



Long-time contractivity estimates for kinetic Kolmogorov–Fokker–Planck equations

Nicolò Forcillo¹ · Alessio Porretta²

Received: 20 October 2025 / Revised: 22 March 2026 / Accepted: 14 April 2026

© The Author(s) 2026

Abstract

We prove long-time contractivity estimates and exponential rate of convergence to equilibrium for solutions of hypoelliptic diffusion equations, which include the well-known Kolmogorov equation and similar kinetic Fokker–Planck equations in \mathbb{R}^d . Compared to the existing literature, our proof exploits a different approach, elementary and self-contained, based on oscillation estimates for the adjoint problem. We first prove contractivity in Wasserstein distances through doubling-variable (coupling) methods. Next, we upgrade the estimate to weighted L^1 - (or total variation) norms, thanks to short-time hypocoercivity gradient estimates.

Mathematics Subject Classification 35Q84 · 35K65 · 35H10

Contents

1	Introduction
2	Setting of the problem
2.1	Notation

Partially supported by the MIUR Excellence Department Project Math@TOV awarded to the Department of Mathematics, University of Rome Tor Vergata. The author was also partially supported by the “Problemi variazionali/nonvariazionali: interazione tra metodi integrali e principi del massimo” 2023–2024 INDAM-GNAMPA project. Partially supported by the Excellence Project MatMod@TOV of the Department of Mathematics of the University of Rome Tor Vergata and Italian (EU Next Gen) PRIN project 2022W58BJ5 (*PDEs and optimal control methods in mean field games, population dynamics and multi-agent models*, CUP E53D23005910006). The author is also a member of a GNAMPA research group of Indam.

✉ Alessio Porretta
porretta@mat.uniroma2.it
Nicolò Forcillo
forcill1@msu.edu

¹ Department of Mathematics, Michigan State University, 619 Red Cedar Road, East Lansing, MI 48824, USA

² Department of Mathematics, University of Rome Tor Vergata, Via della Ricerca Scientifica 1, 00133 Rome, Italy

2.2 Assumptions and well-posedness of the differential problems

3 Oscillation decay estimates

3.1 Exponential decay of oscillation seminorms

3.2 Decay of Wasserstein distances for the kinetic Fokker–Planck equation

4 Hypocoercivity and L^∞ -Lipschitz smoothing

5 Improved long-time decay

5.1 Examples

6 Appendix

References

1 Introduction

This paper concerns the long-time convergence of solutions of kinetic-type Fokker–Planck equations, such as

$$\begin{cases} \partial_t m - \Delta_v m + \operatorname{div}_y(H(v, y)m) = \operatorname{div}_v(B(v, y)m), & (v, y) \in \mathbb{R}^{2d}, \\ m(0, v, y) = m_0(v, y), \end{cases} \tag{1.1}$$

where m_0 is a probability measure on \mathbb{R}^{2d} , and $B, H : \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$ are functions such that $D_v H$ is positive definite. The simplest and most famous example is given by the Kolmogorov equation (corresponding to $H(v, y) = v$)

$$\partial_t m - \Delta_v m + v \cdot \nabla_y m = \operatorname{div}_v(B(v, y)m), \tag{1.2}$$

which is the prototype of degenerate convection–diffusion equations that exhibit a hypoelliptic behavior. We recall that the solution of (1.2) describes the law of the Langevin process

$$\begin{cases} dy_t = v_t dt, \\ dv_t = -B(v_t, y_t) dt + \sqrt{2} dW_t, \end{cases}$$

where W_t is a Brownian motion in \mathbb{R}^d . It is well known that solutions of (1.2) converge exponentially fast, as t goes to infinity, towards the stationary state, if either y is in a compact setting (e.g., on the torus) or B is a suitable confining vector field. The occurrence of this exponential stability is a subtle combination of diffusion and transport, and it is strictly related to the underlying hypoellipticity of the operator.

There exist many different methods to show that exponential stabilization occurs. Originally, in the model case, the explicit knowledge of the stationary state and/or of the fundamental solution was used to derive this kind of result. However, an approach that avoids the use of explicit solutions appears preferable for tackling more general problems with a possibly similar structure. In [29], Villani introduced the notion of hypocoercivity to frame the question of long-time stability, showing that an exponential contractivity estimate, such as

$$\|m_1(t) - m_2(t)\|_X \leq C e^{-\omega t} \|m_{01} - m_{02}\|_X,$$

could be proved in the Hilbert setting of suitable energy spaces, being X , typically, a weighted H^1 - or L^2 -space. This is also related to spectral gap methods, already used before to quantify regularizing effects and long-time rate of convergence, see, e.g., [11, 18–20]. Later on, those approaches have been refined to a large extent and merged with entropy methods, geometric Poincaré inequalities, and different tools from semigroup theory [3, 4, 9, 12, 14, 16, 25, 31]. Different techniques, more directly based on Harris’ recurrence theorem for Markov semigroups [17], or the so-called Meyn–Tweedy approach [13, 24], were employed in other works, see, for instance, [6, 7, 23]. Their connection with our present article will be further commented below.

The goal of this paper is to introduce a new method to prove the exponential rate of convergence, extending to degenerate Kolmogorov operators the techniques recently developed in [26] for convection-diffusion equations with uniform or fractional diffusion. This approach derives the exponential decay by duality from oscillation estimates of the adjoint problem. Precisely, the focus is moved from the conservative equation (1.1) to the dual advection–diffusion problem

$$\partial_t u - \Delta_v u - H(v, y) \cdot \nabla_y u + B(v, y) \cdot \nabla_v u = 0, \tag{1.3}$$

where the decay of u is proved in suitably weighted seminorms (see estimate (3.5)).

This idea has its roots in the coupling methods developed in probability [8, 22] and, in parallel, in the doubling-variable techniques of the PDE community [10, 21], in the same spirit as [27]. The outcome is a self-contained, elementary approach, where pointwise oscillation estimates for the advection–diffusion equation (1.3) lead to the exponential decay of (1.1) bypassing any spectral analysis, Poincaré inequality, or integral Harnack inequality. Somehow, our work answers a natural question (mentioned by Villani in [30, Chapter 2]), namely, whether the exponential stabilization of the (hypoelliptic) kinetic Fokker–Planck equation could be proved through a simple coupling method, as in the diffusive case. A positive answer was already given in [15] by probabilistic arguments, showing the exponential contractivity in Wasserstein distances for solutions of (1.2). Here, we complement and extend [15] from a PDE perspective. Our contribution is twofold. First, we provide an entirely PDE-based proof of the decay in Wasserstein distances, only based on maximum principle, which gives conclusions for the general problem (1.1) similar to those obtained in [15] for the Kolmogorov equation. Next, we improve the result by achieving the decay in total variation norm, combining the oscillation estimates with short-time $L^\infty \rightarrow W^{1,\infty}$ smoothing effect. This latter tool seems interesting in its own, and should be considered as the L^∞ -version, on the adjoint problem, of the $L^2 \rightarrow H^1$ smoothing effect lying at the heart of Villani’s hypocoercivity estimate. The final outcome can be summarized in the following contractivity estimate, when the initial data are probability measures with finite first moments, whose space is denoted by $\mathcal{P}_1(\mathbb{R}^{2d})$.

Theorem 1.1 *Assume that H, B satisfy (2.5) and (2.6)–(2.7), and there exists a Lyapunov function φ (cf. Definition 3.1) fulfilling conditions (3.4), (4.8), and (5.1). Then, there exist $\omega, K > 0$ such that, for every initial data $m_{01}, m_{02} \in \mathcal{P}_1(\mathbb{R}^{2d})$ for which $\varphi \in L^1(dm_{0i}), i = 1, 2$, the corresponding solutions m_1, m_2 of (1.1) satisfy*

$$\|m_1(t) - m_2(t)\|_{TV_\varphi} \leq K e^{-\omega t} \|m_{01} - m_{02}\|_{TV_\varphi}, \quad (1.4)$$

where $\|\cdot\|_{TV_\varphi}$ is the norm of the total variation weighted by φ .

Let us spend a few words on the hypotheses of the above result.

A crucial role in Theorem 1.1 is played by the assumption that $D_v H$ is positive definite (cf. (2.6)). Roughly speaking, this means that the v -linearization of the drift field H is nondegenerate. The case in which $D_v H$ is negative definite would, of course, be the same. When H is linear with constant coefficients, our assumptions can be reduced to one form of the Hörmander condition (though not the most general one), in particular, the Hörmander condition of first order (say, only using first-order commutators). We discuss this issue in Remark 2.1. It has to be mentioned that, in the special case of linear drifts with constant coefficients, the exponential convergence under the general Hörmander condition is proved in [3] by using entropy methods and the explicit form of the steady state. We also hope to extend our approach to incorporate the case of general Hörmander operators. So far, our assumptions include and generalize the case of the Kolmogorov equation, appearing as the first step in the direction of treating general nonlinear and inhomogeneous first-order terms.

The requirement, in Theorem 1.1, that a suitable Lyapunov function exists should not be surprising, since the state space is noncompact, and thus some confinement condition is necessary for long-time stabilization. This is a classical ingredient used to show the ergodicity property of the Markov process associated with the operator (1.3). Explicit examples of Lyapunov functions are given at the end of the paper, precisely in Sect. 5.1. Similar forms of these functions already appear in previous works (see, e.g., [4, 6, 7, 15, 23]) with either analytic or probabilistic approaches. Some of the aforementioned papers rely on the version of Harris' theorem proved in [17], whose cornerstone is indeed the existence of a Lyapunov function. We stress again, however, that our method departs from this literature because we bypass the use of Harnack inequalities for the Fokker–Planck equation in favor of global oscillation estimates proved on the adjoint problem (1.3). In this context, the contractivity estimate is reduced to a maximum principle argument on the adjoint problem, where coupling methods naturally come into play.

Let us remark that the estimate of Theorem 1.1 holds for weak (distributional) solutions of (1.1). The key point indeed is to show that weak solutions are dual to viscosity solutions of the adjoint problem (1.3), which are sufficiently robust for the main estimate. This duality, already exploited in [26] for other drift-diffusion operators, is also very natural to handle sign-changing solutions. In this direction, the estimate (1.4) can alternatively be reformulated as the decay of zero-average solutions, see Theorem 5.3. Lastly, all our estimates are shown to be independent of possible additional diffusivity in the y -variable, which makes the results directly applicable to the framework of vanishing viscosity approximation for degenerate equations.

As mentioned before, in the intermediate steps needed to prove Theorem 1.1, we also establish other results that may have interest in their own. They can be found in the next sections, so it is worth giving a quick plan of the paper. In Sect. 2, we present some notation and the setting of our assumptions, and we discuss the notion of solution to (1.1) and (1.3) that we are going to use. We also state an existence and uniqueness

result (Theorem 2.3) for both the adjoint problems, which is the natural background of our estimates. Section 3 is devoted to proving the main estimate for problem (1.3), namely, the exponential decay of suitable oscillation seminorms, see Theorem 3.2. An immediate byproduct is the decay in Wasserstein distances of solutions of (1.1), see Theorem 3.5. In Sect. 4.3, we show the $L^\infty \rightarrow W^{1,\infty}$ smoothing effect for the Eq. (1.3), by suitably rephrasing Villani’s hypocoercivity argument to obtain pointwise gradient bounds through the classical Bernstein method, see Proposition 4.2. A weighted version, including the Lyapunov function growth into the estimate, is also needed for our purposes. Finally, in Sect. 5, we prove our main results, Theorem 5.2 (which includes the above Theorem 1.1) and 5.3, and we give some examples of application, where confinement properties of the drift ensure the existence of Lyapunov functions satisfying the desired conditions.

2 Setting of the problem

2.1 Notation

Let us start with some notation. In the following, $\mathbb{R}^d, d \geq 1$, denotes the d -dimensional Euclidean space, where dy represents the Lebesgue measure. I_d is the d -dimensional identity matrix, and $|y|$ is the usual Euclidean norm, for $y \in \mathbb{R}^d$. Given two real numbers a, b , we use the standard notation $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For vectors $z, w \in \mathbb{R}^d, z \otimes w$ is the matrix such that $(z \otimes w)_{ij} = z_i w_j$. For any $d \times d$ (real) matrix $A, \text{tr}(A)$ denotes its trace.

The space $C_b(U)$ is the set of continuous and bounded functions on an open set U , and $C_c(U)$ is the set of continuous and compactly supported functions. The space $\mathcal{M}(\mathbb{R}^d)$ denotes the set of finite (signed) Borel measures m , whose total variation norm is $\|m\|_{TV} := |m|(\mathbb{R}^d)$. Here, $|m| = m^+ + m^-$, where m^\pm are positive measures giving the Hahn decomposition of m . The subset of probability measures is represented by $\mathcal{P}(\mathbb{R}^d)$, endowed with the topology of narrow convergence. This is equivalent to the weak- $*$ convergence of measures in $\mathcal{M}(\mathbb{R}^d)$, induced by duality with the uniform convergence in $C_b(\mathbb{R}^d)$, see, e.g., [1, Chapter 5]. We refer to $\mathcal{M}_k(\mathbb{R}^d)$ as the subset of $\mathcal{M}(\mathbb{R}^d)$ consisting of measures m with finite k th moment, i.e., $|v|^k \in L^1(\mathbb{R}^d, d|m|)$. In particular, the set $\mathcal{P}_1(\mathbb{R}^d)$ denotes the space of probability measures with finite first moment, where we consider the Kantorovich–Rubinstein distance

$$\mathcal{W}_1(m, \tilde{m}) := \sup \left\{ \int_{\mathbb{R}^d} \zeta d(m - \tilde{m}) : \zeta \text{ Lipschitz, } \|D\zeta\|_\infty \leq 1 \right\}. \tag{2.1}$$

We fix a positive, continuous function φ on \mathbb{R}^d . $L^1(\varphi(x))$ represents the usual Lebesgue space $L^1(\mathbb{R}^d, \mu)$, defined in terms of the measure $\mu = \varphi(x)dx$, and $L^\infty(\varphi(x))$ is the space of measurable functions $u : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $u\varphi$ is essentially bounded (namely, $u\varphi \in L^\infty(\mathbb{R}^d)$). Similarly, we denote by $\mathcal{M}_\varphi(\mathbb{R}^d)$ the subset of measures m such that $\varphi \in L^1(\mathbb{R}^d, d|m|)$, and we set $\|m\|_{TV_\varphi} := \int_{\mathbb{R}^d} \varphi d|m|$.

For given $T > 0$, we set $Q_T := (0, T) \times \mathbb{R}^d$. We denote by $C^{1,2}(Q_T)$ the space of functions, defined on Q_T , that are C^1 in t and C^2 in x . We refer to $C([0, T]; \mathcal{M}(\mathbb{R}^d)^*)$

as the space of functions m from $[0, T]$ into $\mathcal{M}(\mathbb{R}^d)$, which are continuous with respect to the weak*-convergence of measures. Similarly, we use $C([0, T]; \mathcal{P}(\mathbb{R}^d))$ when working with probability measures, and $C([0, T]; \mathcal{P}_1(\mathbb{R}^d))$ for continuous functions m from $[0, T]$ into $\mathcal{P}_1(\mathbb{R}^d)$, endowed with the \mathcal{W}_1 -distance.

2.2 Assumptions and well-posedness of the differential problems

Let $v, y \in \mathbb{R}^d$ and denote by $\nabla_v, \operatorname{div}_v, \Delta_v$, etc..., the corresponding differential operators in the v -variable (the same for y). Given $\kappa \geq 0$, we consider the problem

$$\begin{cases} \partial_t m - \Delta_v m - \kappa \Delta_y m + \operatorname{div}_y(H(v, y)m) = \operatorname{div}_v(B(v, y)m), & t \in (0, T), \quad (v, y) \in \mathbb{R}^d \times \mathbb{R}^d, \\ m(0, v, y) = m_0(v, y), & (v, y) \in \mathbb{R}^d \times \mathbb{R}^d, \end{cases} \tag{2.2}$$

where $m_0 \in \mathcal{M}(\mathbb{R}^{2d})$. We recall that this equation is associated with the stochastic system

$$\begin{cases} dy_t = H(v_t, y_t) dt + \kappa \sqrt{2} dW_t \\ dv_t = -B(v_t, y_t) dt + \sqrt{2} d\tilde{W}_t \end{cases}$$

for two independent d -dimensional Brownian motions W_t, \tilde{W}_t , since $m(t)$ can be interpreted as the law of the process (v_t, y_t) in \mathbb{R}^{2d} , see, e.g., [2]. The case in which y is taken in the flat torus could also be considered, as well as the case $v \in \mathbb{R}^d, y \in \mathbb{R}^k$ with $k \neq d$. We will comment those two possible extensions later, see Remark 3.4. Here, we focus on the symmetric case, which includes the Langevin equation of Brownian particles when $\kappa = 0, H = v$, and, formally, $\ddot{y}_t = -B(\dot{y}_t, y_t) + \sqrt{2} \frac{dB_t}{dt}$.

We introduce the linear operator

$$\mathcal{L}[\varphi] := -\Delta_v \varphi - \kappa \Delta_y \varphi - H(v, y) \cdot \nabla_y \varphi + B(v, y) \cdot \nabla_v \varphi, \tag{2.3}$$

which defines the adjoint problem to (2.2). Indeed, we will also consider the following adjoint PDE (put in forward time direction)

$$\begin{cases} \partial_t u - \Delta_v u - \kappa \Delta_y u - H(v, y) \cdot \nabla_y u + B(v, y) \cdot \nabla_v u = 0, & t \in (0, T), \quad (v, y) \in \mathbb{R}^d \times \mathbb{R}^d, \\ u(0, v, y) = u_0(v, y), & (v, y) \in \mathbb{R}^d \times \mathbb{R}^d. \end{cases} \tag{2.4}$$

In (2.2) and (2.4), we assume that $B : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfies the following Lipschitz condition

$$|B(v_1, y_1) - B(v_2, y_2)| \leq \ell_B(|v_1 - v_2| + |y_1 - y_2|), \tag{2.5}$$

for some $\ell_B > 0$. Regarding the function $H : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, we suppose that it holds

$$\begin{aligned} D_v H(v, y) &\geq \gamma I_d \quad \forall (v, y) \in \mathbb{R}^{2d}, & (2.6) \\ |H(v_1, y_1) - H(v_2, y_2)| &\leq \ell_H(|v_1 - v_2| + |y_1 - y_2|) \quad \forall (v_1, y_1), (v_2, y_2) \in \mathbb{R}^{2d}, & (2.7) \end{aligned}$$

for some $\gamma, \ell_H > 0$. Note, incidentally, that (2.6) and (2.7) imply $\ell_H \geq \gamma$.

As mentioned in the Introduction, the simplest case satisfying (2.6)–(2.7) is given by $H(v, y) = v$, as in the Kolmogorov equation. For a more general class of examples, one can take $H(v, y) = vh(|v|, y)$, where h is any Lipschitz function increasing in $|v|$ and bounded below by a positive constant.

Remark 2.1 The nondegeneracy condition (2.6) plays a crucial role in what follows, ensuring the hypoellipticity of the drift-diffusion operator.

