Distribution dependent SDEs driven by additive continuous noise*

Lucio Galeati[†]   Fabian A. Harang[‡]   Avi Mayorcas[§]

Abstract

We study distribution dependent stochastic differential equation driven by a continuous process, without any specification on its law, following the approach initiated in [17]. We provide several criteria for existence and uniqueness of solutions which go beyond the classical globally Lipschitz setting. In particular we show well-posedness of the equation, as well as almost sure convergence of the associated particle system, for drifts satisfying either Osgood-continuity, monotonicity, local Lipschitz or Sobolev differentiability type assumptions.

1 Introduction

In this work we consider distribution dependent SDEs (henceforth DDSDEs) of the form

$$X_t = \int_0^t B_s(X_s, \mu_s)\, \mathrm{d}s + Y_t, \quad \mu_t = \mathcal{L}(X_t) \tag{1.1}$$

where $B : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, X is the unknown and Y is a continuous stochastic process with prescribed law; $\mathcal{L}(X_t)$ stands for law of the random variable X_t.

In the literature the process Y is typically sampled as $Y_t = \xi + \sqrt{2\varepsilon}\, W_t$, where $\varepsilon \geq 0$, W is a standard Brownian motion and ξ is an $\mathbb{R}^d$-valued random variable independent

[†]Institute of Applied Mathematics, University of Bonn, 53115 Endenicher Allee 60, Bonn, Germany
[‡]Department of Mathematics, University of Oslo, P.O. box 1053, Blindern, 0316, Oslo, Norway.
 E-mail: fabianah@math.uio.no
[§]Centre for Mathematical Sciences, Wilberforce Rd, Cambridge CB3 0WA, UK.
 E-mail: am3015@cam.ac.uk

of W, encoding the initial condition. In this case the DDSDE is also called a McKean–Vlasov SDE, after the pioneering work [39] where it was first introduced (setting $\varepsilon = 0$ corresponds to the Vlasov equation).

One reason for the importance of McKean–Vlasov equations is their connection to systems of N particles subject to a mean field interaction of the form

$$X_t^{i,N} = \int_0^t B_s\big(X_s^{i,N}, L^N(X_s^{(N)})\big)\,\mathrm{d}s + Y_t^i, \quad L^N(X_t^{(N)}) := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}} \qquad (1.2)$$

for all $i = 1, \ldots, N$; where $L^N(X_t^{(N)})$ stands for the empirical measure of the system at time t and Y^i are usually sampled as i.i.d. copies of Y. One expects the DDSDE (1.1) to be the mean field limit of (1.2) in the sense that, as N goes to infinity, $L^N(X_t^{(N)})$ converges weakly to $\mathcal{L}(X_t)$ with probability 1. Moreover for $Y = \xi + \sqrt{2\varepsilon}\,W$, by Itô calculus, the evolution of the marginal $\mathcal{L}(X_t)$ is given by the nonlinear Fokker–Planck equation (also called McKean–Vlasov equation)

$$\partial_t \rho + \nabla \cdot ((B_t(\,\cdot\,, \rho)\,\rho) = \varepsilon \Delta \rho, \quad \rho_0 = \mathcal{L}(\xi). \qquad (1.3)$$

In particular, both (1.1) and (1.3) provide a macroscopic, compact description of system (1.2), reducing its complexity. For this reason, DDSDEs have found applications in numerous fields, including galaxy formation, plasma physics, 2D fluid dynamics, bacteria chemotaxis, agent-based modelling, neuroscience, flocking and swarming dynamics; we refer the interested reader to [17], the review [35] and the references therein. In recent years they have also attracted a lot of attention due to their connection to mean-field games [38].

Classical results concerning the mean-field limit property go back to Dobrushin [22] for $\varepsilon = 0$ and to Sznitman [50] for general $\varepsilon \geq 0$, under Lipschitz continuity of B; let us also mention the work by Gärtner [30] which remains among the most general on the topic.

Recently the field has received a lot of attention both from the analytic and probabilistic communities. On one hand, new methods based on entropy inequalities [28, 36, 9] and modulated energy [47, 48] have allowed for the rigorous derivation of mean field limits for fairly singular B; on the other, DDSDEs with very irregular drifts are related to the flourishing field of regularization by noise phenomena, see [3, 42, 45, 14, 37, 15].

Most of the works mentioned above rely on exploiting either the connection to the PDE (1.3) or stochastic analysis tools related to the Brownian motion W. In this work we instead focus on studying the DDSDE (1.1) with as minimal assumptions on the noise Y as possible, usually merely continuity and finiteness of moments. We are strongly inspired by the recent work [17], where the authors show that in this setting, for globally Lipschitz drifts B, (1.1) is still the mean field limit of (1.2), although the connection to the PDE (1.3) breaks down; they present many additional results including the use of common noise, large deviations and a central limit theorem. Let us stress that the additive structure of the noise Y is such that the DDSDE (1.1) in integral form is pathwise meaningful, without any need of stochastic integration; in particular the results apply to naturally non-Markovian, non-martingale choices of Y, e.g. fractional Brownian motion.

A fundamental feature of [17] is the revisitation of a technique first introduced by Tanaka [51]; it allows, by a clever transformation of the underlying probability space, to transfer any stability estimate available for (1.1) into a bound for the corresponding empirical measure $L^N(X^{(N)})$ of the system (1.2). The argument is quite robust and allows one to deduce the mean field convergence property under very mild assumptions,

even when the family $\{Y^i\}_i$ is not i.i.d. but only requiring

$$L^N(Y^{(N)}) = \frac{1}{N}\sum_{i=1}^N \delta_{Y^i} \rightharpoonup \mathcal{L}(Y) \quad \mathbb{P}\text{-a.s.}$$

together with some finiteness of moments. The i.i.d. assumption indeed is quite strong, as it implies exchangeability of the system (see [50, Section I.2] for a detailed discussion) and sometimes one would prefer to avoid it, see [20, Example II]. In this regard, like [17], our work can be seen as another attempt to advertise the power of Tanaka's idea, which has found applications also in the rough path setting in [13, 2].

As the global Lipschitz assumption considered in [17] is quite restrictive, our main goal is conduct a deeper analysis of the DDSDE (1.1) providing sufficient conditions for existence, uniqueness and stability, which will also imply the mean field limit. As our setting includes in particular the typical case $Y = \xi + \sqrt{2\varepsilon}\,W$, it also serves as a review of many standard techniques used in McKean–Vlasov SDEs, including Yamada–Watanabe principles, compactness arguments and monotonicity assumptions; we believe one of the merits of this work is to highlight their extreme robustness, as they apply for any given Y. On the downside, since our results all apply in the case $\varepsilon = 0$, our noise can be very degenerate and we can never go beyond the classical Vlasov setting.

In this regard, another motivation for the current work is to provide a family of "baseline results" to be compared to the regularizing features of suitable nondegenerate choices of Y; in our second article [29] we will study in detail the well-posedness of (1.1) for highly irregular B when the noise is sampled as a fractional Brownian motion.

However, we stress that this work is more than just a survey of known facts; our aim is also to provide state-of-the-art results.

To illustrate this, for the rest of the introduction we will present a specific result from Section 4, concerning mean field convergence for McKean–Vlasov equations with potentially degenerate noise and convolutional drift. Namely, the drift B is assumed to be of the form $B_t(x,\mu) = (b_t * \mu)(x)$ and the noise $Y = \xi + W$ is given by independent variables (ξ, W) such that $\mathcal{L}(\xi)(\mathrm{d}x) = \rho(x)\,\mathrm{d}x$ for some density function ρ; we make no assumption on W other than continuity. The corresponding DDSDE becomes

$$X_t = \xi + \int_0^t \int_{\mathbb{R}^d} b_s(X_s - y)\,\mathcal{L}(X_s)(\mathrm{d}y) + W_t. \tag{1.4}$$

In the next statement we let $\mathcal{P}_r(C_T)$ denote the space of probability measures on the space of continuous functions C_T with finite $r < \infty$ moment, and refer to Section 1.1 below for details of some other notation used.

Theorem 1.1. *Assume* $b \in L_T^1 W_x^{1,q}$, $\mathrm{div}\,b \in L_T^1 L_x^\infty$ *and* $\rho \in L_x^p$ *for* $(p,q) \in (1,\infty)$ *satisfying*

$$q > d, \quad \frac{1}{p} + \frac{1}{q} < 1;$$

then there exists a unique solution to (1.4)*, which moreover satisfies*

$$\sup_{t\in[0,T]} \|\mathcal{L}(X_t)\|_{L_x^p} < \infty. \tag{1.5}$$

Suppose additionally that $\mathcal{L}(\xi + W) \in \mathcal{P}_r(C_T)$ *for all* $r < \infty$ *and that we are given a sequence* $\{Y^i, i \in \mathbb{N}\}$ *of* C_T*-valued random variables such that, for any* $r < \infty$,

$$\limsup_{N\to\infty} \|L^N(Y^{(N)})(\omega)\|_r < \infty, \quad L^N(Y^{(N)})(\omega) \rightharpoonup \mathcal{L}(\xi + W) \quad \text{for } \mathbb{P}\text{-a.e. }\omega;$$

then any solution $\{X^{i,N}\}$ *to* (1.2) *associated to* $\{Y^i\}_i$ *satisfies*

$$L^N(X^{(N)})(\omega) \rightharpoonup \mathcal{L}(X) \quad \text{for } \mathbb{P}\text{-a.e. }\omega.$$

The above statement applies in particular for the choice $W = 0$, i.e. when Y^i is taken as i.i.d. copies of ξ (see Lemma 5.5) and the result becomes a mean field limit for the corresponding Vlasov equation. Although singular drifts have already been considered in such settings, see [32, 33], to the best of our knowledge no general statements for $b \in L^1_T W^{1,q}_x$ are available in the literature; in any case, our proof technique is novel and quite simple, as we illustrate below.

