the a priori estimate $\mathbb{E}[|\Theta_t^\varepsilon|^{p_*}] \leq \varepsilon^2 K t^2 =: \varepsilon^2 R_0(t)$. Inserting the a priori estimate in (B.33) and using the Hölder inequality for $p_* > 2$ we obtain

$$\begin{aligned} \mathbb{E}[|\Theta_t^\varepsilon|^{p_*}] & \leq -\delta p_* \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}] ds + \varepsilon^2 K \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*-2}] ds \\ & \leq -\delta p_* \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}] ds + \varepsilon^2 K \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|^{p_*}]^{\frac{p_*-2}{p_*}} ds \\ & \leq \varepsilon^{2+2\frac{p_*-2}{p_*}} K^{1+\frac{p_*-2}{p_*}} \int_0^t s^{2\frac{p_*-2}{p_*}} ds =: \varepsilon^{2+2\frac{p_*-2}{p_*}} R_1(t). \end{aligned}$$

By induction we deduce after the i -th iterations of the bootstrap the estimate

$$\mathbb{E}[|\Theta_t^\varepsilon|^{p_*}] \leq \varepsilon^{2\sum_{j=0}^i (\frac{p_*-2}{p_*})^j} R_i(t)$$

for a polynomial order function $R_i(t)$. Clearly, $\lim_{i \rightarrow \infty} 2\sum_{j=0}^i (\frac{p_*-2}{p_*})^j = p_*$ and therefore for any $0 < p < p_*$ there is an iteration $i_0 = i_0(p_*, p)$ such that we obtain $\mathbb{E}[|\Theta_t^\varepsilon|^{p_*}] \leq \varepsilon^p R_{i_0}(t)$. This finishes the proof of the lemma. $\square$

Lemma B.5 *Let $p_* > 0$. Then for any $p \in (0, 2 \wedge p_*)$ there exists a positive constant C_p such that for any $t \geq 0$ and $\varepsilon > 0$ we have*

$$\mathbb{E}[|X_t^\varepsilon(x) - X_t^0(x)|^p] \leq \varepsilon^p (1 + C_p \cdot t).$$

Proof Without loss of generality let $p_* \in (0, 2]$. Itô's formula yields for $\Theta_t^\varepsilon = X_t^\varepsilon(x) - X_t^0(x)$ and the function $G(z) = |z|_c^p$

$$\begin{aligned} d|\Theta_t^\varepsilon|_c^p &= -p|\Theta_t^\varepsilon|_c^{p-2} \langle \Theta_t^\varepsilon, b(X_t^\varepsilon(x)) - b(X_t^0(x)) \rangle dt + p|\Theta_t^\varepsilon|_c^{p-2} \langle \Theta_t^\varepsilon, \varepsilon \Sigma^{1/2} dB_t \rangle \\ &\quad + \frac{\varepsilon^2}{2} \text{trace}(\Sigma^{1/2} \text{Hess} G(\Theta_t^\varepsilon) (\Sigma^{1/2})^*) dt \\ &\quad + \int_{\mathbb{R}^d} (|\Theta_t^\varepsilon + \varepsilon z|_c^p - |\Theta_t^\varepsilon|_c^p - p|\Theta_t^\varepsilon|_c^{p-2} \langle \Theta_t^\varepsilon, \varepsilon z \rangle \mathbf{1}_{\{|z| \leq 1\}}) \nu(dz) dt \\ &\quad + \int_{\mathbb{R}^d} (|\Theta_t^\varepsilon + \varepsilon z|_c^p - |\Theta_t^\varepsilon|_c^p) \tilde{N}(dt, dz). \end{aligned}$$

Taking expectation and using Hypothesis 1 we have

$$\begin{aligned} \mathbb{E}[|\Theta_t^\varepsilon|_c^p] &\leq c^p - p\delta \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|_c^{p-2} |\Theta_s^\varepsilon|^2] ds + \varepsilon^2 \int_0^t \text{trace}(\Sigma^{1/2} \text{Hess} G(\Theta_s^\varepsilon) (\Sigma^{1/2})^*) ds \\ &\quad + \int_0^t \int_{\mathbb{R}^d} \mathbb{E}[|\Theta_s^\varepsilon + \varepsilon z|_c^p - |\Theta_s^\varepsilon|_c^p - p|\Theta_s^\varepsilon|_c^{p-2} \langle \Theta_s^\varepsilon, \varepsilon z \rangle \mathbf{1}_{\{|z| \leq 1\}}] \nu(dz) ds. \end{aligned}$$