Let us discuss this condition in the linear case, i.e., if $H(v, y) = Cv + Dy$, where C, D are constant matrices in \mathbb{R}^d . There, assumption (2.6) corresponds to requiring C to be a positive definite matrix. To fix the ideas, we assumed $C > 0$, but, of course, the case $C < 0$ is equally treated up to changing y into $-y$ in the equation. We also stress that if $\kappa = 0$, then the change of variable $My = y'$ would turn equation (2.2) into a similar one where C is replaced by MC . So, actually, we only need the condition $MC > 0$ for some invertible matrix M . This is the case, for instance, if C is just assumed to be invertible, rather than positive definite.

To sum up, when $\kappa = 0$ and the drift is linear, our assumptions are equivalent to requiring the so-called first-order Hörmander condition for linear drift-diffusion operators $-\text{Tr}(AD^2(\cdot)) + FD(\cdot)$, namely, that no subspace of $\text{Ker}(A)$ is mapped into $\text{Ker}(A)$ by F^* (in our setting, if $H(v, y) = Cv + Dy$, then this would mean $\text{Ker}(C) \neq 0$). As pointed out, for instance, in [3, Section 2], the full Hörmander condition is actually more general, only asking that no subspace of $\text{Ker}(A)$ is left-invariant by F^* . However, this more general condition requires to use commutators of order higher than one. We expect to exploit this further generality in the future. By the way, we stress once more that our setting is neither confined to linear drifts nor to constant coefficients, obviously.

Let us now specify the notion of solution we use for the above initial value problems. As for (2.2), we work with the standard notion of weak (distributional) solution.

Definition 2.2 Let $m_0 \in \mathcal{M}(\mathbb{R}^{2d})$. A function $m \in C([0, T]; \mathcal{M}(\mathbb{R}^{2d})^*)$ is a weak solution to (2.2) if

$$\int_0^T \int_{\mathbb{R}^{2d}} (-\partial_t \phi + \mathcal{L}[\phi]) \, dm(\tau) \, d\tau = \int_{\mathbb{R}^{2d}} \phi(0, v, y) \, dm_0 \quad \forall \phi \in C_c^{1,2}([0, T] \times \mathbb{R}^{2d}).$$

Concerning problem (2.4), which is a degenerate second-order problem, it seems very natural to exploit the general concept of viscosity solutions, nowadays classical because of [10]. This notion preserves the maximum principle (the only ingredient needed for the estimates) and is stable for vanishing viscosity (i.e., as $\kappa \rightarrow 0$ above). As a byproduct of the estimates we prove later, we get the following existence and uniqueness result, whose proof is postponed to the Appendix.

Theorem 2.3 Assume that (2.5) and (2.6)–(2.7) hold, and $\kappa \geq 0$. For any $T > 0$, we have:

- (i) Given $u_0 \in C_b(\mathbb{R}^{2d})$, there exists a unique viscosity solution $u \in C_b(\overline{Q}_T)$ of (2.4). Moreover, for every $t > 0$, $u(t)$ is Lipschitz continuous and satisfies estimate (4.3).

(ii) Given $m_0 \in \mathcal{P}_1(\mathbb{R}^{2d})$, there exists a unique weak solution $m \in C([0, T]; \mathcal{P}(\mathbb{R}^{2d}))$ of (2.2). In addition, $m \in C([0, T]; \mathcal{P}_1(\mathbb{R}^{2d}))$ and

$$\int_{\mathbb{R}^{2d}} \zeta \, dm(t) = \int_{\mathbb{R}^{2d}} u(0, v, y) \, dm_0$$

$$\forall t \in (0, T), \zeta \in C_b(\mathbb{R}^{2d}), \text{ and } u \in C_b([0, t] \times \mathbb{R}^{2d}) \text{ such that}$$

$$u \text{ is a viscosity solution of } \mathbf{(B)} \begin{cases} -\partial_t u + \mathcal{L}[u] = 0 & \text{in } (0, t) \times \mathbb{R}^{2d}, \\ u(t, v, y) = \zeta(v, y) & \text{in } \mathbb{R}^{2d}. \end{cases} \tag{2.8}$$

For general (possibly signed) measures $m_0 \in \mathcal{M}_1(\mathbb{R}^{2d})$, there exists a unique weak solution m of (2.2) such that $\sup_{t \in (0, T)} \int_{\mathbb{R}^{2d}} |z| \, d|m|(t) < \infty$; in addition, m satisfies (2.8).

Remark 2.4 There is a slight difference in the above statement concerning the uniqueness of solutions of (2.2), depending on whether $m(t)$ is a probability measure or a general (possibly signed) measure. In this latter case, we need to require a priori that $|m|(t)$ has finite first moments, bounded in $(0, T)$. By contrast, for positive measures, this estimate can be deduced directly from the formulation. This subtle distinction between probability solutions (nonnegative and mass-preserving) and general solutions (possibly signed and not a priori mass-preserving) is classical for Fokker–Planck equations, see e.g. [5]. However, the two notions coincide here, provided that the first moments are controlled.

Let us comment further on the statements of Theorem 2.3. The well-posedness of viscosity solutions of (2.4) appears as a variant of the general theory. However, we could not find a result directly applied to (2.4), so we will give a proof here for completeness. Uniqueness of merely *weak solutions* to (2.2) is not trivial in general, although it is not surprising here due to the Lipschitz growth of the drift terms. For nondegenerate diffusions, this is well known, see, for instance, [5, 28]. For the Kolmogorov equation, a similar result can be found in [2]. Consequently, the most interesting outcome of the previous theorem is the duality equality (2.8) between weak and viscosity solutions of the two adjoint problems, as well as the consistency of the problems with the vanishing viscosity limit. As observed in [26], the dual representation (2.8) can eventually serve as a definition of robust solutions of Fokker–Planck equations, whenever the ordinary theory does not apply anymore. As a matter of fact, the formulation (2.8) suits possibly signed measures, and it guarantees positivity and mass preservation. A statement in this sense is the following corollary.

Corollary 2.5 *Under the assumptions of Theorem 2.3, let $m_0 \in \mathcal{M}_k(\mathbb{R}^{2d})$ for some $k > 0$. Then, there exists a unique $m \in C([0, T]; \mathcal{M}(\mathbb{R}^{2d})^*)$ satisfying (2.8). In addition, $m_0 \geq 0$ implies $m(t) \geq 0$ for all $t \in (0, T]$, and $\int_{\mathbb{R}^{2d}} dm(t) = \int_{\mathbb{R}^{2d}} dm_0$ for every $t > 0$.*

3 Oscillation decay estimates

Together with the hypoellipticity condition (2.6), a crucial role will also be played by Lyapunov functions since we are going to investigate the long-time stabilization for a noncompact state space.

Definition 3.1 We say that $\varphi(v, y)$ is a Lyapunov function for the operator \mathcal{L} if it satisfies

$$\varphi \in C^2(\mathbb{R}^d \times \mathbb{R}^d), \varphi(v, y) \xrightarrow{|(v,y)| \rightarrow \infty} \infty, \text{ and } \exists \omega_0 > 0 : \liminf_{|(v,y)| \rightarrow \infty} \frac{\mathcal{L}[\varphi]}{\varphi} \geq \omega_0. \tag{3.1}$$

Without loss of generality, we will also suppose $\varphi \geq 1$ in $\mathbb{R}^d \times \mathbb{R}^d$.

Let us mention that a condition such as (3.1) is satisfied under suitable coercivity assumptions on the drift terms, for which we refer to Sect. 5.1.

In connection with the Lyapunov function φ , we will consider (viscosity) solutions u of (2.4) so that $u = o(\varphi)$ as $|(v, y)| \rightarrow \infty$ (uniformly in $t \in (0, T)$), namely,

$$\lim_{|(v,y)| \rightarrow \infty} \left[\sup_{t \in (0,T)} \frac{u(t, v, y)}{\varphi(v, y)} \right] = 0. \tag{3.2}$$

Of course, this condition is trivially fulfilled for bounded solutions, but it seems convenient (and costless!) to consider the broader setting of possibly unbounded solutions, like, for instance, unbounded Lipschitz solutions.

3.1 Exponential decay of oscillation seminorms

In this subsection, we prove the exponential decay for a weighted oscillation seminorm of the solution. For $\theta \in (0, 1]$ and $w \in C(\mathbb{R}^{2d}) \cap L^\infty(\varphi^{-1})$, we define

$$[w]_\theta := \sup_{(v,y) \neq (\tilde{v}, \tilde{y})} \frac{|w(v, y) - w(\tilde{v}, \tilde{y})|}{(\varphi(v, y) + \varphi(\tilde{v}, \tilde{y}))(d^\theta \wedge 1)}, \quad d = |(v, y) - (\tilde{v}, \tilde{y})|. \tag{3.3}$$

Then, we have the following result.

Theorem 3.2 *Let u be a viscosity solution of (2.4) such that $u(t)$ is Lipschitz continuous in \mathbb{R}^{2d} (with a uniform Lipschitz constant for $t \in [0, T)$) and (3.2) is true, for some Lyapunov function φ associated to the operator \mathcal{L} . Assume that (2.5) and (2.6)-(2.7) hold, and, for some $\ell_\varphi > 0$,*

$$|\nabla_v \varphi(v, y) - \nabla_v \varphi(\tilde{v}, \tilde{y})| \leq \ell_\varphi [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})](|v - \tilde{v}| + |y - \tilde{y}|). \tag{3.4}$$

Then, for every $\theta \in (0, 1]$, there exist $\omega, K > 0$ such that

$$[u(t)]_\theta \leq K e^{-\omega t} [u_0]_\theta \tag{3.5}$$

for every $t > 0$, where ω, K depend on $\gamma, \ell_H, \ell_B, \varphi$, and θ .

Proof For the sake of clarity, we divide the proof into several steps. We first give the proof for any $\theta < 1$. In the end, we explain the modifications needed to handle the limiting case $\theta = 1$.

Step 1. Here we describe the geometric setting, the test function, and the strategy used to obtain (3.4).

We introduce the rotated norm

$$\|(v, y)\| := |v + \mu y| + \lambda |y|, \quad \text{where } \mu = \frac{2\ell_H}{\gamma}, \quad \lambda = \frac{4}{\gamma}((\mu + 1)\ell_H + 2\ell_B). \tag{3.6}$$

Note that μ and λ only depend on the constants γ, ℓ_H , and ℓ_B appearing in (2.5), (2.6), and (2.7). The precise value of μ and λ is not specially relevant and could have been fixed later in the proof. However, we observe since now that $\mu \geq 2$ and $\lambda > 4(\mu + 1)$, being $\ell_H \geq \gamma$, as noticed before.

Of course, the norm (3.6) is equivalent to the standard Euclidean norm in \mathbb{R}^{2d} . In particular, since $\lambda \geq 1 + \mu$, we have

$$|v| + |y| \leq \|(v, y)\| \leq |v| + (\lambda + \mu)|y|.$$

We regularize $\|\cdot\|$ as

$$\|(v, y)\|_\varepsilon := |v + \mu y|_\varepsilon + \lambda |y|_\varepsilon, \quad \text{with } |z|_\varepsilon := \sqrt{\varepsilon^2 + |z|^2} - \varepsilon.$$

For later convenience, we denote by

$$\xi := v + \mu y. \tag{3.7}$$

Next, we define the function ψ as

$$\psi(r) := 1 - e^{-C_2 r^\theta}, \tag{3.8}$$

for $\theta \in (0, 1)$ and some $C_2 \geq 1$ to be determined later. We also fix a generic time horizon $T > 0$. We claim that, for some C_2, ω, K , and L to be chosen later, we have

$$u(t, v, y) - u(t, \tilde{v}, \tilde{y}) \leq e^{-\omega t} (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho) + \frac{\eta}{T - t}, \tag{3.9}$$

for every $t \in (0, T)$, $v, y, \tilde{v}, \tilde{y} \in \mathbb{R}^d$ and η sufficiently small, where

$$\rho = \|(v - \tilde{v}, y - \tilde{y})\|.$$

In order to prove (3.9), we argue by contradiction. We define

$$\mathbf{m} := \sup_{t, (v, y), (\tilde{v}, \tilde{y})} \left\{ u(t, v, y) - u(t, \tilde{v}, \tilde{y}) - e^{-\omega t} (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho) - \frac{\eta}{T - t} \right\}$$

and suppose that $\mathbf{m} > 0$. We are going to show that this leads to a contradiction up to a suitable choice of K, L, C_2 (sufficiently large) and ω sufficiently small.

We first notice that \mathbf{m} is finite and is a global maximum attained at some point. Precisely, we have

$$\psi(\rho) \geq \frac{1}{2}(C_2\rho^\theta \wedge 1) \geq \frac{1}{2}(C_2\rho \wedge 1) \tag{3.10}$$

since $1 - e^{-t} \geq \frac{1}{2}(t \wedge 1)$. Due to assumption (3.2), for any $\sigma > 0$ there exists C_σ (depending on u) such that $|u(t, v, y)| \leq \sigma\varphi(v, y) + C_\sigma$ in $(0, T) \times \mathbb{R}^{2d}$. Choosing $\sigma = e^{-\omega T} \frac{K}{4}$, we deduce that $|u(t, v, y)| \leq e^{-\omega t} \frac{K}{4}\varphi(v, y) + C_{K,\omega,T}$, for some constant $C_{K,\omega,T}$ depending on K, ω, T . In addition, $u(t, v, y) - u(t, \tilde{v}, \tilde{y}) \leq \ell_u^T \rho$ for some $\ell_u^T > 0$, due to the Lipschitz continuity of u and the equivalence between $\|\cdot\|$ and the standard Euclidean norm in \mathbb{R}^{2d} . Hence, we can estimate

$$u(t, v, y) - u(t, \tilde{v}, \tilde{y}) \leq e^{-\omega t} \left(\frac{K}{4}[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + \tilde{C}_T \right) (C_2\rho \wedge 1),$$

for some constant \tilde{C}_T (possibly depending on u), which yields, by (3.10),

$$\begin{aligned} &u(t, v, y) - u(t, \tilde{v}, \tilde{y}) - e^{-\omega t} (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho) \\ &\leq e^{-\omega t} \left(-\frac{K}{2}[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + 2\tilde{C}_T - L \right) \psi(\rho). \end{aligned}$$

Since φ blows-up at infinity, we deduce that

$$\limsup_{|(v,y)| \vee |(\tilde{v},\tilde{y})| \rightarrow \infty} \left\{ u(t, v, y) - u(t, \tilde{v}, \tilde{y}) - e^{-\omega t} (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho) - \frac{\eta}{T-t} \right\} \leq 0, \tag{3.11}$$

uniformly for $t \in (0, T)$. Therefore, assuming $\mathbf{m} > 0$ implies (by continuity of u) that the sup is attained at some point in $[0, T] \times \mathbb{R}^{2d}$. Note that this happens for any given choice of K, L, η and T , being just a consequence of the regularity of u and of the properties of the Lyapunov function φ .

Now, for $\varepsilon > 0$ small, we consider the regularized distance $\rho_\varepsilon = \|(v - \tilde{v}, y - \tilde{y})\|_\varepsilon$ and, correspondingly, the value

$$\mathbf{m}_\varepsilon := \sup_{t, (v,y), (\tilde{v},\tilde{y})} \left\{ u(t, v, y) - u(t, \tilde{v}, \tilde{y}) - e^{-\omega t} (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho_\varepsilon) - \frac{\eta}{T-t} \right\}. \tag{3.12}$$

Notice that $|z|_\varepsilon \leq |z| \leq |z|_\varepsilon + \varepsilon$, and so

$$\rho_\varepsilon \leq \rho \leq \rho_\varepsilon + (1 + \lambda)\varepsilon. \tag{3.13}$$

In particular, being ψ increasing, we have $\mathbf{m}_\varepsilon \geq \mathbf{m} > 0$. Moreover, we claim that (3.11) even holds with ρ replaced by ρ_ε . Specifically, repeating the argument exploited above, for any small $\sigma > 0$ there exists a constant C_σ such that

$$u(t, v, y) - u(t, \tilde{v}, \tilde{y}) \leq [\sigma(\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})) + C_\sigma] \mathbb{1}_{\{C_2\rho_\varepsilon^\theta > 1\}} + \ell_u^T \rho \mathbb{1}_{\{C_2\rho_\varepsilon^\theta \leq 1\}},$$

for some $\ell_u^T > 0$. So, from (3.13), we infer (recalling that $C_2 \geq 1$ and $\theta \in (0, 1)$)

$$\begin{aligned} u(t, v, y) - u(t, \tilde{v}, \tilde{y}) &\leq [\sigma(\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})) + C_\sigma] \mathbb{1}_{\{C_2 \rho_\varepsilon^\theta > 1\}} + \ell_u^T \rho_\varepsilon \mathbb{1}_{\{C_2 \rho_\varepsilon^\theta \leq 1\}} + \ell_u^T (1 + \lambda)\varepsilon \\ &\leq [\sigma(\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})) + C_\sigma] \mathbb{1}_{\{C_2 \rho_\varepsilon^\theta > 1\}} + \ell_u^T C_2 \rho_\varepsilon^\theta \mathbb{1}_{\{C_2 \rho_\varepsilon^\theta \leq 1\}} + \ell_u^T (1 + \lambda)\varepsilon \\ &\leq [\sigma(\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})) + C_\sigma + \ell_u^T] 2 \psi(\rho_\varepsilon) + \ell_u^T (1 + \lambda)\varepsilon, \end{aligned}$$

where we have also used (3.10) in the last step.

Hence, we get, for any $t \in (0, T)$,

$$\begin{aligned} u(t, v, y) - u(t, \tilde{v}, \tilde{y}) - e^{-\omega t} (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho_\varepsilon) - \frac{\eta}{T - t} \\ \leq \left\{ [C_\sigma + \ell_u^T] 2 - e^{-\omega T} \left((K - 2\sigma e^{\omega T})[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L \right) \right\} \psi(\rho_\varepsilon) \\ + \ell_u^T (1 + \lambda)\varepsilon - \frac{\eta}{T - t}. \end{aligned}$$

Taking $\sigma < e^{-\omega T} K/2$ and ε such that $\varepsilon \ell_u^T (1 + \lambda) < \frac{\eta}{T}$, we obtain (3.11) with ρ replaced by ρ_ε . Consequently, \mathbf{m}_ε is a global maximum as well, attained at some point $(t_\varepsilon, v_\varepsilon, y_\varepsilon, \tilde{v}_\varepsilon, \tilde{y}_\varepsilon)$. We also notice that obviously $t_\varepsilon < T$, and moreover $t_\varepsilon > 0$, provided

$$u_0(v, y) - u_0(\tilde{v}, \tilde{y}) \leq (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho_\varepsilon) + \frac{\eta}{T} \tag{3.14}$$

for every $v, y, \tilde{v}, \tilde{y}$ and ε sufficiently small. This is the case, for instance, if (recall (3.3))

$$K \geq 2[u_0]_\theta \tag{3.15}$$

holds. Indeed, recall that $u(t)$ is Lipschitz continuous up to $t = 0$, by assumption. Thus, using (3.10) and the fact that, if $\rho \geq 2(1 + \lambda)\varepsilon$, then we have $\rho \leq 2\rho_\varepsilon$ (due to (3.13)), taking $C_2 \geq 2$ we get

$$\begin{aligned} u_0(v, y) - u_0(\tilde{v}, \tilde{y}) &\leq [u_0]_\theta [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] ((2\rho_\varepsilon)^\theta \wedge 1) + \ell_{u_0} \rho \mathbb{1}_{\{\rho < 2(1+\lambda)\varepsilon\}} \\ &\leq [u_0]_\theta [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] ((C_2 \rho_\varepsilon)^\theta \wedge 1) + \ell_{u_0} 2(1 + \lambda)\varepsilon \\ &\leq [u_0]_\theta [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] 2\psi(\rho_\varepsilon) + \ell_{u_0} 2(1 + \lambda)\varepsilon. \end{aligned}$$

Therefore, (3.14) is satisfied as long as $K \geq 2[u_0]_\theta$ and $2(1 + \lambda)\varepsilon \ell_{u_0} \leq \frac{\eta}{T}$. We conclude that $(t_\varepsilon, v_\varepsilon, y_\varepsilon, \tilde{v}_\varepsilon, \tilde{y}_\varepsilon)$ is a local (and global) maximum point, where \mathbf{m}_ε is attained.