As typical of transport PDEs, the assumption $\mathrm{div}\, b \in L^1_T L^\infty_x$ is used to propagate the regularity $\mathcal{L}(X_0) \in L^p_x$ and construct a solution X to (1.4) satisfying (1.5) by compactness arguments; the regularity $b \in L^1_T W^{1,q}_x$ for $q > d$, together with the one-sided Lipschitz estimates recently obtained in [11], is then used to show that this solution X is stable around *any other* solution $\tilde{X}$, not necessarily satisfying (1.5). Finally, Tanaka's argument allows to translate this information into an estimate for the corresponding particle system and obtain the conclusion.

We conclude this introduction by outlining the structure of the paper. In Section 1.1 we briefly recall some notation, conventions and well-known facts used throughout; Section 2 shortly revisits the theory of DDSDEs with Lipschitz drifts, in order to derive useful estimates for the sequel. Section 3 constitutes the main body of the paper, establishing sufficient conditions for existence and uniqueness in several situations for general B. Section 4 provides some more refined results in the more specific convolutional case $B = b * \mu$. Finally, in Section 5 we present Tanaka's idea and apply it to all the cases previously considered, deriving mean field limit results for each one. The appendices A and B contain analytic tools used throughout the paper.

1.1 Notations, conventions and well-known facts

Throughout the article we will always work on a finite time interval $[0, T]$, that may be arbitrarily large; in particular we do not investigate the long-time behaviour of solutions nor convergence to equilibrium. We will use the convention that whenever there exists a positive constant C such that $a \leq Cb$ we write $a \lesssim b$; if the constant C depends on a significant parameter p, we write $a \lesssim_p b$.

For a Banach space E we write $C_T E := C([0, T]; E)$ for the space of continuous E-valued maps $f : [0, T] \to E$. Similarly, for $p \geq 1$ we write $L^p_T E := L^p(0, T; E)$ for Bochner–Lebesgue space of maps of the same form. When $E = \mathbb{R}^d$ we simply write C_T, L^p_T etc.

Given a map $f : \mathbb{R}^d \to \mathbb{R}^m$ we write Df for the $d \times m$ matrix of first derivatives. For $p \in [1, \infty]$, we denote by $L^p(\mathbb{R}^d; \mathbb{R}^m)$, $W^{1,p}(\mathbb{R}^d; \mathbb{R}^m)$ (or simply L^p_x, $W^{1,p}_x$ when it does not cause confusion) the classical Lebesgue and Sobolev spaces. We treat the space $W^{1,\infty}_x$ as synonymous with the Lipschitz continuous functions. Whenever $p \in [1, \infty]$ and where it will not cause confusion, we write p' to denote the dual exponent to p, that is $1/p + 1/p' = 1$, with the interpretation $p = 1 \iff p' = \infty$. Similarly, we write C^0_b for the space of bounded, continuous functions from $\mathbb{R}^d$ to $\mathbb{R}^m$; $f \in C^1_b$ if f and Df belong to C^0_b.

Given a separable Banach space E, we denote by $\mathcal{P}(E)$ the set of probability measures over E; we write $\mu_n \rightharpoonup \mu$ for weak convergence of measures; i.e. convergence against continuous bounded functions.

Given $\mu, \nu \in \mathcal{P}(E)$, $\Pi(\mu, \nu)$ stands for the set of all possible couplings of (μ, ν), i.e. the subset of $\mathcal{P}(E \times E)$ with first and second marginals given respectively by μ and ν. For any $p \in [1, \infty)$, we define

$$d_p(\mu, \nu) := \inf_{m \in \Pi(\mu, \nu)} \left(\int_{E \times E} \|x - y\|^p_E \, m(\mathrm{d}x, \mathrm{d}y) \right)^{1/p}$$

which is a well defined quantity (possibly taking value $+\infty$). By [52, Theorem 4.1], an optimal coupling $\bar{m} \in \Pi(\mu, \nu)$ realizing the above infimum always exists. Noting that we

may equally describe the set $\Pi(\mu, \nu)$ as the collection of all E-valued random variables X, Y defined on the same probability space, such that $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$, we obtain the frequently used inequality

$$d_p(\mu, \nu) \leq \mathbb{E}[\|X - Y\|_E^p]^{1/p} \quad \forall (X, Y) \in \Pi(\mu, \nu).$$

We define $\mathcal{P}_p(E)$ to be set of p-integrable probability measures; that is, $\mu \in \mathcal{P}(E)$ s.t.

$$\|\mu\|_p := \left(\int_E \|x\|_E^p \, \mu(\mathrm{d}x) \right)^{1/p} < \infty.$$

It is well known that $d_p(\mu, \nu) < \infty$ for $\mu, \nu \in \mathcal{P}_p(E)$ and that $(\mathcal{P}_p(E), d_p)$ is a complete metric space, usually referred to as the p-Wasserstein space on E; let us stress however that our definition of $d_p(\mu, \nu)$ holds for all $\mu, \nu \in \mathcal{P}(E)$. We recall that, given a sequence $\{\mu_n\}_n \subset \mathcal{P}_p(E)$, $d_p(\mu_n, \mu) \to 0$ is equivalent to $\mu_n \rightharpoonup \mu$ weakly and $\|\mu_n\|_p \to \|\mu\|_p$, see [52, Theorem 6.9].

Given $\mu \in \mathcal{P}(\mathbb{R}^d)$, with a slight abuse of notation we will write $\mu \in L^q(\mathbb{R}^d)$ (or simply L_x^q) for $q \in [1, \infty]$ to indicate that μ admits a density $\mu(\mathrm{d}x) = \rho(x) \, \mathrm{d}x$ with respect to the d-dim. Lebesgue measure, such that $\rho \in L^q(\mathbb{R}^d)$.

Whenever it doesn't create confusion, we will use the bracket $\langle \cdot, \cdot \rangle$ to denote both scalar product in $\mathbb{R}^d$, $\langle x, y \rangle = \sum_i x_i y_i$, and duality pairing between functions and measures, $\langle f, \mu \rangle = \int f(z) \mu(\mathrm{d}z)$.

Throughout the article, whenever not mentioned explicitly, we will consider an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where the σ-algebra $\mathcal{F}$ is $\mathbb{P}$-complete, and $\mathcal{L}(X)$ is the push-forward measure of a random variable X on this probability space, i.e. $\mathcal{L}_{\mathbb{P}}(X) = \mathcal{L}(X) = \mathbb{P} \circ X^{-1} =: X \# \mathbb{P}$. For $p \geq 1$, we will frequently consider E-valued random variables in the space $L^p(\Omega, \mathcal{F}, \mathbb{P}; E)$ which for notational simplicity we will denote by $L_\Omega^p E$. This applies for instance to $L_\Omega^p C_T$, $L_\Omega^p \mathbb{R}^d$, etc. We say that a sequence of random variables $\{X^n\}_n$ converges to X in law if $\mathcal{L}(X^n) \rightharpoonup \mathcal{L}(X)$.

2 Well-posedness under Lipschitz assumptions

We start by treating DDSDEs with Lipschitz continuous drifts. Although this case was already treated in [17], we revisit it in order to derive useful a priori estimates for the sequel; to avoid repetitiveness, we will also present alternative proofs to those of [17].

In this section we consider the DDSDE

$$X_t = \int_0^t B_s(X_s, \mathcal{L}(X_s)) \, \mathrm{d}s + Y_t \tag{2.1}$$

under the assumption that $B : [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ satisfies

$$|B_t(x, \mu) - B_t(x', \nu)| \leq g_t (|x - x'| + d_p(\mu, \nu)), \quad |B_t(x, \mu)| \leq h_t (1 + |x| + \|\mu\|_p) \tag{2.2}$$

for some $p \in [1, \infty)$ and some $g, h \in L^1(0, T; \mathbb{R}^d)$.

Although the functions g, h could be unified, we keep them separate to highlight the differing roles they play in the following estimates. Let us also point out that if B is of the form $B_t(x, \mu) = \int b_t(x, y) \mu(\mathrm{d}y)$, then sufficient conditions for (2.2) to hold for $p = 1$ (and thus for any $p \in [1, \infty)$) are

$$|b_t(x, y) - b_t(x', y')| \leq g_t(|x - x'| + |y - y'|), \quad |b_t(x, y)| \leq h_t(1 + |x| + |y|).$$

We refrain for now from providing a rigorous definition of solution to (2.1), which will be discussed in detail in Section 3; for now we note that thanks to the assumptions on B, the integral appearing in (2.1) is pathwise meaningful and thus, given $Y \in L_\Omega^p C_T$ for some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that $X \in L_\Omega^p C_T$ solves (2.1) if the identity holds $\mathbb{P}$-a.s.

Proposition 2.1. *Assume B satisfies (2.2); then for any $Y \in L^p_\Omega C_T$ there exists a unique $X \in L^p_\Omega C_T$ solving (2.1). Moreover given Y^1, Y^2 as above, denoting by X^1, X^2 the associated solutions, it holds that*

$$\mathbb{E}[\|X^1 - X^2\|^p_{C_T}]^{1/p} \leq e^{2\|g\|_{L^1_T}} \mathbb{E}[\|Y^1 - Y^2\|^p_{C_T}]^{1/p}. \tag{2.3}$$

Proof. Given Y as above, we define a map $\mathcal{I}$ on $L^p_\Omega C_T$, by

$$\mathcal{I}(X)_\cdot := \int_0^\cdot B_s(X_s, \mathcal{L}(X_s)) \, \mathrm{d}s + Y_\cdot \, ;$$

it is immediate to check, using assumption (2.2), that $\mathcal{I}$ maps $L^p_\Omega C_T$ into itself. Let us endow C_T with the equivalent norm

$$\|x\|\tilde{}_{C_T} := \sup_{t \in [0,T]} \left\{ e^{-4 \int_0^t g_r \, \mathrm{d}r} |x_s| \right\},$$

and by extension $L^p_\Omega C_T$ with the norm $\| \cdot \|\tilde{}_{L^p C_T}$. We claim that $\mathcal{I}$ is a contraction on $(L^p_\Omega C_T, \| \cdot \|\tilde{}_{L^p C_T})$, which implies existence and uniqueness of a solution X by the Banach–Caccioppoli theorem.