Since $|x|^2 = |x|_c^2 - c^2$, we obtain

$$\begin{aligned} \mathbb{E}[|\Theta_t^\varepsilon|_c^p] &\leq c^p - p\delta \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|_c^p] ds + p\delta c^p t + \varepsilon^2 c^{p-2} t C(p, d) \text{trace}(\Sigma^{1/2} (\Sigma^{1/2})^*) \\ &\quad + \int_0^t \int_{\mathbb{R}^d} \mathbb{E}[|\Theta_s^\varepsilon + \varepsilon z|_c^p - |\Theta_s^\varepsilon|_c^p - p|\Theta_s^\varepsilon|_c^{p-2} \langle \Theta_s^\varepsilon, \varepsilon z \rangle \mathbf{1}_{\{|z| \leq 1\}}] \nu(dz) ds. \end{aligned} \quad (\text{B.34})$$

In the sequel we estimate the second order term for small increments with the help of (B.4) by

$$\begin{aligned} \int_0^t \int_{|z| \leq 1} \mathbb{E}[|\Theta_s^\varepsilon + \varepsilon z|_c^p - |\Theta_s^\varepsilon|_c^p - p|\Theta_s^\varepsilon|_c^{p-2} \langle \Theta_s^\varepsilon, \varepsilon z \rangle] \nu(dz) ds \\ \leq C(p, d) \varepsilon^2 c^{p-2} t \int_{|z| \leq 1} |z|^2 \nu(dz) =: K_1 \varepsilon^2 c^{p-2} t. \end{aligned} \quad (\text{B.35})$$

For the large increments, we use the mean value theorem and obtain

$$\int_0^t \int_{|z| > 1} \mathbb{E}[|\Theta_s^\varepsilon + \varepsilon z|_c^p - |\Theta_s^\varepsilon|_c^p] \nu(dz) ds = p\varepsilon \int_0^t \int_{|z| > 1} \int_0^1 \mathbb{E}[|\Theta_s^\varepsilon + \theta \varepsilon z|_c^{p-1}] |z| d\theta \nu(dz) ds.$$

For $p \in (0, 1]$, note that $|x + y|_c^p \leq |x|_c^p + |y|_c^p + c^p$ for all $x, y \in \mathbb{R}^d$. Then we have for all $t \geq 0$

$$\begin{aligned} \int_0^t \int_{|z| > 1} \mathbb{E}[|\Theta_s^\varepsilon + \varepsilon z|_c^p - |\Theta_s^\varepsilon|_c^p] \nu(dz) ds &\leq \int_0^t \int_{|z| > 1} (\varepsilon^p |z|^p + c^p) \nu(dz) ds \\ &= t \varepsilon^p \int_{|z| > 1} |z|^p \nu(dz) + t c^p \nu(\{|z| > 1\}). \end{aligned} \quad (\text{B.36})$$

For $p > 1$, due to $|x + y|_c^{p-1} \leq |x|^{p-1} + |y|^{p-1} + c^{p-1}$ for all $x, y \in \mathbb{R}^d$, we split the intermediate value as follows

$$\begin{aligned}
 & p\varepsilon \int_0^t \int_{|z|>1} \int_0^1 \mathbb{E}[|\Theta_s^\varepsilon + \theta \varepsilon z|_c^{p-1}] |z| d\theta v(dz) ds \\
 & \leq p\varepsilon \int_0^t \int_{|z|>1} \mathbb{E}[|\Theta_s^\varepsilon|^{p-1}] |z| dv(dz) ds \\
 & \quad + p\varepsilon \int_0^t \int_{|z|>1} |\varepsilon z|^{p-1} |z| v(dz) ds + p\varepsilon c^{p-1} t v(\{|z| > 1\}) \\
 & = p\mathbb{E} \left[\int_0^t \varepsilon |\Theta_s^\varepsilon|^{p-1} ds \right] \int_{|z|>1} |z| dv(dz) + p\varepsilon^p t \int_{|z|>1} |z|^p v(dz) + p\varepsilon c^{p-1} t v(\{|z| > 1\}) \\
 & \leq t p (1/K_3)^p \varepsilon^p + \frac{p\delta}{2} \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|_c^p] ds + p\varepsilon^p t \int_{|z|>1} |z|^p v(dz) + p\varepsilon c^{p-1} t v(\{|z| > 1\}),
 \end{aligned} \tag{B.37}$$

where we have used in the last line the following weighted Young inequality

$$\begin{aligned}
 \int_0^t K_2 \varepsilon |\Theta_s^\varepsilon|^{p-1} ds & \leq (1/K_3)^p t K_2^p \varepsilon^p + K_3^{p/(p-1)} \int_0^t |\Theta_s^\varepsilon|^p ds \\
 & \leq t (1/K_3)^p \varepsilon^p + \frac{\delta}{2} \int_0^t |\Theta_s^\varepsilon|_c^p ds
 \end{aligned}$$

with $K_2 = \int_{|z|>1} |z| dv(dz) + 1$ and $K_3 = (\delta/2)^{p/(p-1)}$ followed by $|y| \leq |y|_c$. Combining (B.35) with (B.37) for $p \geq 1$, and (B.36) with (B.37) for $p < 1$, respectively, in (B.34) we obtain