Step 2. In this step, we exploit the maximum principle (for viscosity solutions) together with the coupling method.

To lighten the notation, henceforth we drop the index ε from the maximum point found above, where m_ε in (3.12) is attained. At such $t, (v, y), (\tilde{v}, \tilde{y})$, we apply the classical argument in viscosity solutions theory [10], which provides with suitable

replacements of $Du(t, v, y)$, $D^2u(t, v, y)$ and $Du(t, \tilde{v}, \tilde{y})$, $D^2u(t, \tilde{v}, \tilde{y})$ in Eq. (2.4). For a proper formulation in this context, we refer e.g. to [27, Thm 2.2 and Rmk 2.3] or [26, Thm 2.3]. Namely, since $t, (v, y), (\tilde{v}, \tilde{y})$ is a local maximum in (3.12), there exist matrices $V, \tilde{V}, Y, \tilde{Y}$ such that

$$\begin{pmatrix} V & \dots & 0 & 0 \\ \dots & Y & 0 & 0 \\ 0 & 0 & -\tilde{V} & \dots \\ 0 & 0 & \dots & -\tilde{Y} \end{pmatrix} \leq e^{-\omega t} D_{(v,y),(\tilde{v},\tilde{y})}^2 \zeta, \tag{3.16}$$

where

$$\zeta := (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho_\epsilon),$$

and, subtracting the equations of $u(t, v, y)$ and $u(t, \tilde{v}, \tilde{y})$,

$$\begin{aligned} & \frac{\eta}{(T-t)^2} - \omega e^{-\omega t} \zeta - e^{-\omega t} (H(v, y) \cdot D_y \zeta + H(\tilde{v}, \tilde{y}) \cdot D_{\tilde{y}} \zeta) \\ & + e^{-\omega t} (B(v, y) \cdot \nabla_v \zeta + B(\tilde{v}, \tilde{y}) \cdot \nabla_{\tilde{v}} \zeta) \leq \text{tr}(V - \tilde{V}) + \kappa \text{tr}(Y - \tilde{Y}). \end{aligned} \tag{3.17}$$

To handle the diffusion terms, we introduce the coupling matrix

$$\begin{pmatrix} I_d & 0 & C & 0 \\ 0 & \kappa I_d & 0 & \kappa I_d \\ C^* & 0 & I_d & 0 \\ 0 & \kappa I_d & 0 & \kappa I_d \end{pmatrix}, \tag{3.18}$$

which we assume to be nonnegative. We also suppose that C is symmetric to ease computations. Note that we are only using the synchronous coupling in the y -variable, since in the spirit of our result the y -diffusion is meant to be vanishing.

Multiplying (3.16) by the above matrix and taking the traces, we get

$$\begin{aligned} \text{tr}(V - \tilde{V}) + \kappa \text{tr}(Y - \tilde{Y}) & \leq e^{-\omega t} \text{tr} \left(D_{vv}^2 \zeta + 2C D_{\tilde{v}\tilde{v}}^2 \zeta + D_{\tilde{v}\tilde{v}}^2 \zeta \right) \\ & + \kappa e^{-\omega t} \text{tr} \left(D_{yy}^2 \zeta + 2D_{\tilde{y}\tilde{y}}^2 \zeta + D_{\tilde{y}\tilde{y}}^2 \zeta \right), \end{aligned}$$

so that (3.17) implies

$$\begin{aligned} & \frac{\eta}{(T-t)^2} - \omega e^{-\omega t} \zeta - e^{-\omega t} (H(v, y) \cdot D_y \zeta + H(\tilde{v}, \tilde{y}) \cdot D_{\tilde{y}} \zeta) \\ & + e^{-\omega t} (B(v, y) \cdot \nabla_v \zeta + B(\tilde{v}, \tilde{y}) \cdot \nabla_{\tilde{v}} \zeta) \\ & \leq e^{-\omega t} \text{tr} \left(D_{vv}^2 \zeta + 2C D_{\tilde{v}\tilde{v}}^2 \zeta + D_{\tilde{v}\tilde{v}}^2 \zeta \right) + \kappa e^{-\omega t} \text{tr} \left(D_{yy}^2 \zeta + 2D_{\tilde{y}\tilde{y}}^2 \zeta + D_{\tilde{y}\tilde{y}}^2 \zeta \right). \end{aligned}$$

Using the explicit form of the test function ζ , dropping the first term, and dividing by $e^{-\omega t}$, it follows that, in view of (2.3),

$$\begin{aligned} &\psi(\rho_\varepsilon)K \{ \mathcal{L}[\varphi](v, y) - \omega\varphi(v, y) + \mathcal{L}[\varphi](\tilde{v}, \tilde{y}) - \omega\varphi(\tilde{v}, \tilde{y}) \} - \omega L\psi(\rho_\varepsilon) \\ &\quad + (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi'(\rho_\varepsilon) \\ &\quad \times (- (H(v, y) - H(\tilde{v}, \tilde{y})) \cdot \nabla_y \rho_\varepsilon + (B(v, y) - B(\tilde{v}, \tilde{y})) \cdot \nabla_v \rho_\varepsilon) \\ &\leq (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \times 2\text{tr}(I_d - C) D_{vv}^2(\psi(\rho_\varepsilon)) \\ &\quad + 2K \psi'(\rho_\varepsilon) (I_d - C) \nabla_v \rho_\varepsilon \cdot (\nabla\varphi(v, y) - \nabla\varphi(\tilde{v}, \tilde{y})). \end{aligned} \tag{3.19}$$

Let us now specify our choice of C . For ξ given by (3.7) and $\tau \geq 0$, we take

$$C = I - \tau \frac{\xi - \tilde{\xi}}{|\xi - \tilde{\xi}|} \otimes \frac{\xi - \tilde{\xi}}{|\xi - \tilde{\xi}|}. \tag{3.20}$$

With this, we observe that the matrix in (3.18) is nonnegative provided $0 \leq \tau \leq 2$. This selection of coupling includes what are called the synchronous coupling ($\tau = 0$) and the reflection coupling ($\tau = 2$), and we will choose one of these two options depending on the distance ρ_ε between the points (v, y) and (\tilde{v}, \tilde{y}) . Notice, indeed, that we need to take $\tau = 0$ whenever $\xi = \tilde{\xi}$. Otherwise, C would be singular.

Let us further elaborate the second-order terms. According to the definition of ρ_ε , we compute

$$\begin{aligned} D_{vv}^2\psi(\rho_\varepsilon) &= \psi''(\rho_\varepsilon)\nabla_v \rho_\varepsilon \otimes \nabla_v \rho_\varepsilon + \psi'(\rho_\varepsilon)D_{vv}^2\rho_\varepsilon \\ &= \psi''(\rho_\varepsilon)\left(\frac{\xi - \tilde{\xi}}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \otimes \frac{\xi - \tilde{\xi}}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon}\right) \\ &\quad + \psi'(\rho_\varepsilon)\frac{1}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon}\left(I_d - \frac{\xi - \tilde{\xi}}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \otimes \frac{\xi - \tilde{\xi}}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon}\right). \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} 2\tau \left((I_d - C) D_{vv}^2(\psi(\rho_\varepsilon)) \right) &= 2\tau \psi''(\rho_\varepsilon) \left(\frac{|\xi - \tilde{\xi}|}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \right)^2 \\ &\quad + 2\tau \frac{\psi'(\rho_\varepsilon)}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \frac{\varepsilon^2}{(|\xi - \tilde{\xi}|_\varepsilon + \varepsilon)^2}. \end{aligned} \tag{3.21}$$

Step 3. We now show that (3.19), together with (3.21), leads to a contradiction, with suitable choices of the parameters. We analyze three different cases, depending on the distance between the points (v, y) and (\tilde{v}, \tilde{y}) .

Case 1 (ρ_ε large): We first analyze the case that $\rho_\varepsilon \geq R_1$, for some large $R_1 > 1$ to be fixed later.

We estimate, by virtue of (2.5) and (2.7), since $|\nabla_v \rho_\varepsilon| \leq 1$, $|\nabla_y \rho_\varepsilon| \leq \mu + \lambda$, and recalling that $|v - \tilde{v}| + |y - \tilde{y}| \leq \rho$,

$$\begin{aligned} & |(H(v, y) - H(\tilde{v}, \tilde{y})) \cdot \nabla_y \rho_\varepsilon| + |(B(v, y) - B(\tilde{v}, \tilde{y})) \cdot \nabla_v \rho_\varepsilon| \\ & \leq (\ell_H(\mu + \lambda) + \ell_B)(|v - \tilde{v}| + |y - \tilde{y}|) \\ & \leq (\ell_H(\mu + \lambda) + \ell_B)(\rho_\varepsilon + (1 + \lambda)\varepsilon) \leq C\rho_\varepsilon, \end{aligned} \tag{3.22}$$

where we have also exploited (3.13) and the fact that ρ_ε is large enough. Here, C depends on μ, λ, ℓ_B , and ℓ_H (hence, it only depends on ℓ_B, ℓ_H , and γ). Next, we take $\tau = 0$ in the choice of the coupling (3.20) (i.e., $C = I_d$). Hence, we obtain, from (3.19),

$$\begin{aligned} & \psi(\rho_\varepsilon)K \{ \mathcal{L}[\varphi](v, y) - \omega\varphi(v, y) + \mathcal{L}[\varphi](\tilde{v}, \tilde{y}) - \omega\varphi(\tilde{v}, \tilde{y}) \} - \omega L\psi(\rho_\varepsilon) \\ & - (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) C\psi'(\rho_\varepsilon)\rho_\varepsilon \leq 0. \end{aligned} \tag{3.23}$$

Since (3.8) implies

$$\psi'(r)r = \theta C_2 r^\theta e^{-C_2 r^\theta}, \tag{3.24}$$

it results that $\psi'(r)r$ is decreasing for $C_2 r^\theta \geq 1$, so, a fortiori, if $r \geq 1$. Then, we have $\psi'(\rho_\varepsilon)\rho_\varepsilon \leq \psi'(R_1)R_1$ for $\rho_\varepsilon \geq R_1 \geq 1$. Being $C_2 \geq 1$, it also holds $1/2 \leq \psi(\rho_\varepsilon) \leq 1$ for $\rho_\varepsilon \geq R_1$. Consequently, we get

$$C\psi'(\rho_\varepsilon)\rho_\varepsilon \leq 2C\psi(\rho_\varepsilon)\psi'(R_1)R_1.$$

Renaming $\tilde{C} = 2C$, it follows, in view of (3.23),

$$\begin{aligned} & \psi(\rho_\varepsilon) \left\{ K(\mathcal{L}[\varphi](v, y) - (\omega + \tilde{C}\psi'(R_1)R_1)\varphi(v, y) + \mathcal{L}[\varphi](\tilde{v}, \tilde{y}) \right. \\ & \left. - (\omega + \tilde{C}\psi'(R_1)R_1)\varphi(\tilde{v}, \tilde{y})) - L(\omega + \tilde{C}\psi'(R_1)R_1) \right\} \leq 0. \end{aligned} \tag{3.25}$$

We recall that $\rho_\varepsilon \leq \rho \leq \|(v, y)\| + \|(\tilde{v}, \tilde{y})\|$. Therefore, $\rho_\varepsilon \geq R_1$ implies that $\|(v, y)\| \geq R_1/2$ or $\|(\tilde{v}, \tilde{y})\| \geq R_1/2$. Without loss of generality, we assume the first holds. Up to choosing ω small and R_1 large so that

$$\omega + \tilde{C}\psi'(R_1)R_1 < \frac{\omega_0}{2}, \tag{3.26}$$

where ω_0 is given by (3.1), we have that $\mathcal{L}[\varphi] - (\omega + \tilde{C}\psi'(R_1)R_1)\varphi \geq -k_0$ for some $k_0 > 0$. Hence, we estimate, according to (3.1) and (3.26) again,

$$\begin{aligned} & \mathcal{L}[\varphi](v, y) - (\omega + \tilde{C}\psi'(R_1)R_1)\varphi(v, y) + \mathcal{L}[\varphi](\tilde{v}, \tilde{y}) - (\omega + \tilde{C}\psi'(R_1)R_1)\varphi(\tilde{v}, \tilde{y}) \\ & \geq \mathcal{L}[\varphi](v, y) - \frac{\omega_0}{2}\varphi(v, y) - k_0 \geq \frac{\omega_0}{4}\varphi(v, y), \end{aligned}$$

as long as R_1 is sufficiently large, only depending on ω_0 and φ (and μ and λ). We thus deduce, from (3.25), that if $\rho_\varepsilon \geq R_1$, then

$$K \frac{\omega_0}{4} \left(\inf_{\{|(v,y)| \geq R_1/2\}} \varphi \right) \leq L(\omega + \tilde{C}\psi'(R_1)R_1).$$

However, this is impossible if we choose K as

$$K = L \frac{8(\omega + \tilde{C}\psi'(R_1)R_1)}{\omega_0\varphi_{R_1}}, \quad \varphi_{R_1} := \inf\{\varphi(v, y) : \|(v, y)\| \geq R_1/2\}. \tag{3.27}$$

Case 2 ($|\xi - \tilde{\xi}| > \sqrt{\varepsilon}$) We now discuss the case $\rho_\varepsilon < R_1$ and $|\xi - \tilde{\xi}| > \sqrt{\varepsilon}$. If this occurs, then we have, for ε small,

$$\rho_\varepsilon \geq |\xi - \tilde{\xi}|_\varepsilon \geq |\xi - \tilde{\xi}| - \varepsilon \geq \frac{1}{2}\sqrt{\varepsilon},$$

thus, from (3.13),

$$|v - \tilde{v}| + |y - \tilde{y}| \leq \rho \leq \rho_\varepsilon + (1 + \lambda)\varepsilon \leq C \rho_\varepsilon. \tag{3.28}$$

Then, we can estimate the drift terms in (3.19) exactly as in (3.22), using the Lipschitz character of H and B . Similarly, we treat the last term in (3.19) using $|\nabla_v \rho_\varepsilon| \leq 1$ and assumption (3.4), which yields

$$2K\psi'(\rho_\varepsilon)(I_d - C)\nabla_v \rho_\varepsilon \cdot (\nabla\varphi(v, y) - \nabla\varphi(\tilde{v}, \tilde{y})) \leq CK[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})]\psi'(\rho_\varepsilon)\rho_\varepsilon.$$

Hereafter, we denote by C possibly different constants only depending on γ, ℓ_H, ℓ_B , and ℓ_φ . Finally, we use (3.21) with $\tau = 2$ (note that $\tau = 1$ would equally work) and $\mathcal{L}[\varphi] - \omega\varphi \geq -k_0$. Overall, we obtain from (3.19),

$$\begin{aligned} & -2\psi(\rho_\varepsilon)k_0K - \omega L\psi(\rho_\varepsilon) - (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L)C\psi'(\rho_\varepsilon)\rho_\varepsilon \\ & \leq (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \\ & \quad \times \left\{ 4\psi''(\rho_\varepsilon) \left(\frac{|\xi - \tilde{\xi}|}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \right)^2 + 4 \frac{\psi'(\rho_\varepsilon)}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \frac{\varepsilon^2}{(|\xi - \tilde{\xi}|_\varepsilon + \varepsilon)^2} \right\}. \end{aligned} \tag{3.29}$$

We now observe that, due to (3.27), there exists $c > 0$ (independent of C_2, R_1 , and L) such that

$$K\psi(\rho_\varepsilon) \leq cL \left(\frac{\omega\psi(\rho_\varepsilon) + \psi'(\rho_\varepsilon)\rho_\varepsilon}{\varphi_{R_1}} \right) \quad \forall \rho_\varepsilon < R_1. \tag{3.30}$$

Indeed, if $C_2\rho_\varepsilon^\theta \geq 1$, then $\psi'(\rho)\rho$ is decreasing for $\rho \geq \rho_\varepsilon$, so (3.27) implies, for $\rho_\varepsilon < R_1$,

$$K\psi(\rho_\varepsilon) \leq L \frac{8(\omega\psi(\rho_\varepsilon) + \tilde{C}\psi'(\rho_\varepsilon)\rho_\varepsilon)}{\omega_0\varphi_{R_1}},$$

which yields (3.30). Alternatively, we have $C_2\rho_\varepsilon^\theta < 1$, and then $\psi(\rho_\varepsilon) \leq \frac{e}{\theta} \psi'(\rho_\varepsilon)\rho_\varepsilon$ (using (3.24) and that $1 - e^{-t} \leq e t e^{-t}$ for $t \leq 1$). Hence, $K\psi(\rho_\varepsilon) \leq c K \psi'(\rho_\varepsilon)\rho_\varepsilon$ and (3.30) immediately follows because $K \leq \frac{4}{\varphi_{R_1}}L$, according to (3.26) and (3.27). Therefore, (3.30) holds for every $\rho_\varepsilon < R_1$ and we have

$$2\psi(\rho_\varepsilon)k_0 K + \omega L\psi(\rho_\varepsilon) \leq \frac{2k_0c}{\varphi_{R_1}}\psi'(\rho_\varepsilon)\rho_\varepsilon L + \left(1 + \frac{2k_0c}{\varphi_{R_1}}\right)\omega L\psi(\rho_\varepsilon).$$

Note that φ_{R_1} is uniformly bounded below (since $\varphi \geq 1$), and eventually it can be taken large up to increasing the value of R_1 . In particular, we will assume henceforth that R_1 is sufficiently large so that

$$\frac{2k_0c}{\varphi_{R_1}} \leq \frac{\gamma}{4},$$

where we recall that γ is given in (2.6). With this in hand, we refine the above estimate as

$$2\psi(\rho_\varepsilon)k_0 K + \omega L\psi(\rho_\varepsilon) \leq \frac{\gamma}{4}\psi'(\rho_\varepsilon)\rho_\varepsilon L + \left(1 + \frac{\gamma}{4}\right)\omega L\psi(\rho_\varepsilon) \tag{3.31}$$

whenever $\rho_\varepsilon < R_1$. Inserting this into (3.29), we obtain

$$\begin{aligned} & (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \left\{ -4\psi''(\rho_\varepsilon) \left(\frac{|\xi - \tilde{\xi}|}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \right)^2 \right. \\ & \left. - \frac{4\psi'(\rho_\varepsilon)}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \frac{\varepsilon^2}{(|\xi - \tilde{\xi}|_\varepsilon + \varepsilon)^2} - C\psi'(\rho_\varepsilon)\rho_\varepsilon \right\} \\ & \leq \left(1 + \frac{\gamma}{4}\right)\omega L\psi(\rho_\varepsilon). \end{aligned} \tag{3.32}$$