By the definition of $\| \cdot \|\tilde{}_{C_T}$, for any $X^1, X^2 \in L^p_\Omega C_T$,

$$d_p(\mathcal{L}(X^1_t), \mathcal{L}(X^2_t)) \leq \mathbb{E}[|X^1_t - X^2_t|^p]^{1/p} \leq e^{4 \int_0^t g_r \, \mathrm{d}r} \|X^1 - X^2\|\tilde{}_{L^p C_T} \, ;$$

therefore for any X^1, X^2, by assumption (2.2) it holds that

$$e^{-4 \int_0^t g_r \, \mathrm{d}r} |\mathcal{I}(X^1)_t - \mathcal{I}(X^2)_t| \leq e^{-4 \int_0^t g_s \, \mathrm{d}s} \int_0^t |B_s(X^1_s, \mathcal{L}(X^1_s)) - B_s(X^2_s, \mathcal{L}(X^2_s))| \, \mathrm{d}s$$

$$\leq e^{-4 \int_0^t g_r \, \mathrm{d}r} \int_0^t g_s \left[|X^1_s - X^2_s| + d_p(\mathcal{L}(X^1_x), \mathcal{L}(X^2_s)) \right] \mathrm{d}s$$

$$\leq e^{-4 \int_0^t g_r \, \mathrm{d}r} \left(\int_0^t g_s \, e^{4 \int_0^s g_r \, \mathrm{d}r} \, \mathrm{d}s \right) [\|X^1 - X^2\|\tilde{}_{C_T}$$

$$+ \|X^1 - X^2\|\tilde{}_{L^p C_T}]$$

$$\leq \frac{1}{4} [\|X^1 - X^2\|\tilde{}_{C_T} + \|X^1 - X^2\|\tilde{}_{L^p C_T}].$$

Taking first the supremum over t on the l.h.s., then the L^p_Ω-norm on both sides, we obtain

$$\|\mathcal{I}(X^1) - \mathcal{I}(X^2)\|\tilde{}_{L^p C_T} \leq \frac{1}{2} \|X^1 - X^2\|\tilde{}_{L^p C_T}$$

showing that $\mathcal{I}$ is a contraction.

Now let X^i be solutions associated to Y^i for $i = 1, 2$. Their difference satisfies

$$|X^1_t - X^2_t| \leq \int_0^t |B_s(X^1_s, \mathcal{L}(X^1_s)) - B_s(X^2_s, \mathcal{L}(X^2_s))| \, \mathrm{d}s + |Y^1_t - Y^2_t|$$

$$\leq \int_0^t g_s \left(\sup_{r \leq s} |X^1_r - X^2_r| + \mathbb{E}\left[\sup_{r \leq s} |X^1_r - X^2_r|^p \right]^{1/p} \right) \mathrm{d}s + \|Y^1 - Y^2\|_{C_T} \, ;$$

taking the L^p_Ω-norm on both sides, applying Minkowski's integral inequality, we arrive at

$$\mathbb{E}\left[\sup_{r \leq t} |X^1_r - X^2_r|^p \right]^{1/p} \leq \int_0^t 2g_s \, \mathbb{E}\left[\sup_{r \leq s} |X^1_r - X^2_r|^p \right]^{1/p} \mathrm{d}s + \|Y^1 - Y^2\|_{C_T}.$$

Inequality (2.3) then follows from an application of Grönwall's lemma. $\qquad\square$

So far we have regarded X, Y as C_T-valued random variables on an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$; as the result holds on any such space, we may work with $(C_T, \mathcal{B}(C_T), \mu^Y)$, where $\mu^Y = \mathcal{L}(Y)$ and Y is realized as the canonical process.

Let us define the drift $\bar{b}(s, x) := B_s(x, \mathcal{L}(X_s))$; by assumption (2.2), $\bar{b}$ also satisfies a standard Lipschitz-type condition, so by classical ODE theory, for any $\omega \in C_T$ there exists a unique solution $x \in C_T$ to

$$x_{\cdot} = \int_0^{\cdot} \bar{b}(s, x_s) \, \mathrm{d}s + \omega.$$

which can be expressed as $x_t = \Phi(t, \omega)$ for some continuous map $\Phi : [0, T] \times C_T \to C_T$.

Since X is a solution to (2.1), it satisfies

$$X_{\cdot}(\omega) = \int_0^{\cdot} \bar{b}(s, X_s(\omega)) \, \mathrm{d}s + Y_{\cdot}(\omega) \quad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega$$

implying that it satisfies $X(\omega) = \Phi(\cdot, Y(\omega))$. In particular, since X can be expressed as a measurable function of Y, its law $\mathcal{L}(X) = \mu^X = \Phi \# \mu^Y$ on C_T is entirely determined by μ^Y. As a consequence, we can regard the solution map S^B coming from Proposition 2.1 not only as a map $L^p_\Omega C_T$ to itself, but also as a map from $\mathcal{P}_p(C_T)$ to itself, $S^B(\mu^Y) = \mu^X$, thus independent of the specific probability space in consideration.

We then obtain the following estimate, recovering the stability results from [17].

Corollary 2.2. *Assume B satisfies (2.2) and let S be defined as above, then*

$$d_p(S(\mu^1), S(\mu^2)) \le e^{2\|g\|_{L^1_T}} d_p(\mu^1, \mu^2) \quad \forall \mu^1, \mu^2 \in \mathcal{P}_p(C_T).$$

Proof. Given any coupling $m \in \mathcal{P}_p(C_T \times C_T)$ of μ^1, μ^2, considering Y^1, Y^2 as the canonical variables on $C_T \times C_T$, by Proposition 2.1 we have

$$d_p(S(\mu^1), S(\mu^2)) \le \mathbb{E}[\|X^1 - X^2\|^p_{C_T}]^{1/p} \le e^{2\|g\|_{L^1_T}} \mathbb{E}[\|Y^1 - Y^2\|^p_{C_T}]^{1/p};$$

the conclusion follows by choosing m to be the optimal coupling for $d_p(\mu^1, \mu^2)$. $\square$

By the above discussion, in what follows we could always deal with measures $\mu \in \mathcal{P}_p(C_T)$; however for the sake of computations we prefer to keep working with random variables X, Y defined on an abstract probability space (which can be taken as the canonical one if needed).

Given $h \in L^1(0, T; \mathbb{R}^d)$, we denote by f_h the associated modulus of continuity, namely for any $\delta > 0$,

$$f_h(\delta) := \sup_{s, t \in [0, T], |t-s| < \delta} \int_s^t |h_r| \, \mathrm{d}r;$$

let us also define, for any $x \in C_T$, the associate seminorm

$$[\![x]\!]_h := \sup_{s, t \in [0, T], s \ne t} \frac{|x_t - x_s|}{f_h(|t-s|)}$$

Lemma 2.3. *Let B satisfy (2.2), $Y \in L^p_\Omega C_T$; then there exist a constant C, only depending on $\|h\|_{L^1_T}$, such that the solution X to (2.1) satisfies*

$$\mathbb{E}[\|X\|^p_{C_T}]^{1/p} \le C(1 + \mathbb{E}[\|Y\|^p_{C_T}]^{1/p}) \tag{2.4}$$

as well as the pathwise bound

$$\|X(\omega)\|_{C_T} \le C\left(1 + \|Y(\omega)\|_{C_T}\right) \quad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{2.5}$$

Moreover, $X = Z + Y$ for some $Z \in L^p_\Omega C_T$ satisfying $Z_0 = 0$ and

$$\mathbb{E}[[\![Z]\!]^p_h]^{1/p} \le C \, \mathbb{E}[\|Y\|^p_{C_T}]^{1/p}. \tag{2.6}$$

Proof. Let X be a solution, then by assumption (2.2)

$$
\begin{aligned}
\sup_{r \le t} |X_r| &\le \int_0^t h_s (1 + |X_s| + \|\mathcal{L}(X_s)\|_p)\, \mathrm{d}s + |Y_t| \\
&\le \int_0^t h_s \left(1 + \sup_{r \le s} |X_r| + \mathbb{E}\left[\sup_{r \le s} |X_r|^p \right]^{1/p} \right) \mathrm{d}s + \|Y\|_{C_T}.
\end{aligned}
\tag{2.7}
$$

Taking the L_Ω^p-norm on both sides, applying Minkowski's integral inequality, we find

$$
\mathbb{E}\left[\sup_{r \le t} |X_r|^p \right]^{1/p} \le \int_0^t h_t \left(1 + 2\mathbb{E}\left[\sup_{r \le s} |X_r|^p \right]^{1/p} \right) \mathrm{d}s + \mathbb{E}[\|Y\|_{C_T}^p]^{1/p};
$$

inequality (2.4) then follows by Grönwall's lemma, with $C = \exp(2\|h\|_{L_T^1})(1 + \|h\|_{L_T^1})$.

On the other hand, we can apply Grönwall's lemma at a pathwise level to (2.7) to find

$$
\|X(\omega)\|_{C_T} \le e^{\|h\|_{L_T^1}} \left(\|h\|_{L_T^1} + \mathbb{E}[\|X\|_{C_T}^p]^{1/p} + \|Y(\omega)\|_{C_T} \right)
$$

which together with (2.4) implies (2.5), for a suitable choice of C.

Finally, since $Z_t = \int_0^t B_s(X_s, \mathcal{L}(X_s))$, applying again (2.2) we find

$$
|Z_t - Z_s| \le \int_s^t |h_r|\, \mathrm{d}r (1 + \|X\|_{C_T} + \mathbb{E}[\|X\|_{C_T}^p]^{1/p});
$$

combined with (2.4), (2.5) this implies the pathwise bound

$$
[\![Z(\omega)]\!]_h \le 1 + \|X(\omega)\|_{C_T} + \mathbb{E}[\|X\|_{C_T}^p]^{1/p} \lesssim_h 1 + \|Y(\omega)\|_{C_T} + \mathbb{E}[\|Y\|_{C_T}^p]^{1/p}.
$$

Inequality (2.6) readily follows taking the L_Ω^p-norm on both sides. $\qquad \square$

3 Refined criteria for existence and uniqueness

In many applications one cannot expect the drift B to be globally Lipschitz, thus making the results of Section 2 not applicable. Motivated by this fact, in this section we present several alternative conditions to establish existence and/or uniqueness of solutions to (2.1).