$$\begin{aligned}
 \mathbb{E}[|\Theta_t^\varepsilon|_c^p] & \leq c^p - \frac{p\delta}{2} \int_0^t \mathbb{E}[|\Theta_s^\varepsilon|_c^p] ds + p\delta c^p t + K_0 \varepsilon^2 c^{p-2} t + K_1 \varepsilon^2 c^{p-2} t \\
 & \quad + t p (1/K_3)^p \varepsilon^p \cdot \mathbf{1}_{\{p \geq 1\}} + p\varepsilon^p t \int_{|z|>1} |z|^p v(dz) + p\varepsilon c^{p-1} t v(\{|z| > 1\}),
 \end{aligned}$$

where $K_0 = C(p, d) \text{trace}(\Sigma^{1/2}(\Sigma^{1/2})^*)$. Since $|x|^p \leq |x|_c^p$, the choice $c = c_\varepsilon = \varepsilon$ yields for all $t \geq 0$ $\mathbb{E}[|\Theta_t^\varepsilon|^p] \leq \varepsilon^p (1 + Ct)$ for some constant $C = C(p, \delta)$. This completes the proof of the lemma. $\square$

References

1. Applebaum, D.: Lévy Processes and Stochastic Calculus, 2nd edn. Cambridge University Press, Cambridge (2009)
2. Barrera, G., Högele, M.A., Pardo, J.C.: Cutoff thermalization for Ornstein–Uhlenbeck systems with small Lévy noise in the Wasserstein distance. J. Stat. Phys. **184**(27) (2021)
3. Barrera, G., Högele, M.A., Pardo, J.C.: The cutoff phenomenon in total variation for nonlinear Langevin systems with small layered stable noise. Electron. J. Probab. **26**, 1–76 (2021)
4. Barrera, G., Jara, M.: Abrupt convergence of stochastic small perturbations of one dimensional dynamical systems. J. Stat. Phys. **163**(1), 113–138 (2016)
5. Barrera, G., Jara, M.: Thermalisation for small random perturbation of dynamical systems. Ann. Appl. Probab. **30**(3), 1164–1208 (2020)
6. Barrera, G., Pardo, J.C.: Cut-off phenomenon for Ornstein–Uhlenbeck processes driven by Lévy processes. Electron. J. Probab. **25**(15), 1–33 (2020)
7. Da Prato, G., Gatarek, D., Zabczyk, J.: Invariant measures for semilinear stochastic equations. Stoch. Anal. Appl. **10**(4), 387–408 (1992)

8. Hartman, P.: On local homeomorphisms of Euclidean spaces. *Bol. Soc. Mat. Mexicana* **2**(5), 220–241 (1960)
9. Kallianpur, G., Sundar, P.: *Stochastic Analysis and Diffusion Processes*. Oxford University Press, Oxford (2014)
10. Majka, M.: A note on existence of global solutions and invariant measures for jump SDEs with locally one-sided Lipschitz drift. *Probab. Math. Stat.* **40**(1), 37–55 (2020)
11. Mikami, T.: Asymptotic expansions of the invariant density of a Markov process with a small parameter. *Ann. Inst. H. Poincaré Probab. Stat.* **24**(3), 403–424 (1988)
12. Panaretos, V., Zemel, Y.: *An Invitation to Statistics in Wasserstein Space*. Springer (2020)
13. Perko, L.: *Differential Equations and Dynamical Systems*. Texts in Applied Mathematics, vol. 7, 3rd edn. Springer, New York (2001)
14. Protter, P.: *Stochastic Integration and Differential Equations*, 2nd edn. Springer-Verlag, Berlin (2004)
15. Saint Loubert Bié, E.: Étude d'une EDPS conduite par un bruit poissonnien. *Probab. Theory Relat. Fields* **111**(2), 287–321 (1998)
16. Sato, K.: *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge (1999)
17. Siorpaes, P.: Applications of pathwise Burkholder–Davis–Gundy inequalities. *Bernoulli* **24**(4B), 3222–3245 (2018)
18. Situ, R.: *Theory of Stochastic Differential Equations with Jumps and Applications*. Springer, New York (2005)
19. Tudoran, R.M.: On the coercivity of continuously differentiable vector fields. *Qual. Theory Dyn. Syst.* **19**(2), Paper No. 58, 1–7 (2020)
20. Villani, C.: *Optimal Transport. Old and New*. Springer-Verlag, Berlin (2009)
21. Wang, J.: Regularity of semigroups generated by Lévy type operators via coupling. *Stoch. Process. Appl.* **120**(9), 1680–1700 (2010)
22. Watanabe, S., Ikeda, N.: *Stochastic Differential Equations and Diffusion Processes*. North-Holland Publishing Co., Amsterdam-New York, Kodansha Ltd., Tokyo (1981)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.