From the definition of the function ψ , it holds

$$\begin{aligned} & \left\{ -4\psi''(\rho_\varepsilon) \left(\frac{|\xi - \tilde{\xi}|}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \right)^2 - \frac{4\psi'(\rho_\varepsilon)}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \frac{\varepsilon^2}{(|\xi - \tilde{\xi}|_\varepsilon + \varepsilon)^2} - C\psi'(\rho_\varepsilon)\rho_\varepsilon \right\} \\ & = C_2\theta \rho_\varepsilon^{\theta-2} e^{-C_2\rho_\varepsilon^\theta} \left(4[(1-\theta) + C_2\theta\rho_\varepsilon^\theta] \left(\frac{|\xi - \tilde{\xi}|}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \right)^2 - 4\rho_\varepsilon \frac{\varepsilon^2}{(|\xi - \tilde{\xi}|_\varepsilon + \varepsilon)^3} - C\rho_\varepsilon^2 \right) \\ & \geq C_2\theta \rho_\varepsilon^{\theta-2} e^{-C_2\rho_\varepsilon^\theta} \left(2(1-\theta) + 2C_2\theta\rho_\varepsilon^\theta - 4\sqrt{\varepsilon}\rho_\varepsilon - C\rho_\varepsilon^2 \right), \end{aligned} \tag{3.33}$$

where we have used that $|\xi - \tilde{\xi}| > \sqrt{\varepsilon}$ gives

$$\left(\frac{|\xi - \tilde{\xi}|}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \right)^2 \geq 1 - \varepsilon \geq \frac{1}{2} \quad \text{and} \quad \frac{\varepsilon^2}{(|\xi - \tilde{\xi}|_\varepsilon + \varepsilon)^3} \leq \sqrt{\varepsilon}. \tag{3.34}$$

Now, we fix C_2 , in terms of R_1 , by imposing

$$2C_2\theta r^\theta - 4\sqrt{\varepsilon}r - Cr^2 \geq 0 \quad \forall r \leq R_1. \tag{3.35}$$

Thus, we deduce, from (3.32) and (3.33),

$$(K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) C_2\theta \rho_\varepsilon^{\theta-2} e^{-C_2\rho_\varepsilon^\theta} 2(1 - \theta) \leq \left(1 + \frac{\gamma}{4}\right) \omega L\psi(\rho_\varepsilon),$$

which yields, since $\rho_\varepsilon \leq R_1$, $\theta \in (0, 1)$, and $\psi \leq 1$,

$$C_2\theta R_1^{\theta-2} e^{-C_2R_1^\theta} 2(1 - \theta) \leq \left(1 + \frac{\gamma}{4}\right) \omega.$$

Nevertheless, this is impossible if we choose ω sufficiently small, namely,

$$\omega < \left(1 + \frac{\gamma}{4}\right)^{-1} C_2\theta R_1^{\theta-2} e^{-C_2R_1^\theta} 2(1 - \theta). \tag{3.36}$$

Case 3 ($|\xi - \tilde{\xi}| \leq \sqrt{\varepsilon}$) By virtue of the above choices and conditions for the parameters, we are only left with the possibility that $\rho_\varepsilon < R_1$ and $|\xi - \tilde{\xi}| \leq \sqrt{\varepsilon}$. This scenario corresponds to $|\xi - \tilde{\xi}|$ near zero (or possibly zero), where we cannot use the coupling by reflection for the variable ξ . We indeed take $\mathcal{C} = I_d$. Moreover, we exploit $\mathcal{L}[\varphi] - \omega\varphi \geq -k_0$. Hence, we achieve, according to (3.19),

$$\begin{aligned} & -2\psi(\rho_\varepsilon)k_0K - \omega L\psi(\rho_\varepsilon) \\ & + (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi'(\rho_\varepsilon) \{- (H(v, y) - H(\tilde{v}, \tilde{y})) \\ & \cdot \nabla_y \rho_\varepsilon + (B(v, y) - B(\tilde{v}, \tilde{y})) \cdot \nabla_v \rho_\varepsilon\} \leq 0. \end{aligned} \tag{3.37}$$

As for the drift terms, recalling that $v = \xi - \mu y$ and $|\nabla_y \rho_\varepsilon| \leq (\mu + \lambda)$, we have

$$\begin{aligned} & - (H(v, y) - H(\tilde{v}, \tilde{y})) \cdot \nabla_y \rho_\varepsilon \\ & = - \int_0^1 D_v H(sv + (1-s)\tilde{v}, sy + (1-s)\tilde{y})(v - \tilde{v}) \cdot \nabla_y \rho_\varepsilon ds \\ & \quad - \int_0^1 D_y H(sv + (1-s)\tilde{v}, sy + (1-s)\tilde{y})(y - \tilde{y}) \cdot \nabla_y \rho_\varepsilon ds \\ & \geq \mu \int_0^1 D_v H(sv + (1-s)\tilde{v}, sy + (1-s)\tilde{y})(y - \tilde{y}) \cdot \nabla_y \rho_\varepsilon \\ & \quad - (\mu + \lambda)\ell_H (|\xi - \tilde{\xi}| + |y - \tilde{y}|) \\ & \geq \mu \gamma \lambda \frac{|y - \tilde{y}|^2}{\sqrt{\varepsilon^2 + |y - \tilde{y}|^2}} \\ & \quad - \mu^2 \ell_H |y - \tilde{y}| - (\mu + \lambda)\ell_H (|\xi - \tilde{\xi}| + |y - \tilde{y}|), \end{aligned}$$

where we have used $D_v H \geq \gamma I_d$, $\|D_v H\|, \|D_y H\| \leq \ell_H$, and $|\nabla_y \rho_\varepsilon - \lambda \frac{(y-\tilde{y})}{\sqrt{\varepsilon^2 + |y-\tilde{y}|^2}}| \leq \mu$. Being

$$\frac{|y - \tilde{y}|^2}{\sqrt{\varepsilon^2 + |y - \tilde{y}|^2}} \geq |y - \tilde{y}|_\varepsilon \geq |y - \tilde{y}| - \varepsilon,$$

in view of the Lipschitz character of B we get

$$\begin{aligned} & - (H(v, y) - H(\tilde{v}, \tilde{y})) \cdot \nabla_y \rho_\varepsilon + (B(v, y) - B(\tilde{v}, \tilde{y})) \cdot \nabla_v \rho_\varepsilon \\ & \geq \mu \gamma \lambda |y - \tilde{y}| - \mu^2 \ell_H |y - \tilde{y}| - (\mu + \lambda) \ell_H (|\xi - \tilde{\xi}| \\ & \quad + |y - \tilde{y}|) - \mu \gamma \lambda \varepsilon - \ell_B (|v - \tilde{v}| + |y - \tilde{y}|) \\ & \geq (\mu \gamma \lambda - \mu^2 \ell_H - (\mu + \lambda) \ell_H - \ell_B (\mu + 1)) |y - \tilde{y}| \\ & \quad - ((\mu + \lambda) \ell_H + \ell_B) |\xi - \tilde{\xi}| - \mu \gamma \lambda \varepsilon. \end{aligned}$$

We recall that $\mu = \frac{2\ell_H}{\gamma} \geq 2$, and then $\mu \gamma \lambda - \lambda \ell_H = \frac{1}{2} \mu \gamma \lambda$. Also using $|\xi - \tilde{\xi}| \leq \sqrt{\varepsilon}$, it holds

$$\begin{aligned} & (\mu \gamma \lambda - \mu^2 \ell_H - (\mu + \lambda) \ell_H - \ell_B (\mu + 1)) |y - \tilde{y}| - ((\mu + \lambda) \ell_H + \ell_B) |\xi - \tilde{\xi}| - \mu \gamma \lambda \varepsilon \\ & \geq \mu \left(\frac{1}{2} \gamma \lambda - (\mu + 1) \ell_H - 2\ell_B \right) |y - \tilde{y}| - C \sqrt{\varepsilon} = \frac{1}{4} \mu \gamma \lambda |y - \tilde{y}| - C \sqrt{\varepsilon}, \end{aligned}$$

by the definition of λ in (3.6). According to $\mu \geq 2$ and $\lambda |y - \tilde{y}| \geq \rho - \sqrt{\varepsilon} \geq \rho_\varepsilon - \sqrt{\varepsilon}$, we conclude that

$$- (H(v, y) - H(\tilde{v}, \tilde{y})) \cdot \nabla_y \rho_\varepsilon + (B(v, y) - B(\tilde{v}, \tilde{y})) \cdot \nabla_v \rho_\varepsilon \geq \frac{\gamma}{2} \rho_\varepsilon - C \sqrt{\varepsilon}.$$

Inserting the above estimate in (3.37), we obtain

$$\begin{aligned} & (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi'(\rho_\varepsilon) \left\{ \frac{\gamma}{2} \rho_\varepsilon - C \sqrt{\varepsilon} \right\} \\ & \leq \psi(\rho_\varepsilon) K 2k_0 + \omega L \psi(\rho_\varepsilon) \\ & \leq \frac{\gamma}{4} \psi'(\rho_\varepsilon) \rho_\varepsilon L + \left(1 + \frac{\gamma}{4} \right) \omega L \psi(\rho_\varepsilon), \end{aligned}$$

in view of (3.31). Therefore, we infer that

$$\psi'(\rho_\varepsilon) \left\{ \frac{\gamma}{4} \rho_\varepsilon - C \sqrt{\varepsilon} \right\} \leq \left(1 + \frac{\gamma}{4} \right) \omega \psi(\rho_\varepsilon). \tag{3.38}$$

Finally, we can choose ω so that

$$\left(1 + \frac{\gamma}{4} \right) \omega < \frac{\gamma}{8} \theta e^{-C_2 R_1^\rho}, \tag{3.39}$$

in addition to (3.36). Since $\rho_\varepsilon \leq R_1$, this yields

$$\left(1 + \frac{\gamma}{4}\right) \omega \psi(\rho_\varepsilon) \leq \frac{\gamma}{8} \theta e^{-C_2 \rho_\varepsilon^\theta} \psi(\rho_\varepsilon) \leq \frac{\gamma}{8} \psi'(\rho_\varepsilon) \rho_\varepsilon,$$

where we have exploited $\psi(r) \leq C_2 r^\theta$ and (3.24). Then, by virtue of this inequality and (3.38), we achieve

$$\frac{\gamma}{8} \rho_\varepsilon \leq C \sqrt{\varepsilon}. \tag{3.40}$$

On the other hand, since $u(t)$ is Lipschitz continuous, with some constant ℓ_u^T , uniformly in $(0, T)$, by definition (3.12), and using (3.28) and (3.40), we have

$$\mathbf{m} \leq \mathbf{m}_\varepsilon \leq u(t, v, y) - u(t, \tilde{v}, \tilde{y}) \leq \ell_u^T (\rho_\varepsilon + C\varepsilon) \leq \ell_u^T (8C\gamma^{-1}\sqrt{\varepsilon} + C\varepsilon). \tag{3.41}$$

Letting $\varepsilon \rightarrow 0$, we reach a contradiction because $\mathbf{m} > 0$.

We have thus proved that (3.9) holds for K given by (3.27), with a choice of R_1, ω , and C_2 only depending on ℓ_B, ℓ_H, γ and the Lyapunov function φ . The constant L is finally chosen in a way that K , as defined by (3.27), satisfies (3.15). This means taking $L = n [u_0]_\theta$, for some $n \geq \frac{\omega_0 \varphi_{R_1}}{4(\omega + \tilde{C} \psi'(R_1) R_1)}$. Letting $\eta \rightarrow 0$ in (3.9), we get

$$\begin{aligned} u(t, v, y) - u(t, \tilde{v}, \tilde{y}) &\leq e^{-\omega t} (K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] + L) \psi(\rho) \\ &\leq e^{-\omega t} (K + L)[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \psi(\rho). \end{aligned}$$

Since $\psi(\rho) \leq (C_2 \rho^\theta) \wedge 1$ and ρ is equivalent to the Euclidean distance, we deduce that $[u(t)]_\theta \leq C e^{-\omega t} (K + L)$, according to (3.3). Recalling that K is proportional to L , where $L = n [u_0]_\theta$, we conclude with the desired estimate (3.5) (up to renaming K). Hence, the case $\theta \in (0, 1)$ is completed.

Step 4. We conclude by explaining how to get the result for the limiting case $\theta = 1$.

First of all, note that we just need to refine the previous estimate when ρ is small. So, we only consider the case that $\rho < \delta$, for a small $\delta < 1$ (to be fixed later). First, we modify the function ψ used before. To this end, we fix any $\theta < 1$, take a constant β (to be fixed later, depending on δ) and define

$$\tilde{\psi}(r) := r - \frac{\beta}{1 + \theta} r^{1+\theta}.$$

We suppose, for now, that

$$\beta \delta^\theta < \frac{1}{2}, \tag{3.42}$$

so that

$$\frac{1}{2} r \leq \tilde{\psi}(r) \leq r, \quad \frac{1}{2} \leq \tilde{\psi}'(r) \leq 1 \quad \forall r \leq \delta. \tag{3.43}$$

We claim that, for a suitable choice of ω and K , it holds

$$\begin{aligned} &u(t, v, y) - u(t, \tilde{v}, \tilde{y}) \\ &\leq e^{-\omega t} K [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \tilde{\psi}(\rho) + \frac{\eta}{T - t} \quad \forall t \in (0, T), \forall v, y, \tilde{v}, \tilde{y} \in \mathbb{R}^d : \rho < \delta, \end{aligned} \tag{3.44}$$

for every η sufficiently small. Here, φ and ρ are the same quantities used above, except that the parameters μ and λ in (3.6) might be chosen differently (as precised later), yet only depending on γ , ℓ_H , and ℓ_B . Following the same approach as before, (3.44) is proved arguing by contradiction. We suppose that

$$\mathbf{m} := \sup_{\rho < \delta} \left\{ u(t, v, y) - u(t, \tilde{v}, \tilde{y}) - e^{-\omega t} K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})]\tilde{\psi}(\rho) - \frac{\eta}{T-t} \right\} > 0.$$

If ω_θ and K_θ denote the constants given by (3.5), here we choose $\omega < \omega_\theta$ and K such that

$$K \geq 2 \max \left(1, \delta^{\theta-1} K_\theta \right) [u_0]_1.$$

Thus, according to this and (3.43), when $t = 0$, we have

$$u_0(v, y) - u_0(\tilde{v}, \tilde{y}) \leq [u_0]_1 [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \rho \leq K [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \tilde{\psi}(\rho) \quad \forall \rho \leq \delta,$$

and, if $\rho = \delta$, using what proved for $\theta < 1$,

$$\begin{aligned} u(t, v, y) - u(t, \tilde{v}, \tilde{y}) &\leq e^{-\omega t} K_\theta [u_0]_\theta [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \delta^\theta \\ &\leq e^{-\omega t} K_\theta \delta^\theta [u_0]_1 [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \\ &\leq e^{-\omega t} K \frac{\delta}{2} [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \\ &\leq e^{-\omega t} K [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \tilde{\psi}(\delta). \end{aligned}$$

Consequently, if \mathbf{m} is positive, then it can be attained neither at $t = 0$ nor for $\rho = \delta$. Excluding the case $t = T$ as before, it follows that \mathbf{m} is a local maximum. Then, we proceed exactly as in the first part of the proof. Specifically, we introduce the parameter ε and the corresponding \mathbf{m}_ε , which, for ε sufficiently small, is attained at some $t \in (0, T)$, and $v, y, \tilde{v}, \tilde{y}$ with $\rho_\varepsilon < \delta$. With the notation already used above, if $|\xi - \tilde{\xi}| > \sqrt{\varepsilon}$, then we employ the coupling by reflection in the ξ -variable and obtain the analogue of (3.29) (with $L = 0$ in the present choice), namely,

$$\begin{aligned} &-2\tilde{\psi}(\rho_\varepsilon)k_0 K - K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})]C \tilde{\psi}'(\rho_\varepsilon)\rho_\varepsilon \\ &\leq K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \left\{ 4\tilde{\psi}''(\rho_\varepsilon) \left(\frac{|\xi - \tilde{\xi}|}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \right)^2 + 4 \frac{\tilde{\psi}'(\rho_\varepsilon)}{|\xi - \tilde{\xi}|_\varepsilon + \varepsilon} \frac{\varepsilon^2}{(|\xi - \tilde{\xi}|_\varepsilon + \varepsilon)^2} \right\} \\ &\leq K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \left\{ 2\tilde{\psi}''(\rho_\varepsilon) + 4\sqrt{\varepsilon}\tilde{\psi}'(\rho_\varepsilon) \right\}, \end{aligned}$$

by virtue of (3.34) and the concavity of $\tilde{\psi}$ again. We recall that C only depends on γ, ℓ_H, ℓ_B , and ℓ_φ . Using (3.43), we deduce

$$0 \leq K[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \left\{ 2\tilde{\psi}''(\rho_\varepsilon) + 4\sqrt{\varepsilon} + (2k_0 + C)\rho_\varepsilon \right\},$$

whence, computing $\tilde{\psi}''$ and being $|\xi - \tilde{\xi}| > \sqrt{\varepsilon}$,

$$2\theta\beta\rho_\varepsilon^{\theta-1} \leq 4\sqrt{\varepsilon} + (2k_0 + C)\rho_\varepsilon \leq \tilde{C}\rho_\varepsilon,$$

for some \tilde{C} only depending on γ, ℓ_H, ℓ_B , and φ . We now choose $\beta = \frac{\tilde{C}}{\theta}\delta^{2-\theta}$. In this way, the previous inequality cannot occur for $\rho_\varepsilon \leq \delta$. Hence, the case $|\xi - \tilde{\xi}| > \sqrt{\varepsilon}$ is not possible. Observe that, with this choice of β , (3.42) is satisfied as long as δ is sufficiently small. It remains to deal with $|\xi - \tilde{\xi}| \leq \sqrt{\varepsilon}$. We reason as in Case 3 above. In particular, we estimate

$$\begin{aligned} & - (H(v, y) - H(\tilde{v}, \tilde{y})) \cdot \nabla_y \rho_\varepsilon + (B(v, y) - B(\tilde{v}, \tilde{y})) \cdot \nabla_v \rho_\varepsilon \\ & \geq (\mu\gamma\lambda - \mu^2\ell_H - (\mu + \lambda)\ell_H - \ell_B(\mu + 1))|y - \tilde{y}| \\ & \quad - ((\mu + \lambda)\ell_H + \ell_B)|\xi - \tilde{\xi}| - \mu\gamma\lambda\varepsilon \\ & \geq (\mu\gamma\lambda - \mu^2\ell_H - (\mu + \lambda)\ell_H - \ell_B(\mu + 1))|y - \tilde{y}| - C\sqrt{\varepsilon}, \end{aligned}$$

which yields, from $\rho_\varepsilon + \lambda\sqrt{\varepsilon} \geq \lambda|y - \tilde{y}| \geq \rho_\varepsilon - \sqrt{\varepsilon}$,

$$\begin{aligned} & - (H(v, y) - H(\tilde{v}, \tilde{y})) \cdot \nabla_y \rho_\varepsilon + (B(v, y) - B(\tilde{v}, \tilde{y})) \cdot \nabla_v \rho_\varepsilon \\ & \geq (\mu\gamma - \ell_H)\rho_\varepsilon - \frac{1}{\lambda}(\mu^2\ell_H + \mu\ell_H + \ell_B(\mu + 1))\rho_\varepsilon - C\sqrt{\varepsilon}, \end{aligned}$$

for a possibly different C , potentially depending on μ and λ . The analogue of (3.37), in this case, leads to

$$\begin{aligned} (\mu\gamma - \ell_H)\rho_\varepsilon - \frac{1}{\lambda}(\mu^2\ell_H + \mu\ell_H + \ell_B(\mu + 1))\rho_\varepsilon & \leq 2\frac{\psi(\rho_\varepsilon)}{\psi'(\rho_\varepsilon)}k_0 + C\sqrt{\varepsilon} \\ & \leq 4\rho_\varepsilon k_0 + C\sqrt{\varepsilon}, \end{aligned}$$

also using (3.43). Now, we fix the values of μ and λ . First, we choose $\mu = \frac{4k_0 + 2\ell_H}{\gamma}$ so that the previous inequality yields

$$\ell_H\rho_\varepsilon - \frac{1}{\lambda}(\mu^2\ell_H + \mu\ell_H + \ell_B(\mu + 1))\rho_\varepsilon \leq C\sqrt{\varepsilon}.$$