The first part of this section is focused on the different concepts of solution and the relations between them; Section 3.1 provides a general result of existence of weak solutions under continuity and linear growth assumptions; Section 3.2 deals with several refinements of the globally Lipschitz condition which are still sufficient to establish uniqueness. While we treat existence and uniqueness as separate problems, we provide examples of well-posedness results that combine the two. We leave open the possibility that ad-hoc existence or uniqueness results may be found in more specific cases, complementing the general results contained here.

The equations considered here are of the general form

$$
X_t = \int_0^t B_s(X_s, \mathcal{L}(X_s))\, \mathrm{d}s + Y_t \quad \forall t \in [0, T]
\tag{3.1}
$$

where $B : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ is a measurable map, for $p \in [1, \infty)$, and Y is a continuous process with finite p-moment. We detail a variety of assumptions on B below.

Definition 3.1 (Weak Solutions I). Given a drift B as above and $\mu^Y \in \mathcal{P}_p(C_T)$, we say that a tuple $(\Omega, \mathcal{F}, \mathbb{P}; X, Y)$, consisting of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable map $(X, Y) : \Omega \to C_T \times C_T$, is a weak solution to the DDSDE (3.1) on $[0, T]$ if:

1. $\mathcal{L}(Y) = \mu^Y$;

2. $\sup_{t \in [0,T]} \|\mathcal{L}(X_t)\|_p < \infty$;

3. $\mathbb{P}$-a.s. $\int_0^T |B_t(X_t, \mathcal{L}(X_t))| \, \mathrm{d}t < \infty$;

4. the relation (3.1) holds $\mathbb{P}$-a.s., the integral being interpreted in the Lebesgue sense.

We will refer to both μ^Y and Y as the input data of the DDSDE.

We readapt here the classical concepts of existence and uniqueness for SDEs to our setting.

Definition 3.2 (Uniqueness in Law). Given $\mu^Y \in \mathcal{P}(C_T)$, we say that weak uniqueness holds for the DDSDE (3.1) if for any pair of weak solutions $(\Omega^i, \mathcal{F}^i, \mathbb{P}^i; X^i, Y^i)$, $i = 1, 2$, such that $\mathcal{L}_{\mathbb{P}^i}(Y^i) = \mu^Y$, one has $\mathcal{L}_{\mathbb{P}^1}(X^1) = \mathcal{L}_{\mathbb{P}^2}(X^2) \in \mathcal{P}(C_T)$.

Definition 3.3 (Pathwise Uniqueness). Given $\mu^Y \in \mathcal{P}(C_T)$, we say that pathwise uniqueness holds for the DDSDE (3.1) if for any two solutions X^1, X^2 defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and w.r.t. to the same input process Y, it holds that $X^1 = X^2$ $\mathbb{P}$-a.s.

Definition 3.4 (Strong Existence). Given $\mu^Y \in \mathcal{P}(C_T)$, we say that strong existence holds for the DDSDE associated to μ^Y if there exists a solution $(\Omega, \mathcal{F}, \mathbb{P}, X, Y)$ such that the process X is adapted to the filtration $\mathcal{F}_t^Y = \sigma\{Y_s : s \in [0, t]\}$.

It follows from the definition that if strong existence holds, we can construct a solution X on *any* given probability space $(\Omega, \mathcal{F}, \mathbb{P}; Y)$ on which a process with $\mathcal{L}_{\mathbb{P}}(Y) = \mu^Y$ is defined.

We now provide several links between the various concepts of existence and uniqueness for DDSDEs; we start with a result in the style of the Yamada–Watanabe theorem [54, Prop. 1], showing that pathwise uniqueness implies uniqueness in law.

Proposition 3.5. *Let $p \in [1, \infty)$, $\mu^Y \in \mathcal{P}_p(C_T)$, $B : [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ measurable; if pathwise uniqueness holds for the DDSDE, then uniqueness in law holds as well.*

Proof. Let $(\Omega^i, \mathcal{F}^i, \mathbb{P}^i; X^i, Y^i)$, $i = 1, 2$ be two weak solutions to (3.1) with $\mathcal{L}(Y^1) = \mathcal{L}(Y^2) = \mu^Y$. In order to prove uniqueness in law, it suffices to construct a coupling $(\tilde{X}^1, \tilde{X}^2)$ of (X^1, X^2) such that $(\tilde{X}^1, \tilde{X}^2)$ solves (3.1) on the same probability space and with same input $\tilde{Y}$; indeed once this is done, pathwise uniqueness implies $\mathcal{L}_{\mathbb{P}^1}(X^1) = \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{X}^1) = \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{X}^2) = \mathcal{L}_{\mathbb{P}^2}(X^2)$.

To construct the aforementioned coupling, we follow the proof of [54, Prop. 1] based on disintegration of measures, for which we refer the reader to [1, Thm. 5.3.1].

Set $\mu^i := \mathcal{L}(X^i, Y^i) \in \mathcal{P}(C_T \times C_T)$ for $i = 1, 2$; since C_T is Polish, we can disintegrate μ^i as $\{\mu_y^i\}_{y \in C_T} \subset \mathcal{P}(C_T)$, μ_y^i representing the law of X^i given $Y^i = y$. Define a measure $\bar{\mu} \in \mathcal{P}(C_T \times C_T \times C_T)$ by setting

$$\mathrm{d}\bar{\mu}(x^1, x^2, y) := \mathrm{d}\mu_y^1(x^1) \, \mathrm{d}\mu_y^2(x^2) \, \mathrm{d}\mu^Y(y)$$

and consider the probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mu})$ given by $\Omega = C_T \times C_T \times C_T$, $\bar{\mathcal{F}}$ the $\bar{\mu}$-completion of $\mathcal{B}(C_T \times C_T \times C_T)$. We also define the projection maps $\pi_{x^1, x^2}(x^1, x^2, y) = (x^1, x^2)$, and similarly $\pi_{x^i, y}$ for $i = 1, 2$.

We claim that $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mu}; \pi_{x^1, x^2})$ is the desired coupling. By construction,

$$\mathcal{L}_{\bar{\mathbb{P}}}(\pi_{x^i, y}) = \mathrm{d}\mu_y^i(x^i) \, \mathrm{d}\mu^Y(y) = \mathcal{L}_{\mathbb{P}^i}(X^i, Y^i) \quad \text{for } i = 1, 2.$$

Moreover the DDSDE relation (3.1) can be written as $Y^i = F^i(X^i)$ for some measurable maps $F^i : C_T \to C_T$ given by $F^i(X^i)_\cdot := X^i - \int_0^\cdot B_s(X_s^i, \mathcal{L}_{\mathbb{P}^i}(X_s^i)) \, \mathrm{d}s$; the same relations must then hold for $\bar\mu$-a.e. (x^1, x^2, y), namely

$$x_t^i = \int_0^t B_s(x_s^i, \mathcal{L}_{\mathbb{P}^i}(X_t^i)) \, \mathrm{d}s + y_t = \int_0^t B_s(x_s^i, \mathcal{L}_{\bar\mu}(\pi_{x_t^i})) \, \mathrm{d}s + y_t \quad \text{for } i = 1, 2$$

where the property $\mathcal{L}_{\bar\mu}(\pi_{x_t^i}) = \mathcal{L}_{\mathbb{P}^i}(X_t^i)$ comes from the identity $\mathcal{L}_{\bar\mu}(\pi_{x^i, y}) = \mathcal{L}_{\mathbb{P}^i}(X^i, Y^i)$. Overall this shows that π_{x^1, x^2} is a coupling of X^1, X^2 solving the DDSDE (3.1) with same input π_y, $\mathcal{L}_{\bar\mu}(\pi_y) = \mu^Y$. $\qquad\square$

Remark 3.6. The proof above also contains the nontrivial information that, given any two weak solutions $(\Omega^i, \mathcal{F}^i, \mathbb{P}^i; X^i, Y^i)$, we may construct a coupling of them. This is also true in the case $\mu^{Y^1} = \mathcal{L}_{\mathbb{P}^1}(Y^1) \neq \mathcal{L}_{\mathbb{P}^2}(Y^2) = \mu^{Y^2}$, assuming we already have a coupling $m \in \mathcal{P}(C_T \times C_T)$ of (μ^{Y^1}, μ^{Y^2}) available. Following the notation of the proof, in this case one can take

$$\mathrm{d}\bar\mu(x^1, x^2, y^1, y^2) := \mathrm{d}\mu_{y^1}^1(x^1) \, \mathrm{d}\mu_{y^2}^2(x^2) m(\mathrm{d}y^1, \mathrm{d}y^2)$$

which defines $\bar\mu \in \mathcal{P}(C_T \times C_T \times C_T \times C_T)$; $(\pi_{x^1, y^1}, \pi_{x^2, y^2})$ is then a coupling of (X^1, Y^1), (X^2, Y^2).

The next proposition is in the style of [53, Theorem 6.3].

Proposition 3.7. *Let $p \in [1, \infty)$, $B : [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ measurable. Suppose that for any $\mu^Y \in \mathcal{P}_p(C_T)$ and any measurable $\mu : [0, T] \to \mathcal{P}_p(\mathbb{R}^d)$, strong existence and both pathwise uniqueness and uniqueness in law hold for the SDE*

$$X_t = \int_0^t B_s(X_s, \mu_s) \, \mathrm{d}s + Y_t. \tag{3.2}$$

Then weak existence (resp. uniqueness in law) holds for the DDSDE (3.1) *if and only if strong existence (resp. pathwise uniqueness) holds for it.*

Proof. Strong existence always implies weak existence and by Proposition 3.5, pathwise uniqueness implies weak uniqueness.