Next, we choose λ large enough to reach that $\rho_\varepsilon \leq \hat{C}\sqrt{\varepsilon}$, for some different \hat{C} . From here, we obtain (3.41) and we conclude the contradiction argument as before. \square

Remark 3.3 The exponential rate ω in (3.5) heavily depends on the Lyapunov function φ , which makes it hard to provide explicit quantifications. We just observe that, in the above proof, we have $C_2 \simeq R_1^{2-\theta}$ from (3.35), hence the conditions (3.36) and (3.39) are essentially equivalent. Together with (3.26), these will impose some condition on ω in terms of ω_0 (appearing in (3.1)) and R_1 , which also depends on φ . Typically, the selection of the Lyapunov function involves conditions depending on the dimension d , the confining potentials, and, possibly, the diffusion strength. However, even in the

simplest case of the confined Kolmogorov equation, where the Lyapunov function is sufficiently explicit (cf. Sect. 5.1), our method does not likely provide sharp exponential rates, since it also depends on the auxiliary function ψ , which is just a simple choice of a concave bounded function and is not meant to be optimal at all.

For the interested reader, we refer to [15] for a detailed discussion of the exponential rate that could come out from similar coupling methods, though this is quite involved even in model examples. Our purpose was actually rather opposite, namely, to give a quite general approach to exponential decay, aimed at covering a broader class of equations, at the expense of any sharpness of the rate.

Remark 3.4 The approach developed in the above proof can be readily adapted to slightly different settings. For instance, if the state variable y belongs to a compact space, then it is enough to use a Lyapunov function φ only depending on the v -variable.

In particular, a similar result to Theorem 3.2 can be proved if $y \in \mathbb{T}^d$, where $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ is the flat d -dimensional torus. In this case, we consider solutions u of (2.4) such that $u(t, v, y)$ is \mathbb{Z}^d -periodic in y , and we suppose the existence of a Lyapunov function $\varphi = \varphi(v)$ (or alternatively some $\varphi(v, y)$, periodic in y), which satisfies (3.1) as $|v| \rightarrow \infty$. In order to follow the above proof, we simply remark that, using the periodicity of u and the increasing character of ψ , we still have that

$$\sup_{t, (v, y), (\tilde{v}, \tilde{y})} \Phi := \left\{ u(t, v, y) - u(t, \tilde{v}, \tilde{y}) - e^{-\omega t} (K[\varphi(v) + \varphi(\tilde{v})] + L) \psi(\rho) - \frac{\eta}{T - t} \right\} \tag{3.45}$$

is actually a maximum attained at some point.

Indeed, for any $y \in \mathbb{R}^d$, denote by $[y] \in \mathbb{Z}^d$ the period such that $y - [y] \in Q_1 := [0, 1]^d$, which is the reference unit cube, and set $\hat{y} := y - [y]$. Observe that

$$|v - \tilde{v} + \mu(y - \tilde{y})| + \lambda|y - \tilde{y}| \geq |v - \tilde{v} + \mu(\hat{y} - \hat{\tilde{y}})| + \lambda|\hat{y} - \hat{\tilde{y}}| + (\lambda - \mu)|y - \tilde{y}| - (\lambda + \mu)c_d,$$

for some constant c_d only depending on the dimension. Hence, if we set $C_{\lambda, \mu} = \frac{(\lambda + \mu)c_d}{\lambda - \mu}$, we have

$$\|(v - \tilde{v}, y - \tilde{y})\| \geq \|(v - \tilde{v}, \hat{y} - \hat{\tilde{y}})\| \quad \text{whenever } |y - \tilde{y}| \geq C_{\lambda, \mu}.$$

Then, due to the periodicity of u , and since $\psi(\cdot)$ is increasing, we deduce that $\Phi(t, v, y, \tilde{v}, \tilde{y}) \leq \Phi(t, v, \hat{y}, \tilde{v}, \hat{\tilde{y}})$ whenever $|y - \tilde{y}| \geq C_{\lambda, \mu}$. This means that the sup in (3.45) is restricted to the set $|y - \tilde{y}| \leq C_{\lambda, \mu}$ (note that we can assume $C_{\lambda, \mu} \geq \text{diam}(Q_1)$ up to increasing c_d). In addition, exactly as in (3.11), we have

$$\limsup_{|v| \vee |\tilde{v}| \rightarrow \infty} \left\{ u(t, v, y) - u(t, \tilde{v}, \tilde{y}) - e^{-\omega t} (K[\varphi(v) + \varphi(\tilde{v})] + L) \psi(\rho) - \frac{\eta}{T - t} \right\} \leq 0, \tag{3.46}$$

using the Lyapunov function φ , and the facts that u is Lipschitz and $u = o(\varphi)$ as $|v| \rightarrow \infty$. The above two conditions ensure that, if the sup in (3.45) is positive, then any maximizing sequence is relatively compact, up to normalization in y . Namely, suppose that

$$\Phi(t_n, v_n, y_n, \tilde{v}_n, \tilde{y}_n) \rightarrow \sup_{t, (v, y), (\tilde{v}, \tilde{y})} \Phi > 0.$$

Being $|y_n - \tilde{y}_n| \leq C_{\lambda, \mu}$, up to subsequences we can assume that $\hat{y}_n \rightarrow y_0$ and $y_n - \tilde{y}_n \rightarrow w_0$, and it holds

$$\begin{aligned} \Phi(t_n, v_n, y_n, \tilde{v}_n, \tilde{y}_n) &= u(t_n, v_n, \hat{y}_n) - u(t_n, \tilde{v}_n, \tilde{y}_n - [y_n]) \\ &\quad - e^{-\omega t_n} (K[\varphi(v_n) + \varphi(\tilde{v}_n)] + L) \psi(\rho_n) - \frac{\eta}{T - t_n}, \end{aligned}$$

where $\rho_n = \|(v_n - \tilde{v}_n, y_n - \tilde{y}_n)\|$. We also know that v_n, \tilde{v}_n are bounded due to (3.46), and so relatively compact, as well as ρ_n . Since $\hat{y}_n \rightarrow y_0, \tilde{y}_n - [y_n] \rightarrow y_0 - w_0$, and t_n converges up to subsequences, we conclude that Φ achieves a maximum at some point $(t_0, v, y_0, \tilde{v}, y_0 - w_0)$. Once we have justified the existence of a maximum point for Φ , the rest of the proof of Theorem 3.2 remains unchanged, with the minor difference that the case where ρ is large is now reduced to having $|v - \tilde{v}|$ large, which explains why confining is only needed in the v -variable.

We conclude by mentioning that another extension, which is doable up to suitable modifications, could be achieved if the variable y belongs to a space with a different dimension, say $y \in \mathbb{R}^k, k < d$. There, we need to change the scalar μ to a suitable matrix $M \in \mathcal{M}_{k, n}$ in the definition of the rotated norm, and the nondegeneracy condition on $D_v H$ to a full-rank condition. This variant is to be better analyzed within the more general picture of degenerate drift-diffusion operators under Hormander’s condition.

3.2 Decay of Wasserstein distances for the kinetic Fokker–Planck equation

An immediate corollary is a decay result in Wasserstein distances for solutions of (2.2). Let us introduce the following weighted distance for $\mu, \nu \in \mathcal{M}_\varphi(\mathbb{R}^{2d}) \cap \mathcal{P}_1(\mathbb{R}^{2d})$:

$$d_{1, \varphi}(\mu, \nu) := \sup \left\{ \int_{\Omega} \zeta d(\mu - \nu), \quad \zeta \in L^\infty(\varphi^{-1}) : [\zeta]_1 \leq 1 \right\},$$

where $[\cdot]_1$ is defined in (3.3). Then, for every $\zeta \in L^\infty(\varphi^{-1})$, one can estimate

$$\int_{\Omega} \zeta d(\mu - \nu) \leq [\zeta]_1 d_{1, \varphi}(\mu, \nu). \tag{3.47}$$

Note that if φ has at least linear growth at infinity, then we have

$$\frac{|u(\chi) - u(\tilde{\chi})|}{(\varphi(\chi) + \varphi(\tilde{\chi}))(|\chi - \tilde{\chi}| \wedge 1)} \leq C \frac{|u(\chi) - u(\tilde{\chi})|}{|\chi - \tilde{\chi}|}, \quad \chi = (v, y), \quad \tilde{\chi} = (\tilde{v}, \tilde{y}),$$

for some $C > 0$ independent of u . Hence, we get

$$\mathcal{W}_1(\mu, \nu) \leq C d_{1,\varphi}(\mu, \nu), \tag{3.48}$$

where \mathcal{W}_1 is the standard 1-Wasserstein distance, recall (2.1). In particular, in this case, the decay of $d_{1,\varphi}$ implies the decay of \mathcal{W}_1 .

Theorem 3.5 *Under the assumptions of Theorem 3.2, let μ_1, μ_2 be two solutions of (2.2) corresponding to initial data $\mu_{01}, \mu_{02} \in \mathcal{P}_1(\mathbb{R}^{2d}) \cap \mathcal{M}_\varphi(\mathbb{R}^{2d})$. Then, there exist $K, \omega > 0$ such that*

$$d_{1,\varphi}(\mu_1(t), \mu_2(t)) \leq K e^{-\omega t} d_{1,\varphi}(\mu_{01}, \mu_{02}). \tag{3.49}$$

In particular, if φ grows at least linearly at infinity, then

$$\mathcal{W}_1(\mu_1(t), \mu_2(t)) \leq C K e^{-\omega t} d_{1,\varphi}(\mu_{01}, \mu_{02})$$

for some $C > 0$.

Proof Let $\zeta \in L^\infty(\varphi^{-1})$ be such that $[\zeta]_1 \leq 1$. This means that ζ is a locally Lipschitz function and (see Corollary 4.6) there exists a viscosity solution u of the backward problem

$$\begin{cases} -\partial_t u + \mathcal{L}[u] = 0 & \text{in } (0, t) \times \mathbb{R}^{2d}, \\ u(t) = \zeta & \text{in } \mathbb{R}^{2d}. \end{cases}$$

By (2.8) applied to μ_1 and μ_2 , it holds

$$\int_{\Omega} \zeta d(\mu_1(t) - \mu_2(t)) = \int_{\Omega} u(0) d(\mu_{01} - \mu_{02}) \leq K e^{-\omega t} d_{1,\varphi}(\mu_{01}, \mu_{02}),$$

where we have used (3.47) and (3.5) (applied backwards, and with $\theta = 1$) to estimate $[u(0)]_1$. This gives (3.49). The last statement immediately follows from (3.48). \square

We conclude by pointing out that Theorem 3.5 provides the existence (and uniqueness) of a stationary measure \bar{m} solving (2.2) and attracting all other solutions in long time. As detailed in [26, Section 5], the exponential decay rate enables building such a solution, whose uniqueness and exponential stability follow from the same contractivity estimate.

4 Hypocoercivity and L^∞ -Lipschitz smoothing

In this section, we prove short-time smoothing estimates for the problem (2.4). To work in a slightly broader framework, we consider the operator

$$\mathcal{L}(u) := -\text{tr}(A(v)D_{vv}^2u) - \kappa \Delta_y u - H(v, y) \cdot \nabla_y u + B(v, y) \cdot \nabla_v u,$$

where $A(v)$, $H(v, y)$, and $B(v, y)$ are C^1 -functions, globally Lipschitz continuous, and $\kappa \geq 0$. In particular, we assume that $A(v)$ is a bounded and elliptic matrix, i.e.,

$$\exists \alpha_0 > 0 : \quad \alpha_0 I_d \leq A(v) \leq \frac{1}{\alpha_0} I_d, \tag{4.1}$$

and also Lipschitz continuous:

$$|A(v) - A(\tilde{v})| \leq \ell_A |v - \tilde{v}| \quad \forall v, \tilde{v} \in \mathbb{R}^d. \tag{4.2}$$

We start with a computational lemma.

Lemma 4.1 *If u is a smooth solution of*

$$\partial_t u + \mathcal{L}(u) = 0,$$

then we have:

(i)

$$\frac{1}{2}(\partial_t + \mathcal{L})(u^2) = -A(v)\nabla_v u \cdot \nabla_v u - \kappa |\nabla_y u|^2;$$

(ii)

$$\begin{aligned} \frac{1}{2}(\partial_t + \mathcal{L})(|\nabla_v u|^2) &= -\sum_k A(v)\nabla_v u_{v_k} \cdot \nabla_v u_{v_k} + \sum_{ijk} (a_{ij})_{v_k} u_{v_k} u_{v_i} u_{v_j} \\ &\quad + D_v H \nabla_y u \cdot \nabla_v u - D_v B \nabla_v u \cdot \nabla_v u - \kappa \sum_{ij} u_{v_i}^2 u_{v_j}; \end{aligned}$$

(iii)

$$\begin{aligned} (\partial_t + \mathcal{L})(\nabla_v u \cdot \nabla_y u) &= \sum_{ijk} ((a_{ij})_{v_k} u_{y_k} u_{v_i} u_{v_j} - 2a_{ij} u_{v_i} u_{v_k} u_{y_k} u_{v_j}) \\ &\quad - 2\kappa \sum_{ij} u_{v_i} u_{y_j} u_{y_i} u_{y_j} + D_v H \nabla_y u \cdot \nabla_y u \\ &\quad + D_y H \nabla_y u \cdot \nabla_v u - D_v B \nabla_v u \cdot \nabla_y u - D_y B \nabla_v u \cdot \nabla_v u; \end{aligned}$$

(iv)

$$\begin{aligned} \frac{1}{2}(\partial_t + \mathcal{L})\left(|\nabla_y u|^2\right) &= -\sum_k A(v)\nabla_v u_{y_k} \cdot \nabla_v u_{y_k} + D_y H \nabla_y u \cdot \nabla_y u \\ &\quad - D_y B \nabla_v u \cdot \nabla_y u - \kappa \sum_{ij} u_{y_i y_j}^2. \end{aligned}$$

Proof Equality (i) immediately follows from the chain rule and the equation $\partial_t u + \mathcal{L}(u) = 0$. To show (ii), we denote the second-order terms in the operator \mathcal{L} by $\mathcal{L}_0(u) := -\text{tr}(A(v)D_{vv}^2 u) - \kappa \Delta_y u$. Then, we have

$$\begin{aligned} \frac{1}{2}(\partial_t + \mathcal{L}_0)\left(|\nabla_v u|^2\right) &= \nabla_v((\partial_t + \mathcal{L}_0)(u)) \cdot \nabla_v u \\ &\quad + \sum_{ijk} \left((a_{ij})_{v_k} u_{v_k} u_{v_i v_j} - a_{ij} u_{v_k v_i} u_{v_k v_j} \right) - \kappa \sum_{ij} u_{v_i y_j}^2 \\ &= \nabla_v \left(H(v, y) \cdot \nabla_y u - B(v, y) \cdot \nabla_v u \right) \cdot \nabla_v u \\ &\quad + \sum_{ijk} \left((a_{ij})_{v_k} u_{v_k} u_{v_i v_j} - a_{ij} u_{v_k v_i} u_{v_k v_j} \right) - \kappa \sum_{ij} u_{v_i y_j}^2, \end{aligned}$$

in view of $\partial_t u + \mathcal{L}(u) = 0$ again. Developing the drift terms, we reach (ii).

Similarly, it holds

$$\frac{1}{2}(\partial_t + \mathcal{L}_0)\left(|\nabla_y u|^2\right) = \nabla_y((\partial_t + \mathcal{L}_0)(u)) \cdot \nabla_y u - \sum_{ijk} a_{ij} u_{y_k v_i} u_{y_k v_j} - \kappa \sum_{ij} u_{y_i y_j}^2,$$

and we obtain (iv) using once more the equation for u .

Finally, to get (iii) we compute

$$\begin{aligned} (\partial_t + \mathcal{L}_0)(\nabla_v u \cdot \nabla_y u) &= \nabla_v((\partial_t + \mathcal{L}_0)(u)) \cdot \nabla_y u + \nabla_y((\partial_t + \mathcal{L}_0)(u)) \cdot \nabla_v u \\ &\quad + \sum_{ijk} \left((a_{ij})_{v_k} u_{y_k} u_{v_i v_j} - 2a_{ij} u_{v_i v_k} u_{v_j y_k} \right) - 2\kappa \sum_{ij} u_{v_i y_j} u_{y_i y_j} \\ &= \nabla_v \left(H(v, y) \cdot \nabla_y u - B(v, y) \cdot \nabla_v u \right) \cdot \nabla_y u + \nabla_y \left(H(v, y) \cdot \nabla_y u - B(v, y) \cdot \nabla_v u \right) \cdot \nabla_v u \\ &\quad + \sum_{ijk} \left((a_{ij})_{v_k} u_{y_k} u_{v_i v_j} - 2a_{ij} u_{v_i v_k} u_{v_j y_k} \right) - 2\kappa \sum_{ij} u_{v_i y_j} u_{y_i y_j} \\ &= H(v, y) \cdot \nabla_y(\nabla_v u \cdot \nabla_y u) - B(v, y) \cdot \nabla_v(\nabla_v u \cdot \nabla_y u) + D_v H \nabla_y u \cdot \nabla_y u + D_y H \nabla_y u \cdot \nabla_v u \\ &\quad - D_v B \nabla_v u \cdot \nabla_y u - D_y B \nabla_v u \cdot \nabla_v u + \sum_{ijk} \left((a_{ij})_{v_k} u_{y_k} u_{v_i v_j} - 2a_{ij} u_{v_i v_k} u_{v_j y_k} \right) \\ &\quad - 2\kappa \sum_{ij} u_{v_i y_j} u_{y_i y_j}. \end{aligned}$$

□

Exploiting the above lemma, we deduce the short-time smoothing effect of Kolmogorov-type operators. We actually employ an L^∞ -version of Villani’s hypocoercivity estimates, obtained via the classical Bernstein method.

Proposition 4.2 *Let $u \in L^\infty((0, T); W^{1,\infty}(\mathbb{R}^{2d}))$ be a smooth solution of the equation*

$$\begin{cases} \partial_t u - \text{tr}(A(v)D_{vv}^2 u) - \kappa \Delta_y u - H(v, y) \cdot \nabla_y u + B(v, y) \cdot \nabla_v u = 0 \\ u(0) = u_0, \end{cases}$$

where $\kappa \geq 0$, $A(v)$ satisfies (4.1)–(4.2), $B(v, y)$ fulfils (2.5), and $H(v, y)$ satisfies (2.6)–(2.7).

Then, there exists a constant $C > 0$, depending on $\alpha_0, \ell_A, \ell_B, \ell_H$ and γ , such that

$$\|\nabla_v u(t)\|_\infty + t \|\nabla_y u(t)\|_\infty \leq C \frac{\|u_0\|_\infty}{\sqrt{t}} \quad \forall t \in (0, 1]. \tag{4.3}$$

Proof We introduce the function

$$W := \frac{1}{2} \left(\lambda u^2 + t |\nabla_v u|^2 - 2\varepsilon t^2 \nabla_v u \cdot \nabla_y u + \delta t^3 |\nabla_y u|^2 \right), \tag{4.4}$$

where λ is a large constant, to be fixed later, and ε, δ satisfy (at least) the condition

$$\varepsilon^2 < \delta, \tag{4.5}$$

so that $W > 0$. The goal is to show that W is a subsolution (in short time).

From Lemma 4.1-(ii), using (2.5) and the ellipticity of $A(v)$, see (4.1), we have

$$\begin{aligned} \frac{1}{2}(\partial_t + \mathcal{L})(t |\nabla_v u|^2) &\leq \left(\frac{1}{2} + t \ell_B\right) |\nabla_v u|^2 - t \alpha_0 \sum_{ik} u_{v_i v_k}^2 + t \sum_{ijk} (a_{ij})_{v_k} u_{v_k} u_{v_i v_j} \\ &\quad + t D_v H \nabla_y u \cdot \nabla_v u - t \kappa \sum_{ij} u_{v_i v_j}^2 \\ &\leq \left(\frac{1}{2} + t(\ell_B + C_A)\right) |\nabla_v u|^2 - t \frac{\alpha_0}{2} \sum_{ik} u_{v_i v_k}^2 \\ &\quad + t \ell_H |\nabla_v u| |\nabla_y u| - t \kappa \sum_{ij} u_{v_i v_j}^2, \end{aligned} \tag{4.6}$$

thanks to (4.2), (2.7), and Young’s inequality. Hereafter, we denote by C_A possibly different constants depending on the $W^{1,\infty}$ -bound of A , namely, on α_0 and ℓ_A .

Similarly, according to Lemma 4.1-(iii), by virtue of (4.1)–(4.2) and Young’s inequality again, it holds

$$\begin{aligned}
 (\partial_t + \mathcal{L}) \left(-\varepsilon t^2 \nabla_v u \cdot \nabla_y u \right) &\leq t \frac{\alpha_0}{4} \sum_{ij} u_{v_i v_j}^2 + C_A \varepsilon^2 t^3 |\nabla_y u|^2 + C_A \varepsilon^2 t^3 \sum_k |\nabla_v u_{y_k}|^2 \\
 &\quad - \varepsilon t^2 (D_v H \nabla_y u \cdot \nabla_y u + D_y H \nabla_y u \cdot \nabla_v u) + \varepsilon t^2 (D_v B \nabla_v u \cdot \nabla_y u + D_y B \nabla_v u \cdot \nabla_v u) \\
 &\quad + 2\kappa \varepsilon t^2 \sum_{ij} u_{v_i y_j} u_{y_i y_j} - 2\varepsilon t \nabla_v u \cdot \nabla_y u.
 \end{aligned}$$

The terms with H and B can be estimated by (2.6)–(2.7) and (2.5). So, we get

$$\begin{aligned}
 (\partial_t + \mathcal{L}) \left(-\varepsilon t^2 \nabla_v u \cdot \nabla_y u \right) &\leq t \frac{\alpha_0}{4} \sum_{ij} u_{v_i v_j}^2 + C_A \varepsilon^2 t^3 |\nabla_y u|^2 + C_A \varepsilon^2 t^3 \sum_k |\nabla_v u_{y_k}|^2 \\
 &\quad - \varepsilon \frac{\gamma}{2} t^2 |\nabla_y u|^2 + C_{B,H} \varepsilon t^2 |\nabla_v u|^2 + 2\kappa \varepsilon t^2 \sum_{ij} u_{v_i y_j} u_{y_i y_j} - 2\varepsilon t \nabla_v u \cdot \nabla_y u,
 \end{aligned}$$

where $C_{B,H}$ only depends on the constants ℓ_B, ℓ_H, γ . Finally, in view of Lemma 4.1-(iv), we also have

$$\begin{aligned}
 \frac{1}{2} (\partial_t + \mathcal{L}) \left(\delta t^3 |\nabla_y u|^2 \right) &\leq -\delta t^3 \alpha_0 \sum_k |\nabla_v u_{y_k}|^2 + \delta t^3 \left(\ell_B |\nabla_v u| |\nabla_y u| + \ell_H |\nabla_y u|^2 \right) \\
 &\quad - \kappa \delta t^3 \sum_{ij} u_{y_i y_j}^2 + \frac{3}{2} \delta t^2 |\nabla_y u|^2.
 \end{aligned}$$

We now add up the previous inequalities, gathering together the similar terms. Hence, we obtain

$$\begin{aligned}
 (\partial_t + \mathcal{L}) \left(W - \frac{\lambda}{2} u^2 \right) &\leq \left(\frac{1}{2} + t (\ell_B + C_A) \right) |\nabla_v u|^2 - t \frac{\alpha_0}{4} \sum_{ik} u_{v_i v_k}^2 + t \ell_H |\nabla_v u| |\nabla_y u| \\
 &\quad - t^2 |\nabla_y u|^2 \left(\varepsilon \frac{\gamma}{2} - \frac{3}{2} \delta - C_A \varepsilon^2 t \right) - t^3 \left(\delta \alpha_0 - C_A \varepsilon^2 \right) \sum_k |\nabla_v u_{y_k}|^2 - 2\varepsilon t \nabla_v u \cdot \nabla_y u \\
 &\quad + C_{B,H} \varepsilon t^2 |\nabla_v u|^2 + \delta t^3 \left(\ell_B |\nabla_v u| |\nabla_y u| + \ell_H |\nabla_y u|^2 \right) \\
 &\quad + \kappa t \left(2\varepsilon t \sum_{ij} u_{v_i y_j} u_{y_i y_j} - \delta t^2 \sum_{ij} u_{y_i y_j}^2 - \sum_{ij} u_{v_i y_j}^2 \right).
 \end{aligned}$$

Recall that $t \leq 1$ and $\varepsilon, \delta \leq 1$. Moreover, thanks to (4.5), the last term is negative and can be dropped. Then, applying Young’s inequality, we infer that there exists $K > 0$

(only depending on $\alpha_0, \ell_A, \ell_B, \ell_H$, and γ) such that

$$\begin{aligned}
 (\partial_t + \mathcal{L}) \left(W - \frac{\lambda}{2} u^2 \right) &\leq K |\nabla_v u|^2 + (\ell_H + 2\varepsilon)t |\nabla_v u| |\nabla_y u| \\
 - t^2 \left(\varepsilon \frac{\gamma}{2} - \frac{3}{2} \delta - C_A \varepsilon^2 t - \ell_B \delta^2 t - \ell_H \delta t \right) &|\nabla_y u|^2 \\
 - t^3 \left(\delta \alpha_0 - C_A \varepsilon^2 \right) \sum_k &|\nabla_v u_{y_k}|^2,
 \end{aligned}$$

which yields, because $(\ell_H + 2\varepsilon)t |\nabla_v u| |\nabla_y u| \leq \frac{(\ell_H + 2\varepsilon)^2}{\varepsilon \gamma} |\nabla_v u|^2 + t^2 \varepsilon \frac{\gamma}{4} |\nabla_y u|^2$,

$$\begin{aligned}
 (\partial_t + \mathcal{L}) \left(W - \frac{\lambda}{2} u^2 \right) &\leq \left(K + \frac{(\ell_H + 2\varepsilon)^2}{\varepsilon \gamma} \right) |\nabla_v u|^2 \\
 - t^2 \left(\varepsilon \frac{\gamma}{4} - \frac{(3 + 2\ell_H)}{2} \delta - C_A \varepsilon^2 - \ell_B \delta^2 \right) &|\nabla_y u|^2 \quad (4.7) \\
 - t^3 \left(\delta \alpha_0 - C_A \varepsilon^2 \right) \sum_k &|\nabla_v u_{y_k}|^2,
 \end{aligned}$$

where we have also used $t \leq 1$ inside the second term on the right. We remark that the last two terms are negative as long as δ, ε are sufficiently small with $2(3 + 2\ell_H)\delta < \gamma \varepsilon$ and $\delta \alpha_0 > C_A \varepsilon^2$. We fix, e.g., $\delta = \frac{\gamma \varepsilon}{4(3 + 2\ell_H)}$ and take ε small enough, ensuring that (4.5) is also satisfied. Then, being $\delta = O(\varepsilon)$, it holds

$$-t^2 \left(\varepsilon \frac{\gamma}{4} - \frac{(3 + 2\ell_H)}{2} \delta - C_A \varepsilon^2 - \ell_B \delta^2 \right) \leq -t^2 \left(\varepsilon \frac{\gamma}{8} + O(\varepsilon^2) \right),$$

and the right-hand side is negative choosing ε sufficiently small, only depending on C_A, ℓ_B and γ . Similarly for the last term in (4.7), since $\delta = O(\varepsilon)$ implies $\delta \alpha_0 > C_A \varepsilon^2$ for small ε . Therefore, we achieve

$$(\partial_t + \mathcal{L}) \left(W - \frac{\lambda}{2} u^2 \right) \leq \left(K + \frac{(\ell_H + 2\varepsilon)^2}{\varepsilon \gamma} \right) |\nabla_v u|^2,$$

and, recalling Lemma 4.1-(i), we reach

$$(\partial_t + \mathcal{L}) W \leq - \left(\lambda \alpha_0 - \left(K + \frac{(\ell_H + 2\varepsilon)^2}{\varepsilon \gamma} \right) \right) |\nabla_v u|^2 \quad \forall t \in (0, 1].$$

Here, we choose λ sufficiently large so that $(\partial_t + \mathcal{L}) W \leq 0$ for $t \leq 1$. Now, W is a bounded subsolution (since $u \in L^\infty((0, T); W^{1,\infty}(\mathbb{R}^d))$), so we can apply the maximum principle to get

$$\sup_{t \in (0, 1]} W(t) \leq \|W(0)\|_\infty \leq \frac{\lambda}{2} \|u_0\|_\infty^2.$$

In view of (4.5), we have $W \geq c_\varepsilon (t|\nabla_v u(t)|^2 + t^3|\nabla_y u(t)|^2)$, from which estimate (4.3) follows. □

Remark 4.3 At different steps, one may use some dissipativity property of $D_v B, D_y B, D_y H$ (which enter in (ii)–(iv) of Lemma 4.1) in an attempt to improve the above estimates. However, here we are just exploiting the smoothing effect in *short-time*, so this seems unnecessary. In parallel, it could be interesting to understand under what conditions on the drift terms an estimate such as (4.3) can hold globally in time.

We now observe that a weighted version of Proposition 4.2 can be obtained similarly.

Proposition 4.4 *Let $u \in L^\infty((0, T); W^{1,\infty}(\mathbb{R}^{2d}))$ be a smooth solution of the equation*

$$\begin{cases} \partial_t u - \text{tr}(A(v)D_{vv}^2 u) - \kappa \Delta_y u - H(v, y) \cdot \nabla_y u + B(v, y) \cdot \nabla_v u = 0 \\ u(0) = u_0, \end{cases}$$

where $A(v)$ satisfies (4.1)–(4.2), $B(v, y)$ fulfils (2.5), and $H(v, y)$ satisfies (2.6)–(2.7), and $\kappa \geq 0$. Assume there exists some function $\varphi \in C^{1,2}([0, T] \times \mathbb{R}^{2d})$ such that $\varphi \geq 1$, $\varphi(t, v, y) \xrightarrow{|(v,y)| \rightarrow \infty} \infty$ uniformly for $t \in (0, T)$, and satisfies

$$(\partial_t + \mathcal{L})(\varphi^2) \geq -k_0, \quad t \in (0, 1), \quad v, y \in \mathbb{R}^d, \tag{4.8}$$

for some $k_0 > 0$. Then, we have $\|u(t)\|_{L^\infty(\varphi^{-1})} \leq e^{\frac{k_0}{2}t} \|u_0\|_{L^\infty(\varphi_0^{-1})}$ for any $t \in (0, 1]$, and there exists $C > 0$, depending on $\alpha_0, \ell_A, \ell_B, \ell_H, \gamma$, and k_0 , such that

$$|\nabla_v u(t, v, y)| + t |\nabla_y u(t, v, y)| \leq C \varphi(t, v, y) \frac{\|u_0\|_{L^\infty(\varphi_0^{-1})}}{\sqrt{t}} \quad \forall t \in (0, 1], \quad v, y \in \mathbb{R}^d, \tag{4.9}$$

where $\varphi_0(\cdot) = \varphi(0, \cdot)$.

Remark 4.5 In this result, the Lyapunov function φ may also depend on time, as we are only concerned with short-time estimates. Indeed, given the linear growth of the drift fields H and B , the function $\varphi = e^{\lambda t} (\langle v \rangle^k + M)$, with $\langle v \rangle = \sqrt{1 + |v|^2}$, satisfies (4.8) and the above conditions, provided that $0 < k \leq 2$ and λ, M are chosen sufficiently large.

Proof Take $L := \|u_0\|_{L^\infty(\varphi_0^{-1})}$, which means that $L = \sup \left(\frac{|u_0|}{\varphi_0} \right)$. We claim that

$$e^{-k_0 t} W - \frac{\lambda}{2} L^2 \varphi^2 \leq 0 \quad \text{for } t \in (0, 1],$$

where W is the function (4.4) built in Proposition 4.2. Of course, the above inequality is true at $t = 0$ by definition of L and W . Next, we point out that

$$\begin{aligned} (\partial_t + \mathcal{L}) \left(e^{-k_0 t} W - \frac{\lambda}{2} L^2 \varphi^2 \right) &= -k_0 e^{-k_0 t} W + e^{-k_0 t} (\partial_t + \mathcal{L}) W - \frac{\lambda}{2} L^2 (\partial_t + \mathcal{L}) \varphi^2 \\ &\leq -k_0 e^{-k_0 t} W + \frac{\lambda}{2} L^2 k_0, \end{aligned}$$

where we have used the subsolution property of W and (4.8). If $e^{-k_0 t} W - \frac{\lambda}{2} L^2 \varphi^2$ has a positive maximum, then we have $e^{-k_0 t} W > \frac{\lambda}{2} L^2 \varphi^2 \geq \frac{\lambda}{2} L^2$ at a maximum point, which yields

$$(\partial_t + \mathcal{L})\left(e^{-k_0 t} W - \frac{\lambda}{2} L^2 \varphi^2\right) \leq -k_0 e^{-k_0 t} W + \frac{\lambda}{2} L^2 k_0 < 0$$

there. This violates the maximum principle, showing that $e^{-k_0 t} W - \frac{\lambda}{2} L^2 \varphi^2 \leq 0$ for $t \in (0, 1]$. Hence, $W \leq e^{k_0 t} \frac{\lambda}{2} L^2 \varphi^2$, which gives the desired conclusions according to the definition of W . \square

To conclude this section, we note that the above estimate allows us to extend the well-posedness result of Theorem 2.3 to possibly unbounded initial data u_0 .

Corollary 4.6 *Assume that (2.5) and (2.6)–(2.7) hold, and $\kappa \geq 0, T > 0$. Let $\varphi(t, v, y)$ be a function satisfying the assumptions of Proposition 4.4. Given $u_0 \in C(\mathbb{R}^{2d}) \cap L^\infty(\varphi_0^{-1})$, there exists a unique viscosity solution $u \in C([0, T] \times \mathbb{R}^{2d}) \cap L^\infty(\varphi^{-1})$ of the problem (2.4). Moreover, for every $t > 0$, $u(t)$ is locally Lipschitz continuous and fulfils estimate (4.9).*

Proof The existence follows by approximation. Precisely, if $u_0 := \xi \varphi_0$, for some bounded ξ , then we consider the viscosity solutions u_n corresponding to the initial data $u_{0n} = \xi(\varphi_0 \wedge n)$. By Proposition 4.4, we have

$$\|u_n(t)\|_{L^\infty(\varphi^{-1})} \leq C_T \|u_{0n}\|_{L^\infty(\varphi_0^{-1})} \leq C_T \|u_0\|_{L^\infty(\varphi_0^{-1})},$$

and then, by (4.9), we deduce uniform bounds for $\sqrt{t} \|\nabla_v u_n(t)\|_{L^\infty(\varphi^{-1})}$ and $t^{\frac{3}{2}} \|\nabla_y u_n(t)\|_{L^\infty(\varphi^{-1})}$. Henceforth, one proceeds as in the proof of Proposition 6.1 to deduce the compactness of u_n (in local uniform topology) and the convergence of a subsequence towards a viscosity solution u . The uniqueness follows as in Proposition 6.1, replacing $|v|^2$ with φ^2 and using the supersolution property (4.8). \square

5 Improved long-time decay

Thanks to Proposition 4.4, we can now improve the long-time decay estimate of Theorem 3.2.

Theorem 5.1 *Assume that (2.5) and (2.6)–(2.7) hold, and $\kappa \geq 0$. Let $\varphi(v, y)$ be a Lyapunov function (cf. Definition 3.1) fulfilling (3.4) and (4.8). In addition, we suppose*

$$\int_0^1 \varphi(sv + (1-s)\tilde{v}, sy + (1-s)\tilde{y}) ds \leq c[\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] \quad \forall v, y, \tilde{v}, \tilde{y} \in \mathbb{R}^d, \tag{5.1}$$

for some $c > 0$. Given $u_0 \in C(\mathbb{R}^{2d}) \cap L^\infty(\varphi^{-1})$, let u be the viscosity solution of (2.4). Then, there exist $\omega, K > 0$ such that

$$|u(t, v, y) - u(t, \tilde{v}, \tilde{y})| \leq K e^{-\omega t} \|u_0\|_{L^\infty(\varphi^{-1})} (\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})) \quad \forall t > 0, v, y, \tilde{v}, \tilde{y} \in \mathbb{R}^d, \tag{5.2}$$

where ω, K depend on ℓ_B, ℓ_H, γ , and φ .