Suppose now that there exists a weak solution $(\Omega, \mathcal{F}, \mathbb{P}; X, Y)$ to the DDSDE, then X is also a solution to the SDE (3.2) for the choice $\mu_t = \mathcal{L}(X_t)$. By hypothesis, we can construct a strong solution $\bar X$ to the SDE on the canonical space $(C_T, \mathcal{B}(C_T), \mu^Y)$; since weak uniqueness holds for the SDE, $\mathcal{L}_{\mathbb{P}}(X) = \mathcal{L}_{\mu^Y}(\bar X)$, which implies that $\bar X$ is a strong solution to the DDSDE.

Finally, assume that weak uniqueness holds and let X^1, X^2 be two solutions defined on the same probability space and with same input Y. Then $\mathcal{L}(X_t^1) = \mathcal{L}(X_t^2) =: \mu_t$ and they solve the same SDE with drift $(t, x) \mapsto B_t(x, \mu_t)$, so by pathwise uniqueness for the SDE it holds $X^1 = X^2$ $\mathbb{P}$-a.s. $\qquad\square$

3.1 Existence

Here we study existence of solutions to (3.1) for continuous drifts B satisfying suitable linear growth conditions.

Assumption 3.8. *Given $p \in [1, \infty)$, $B : [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ is measurable and s.t.:*

i. For any $t \in [0, T]$, $B_t : \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ is uniformly continuous on bounded sets.

ii. There exists $h \in L_T^1$ such that for all $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d)$ one has

$$|B_t(x, \mu)| \leq h_t \left(1 + |x| + \|\mu\|_p\right). \tag{3.3}$$

Remark 3.9. Let δ_0 denote the Dirac delta in 0, which belongs to $\mathcal{P}_p(\mathbb{R}^d)$ for any $p \in [1, \infty)$; it's then easy to check that $\|\mu\|_p = d_p(\mu, \delta_0)$ for any $\mu \in \mathcal{P}_p$. Therefore if B satisfies Assumption 3.8, it also satisfies Assumption A.1 from Appendix A in the metric space $E = \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d)$.

The next result can be seen as an extension of Peano's theorem to DDSDEs.

Proposition 3.10. *Let B satisfy Assumption 3.8 for some $p \in [1, \infty)$ and $\mu^Y \in \mathcal{P}_p(C_T)$; then there exists a weak solution to (3.1) in the sense of Definition 3.1. Moreover, any weak solution $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \tilde{X}, \tilde{Y})$ satisfies the pathwise bound*

$$\|\tilde{X}\|_{C_T} \leq C(1 + \|\tilde{Y}\|_{C_T}) \quad \tilde{\mathbb{P}}\text{-a.s.} \tag{3.4}$$

for some constant $C = C(\|h\|_{L_T^1})$.

Proof. Estimate (3.4) follows in exactly the same manner as (2.5), since it only relies on the linear growth of B and Grönwall's lemma; thus we focus on establishing existence.

Thanks to Assumption 3.8 and Remark 3.9, we can apply Proposition A.3 from Appendix A to B; we then find a sequence of drifts $\{B^n\}_n$ such that each $B^n : [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ satisfies condition (2.2) with the same growth function h_t and B^n is Lipschitz with constant $g^n > 0$ independent of t. Denote by Y the canonical process associated to μ^Y; since $\mu^Y \in \mathcal{P}_p(C_T)$, we may apply Proposition 2.1 to obtain the existence of a sequence $\{X^n\}_n$, each X^n solving (3.1) with B replaced by B^n.

Lemma 2.3 furnishes us with suitable a priori estimates for the family $\{X^n\}_n$: there exists $C = C(\|h\|_{L_T^1})$ such that

$$\mathbb{E}\big[\|X^n\|_{C_T}^p\big]^{1/p} + \mathbb{E}\big[\llbracket Z^n \rrbracket_h^p\big]^{1/p} \leq C\Big(1 + \mathbb{E}[\|Y\|_{C_T}^p]^{1/p}\Big) \tag{3.5}$$

where the estimate is uniform over n and Z^n are given by the relation $X^n = Z^n + Y$; we recall that $\llbracket \cdot \rrbracket_h$ was the seminorm associated to the modulus of continuity f_h defined in Section 2.

By Arzelà–Ascoli, subsets of C_T of the form $\{\omega \in C_T : \omega_0 = 0, \llbracket \omega \rrbracket_h \leq R\}$ are compact; therefore by the moment estimate (3.5), the family of measures $\{\mathcal{L}(Z^n)\}_n$ are tight in C_T. Since $\mathcal{L}(Y)$ is also tight on C_T (being a probability measure on a Polish space), we deduce that the family $\{\mathcal{L}(Y, X^n)\}_n$ of probability measures on $C_T \times C_T$ is tight as well.

We can then apply Prokhorov's theorem to extract a (not relabelled) subsequence such that $\{\mathcal{L}(X^n, Y)\}_n$ converge weakly to a limit measure μ. By Skorokhod's theorem we can construct a new probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, carrying a family of random variables $(\tilde{X}^n, \tilde{Y}^n)$, $(\tilde{X}, \tilde{Y})$, such that $(\tilde{X}^n(\tilde{\omega}), \tilde{Y}^n(\tilde{\omega})) \to (\tilde{X}(\tilde{\omega}), \tilde{Y}(\tilde{\omega}))$ for $\tilde{\mathbb{P}}$-a.e. $\tilde{\omega}$, $\mathcal{L}_{\mathbb{P}}(X^n, Y) = \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{X}^n, \tilde{Y}^n)$ and $\mu = \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{X}, \tilde{Y})$. It remains to show that $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \tilde{X}, \tilde{Y})$ is a weak solution to (3.1).

First of all, observe that $\mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{Y}^n) = \mathcal{L}_{\mathbb{P}}(Y) = \mu^Y$ and $\mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{Y}^n) \rightharpoonup \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{Y})$ so that $\mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{Y}) = \mu^Y$, as well as $\mathbb{E}_{\tilde{\mathbb{P}}}[\|\tilde{Y}^n\|_{C_T}^p] \to \mathbb{E}_{\tilde{\mathbb{P}}}[\|\tilde{Y}\|_{C_T}^p]$; together with the convergence $\tilde{Y}^n \to \tilde{Y}$ $\tilde{\mathbb{P}}$-a.s., we find that $\tilde{Y}^n \to \tilde{Y}$ in $L_{\tilde{\Omega}}^p C_T$. Since $(\tilde{Y}^n, \tilde{X}^n)$ have the same law as (Y, X^n), they are still solutions to the DDSDEs associated to B^n; in particular they satisfy the pathwise bound (2.5). This information together with the convergence $\tilde{Y}^n \to \tilde{Y}$ in $L_{\tilde{\Omega}}^p C_T$ implies that $\{\tilde{X}^n\}_n$ is uniformly p-integrable. However, since we also have $\tilde{X}^n \to \tilde{X}$ $\tilde{\mathbb{P}}$-a.s., by Vitali's convergence theorem $\tilde{X}^n \to \tilde{X}$ in $L_{\tilde{\Omega}}^p C_T$. In particular $d_p(\mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{X}_t^n), \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{X}_t)) \to 0$ and overall we have the $\tilde{\mathbb{P}}$-a.s. convergence $(\tilde{X}_t^n, \mathcal{L}(\tilde{X}_t^n)) \to (\tilde{X}_t, \mathcal{L}(\tilde{X}_t))$ in $\mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d)$ for any $t \in [0, T]$.

The above convergence also implies that $\tilde{X}$ satisfies the pathwise bound (3.4); together with assumption (3.3) this shows that points (2) and (3) of Definition 3.1 are

satisfied. To conclude that identity (3.1) holds $\tilde{\mathbb{P}}$-a.s. it then suffices to show that $\int_0^\cdot B_s^n(\tilde{X}_s^n, \mathcal{L}(\tilde{X}_s^n))\,\mathrm{d}s$ converge to $\int_0^\cdot B_s(\tilde{X}_s, \mathcal{L}(\tilde{X}_s))\,\mathrm{d}s$.

Since B^n converge to B uniformly on compact sets and $(\tilde{X}_t^n, \mathcal{L}(\tilde{X}_t^n)) \to (\tilde{X}_t, \mathcal{L}(\tilde{X}_t))$ $\tilde{\mathbb{P}}$-a.s., it holds $B^n(\tilde{X}_t^n, \mathcal{L}(\tilde{X}_t^n)) \to B(\tilde{X}_t, \mathcal{L}(\tilde{X}_t))$ $\tilde{\mathbb{P}}$-a.s. as well. Moreover since B^n and B both satisfy the growth assumption (3.3), we have

$$\sup_{t\in[0,T]}\left|\int_0^t B_s^n(\tilde{X}_s^n, \mathcal{L}(\tilde{X}_s^n))\,\mathrm{d}s - \int_0^t B_s(\tilde{X}_s, \mathcal{L}(\tilde{X}_s)\,\mathrm{d}s\right|$$
$$\leq \int_0^T |B_s^n(\tilde{X}_s^n, \mathcal{L}(\tilde{X}_s^n)) - B_s(\tilde{X}_s, \mathcal{L}(\tilde{X}_s))|\,\mathrm{d}s$$

with the second integrand satisfying the uniform bound

$$|B_s^n(\tilde{X}_s^n, \mathcal{L}(\tilde{X}_s^n)) - B_s(\tilde{X}_s, \mathcal{L}(\tilde{X}_s))| \leq \|h\|_{L_T^1}(2 + \|\tilde{X}^n\|_{C_T} + \|\mathcal{L}(\tilde{X}^n)\|_p + \|\tilde{X}\|_{C_T} + \|\mathcal{L}(\tilde{X})\|_p)$$
$$\lesssim_h 1 + \|\tilde{Y}^n\|_{C_T} + \|\tilde{Y}\|_{C_T} + \|\mathcal{L}(\tilde{Y})\|_p$$

where we used again the a priori bound (3.4), as well as $\mathcal{L}(\tilde{Y}^n) = \mathcal{L}(\tilde{Y})$. But $\|\tilde{Y}^n\|_{C_T} \to \|\tilde{Y}\|_{C_T}$ $\tilde{\mathbb{P}}$-a.s., thus the previous integrands must be uniformly integrable.

Since $[0,T]$ is finite, we may now apply again Vitali's convergence theorem to give that

$$\int_0^\cdot B^n(s, \tilde{X}_s^n, \mathcal{L}(\tilde{X}_s^n))\,\mathrm{d}s \to \int_0^\cdot B(s, \tilde{X}_s, \mathcal{L}(\tilde{X}_s))\,\mathrm{d}s \quad \tilde{\mathbb{P}}\text{-a.s.}$$

which implies the conclusion. $\qquad\square$

Remark 3.11. Condition *i.* from Assumption 3.8 can be verified in a variety of situations; it suffices to check that for any $t \in [0,T]$, the map $(x,\mu) \mapsto B_t(x,\mu)$ is continuous under a weak topology of $\mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d)$ for which bounded balls are compact (as we can then apply Heine–Cantor theorem).

For instance, if $B_t(\cdot,\cdot)$ is continuous on $\mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d)$, then it's uniformly continuous on bounded balls on $\mathbb{R}^d \times \mathcal{P}_q(\mathbb{R}^d)$ for any $q > p$, due to the compact embedding $\mathcal{P}_q \hookrightarrow \mathcal{P}_p$. Similarly, in the case $p = 1$ it's enough to require

$$|B_t(x,\mu) - B_t(y,\nu)| \lesssim_t F\left(|x-y| + \inf_{m\in\Pi(\mu,\nu)}\int_{\mathbb{R}^{2d}}|x'-y'|^\theta\,m(\mathrm{d}x',\mathrm{d}y')\right)$$

for some $\theta \in (0,1)$ and some continuous F such that $F(0) = 0$. On the use of more abstract weak topologies guaranteeing sequential compactness of bounded balls of $\mathcal{P}(\mathbb{R}^d)$ we also refer to [30, Section B].

Remark 3.12. The reader might wonder if we can replace the growth condition in (3.3) by a monotonicity assumption like $\langle B_t(x,\mu), x\rangle \lesssim 1 + |x|^2$; there are several difficulties in doing so. On a technical side, one would need to control $\int_0^t \langle X_s, \dot{Y}_s\rangle$, which require either Y to be of bounded variation or some more refined integration theory (Itô, rough paths) to be available; recall that here we are not imposing any assumption on Y other than continuity.

More importantly, already in the case of standard ODEs there are examples of finite time blow-up for suitable choices of b and continuous Y, see [18, Section 3.3]; the monotonicity assumption can be replaced by more refined criteria, see [44] and the recent work [8], which however don't seem to transfer easily to the distribution-dependent setting.

3.2 Uniqueness

In this section we provide various conditions for uniqueness of solutions to the DDSDE (3.1), thus extending already known results for ODEs and SDEs to the the distribution dependent case. We do not aim at maximal generality, as the criteria from Sections 3.2.1-3.2.3 could be combined together to create new ones; rather we aim to illustrate how many classical assumptions on B still work perfectly without any need for Y to be Brownian or Markovian.

Whenever possible, we try to keep uniqueness statements separate from their existence counterparts (which in our setting always comes from Proposition 3.10) and only afterwards unite them together in a well-posedness result. Depending on the DDSDE in consideration, ad-hoc existence results which go beyond Assumption 3.8 might be available; they can still be combined with the uniqueness criteria presented here, which are thus of independent interest.

3.2.1 Osgood-type condition

It is well known that classical ODEs admit a unique solution under Osgood type conditions on the drift (originally proposed in [43]); we include here an analogous statement for DDSDEs.

Assumption 3.13. *It holds that*

$$|B_t(x, \mu) - B_t(y, \nu)| \leq h_t\, f(|x - y| + d_1(\mu, \nu)), \tag{3.6}$$

for all $(x, y) \in \mathbb{R}^{2d}$, $(\mu, \nu) \in \mathcal{P}_1(\mathbb{R}^d)^2$ and $t \in [0, T]$, where $h \in L^1_T$ and $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a modulus of continuity satisfying

$$\int_0^\varepsilon \frac{1}{f(u)}\, du = \infty \quad \text{for any } \varepsilon > 0. \tag{3.7}$$

Remark 3.14. By a result of Stechkin which can be found as [40, Lem. A], given any modulus of continuity f, there exists another concave modulus $\tilde{f}$ satisfying the two-sided bound $\tilde{f} \leq f \leq 2\tilde{f}$. Thus we can always assume the modulus f appearing in (3.7) to be concave, which also implies it is increasing and subadditive.

Before proving our uniqueness results, we recall Bihari's inequality, see e.g. [4].

Proposition 3.15. *Let $f : [a, b] \to (0, +\infty)$ be a continuous, monotone, strictly positive function and let x and h be two functions on an interval $[t_0, t_0+T]$ such that $x([t_0, t_0+T]) \subset [a, b]$, $h \geq 0$. Assume that for some $\alpha \in [a, b]$ one has*

$$x_t \leq \alpha + \int_{t_0}^t h_s\, f(x_s)\, \mathrm{d}s, \quad t \in [t_0, t_0 + T];$$

then

$$x_t \leq G^{-1}\left(G(a) + \int_{t_0}^t h_s\, \mathrm{d}s\right), \quad t \in [t_0, t_0 + \tilde{T}], \tag{3.8}$$

where

$$G(u) := \int_{x_0}^u \frac{1}{f(r)}\, \mathrm{d}r \tag{3.9}$$

and $\tilde{T} \leq T$ is chosen such that

$$G(a) + \int_{t_0}^t h(s)\, \mathrm{d}s \in \mathrm{Dom}(G^{-1}), \quad \forall t \in [t_0, t_0 + \tilde{T}].$$

Remark 3.16. It's easy to check, using the definition of G, that functions of the form $a \mapsto G^{-1}(G(a) + \kappa)$ as appearing on the r.h.s. of (3.8) do not depend on the choice of x_0 (as long as a belongs to the corresponding domain). Since any concave modulus of continuity f can grow at most linearly at infinity, under the assumption (3.7) we deduce that G satisfies $\lim_{x \to 0} G(x) = -\infty$, $\lim_{x \to +\infty} G(x) = +\infty$ and $\mathrm{Dom}(G^{-1}) = \mathbb{R}$, so that we can take $\tilde{T} = T$.

We now provide a stability estimate for solutions defined on the same probability space; given two drifts $B^1, B^2 : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}^d$, we introduce the notation

$$\|B^1 - B^2\|_\infty := \sup_{(s,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} |B_s^1(x, \mu) - B_s^2(x, \mu)|.$$

Proposition 3.17. *Let B^1 and B^2 be two drifts satisfying Assumption 3.13 for the same f, h. Assume X^1 and X^2 are two solutions to* (3.1) *defined on the same probability space with input (Y^1, B^1) and (Y^2, B^2) respectively. Then*

$$\mathbb{E}[\|X^1 - X^2\|_{C_T}] \leq G^{-1}\left(G\left(\mathbb{E}[\|Y^1 - Y^2\|_{C_T}] + \|B^1 - B^2\|_\infty T\right) + 2\|h\|_{L_T^1}\right) \quad (3.10)$$

where G is defined in (3.9). *In particular, pathwise uniqueness and uniqueness in law hold for the DDSDE.*

Proof. We begin by observing that

$$X_t^1 - X_t^2 = \int_0^t B_s^1(X_s^1, \mathcal{L}(X_s^1)) - B_s^2(X_s^2, \mathcal{L}(X_s^2)) \, \mathrm{d}s + Y_t^1 - Y_t^2;$$

adding and subtracting $B_s^1(X_s^2, \mathcal{L}(X_s^2))$, applying Assumption 3.13 and the monotonicity of f, we see that

$$\sup_{r \leq t} |X_r^1 - X_r^2| \leq \|Y^1 - Y^2\|_{C_T} + T\|B^1 - B^2\|_\infty$$

$$+ \int_0^t h_s \, f(|X_s^1 - X_s^2| + d_1(\mathcal{L}(X_s^1), \mathcal{L}(X_s^2))) \, \mathrm{d}s$$

$$\leq \|Y^1 - Y^2\|_{C_T} + T\|B^1 - B^2\|_\infty$$

$$+ \int_0^t h_s \, f\left(\sup_{r \leq s} |X_r^1 - X_r^2| + \mathbb{E}\left[\sup_{r \leq s} |X_r^1 - X_r^2|\right]\right) \, \mathrm{d}s$$

Taking expectation on both sides, using subadditivity and concavity of f along with Jensen's inequality, we get that

$$\mathbb{E}\left[\sup_{r \leq t} |X_r^1 - X_r^2|\right] \leq \mathbb{E}[\|Y^1 - Y^2\|_{C_T}] + T\|B^1 - B^2\|_\infty + \int_0^t 2h_s \, f\left(\mathbb{E}\left[\sup_{r \leq s} |X_r^1 - X_r^2|\right]\right) \, \mathrm{d}s.$$

Applying Proposition 3.15 readily gives estimate (3.10).

Regarding the last statement, observe that for $Y^1 = Y^2$, $B^1 = B^2$ estimate (3.10) implies $X^1 \equiv X^2$, proving pathwise uniqueness. Uniqueness in law then follows from Proposition 3.5. $\square$

Combining Assumptions 3.8 and 3.13 we can obtain strong existence and uniqueness of solutions; moreover we can recast the stability estimate (3.10) into a corresponding bound on the Wasserstein distance between the laws of solutions.