Proof The strategy is to apply Theorem 3.2 to $u(t+1)$ and reach the result by estimating $[u(1)]_\theta$ with Proposition 4.4. We first assume that u_0 is bounded, so that $u(t)$ is not only bounded, but also Lipschitz for $t > 0$ (see Theorem 2.3). Next, we apply Proposition 4.4 (for $t = 1$), which gives, according to (5.1),

$$\begin{aligned} u(1, v, y) - u(1, \tilde{v}, \tilde{y}) &= \int_0^1 \{ \nabla_v u(1, sv + (1-s)\tilde{v}, sy + (1-s)\tilde{y}) \cdot (v - \tilde{v}) \\ &\quad + \nabla_y u(1, sv + (1-s)\tilde{v}, sy + (1-s)\tilde{y}) \cdot (y - \tilde{y}) \} ds \\ &\leq C \|u_0\|_{L^\infty(\varphi^{-1})} \int_0^1 \varphi(sv + (1-s)\tilde{v}, sy \\ &\quad + (1-s)\tilde{y}) (|v - \tilde{v}| + |y - \tilde{y}|) ds \\ &\leq \tilde{C} \|u_0\|_{L^\infty(\varphi^{-1})} [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] (|v - \tilde{v}| + |y - \tilde{y}|). \end{aligned}$$

We also notice that if k_0 satisfies $(\partial_t + \mathcal{L})\varphi \geq -k_0$, then $\|u_0\|_{L^\infty(\varphi^{-1})} e^{k_0 t} \varphi$ is a supersolution of the equation since $\varphi \geq 1$. By comparison, we readily get

$$|u(t, v, y)| \leq e^{k_0 t} \|u_0\|_{L^\infty(\varphi^{-1})} \varphi(v, y) \quad \forall t \in (0, 1]. \tag{5.3}$$

Thus, for a possibly different constant \tilde{C} , we have

$$|u(1, v, y) - u(1, \tilde{v}, \tilde{y})| \leq \tilde{C} \|u_0\|_{L^\infty(\varphi^{-1})} [\varphi(v, y) + \varphi(\tilde{v}, \tilde{y})] ((|v - \tilde{v}| + |y - \tilde{y}|) \wedge 1),$$

which implies

$$[u(1)]_1 \leq \tilde{C} \|u_0\|_{L^\infty(\varphi^{-1})}.$$

Finally, once we apply Theorem 3.2 to $u(t+1)$, we get (5.2) for $t \geq 1$. However, up to increasing the value of K , (5.2) also holds for $t \in (0, 1]$ in light of (5.3). The case when u_0 is not bounded is obtained by approximation, as in Corollary 4.6. \square

Consequently, we deduce by duality our main result, which includes Theorem 1.1 for the case $\kappa = 0$.

Theorem 5.2 *Assume that H, B satisfy (2.5), (2.6)–(2.7) and there exists a Lyapunov function φ (cf. Definition 3.1) satisfying conditions (3.4), (4.8) and (5.1). Then, there exist $\omega, K > 0$ such that, for every initial data $m_{01}, m_{02} \in \mathcal{P}_1(\mathbb{R}^{2d})$ for which $\varphi \in L^1(dm_{0i}), i = 1, 2$, the corresponding solutions m_1, m_2 of (2.2) satisfy*

$$\|m_1(t) - m_2(t)\|_{TV_\varphi} \leq K e^{-\omega t} \|m_{01} - m_{02}\|_{TV_\varphi}, \tag{5.4}$$

where $\|\cdot\|_{TV_\varphi}$ is the norm of the total variation weighted by φ .

Proof of Theorem 5.2 (and Theorem 1.1) We first observe that if $\varphi \in L^1(dm_{0i})$, $i = 1, 2$, then we also have $\varphi \in L^1(dm_i(t))$ for all $t > 0$. This is established as in the proof of Theorem 2.3, where we show that if the initial measure has finite k moments, then the corresponding solution enjoys the same property at later times. Setting $m_0 := m_{01} - m_{02}$ and $m := m_1 - m_2$, applying (2.8) to both solutions, and subtracting, we obtain

$$\int_{\Omega} \zeta \, dm(t) = \int_{\Omega} u(0) \, dm_0,$$

for every $\zeta \in C_b(\mathbb{R}^{2d})$, where u is the solution of the backward problem (B) in (2.8). By stability of viscosity solutions, standing on Corollary 4.6, and since $\varphi \in L^1(d|m|(t))$, the equality extends to the case $\zeta \in L^\infty(\varphi^{-1})$. Using that m_0 has zero average, any constant can be added to $u(0)$. Then, it holds

$$\begin{aligned} \int_{\Omega} \zeta \, dm(t) &\leq \|m_0\|_{TV_\varphi} \inf_{c \in \mathbb{R}} \|u(0) + c\|_{L^\infty(\varphi^{-1})} \\ &= \|m_0\|_{TV_\varphi} \sup_{z, \tilde{z}} \frac{|u(0, z) - u(0, \tilde{z})|}{\varphi(z) + \varphi(\tilde{z})} \\ &\leq \|m_0\|_{TV_\varphi} K e^{-\omega t} \|\zeta\|_{L^\infty(\varphi^{-1})}, \end{aligned} \tag{5.5}$$

using Theorem 5.1, with $z = (v, y)$, $\tilde{z} = (\tilde{v}, \tilde{y})$, and $\inf_{c \in \mathbb{R}} \|u(0) + c\|_{L^\infty(\varphi^{-1})} = \sup_{z, \tilde{z}} \frac{|u(0, z) - u(0, \tilde{z})|}{\varphi(z) + \varphi(\tilde{z})}$ follows from, e.g., [17, Lemma 2.1]. By arbitrariness of ζ , this yields $\|m\|_{L^1(d|m|(t))} \leq K e^{-\omega t} \|m_0\|_{TV_\varphi}$, which is the desired conclusion. \square

The result of Theorem 5.2 can also be rephrased as the decay of zero-average solutions. Those solutions are meant in the sense of Corollary 2.5 (possibly signed solutions). Hence, the proof of the result below works by duality, exactly as in (5.5).

Theorem 5.3 *Assume the hypotheses of Theorem 5.2 are valid. Let $m_0 \in \mathcal{M}(\mathbb{R}^{2d})$ be a zero-average measure such that $\varphi \in L^1(d|m_0|)$. Let m be a solution of (2.2) in the sense of Corollary 2.5, i.e., $m \in C([0, T]; \mathcal{M}(\mathbb{R}^{2d})^*)$ and satisfies (2.8). Then, we have*

$$\|m(t)\|_{TV_\varphi} \leq K e^{-\omega t} \|m_0\|_{TV_\varphi} \quad \forall t > 0,$$

with K, ω as in Theorem 5.2.

5.1 Examples

We conclude this section by giving a model class of examples, with a specific Lyapunov function φ , for which the above theorem can be used. Precisely, we assume that

$$H(v, y) = v, \quad B(v, y) = b(v, y) + \nabla \Phi(y),$$

where b is a Lipschitz function satisfying

$$\exists \alpha, c_0 > 0 : \quad b(v, y) \cdot v \geq \alpha |v|^2 - c_0, \quad b(v, y) \cdot y \geq -c_0 |y| (1 + |v|), \quad \forall v, y \in \mathbb{R}^d, \tag{5.6}$$

and $\Phi \in W^{2,\infty}(\mathbb{R}^d)$ fulfils

$$\exists \beta, c_1 > 0 : \quad \nabla\Phi(y) \cdot y \geq \beta |y|^2 - c_1 \quad \forall y \in \mathbb{R}^d. \tag{5.7}$$

Then, we can build a Lyapunov function for (2.4). Note that (2.5)–(2.7) are satisfied by H and B as well.

Lemma 5.4 *Suppose that (5.6) and (5.7) hold. There exist $\varepsilon, \delta > 0$ such that*

$$\varphi := \Phi(y) + \frac{1}{2} \left(|v|^2 + 2\varepsilon v \cdot y + \delta |y|^2 \right)$$

is a Lyapunov function for the operator \mathcal{L} satisfying (3.4), (4.8), and (5.1).

Proof We compute and estimate, by virtue of (5.6) and (5.7),

$$\begin{aligned} \mathcal{L}[\varphi] &= -d - \kappa \delta d - \kappa \Delta \Phi - v \cdot \nabla \Phi(y) - \varepsilon |v|^2 - \delta v \cdot y \\ &\quad + b(v, y) \cdot v + \varepsilon b(v, y) \cdot y + \nabla \Phi(y) \cdot (v + \varepsilon y) \\ &\geq (\alpha - \varepsilon) |v|^2 - \delta v \cdot y + \varepsilon (\beta |y|^2 - c_1) + \varepsilon b(v, y) \\ &\quad \cdot y - [d(1 + \kappa \delta) + c_0 + \kappa \|\Delta \Phi\|_\infty] \\ &\geq \left(\frac{\alpha}{2} - \varepsilon \right) |v|^2 + \varepsilon \left(\frac{\beta}{2} |y|^2 - c_1 \right) - \frac{\delta^2}{2\alpha} |y|^2 \\ &\quad - \varepsilon \frac{c_0^2}{2\beta} (1 + |v|)^2 - [d(1 + \kappa \delta) + c_0 + \kappa \|\Delta \Phi\|_\infty] \\ &\geq \left(\frac{\alpha}{2} - \varepsilon - \varepsilon \frac{c_0^2}{\beta} \right) |v|^2 + \varepsilon \left(\frac{\beta}{2} - \frac{\delta^2}{2\alpha\varepsilon} \right) |y|^2 \\ &\quad - \left(d(1 + \kappa \delta) + c_0 + \varepsilon c_1 + \varepsilon \frac{c_0^2}{\beta} + \kappa \|\Delta \Phi\|_\infty \right). \end{aligned}$$

We now take $\delta = \varepsilon$ and choose $\varepsilon < 1$ sufficiently small (only depending on α, β , and c_0) so that

$$\frac{\alpha}{2} - \varepsilon - \varepsilon \frac{c_0^2}{\beta} \geq \frac{\alpha}{4}, \quad \frac{\beta}{2} - \frac{\varepsilon}{2\alpha} \geq \frac{\beta}{4}.$$

Then, we get

$$\mathcal{L}[\varphi] \geq \frac{\alpha}{4} |v|^2 + \varepsilon \frac{\beta}{4} |y|^2 - C_\varepsilon, \tag{5.8}$$

and, being $\varepsilon < 1$, we also have $\delta - \varepsilon^2 = \varepsilon - \varepsilon^2 > 0$, which implies $\varphi \geq \Phi(y) + c_\varepsilon (|v|^2 + |y|^2)$. Of course, we can assume $\Phi(y) \geq 1$ without loss of generality (anyway, Φ is bounded below due to (5.7)). In addition, we recall that $\nabla\Phi$ is Lipschitz, thus $\varphi \leq C(|v|^2 + |y|^2 + 1)$. Therefore, (5.8) yields that φ is a Lyapunov function, see (3.1). Moreover, (3.4) is satisfied. As for (4.8), it is enough to observe that

$$\frac{1}{2} \mathcal{L}(\varphi^2) = \varphi \mathcal{L}(\varphi) - (|\nabla_v \varphi|^2 + \kappa |\nabla_y \varphi|^2)$$

and the right-hand side diverges at infinity, due to (5.8) and the quadratic growth of φ . Finally, (5.1) holds since $\varphi \in W^{2,\infty}(\mathbb{R}^{2d})$ (being the sum of a quadratic function and $\Phi \in W^{2,\infty}(\mathbb{R}^d)$), hence φ is semiconvex. Then, by using convexity only, (5.1) holds for φ up to adding a quadratic function. Hence, this means that it holds for φ itself because φ dominates a quadratic function. \square

Remark 5.5 We observe that the above lemma admits similar extensions to the case of possibly more general functions $H(v) = \nabla\Psi(v)$, for uniformly convex Ψ , and possibly superquadratic growth of the function $\Phi(y)$.

As a particular case, we can now deduce the exponential decay for the classical Kolmogorov Eq. (1.2) from Theorem 1.1.

Corollary 5.6 *Assume that b satisfies (2.5), (5.6) and $\Phi \in W^{2,\infty}(\mathbb{R}^d)$ fulfils (5.7). Let $m_0 \in \mathcal{M}_2(\mathbb{R}^{2d})$ with $\int m_0 = 0$, and let m be the solution (in the sense of Corollary 2.5) of (1.2) with $B(v, y) = b(v, y) + \Phi(y)$. Then, there exist $\omega, K > 0$ such that*

$$\|m(t)\|_{\mathcal{M}_2} \leq K e^{-\omega t} \|m_0\|_{\mathcal{M}_2}. \quad (5.9)$$

Proof Let φ be defined from Lemma 5.4. At infinity, φ is equivalent to $|v|^2 + |y|^2$, so (5.9) can be rephrased as

$$\|m(t)\|_{TV_\varphi} \leq K e^{-\omega t} \|m_0\|_{TV_\varphi},$$

which is exactly the estimate of Theorem 5.3. \square

Funding Open access funding provided by Università degli Studi di Roma Tor Vergata within the CRUI-CARE Agreement.

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

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/>.

6 Appendix

In this appendix, we provide short proofs of the existence and uniqueness results for problems (2.2) and (2.4). We start with the latter one.

Proposition 6.1 *Assume that (2.5) and (2.6)–(2.7) hold, and $\kappa \geq 0$. For any $u_0 \in C_b(\mathbb{R}^{2d})$, there exists a unique viscosity solution $u \in C_b([0, T] \times \mathbb{R}^{2d})$ of (2.4).*

Proof The existence is given through vanishing viscosity. Indeed, if $\kappa > 0$ in (2.4), then the existence of a unique (classical) solution is well known, due to the global Lipschitz conditions (2.5) and (2.7) on the drift terms. Moreover, if u_0 is bounded and Lipschitz, then the solution is also bounded and Lipschitz, see, e.g., [27, Section 3]. For the general case (possibly degenerate), we consider a Lipschitz and uniformly bounded sequence u_{0n} of initial data, which converges to u_0 locally uniformly, and we take the solutions u_n corresponding to $\kappa + \frac{1}{n}$. By the maximum principle, we have that u_n is uniformly bounded. In addition, being u_n Lipschitz and smooth, according to Proposition 4.2, we achieve that $\nabla_v u_n(t), \nabla_y u_n(t)$ are uniformly bounded for $t > 0$. It follows that the sequence $u_n(t, \cdot)$ is equicontinuous. Now, applying [26, Lemma 6.2] to $u(t + \cdot, \cdot)$, we deduce that, for every compact set $K \subset \mathbb{R}^{2d}$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $\|u_n(t + s) - u_n(t)\|_{L^\infty(K)} \leq \varepsilon$, for any $s \in (0, \delta)$ and any n , where δ only depends on the local bounds of H, B , and u_n , and the equicontinuity of $u_n(t)$. In particular, by Ascoli–Arzelà theorem, the sequence u_n turns out to be relatively compact, in the uniform topology, in $[t, T] \times K$, whatever $t > 0$ and compact set K are given. Selecting compact sets K_j exhausting \mathbb{R}^{2d} and a sequence $t_j \downarrow 0$, by a diagonal procedure, we can extract a subsequence u_{n_k} converging to a continuous function u locally uniformly in $(0, T) \times \mathbb{R}^{2d}$. Of course, from the previous estimates, u is bounded and Lipschitz for $t > 0$. Using again [26, Lemma 6.2], and in view of the local uniform convergence of u_{0n} to u_0 , we also deduce that u is actually continuous up to $t = 0$ and $u(t) \rightarrow u_0$ as $t \rightarrow 0$. Thus, we have $u \in C_b([0, T] \times \mathbb{R}^{2d})$, and, by the stability of the viscosity formulation with respect to local uniform convergence, we also get that u is a viscosity solution of (2.4).

To prove uniqueness, we use classical tools in the theory of viscosity solutions, so we only sketch the argument. First of all, we observe that, due to the linear growth of the drift terms B and H , the function $e^{\lambda t}(|v|^2 + |y|^2 + M)$ is a supersolution of the equation, up to choosing λ and M sufficiently large. Therefore, for any small ε , $u_\varepsilon := u - \varepsilon e^{\lambda t}(|v|^2 + |y|^2 + M)$ and $v_\varepsilon := v + \varepsilon e^{\lambda t}(|v|^2 + |y|^2 + M)$ are a strict subsolution, respectively supersolution, being $u_\varepsilon \leq 0$ and $v_\varepsilon \geq 0$ outside of a suitable compact set K_ε . Thus, up to replacing u, v with $u_\varepsilon, v_\varepsilon$, we are reduced to proving a comparison argument assuming that $\sup(u - v)$, if ever positive, would be attained in some compact set, where the classical doubling method for viscosity solutions allows to conclude. For the interested reader, we refer to [26, Proposition 6.1] for a similar proof in a much more general context, with purely continuous data. \square

The next result shows that weak solutions of the Fokker–Planck equation are in duality with the viscosity solutions constructed above.

Lemma 6.2 *Assume that (2.5) and (2.6)–(2.7) hold, and $\kappa \geq 0$. Let $m_0 \in \mathcal{M}_1(\mathbb{R}^{2d})$, and let $m \in C([0, T]; \mathcal{M}(\mathbb{R}^{2d})^*)$ be a weak solution of (2.2) such that $m(t) \in \mathcal{M}_1(\mathbb{R}^{2d})$ and $\sup_{t \in (0, T)} \int_{\mathbb{R}^{2d}} |z| d|m|(t) < \infty$. Then, m satisfies (2.8).*

Proof In what follows, we denote by $z := (v, y)$ and set $F := (B, -H)$ the drift term in \mathbb{R}^{2d} for the equation satisfied by m . We introduce a sequence $\rho_\delta(z)$ of standard (compactly supported) symmetric mollifiers in \mathbb{R}^{2d} and set

$$m_\delta(t, z) = (m(t) \star \rho_\delta)(z) := \int_{\mathbb{R}^{2d}} m(t, w) \rho_\delta(z - w) dw .$$

We also take a sequence of 1-d mollifiers $\xi_\varepsilon(t)$ such that $\text{supp}(\xi_\varepsilon) \subset (-\varepsilon, 0)$ and set

$$m_{\delta,\varepsilon}(t, z) = (m_\delta \star \xi_\varepsilon)(t, z) := \int_0^T \xi_\varepsilon(s - t) m_\delta(s, z) ds.$$

One can easily verify that $m_{\delta,\varepsilon}$ satisfies the following equation

$$\begin{cases} \partial_t m_{\delta,\varepsilon} - \Delta_v m_{\delta,\varepsilon} - \kappa \Delta_y m_{\delta,\varepsilon} - \text{div}_{(v,y)}((m F)_{\delta,\varepsilon}) = \xi_\varepsilon(-t)(m_0)_\delta, & \text{in } (0, T) \times \mathbb{R}^{2d}, \\ m_{\delta,\varepsilon}(0) = 0 & \text{in } \mathbb{R}^{2d}, \end{cases} \tag{6.1}$$

where $(m F)_{\delta,\varepsilon} = (m F)_\delta \star \xi_\varepsilon$. The fact that $m_{\delta,\varepsilon}(0) = 0$ is due to the decentered choice of the time mollification. Since ξ_ε is compactly supported in $(-\varepsilon, 0)$, it follows that $m_{\delta,\varepsilon}(t)$ vanishes for t in a right neighborhood of zero, and we have

$$\begin{aligned} \int_{\mathbb{R}^{2d}} m_{\delta,\varepsilon}(t, z) \psi dz &= \int_{\mathbb{R}^{2d}} \int_0^T (\xi_\varepsilon(s - t) - \xi_\varepsilon(s)) \psi m_\delta(s, z) ds dz \\ &\leq c_\varepsilon t \|\psi\|_\infty, \quad \forall \psi \in C_b(\mathbb{R}^{2d}), \quad \forall t > 0, \end{aligned} \tag{6.2}$$

for some constant c_ε only depending on ε .

We now consider the vanishing viscosity approximation of the backward problem, namely, the unique classical solution of

$$\begin{cases} -\partial_t u_n + \mathcal{L}[u_n] = \frac{1}{n} \Delta_y u_n & \text{in } (0, t) \times \mathbb{R}^{2d}, \\ u_n(t, v, y) = \zeta(v, y) & \text{in } \mathbb{R}^{2d}. \end{cases}$$

As the first step, we can take $\zeta \in C_b(\mathbb{R}^{2d})$ so that it is also a smooth Lipschitz function. In order to exploit the duality between u_n and $m_{\delta,\varepsilon}$, we need to consider an auxiliary cut-off function $\eta_R(z) := \eta(z/R)$, for some smooth function η supported in $\{|z| \leq 2\}$ and such that $\eta \equiv 1$ in $\{|z| \leq 1\}$. Multiplying the equation of u_n by $m_{\delta,\varepsilon} \eta_R$ and integrating by parts (thanks to η_R), we get

$$\begin{aligned} & - \int_{\mathbb{R}^{2d}} \zeta \eta_R dm_{\delta,\varepsilon}(t) dz + \int_0^t \int_{\mathbb{R}^{2d}} u_n \left[\partial_t - \Delta_v - \left(\kappa + \frac{1}{n} \right) \Delta_y \right] (m_{\delta,\varepsilon} \eta_R) dz d\tau \\ & + \int_0^t \int_{\mathbb{R}^{2d}} F(z) \cdot \nabla_z u_n m_{\delta,\varepsilon} \eta_R dz d\tau = 0, \end{aligned}$$

which yields, using the equation (6.1),

$$\begin{aligned}
 & - \int_{\mathbb{R}^{2d}} \zeta \eta_R dm_{\delta,\varepsilon}(t) dz + \int_0^t \int_{\mathbb{R}^{2d}} F(z) \cdot \nabla_z u_n m_{\delta,\varepsilon} \eta_R dz d\tau \\
 & - \int_0^t \int_{\mathbb{R}^{2d}} (m F)_{\delta,\varepsilon} \cdot \nabla_z (u_n \eta_R) dz d\tau \\
 & + \int_0^t \int_{\mathbb{R}^{2d}} u_n \xi_\varepsilon(-\tau)(m_0)_\delta \eta_R dz d\tau = \frac{1}{n} \int_0^t \int_{\mathbb{R}^{2d}} u_n (\Delta_y m_{\delta,\varepsilon}) \eta_R dz d\tau \\
 & + \int_0^t \int_{\mathbb{R}^{2d}} \left(u_n m_{\delta,\varepsilon} \left[\Delta_v + \left(\kappa + \frac{1}{n} \right) \Delta_y \right] \eta_R dz d\tau \right. \\
 & \left. + 2u_n \left[\nabla_v m_{\delta,\varepsilon} \nabla_v \eta_R + \left(\kappa + \frac{1}{n} \right) \nabla_y m_{\delta,\varepsilon} \nabla_y \eta_R \right] \right) dz d\tau .
 \end{aligned}$$

By letting $R \rightarrow \infty$, we remove the cut-off function. This is justified by Lebesgue’s theorem, using that u_n is bounded and Lipschitz and $F(z)m_{\delta,\varepsilon}(z, t) \in L^1((0, T) \times \mathbb{R}^{2d})$.¹ In particular, all the terms with derivatives of ξ_R vanish as $R \rightarrow \infty$. Then, it holds

$$\begin{aligned}
 & - \int_{\mathbb{R}^{2d}} \zeta dm_{\delta,\varepsilon}(t) dz + \int_0^t \int_{\mathbb{R}^{2d}} F(z) \cdot \nabla_z u_n m_{\delta,\varepsilon} dz d\tau - \int_0^t \int_{\mathbb{R}^{2d}} (m F)_{\delta,\varepsilon} \cdot \nabla_z u_n dz d\tau \\
 & = - \int_0^t \int_{\mathbb{R}^{2d}} u_n \xi_\varepsilon(-\tau)(m_0)_\delta dz d\tau + \frac{1}{n} \int_0^t \int_{\mathbb{R}^{2d}} u_n \Delta_y m_{\delta,\varepsilon} dz d\tau .
 \end{aligned} \tag{6.3}$$

We estimate, using the Lipschitz regularity of F and Proposition 4.2 for ∇u_n ,

$$\begin{aligned}
 & \left| + \int_0^t \int_{\mathbb{R}^{2d}} F(z) \cdot \nabla_z u_n m_{\delta,\varepsilon} dz d\tau - \int_0^t \int_{\mathbb{R}^{2d}} (m F)_{\delta,\varepsilon} \cdot \nabla_z u_n dz d\tau \right| \\
 & = \int_0^t \int_0^s \xi_\varepsilon(s - \tau) \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \text{grad}_z u_n(\tau, z) \\
 & \quad \cdot [F(z) - F(w)] \rho_\delta(z - w) dm(s, w) dw dz ds d\tau \\
 & \leq c \delta \int_0^t \int_{\mathbb{R}^{2d}} \tau^{-\frac{3}{2}} m_{\delta,\varepsilon}(\tau, z) dz d\tau \leq \tilde{c}_\varepsilon \delta ,
 \end{aligned} \tag{6.4}$$

where the last inequality comes from (6.2). Now, we pass to the limit in (6.3). Specifically, we first let $n \rightarrow \infty$ (so that the last term vanishes since $m_{\delta,\varepsilon}$ is smooth), exploiting the uniform convergence of u_n . Secondly, we let $\delta \rightarrow 0$, at fixed ε , using estimate (6.4). We thus obtain

$$\int_{\mathbb{R}^{2d}} \zeta dm_\varepsilon(t) dz = \int_0^t \int_{\mathbb{R}^{2d}} u(\tau) \xi_\varepsilon(-\tau) dm_0 dz d\tau .$$

¹ Thanks to the Lipschitz character of F and $\sup_{t \in (0, T)} \int_{\mathbb{R}^{2d}} |z| d|m|(t) < \infty$. Indeed, we have $|F(z)m_{\delta,\varepsilon}(z)| \leq C((1 + |z|)|m|) * \rho_\delta * \xi_\varepsilon$

Finally, we let $\varepsilon \rightarrow 0$, using the continuity of $u(t)$ (recall that $u(t) \rightarrow u(0)$ locally uniformly), the continuity of $m(t)$, and the fact that m has finite first moment. Therefore, (2.8) holds for any $\zeta \in C_b(\mathbb{R}^{2d})$ and the (unique) viscosity solution corresponding to ζ . \square

In light of the above lemma, we conclude with the proof of Theorem 2.3.

Proof of Theorem 2.3 Item (i) was proved in Proposition 6.1. As for the Fokker–Planck equation, the uniqueness is a straightforward consequence of Lemma 6.2 whenever $m(t) \in \mathcal{M}_1(\mathbb{R}^{2d})$ for $t \in [0, T]$ and $\sup_{t \in (0, T)} \int_{\mathbb{R}^{2d}} |z| dm(t) < \infty$. Indeed, if $m_0 = 0$, then (2.8) implies $\int_{\mathbb{R}^{2d}} \zeta dm(t) = 0$ for any t . Here, any $\zeta \in C_b(\mathbb{R}^{2d})$ is allowed, due to the solvability of the adjoint problem, which is guaranteed by item (i). Hence, $m(t) = 0$.

If $m \in C([0, T]; \mathcal{P}(\mathbb{R}^{2d}))$, then we do not need to assume a priori that m has finite first moment. We now show this fact. Simply using the weak formulation and the time continuity of m , we have

$$\begin{aligned} & \int_{\mathbb{R}^{2d}} \phi(t, v, y) dm(t) + \int_0^t \int_{\mathbb{R}^{2d}} (-\partial_t \phi + \mathcal{L}[\phi]) dm(\tau) d\tau \\ &= \int_{\mathbb{R}^{2d}} \phi(0, v, y) dm_0 \quad \forall \phi \in C^{1,2}(\overline{Q}_t), \quad t > 0. \end{aligned} \tag{6.5}$$

We wish to remove the condition that ϕ has compact support in \mathbb{R}^{2d} . We set $\langle z \rangle := \sqrt{1 + |z|^2}$ and take a real-valued, cut-off function $\xi(s)$, supported in $\{|s| \leq 2\}$ and such that $\xi \equiv 1$ in $\{|s| \leq 1\}$. Then, we replace ϕ with $\phi \xi_R$ as a test function in (6.5), where $\xi_R(z) := \xi\left(\frac{\langle z \rangle}{R}\right)$ and now $\phi \in C^{1,2}(\overline{Q}_t)$ is assumed to be nonnegative, but does not need to be compactly supported in \mathbb{R}^{2d} . We achieve

$$\int_{\mathbb{R}^{2d}} \phi \xi_R dm(t) + \int_0^t \int_{\mathbb{R}^{2d}} (-\partial_t \phi + \mathcal{L}[\phi]) \xi_R dm(\tau) d\tau = \int_{\mathbb{R}^{2d}} \phi(0, v, y) \xi_R dm_0 + o_R(1),$$

where $o_R(1)$ is a quantity converging to zero as $R \rightarrow \infty$, since it involves the derivatives of ξ_R , which vanish uniformly as $R \rightarrow \infty$, while m has finite mass. Notice that if $\phi = e^{\lambda(T-t)}(\langle z \rangle^k + M)$, then one can verify that $(-\partial_t + \mathcal{L})[\phi] \geq c \langle z \rangle^k \geq 0$ in $(0, T)$, by choosing λ and M large enough, and the same holds replacing $\langle z \rangle^k$ with $\langle z \rangle^k \wedge n$, for any fixed $n \in \mathbb{N}$. Thus, as long as ϕ is truncated, we can let $R \rightarrow \infty$ obtaining

$$\int_{\mathbb{R}^{2d}} (\langle z \rangle^k \wedge n) dm(t) \leq C e^{\lambda t} \int_{\mathbb{R}^{2d}} \langle z \rangle^k dm_0.$$

This yields $\int_{\mathbb{R}^{2d}} |z|^k dm(t) \leq C_T \int_{\mathbb{R}^{2d}} (1 + |z|^k) dm_0$ for all $t \in (0, T)$. In particular, we have established that if m_0 has k finite moments, then any nonnegative solution $m(t)$ has k finite and bounded moments in $(0, T)$. This concludes the proof of the uniqueness statement of Theorem 2.3. We also note that, using the above estimate on the moments, one can prove that m is continuous from $[0, T]$ into $\mathcal{P}_1(\mathbb{R}^{2d})$.

We are only left to show the existence of at least one solution, which we construct by vanishing viscosity limit. Assume that $m_0 \in \mathcal{P}(\mathbb{R}^{2d})$ has k finite moments. Let m_n be the unique (classical) solution of (2.2) when replacing κ with $\kappa + \frac{1}{n}$. By the above estimate, we know that

$$\int_{\mathbb{R}^{2d}} \langle z \rangle^k dm_n(t) \leq C_T \int_{\mathbb{R}^{2d}} \langle z \rangle^k dm_0 \leq C, \tag{6.6}$$

which gives uniform bounds for the k th moments of m_n . With standard arguments, see, e.g., [26, Lemma 6.3], one can prove that m_n is relatively compact in $C([0, T]; \mathcal{P}(\mathbb{R}^{2d}))$, and, up to a subsequence, it converges to a weak solution m . In addition, $m(t) \in \mathcal{M}_k(\mathbb{R}^{2d})$ for every $t \geq 0$. For a general signed measure m_0 , of course, we reason by splitting $m_0 = m_0^+ - m_0^-$ and obtaining a solution for both the negative and positive parts. \square

Proof of Corollary 2.5 If $m_0 \in \mathcal{P}(\mathbb{R}^{2d}) \cap \mathcal{M}_k(\mathbb{R}^{2d})$, then it can be approximated by a sequence m_{0n} of initial data belonging to $\mathcal{P}_1(\mathbb{R}^{2d})$ that have k uniformly bounded moments. Reasoning as in the previous proof, the corresponding sequence of solutions, m_n , satisfies a uniform estimate such as (6.6), and it is relatively compact in $C([0, T]; \mathcal{P}(\mathbb{R}^{2d}))$. A subsequence thus converges to some m , and since equality (2.8) holds for m_n , passing to the limit, we get it for m as well. This result extends to any non-negative measure m , and, for a general m_0 , one can reason by splitting $m_0 = m_0^+ - m_0^-$ to construct a solution. Finally, the uniqueness is a straightforward consequence of (2.8) because the adjoint problem is solvable for all $\zeta \in C_b(\mathbb{R}^{2d})$. The preservation of positivity follows from the adjoint problem, and the conservation of mass is obvious being $u = 1$ the only viscosity solution corresponding to $\zeta = 1$. \square

References

1. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lectures in Mathematics ETH Zürich, 2nd edn. Birkhäuser, Basel (2008)
2. Anceschi, F., Ascione, G., Castorina, D., Solombrino, F.: Well-posedness of Kolmogorov–Fokker–Planck equations with unbounded drifts. *J. Math. Anal. Appl.* **543**(2), 128909 (2025)
3. Arnold, A., Erb, J.: Sharp entropy decay of hypocoercive and non-symmetric Fokker–Planck equations with linear drift (2014). [arXiv:1409.5425v2](https://arxiv.org/abs/1409.5425v2)
4. Bakry, D., Cattiaux, P., Guillin, A.: Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254**, 727–759 (2008)
5. Bogachev, V.I., Röckner, M., Shaposhnikov, S.V.: On uniqueness problems related to the Fokker–Planck–Kolmogorov equation for measures. *J. Math. Sci.* **179**, 7–47 (2011)
6. Canizo, J.A., Cao, C., Evans, J., Yoldas, H.: Hypocoercivity of linear kinetic equations via Harris theorem. *Kinet. Relat. Model.* **13**, 97–128 (2020)
7. Cao, C.: The kinetic Fokker–Planck equation with general force. *J. Evol. Equ.* **21**(2), 2293–2337 (2021)
8. Chen, M.F., Li, S.F.: Coupling methods for multidimensional diffusion processes. *Ann. Probab.* **17**, 151–177 (1989)
9. Chiarini, A., Conforti, G., Greco, G., Ren, Z.: Entropic turnpike estimates for the kinetic Schrödinger problem. *Electron. J. Probab.* **27**, 1–32 (2022)
10. Crandall, M.G., Ishii, H., Lions, P.L.: User’s guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc. (N.S.)* **27**, 1–67 (1992)
11. Desvillettes, L., Villani, C.: On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker–Planck equation. *Commun. Pure Appl. Math.* **54**, 1–42 (2001)

12. Dolbeault, J., Mouhot, C., Schmeiser, C.: Hypocoercivity for linear kinetic equations conserving mass. *Trans. AMS.* **367**, 3807–3828 (2015)
13. Down, D., Meyn, S.P., Tweedie, R.L.: Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23**, 1671–1691 (1996)
14. Duan, R.: Hypocoercivity of linear degenerately dissipative kinetic equations. *Nonlinearity* **24**, 2165–2189 (2010)
15. Eberle, A., Guillin, A., Zimmer, R.: Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.* **47**(4), 1982–2010 (2019)
16. Gualdani, M.P., Mischler, S., Mouhot, C.: Factorization of non-symmetric operators and exponential H-theorem. *Mém. Soc. Math. Fr.* **153**, 1–137 (2017)
17. Hairer, M., Mattingly, J.C.: Yet another look at Harris’ ergodic theorem for Markov chains. In: *Seminar on Stochastic Analysis, Random Fields and Applications VI*, *Progress in Probab.*, vol. 63, pp. 109–117 (2011)
18. Helffer, B., Nier, F.: *Hypoelliptic Estimates and Spectral Theory for Fokker–Planck Operators and Witten Laplacians*. *Lecture Notes in Mathematics*, vol. 1862. Springer, Berlin (2005)
19. Hérau, F.: Short and long time behavior of the Fokker–Planck equation in a confining potential and applications. *J. Funct. Anal.* **244**, 95–118 (2007)
20. Hérau, F., Nier, F.: Isotropic hypoellipticity and trend to equilibrium for the Fokker–Planck equation with a high-degree potential. *Arch. Ration. Mech. Anal.* **171**, 151–218 (2004)
21. Ishii, H., Lions, P.L.: Viscosity solutions of fully nonlinear second-order elliptic partial differential equations. *J. Differ. Equ.* **83**, 26–78 (1990)
22. Lindvall, T., Rogers, C.: Coupling of multidimensional diffusions by reflection. *Ann. Probab.* **14**, 860–872 (1986)
23. Mattingly, J.C., Stuart, A.M., Higham, D.J.: Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Process. Appl.* **101**, 185–232 (2002)
24. Meyn, S.P., Tweedie, R.L.: Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.* **25**, 518–548 (1993)
25. Mischler, S., Mouhot, C.: Exponential stability of slowing decaying solutions to the Kinetic-Fokker–Planck equation. *Arch. Ration. Mech. Anal.* **221**(2), 677–723 (2016)
26. Porretta, A.: Decay rates of convergence for Fokker–Planck equations with confining drift. *Adv. Math.* **436**, 109393 (2024)
27. Porretta, A., Priola, E.: Global Lipschitz regularizing effects for linear and nonlinear parabolic equations. *J. Math. Pures Appl.* **100**, 633–686 (2013)
28. Shaposhnikov, S.V.: On the uniqueness of integrable and probability solutions to the Cauchy problem for the Fokker–Planck–Kolmogorov equations. *Dokl. Math.* **84**(1), 565–570 (2011)
29. Villani, C.: Hypocoercivity. *Mem. Am. Math. Soc.* **202**(950), iv+141 (2009)
30. Villani, C.: *Optimal Transport, Old and New*. *Fundamental Principles of Mathematical Sciences*, vol. 338. Springer, Berlin (2009)
31. Zhang, C.: Hypocoercivity and global hypoellipticity for the kinetic Fokker–Planck equation in H^k spaces. *Kinet. Relat. Model.* **17**, 394–435 (2024)

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