Theorem 3.18. *Let B^1 and B^2 satisfy Assumption 3.8, with $p = 1$ and Assumption 3.13 for the same f, h and let μ^{Y^1}, $\mu^{Y^2} \in \mathcal{P}_1(C_T)$ be two input laws. Then for $i = 1, 2$ there*

exist strong solutions, unique both pathwise and in law, with laws μ^{X^i}, associated to (B^i, μ^{Y^i}). Furthermore, there exists a continuous, increasing function $M : \mathbb{R}_+ \to \mathbb{R}_+$, $M(0) = 0$, only depending on f and h, such that

$$d_1(\mu^{X^1}, \mu^{X^2}) \leq M\left(d_1(\mu^{Y^1}, \mu^{Y^2}) + T\|B^1 - B^2\|_\infty\right). \tag{3.11}$$

Proof. Since both B^1 and B^2 satisfy Assumption 3.8, weak existence holds by Proposition 3.10; Assumption 3.13 and Proposition 3.7 then imply strong existence as well. Pathwise uniqueness and uniqueness in law instead are consequences of Proposition 3.17.

Now let (Y^1, Y^2) be an optimal coupling for $d_1(\mu^{Y^1}, \mu^{Y^2})$, which induces a coupling (X^1, X^2) for μ^{X^1}, μ^{X^2} by strong existence and uniqueness in law. Define $M : (0, +\infty) \to (0, +\infty)$ by

$$M(r) := G^{-1}\left(G(r) + 2\|h\|_{L_T^1}\right)$$

for G defined as in (3.9); by the properties $\lim_{x \to 0} G(x) = -\infty$, $\lim_{x \to -\infty} G^{-1}(x) = 0$ we deduce that M extends uniquely and continuously at $r = 0$ with $M(0) = 0$.

But then inequality (3.10) and the choice of optimal coupling (Y^1, Y^2) readily implies

$$d_1(\mu^{X^1}, \mu^{X^2}) \leq \mathbb{E}[\|X^1 - X^2\|_{C_T}] \leq M\left(\mathbb{E}[\|Y^1 - Y^2\|_{C_T}] + T\|B^1 - B^2\|_\infty\right)$$

and the conclusion follows. $\qquad\square$

Let us provide some relevant examples of drifts B satisfying Assumption 3.13 above.

Example 3.19. *Consider the case of a true McKean–Vlasov drift, i.e. B of the form*

$$B_t(x, \mu) = \int_{\mathbb{R}^d} b_t(x, y)\, \mu(\mathrm{d}y)$$

for some measurable function $b : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$. Then Assumption 3.13 is satisfied if there exist $h \in L_T^1$, and a (concave) modulus of continuity f satisfying (3.7) such that

$$|b_t(x, z) - b_t(y, z')| \leq h_t\, f(|x - y| + |z - z'|).$$

Indeed, for any $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and any coupling $m \in \Pi(\mu, \nu)$ it holds

$$\begin{aligned}
|B_t(x, \mu) - B_t(y, \nu)| &= \left| \int_{\mathbb{R}^d} b_t(x, z)\mu(\mathrm{d}z) - \int_{\mathbb{R}^d} b_t(y, z')\nu(\mathrm{d}z') \right| \\
&\leq \int_{\mathbb{R}^{2d}} |b_t(x, z) - b_t(y, z')|\, m(\mathrm{d}z, \mathrm{d}z') \\
&\leq h_t \int_{\mathbb{R}^{2d}} f(|x - y| + |z - z'|)\, m(\mathrm{d}z, \mathrm{d}z') \\
&\leq h_t \left[f(|x - y|) + \int_{\mathbb{R}^{2d}} f(|z - z'|)\, m(\mathrm{d}z, \mathrm{d}z') \right] \\
&\leq h_t \left[f(|x - y|) + f\left(\int_{\mathbb{R}^{2d}} |z - z'|\, m(\mathrm{d}z, \mathrm{d}z') \right) \right]
\end{aligned}$$

where we used monotonicity, subadditivity and concavity of f as well as Jensen's inequality; minimizing over $m \in \Pi(\mu, \nu)$ gives (3.6).

Of particular relevance are convolutional drifts $B_t(x, \mu) = (b_t * \mu)(x) = \int b_t(x - z)\, \mu(\mathrm{d}z)$, in which case the above condition reduces to

$$|b_t(x) - b_t(y)| \leq h_t\, f(|x - y|), \quad \forall (t, x, y) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d.$$

3.2.2 Monotone drifts

Another classical assumption in ODE theory, that goes beyond the Lipschitz setting, is a monotonicity condition (sometimes also referred to as one-sided Lipschitz condition). Similar assumptions in the DDSDE setting have been employed for example in [30, Sec. 5]; see also the more recent works [53] (Assumption (H2) therein) and [23]. Finally, let us mention that monotonicity may be interpreted as a Lyapunov condition on the energy $|\cdot|^2$ and is therefore also close in spirit to the assumptions considered in [41], [31].

We consider the following working assumption for the drift B.

Assumption 3.20. *Given $p \in (1, \infty)$, there exists an $h \in L_T^1$ such that*

$$\langle x - y, B_t(x, \mu) - B_t(y, \nu) \rangle \leq h_t \big[|x - y|^2 + |x - y|\, d_p(\mu, \nu) \big] \tag{3.12}$$

uniformly over $x, y \in \mathbb{R}^d$, and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$.

Proposition 3.21. *Let $B : [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ be a measurable map satisfying Assumption 3.20 for some $p \in (1, \infty)$; then pathwise uniqueness and uniqueness in law hold for (3.1).*

Proof. By virtue of Proposition 3.5, we only need to establish pathwise uniqueness. Let X^1, X^2 be two solutions to (3.1) defined on the same probability space and with same input Y; setting $\mu_t^i = \mathcal{L}(X_t^i)$ for $i = 1, 2$, it holds that

$$X_t^i = \int_0^t B_s(X_s^i, \mu_s^i)\, \mathrm{d}s + Y_t$$

and the difference $X^1 - X^2$ is an absolutely continuous path. By Assumption 3.20 we have that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} |X_t^1 - X_t^2|^p &= p\, |X_t^1 - X_t^2|^{p-2} \langle X_t^1 - X_t^2, B_t(X_t^1, \mu_t^1) - B_t(X_t^2, \mu_t^2) \rangle \\
&\leq p\, h_t \big[|X_t^1 - X_t^2|^p + |X_t^1 - X_t^2|^{p-1}\, d_p(\mu_t^1, \mu_t^2) \big] \\
&\lesssim_p h_t \big(|X_t^1 - X_t^2|^p + d_p(\mu_t^1, \mu_t^2)^p \big)
\end{aligned}$$

where in the last passage we used the basic inequality $ab \lesssim a^{\frac{p}{p-1}} + b^p$. Since $X_0^1 = Y_0 = X_0^2$, integrating and taking expectations we obtain

$$\mathbb{E}\big[|X_t^1 - X_t^2|^p \big] \lesssim \int_0^t h_s \Big(\mathbb{E}[|X_s^1 - X_s^2|^p] + d_p(\mu_s^1, \mu_s^2)^p \Big) \mathrm{d}s \lesssim \int_0^t h_s \mathbb{E}[|X_s^1 - X_s^2|^p]\, \mathrm{d}s$$

and by Grönwall's lemma we deduce that $\mathbb{E}[|X_t^1 - X_t^2|^p] = 0$ for all $t \in [0, T]$; since X^i are continuous paths, pathwise uniqueness follows. $\square$

Remark 3.22. If $p \geq 2$, we can further weaken condition (3.12) by requiring instead

$$\langle x - y, B_t(x, \mu) - B_t(y, \nu) \rangle \leq h_t \big[|x - y|^2 + d_p(\mu, \nu)^2 \big] \tag{3.13}$$

uniformly over t, x, y, μ, ν. Indeed, going through the same computations as in the proof above, gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} |X_t^1 - X_t^2|^p &\leq p\, h_t \big[|X_t^1 - X_t^2|^p + |X_t^1 - X_t^2|^{p-2}\, d_p(\mu_t^1, \mu_t^2)^2 \big] \\
&\lesssim_p h_t \big[|X_t^1 - X_t^2|^p + d_p(\mu_t^1, \mu_t^2)^p \big]
\end{aligned}$$

where we used the basic inequality $ab \lesssim a^{\frac{p}{p-2}} + b^{\frac{p}{2}}$ which holds since $p/2 \geq 1$.

Example 3.23. *It is clear that any Lipschitz B satisfies Assumption 3.20; more generally, one can take $B_t(x, \mu) = F_t(x) + G(x, \mu)$ for globally Lipschitz G and F satisfying*

$$\langle F_t(x) - F_t(y), x - y \rangle \leq h_t |x - y|^2. \tag{3.14}$$

A similar computation holds for $B_t(x, \mu) = F_t(x)G(\mu)$ with F as above, once we additionally impose $\|F_t\|_{L^\infty} \leq h_t$ and $G : \mathcal{P}_p(\mathbb{R}^d) \to \mathcal{P}_p(\mathbb{R}^d)$ is globally Lipschitz and bounded.

In the case $h \equiv C$ for some $C \in \mathbb{R}$, condition (3.14) is satisfied if $F = -\nabla V$ for some $V : \mathbb{R}^d \to \mathbb{R}$ such that $x \mapsto V(x) + \lambda |x|^2/2$ is convex for some $\lambda \geq 0$; interesting cases include $F(x) = -xe^{|x|^2}$ and $F(x) = -|x|^{\gamma-1}x$ for $\gamma \geq 0$ where for $\gamma \in [0, 1)$, F is not even locally Lipschitz.

Although Assumption 3.20 covers a wide range of drifts, sometimes it's natural to further weaken it to an *integrated* monotonicity assumption.

Lemma 3.24. *Suppose $B : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ satisfies*

$$\int_{\mathbb{R}^{2d}} \langle x - y, B_t(x, \mu) - B_t(y, \nu) \rangle \, m(\mathrm{d}x, \mathrm{d}y) \leq h_t \int_{\mathbb{R}^{2d}} |x - y|^2 \, m(\mathrm{d}x, \mathrm{d}y) \tag{3.15}$$

uniformly over all possible $t \in [0, T]$, $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and $m \in \Pi(\mu, \nu)$, for some integrable $h : [0, T] \to \mathbb{R}$; then pathwise uniqueness and uniqueness in law hold for the DDSDE.

Proof. By virtue of Proposition 3.5, we only need to establish pathwise uniqueness. Let X^1, X^2 be two solutions defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with same input Y. It follows that $X^1 - X^2$ is a process of bounded variation satisfying $\mathbb{P}$-a.s.

$$|X_t^1 - X_t^2|^2 = 2 \int_0^t \langle X_t^1 - X_t^2, B_s(X_s^1, \mathcal{L}(X_s^1)) - B_s(X_s^2, \mathcal{L}(X_s^2)) \rangle \, \mathrm{d}s \quad \forall t \in [0, T].$$

Taking expectation and applying assumption (3.15) for

$$m = \mathbb{P} \circ (X_s^1, X_s^2)^{-1} \in \Pi(\mathcal{L}(X_s^1), \mathcal{L}(X_s^2)),$$

we arrive at

$$\mathbb{E}[|X_t^1 - X_t^2|^2] \leq 2 \int_0^t h_s \, \mathbb{E}[|X_s^1 - X_s^2|^2] \, \mathrm{d}s$$

and the conclusion follows by Grönwall's lemma. $\qquad\square$

Example 3.25. *Let $b : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a function satisfying*

$$b(x, x') = -b(x', x), \quad \langle b(x, x') - b(y, y'), x - x' - y + y' \rangle \leq C(|x - y|^2 + |x' - y'|^2)$$

for some constant $C \in \mathbb{R}$, uniformly over $x, x', y, y' \in \mathbb{R}^d$; then the McKean-Vlasov drift $B_t(x, \mu) = \int_{\mathbb{R}^d} b(x, x') \, \mu(\mathrm{d}x')$ satisfies condition (3.15) with $h_t = C$.

Indeed, for any $m \in \Pi(\mu, \nu)$ it holds

$$\int_{\mathbb{R}^{2d}} \langle x - y, B_t(x, \mu) - B_t(y, \nu) \rangle \, m(\mathrm{d}x, \mathrm{d}y)$$

$$= \int_{\mathbb{R}^{2d} \times \mathbb{R}^{2d}} \langle x - y, b(x, x') - b(y, y') \rangle \, m(\mathrm{d}x, \mathrm{d}y) m(\mathrm{d}x', \mathrm{d}y')$$

$$= \frac{1}{2} \int_{\mathbb{R}^{2d} \times \mathbb{R}^{2d}} \langle x - y - x' + y', b(x, x') - b(y, y') \rangle \, m(\mathrm{d}x, \mathrm{d}y) m(\mathrm{d}x', \mathrm{d}y')$$

where the last passage follows by exchanging (x, y) and (x', y') and applying $b(x', x) = -b(x, x')$; by the monotonicity condition on b we readily obtain the conclusion.

For convolutional drifts $B = b * \mu$ it suffices to require

$$b(-x) = -b(x), \quad \langle x - y, b(x) - b(y) \rangle \leq 2C|x - y|^2 \quad \forall x, y \in \mathbb{R}^d;$$

as in Example 3.23, this is true for instance if $b(x) = -\nabla V(x)$ for some even function $V : \mathbb{R}^d \to \mathbb{R}$ such that $x \mapsto V(x) + \lambda|x|^2$ is convex for some $\lambda \geq 0$. This class includes drifts not satisfying the linear growth assumption 3.8, but for which uniqueness still holds.

Let us mention that the special cases $b(x) = -\lambda|x|^{\gamma-1}x$ for $\lambda, \gamma > 0$ are also the drift terms considered in [53, 27, 12]; the setting of these works however, also includes a non-linear diffusion term, associated to kernels of the form $a(x) = |x|^{\gamma-1}(|x|^2 I_d - x \otimes x)$.

It follows from the above results that if B satisfies Assumption 3.8 and one between Assumption 3.20 or conditions (3.13), (3.15), then for any $\mu^Y \in \mathcal{P}_p(C_T)$ there exists a unique solution $\mu^X \in \mathcal{P}_p(C_T)$; this correspondence defines a solution map $S^B : \mathcal{P}_p(C_T) \to \mathcal{P}_p(C_T)$. Unfortunately we cannot establish Lipschitz continuity for S^B, the issue being similar to that of Remark 3.12 and related to controlling $\int_0^t \langle X_s^1 - X_s^2, \mathrm{d}(Y_s^1 - Y_s^2) \rangle$; we can still at least show continuity.

Lemma 3.26. *Let the drift B satisfy Assumption 3.8 and one between Assumption 3.20 or conditions (3.13), (3.15); then the solution map $S^B : \mathcal{P}_p(C_T) \to \mathcal{P}_p(C_T)$ is continuous.*

Proof. Given a sequence $\{\mu^{Y^n}\}_n$ with $\mu^{Y^n} \to \mu^Y$ in $\mathcal{P}_p(C_T)$, we need to show that $\mu^{X^n} = S^B(\mu^{Y^n}) \to S^B(\mu^Y) = \mu^X$. By the assumptions and estimate (3.4) it follows that (μ^{Y^n}, μ^{X^n}) is tight in $\mathcal{P}(C_T \times C_T)$, so by Prokhorov's theorem we can extract a (not relabelled) subsequence s.t. $(\mu^{Y^n}, \mu^{X^n}) \rightharpoonup (\mu^Y, \mu^Z)$ for a suitable $\mu^Z \in \mathcal{P}(C_T)$. By Skorokhod's theorem, we can work on a common probability space where all processes (X^n, Y^n) and (Z, Y) are well defined; using the fact that $Y^n \to Y \in L^p(\Omega; C_T)$, X^n satisfy (3.4) and arguing as in the proof of Proposition 3.10 we deduce that $X^n \to Z$ in $L^p(\Omega; C_T)$ and that Z solves the DDSDE associated to Y. But then by uniqueness $Z = X$ and $\mu^{X^n} \to \mu^X$ in $\mathcal{P}_p(C_T)$; as the reasoning works for any possible subsequence, conclusion follows. $\square$

3.2.3 Locally Lipschitz drifts with growth conditions

In the classical ODE and SDE setting, it is well known that uniqueness holds as soon as the drift is merely locally Lipschitz; the same is not true for DDSDEs, with explicit counterexamples given in [46]. Since global Lipschitz assumptions are often too restrictive to work with, we provide here some uniqueness results for drifts with local Lipschitz constant not growing too fast; the price one has to pay is a strong integrability requirement on the input μ^Y. Similar type of conditions have been proposed in the classical work [7] and more recently in [53], [24].

Assumption 3.27. *Given $\alpha > 0$ and $p \geq 1$; the drift $B : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and input data μ^Y satisfy the following:*

    i. $|B_t(x, \mu)| \leq h_t(1 + |x| + \|\mu\|_p)$,

    ii. $|B_t(x, \mu) - B_t(y, \nu)| \leq g_t(|x - y| + d_p(\mu, \nu))(1 + |x|^\alpha + |y|^\alpha + \|\mu\|_p^\alpha + \|\nu\|_p^\alpha)$,

    iii. $\int_{C_T} \exp(\lambda\|\omega\|_{C_T}^\alpha)\mu^Y(\mathrm{d}\omega) < \infty$ for all $\lambda \in \mathbb{R}$,

for integrable functions $h, g \in L_T^1$.

To obtain a stability estimate under Assumption 3.27, we need the following simple lemma.

Lemma 3.28. *For any $\gamma \in (0,1)$ there exists a $C > 0$ such that the map $\mathbb{R}^d \ni x \mapsto e^{(C+|x|)^\gamma}$ is convex.*

Proof. We reduce to the one dimensional case by arguing component-wise. Then note that it suffices to choose $C = C(\gamma) > 0$ such that $\gamma C^\gamma + \gamma - 1 \geq 0$ and to check that the resulting second derivative is non-negative. $\qquad\square$

Proposition 3.29. *Under Assumption 3.27, strong existence, pathwise uniqueness and uniqueness in law hold for the DDSDE (3.1) associated to μ^Y. Moreover for any $q > p$ there exists a constant $C = C(\|h\|_{L_T^1}, \|g\|_{L_T^1}, \alpha, p, q)$ such that, for any two solutions μ^{X^1} and μ^{X^2} driven by the inputs μ^{Y^1} and μ^{Y^2} respectively, we have the stability estimate:*

$$d_p(\mu^{X^1}, \mu^{X^2}) \leq C \left(\int_{C_T} \exp(C\|\omega\|_{C_T}^\alpha)(\mu^{Y^1} + \mu^{Y^2})(\mathrm{d}\omega) \right) d_q(\mu^{Y^1}, \mu^{Y^2}). \qquad (3.16)$$

Proof. Assumption 3.27 implies that the hypothesis of Proposition 3.10 are met, thus weak existence holds. Moreover any solution X satisfies the a priori estimate (3.4), namely $\|X\|_{C_T} \lesssim_h \|Y\|_{C_T}$.

Strong existence follows from weak existence and Proposition 3.7; by Proposition 3.5, weak uniqueness and pathwise uniqueness are equivalent and will follow from the stability estimate.

Now let (Y^1, Y^2) be an optimal coupling for $d_q(\mu^{Y^1}, \mu^{Y^2})$; by Remark 3.6, given any pair of weak solutions μ^{X^1}, μ^{X^2}, we can construct an associated coupling (X^1, X^2) and thus work with all variables defined on the same probability space. Let us set $\mu_t^i := \mathcal{L}(X_t^i)$.

By point ii. of Assumption 3.27, the difference $X^1 - X^2$ satisfies

$$
\begin{aligned}
|X_t^1 - X_t^2| &\leq |Y_t^1 - Y_t^2| \\
&\quad + \int_0^t g_s \left(|X_s^1 - X_s^2| + d_p(\mu_s^1, \mu_s^2) \right)(1 + |X_s^1|^\alpha + |X_s^2|^\alpha + \|\mu_s^1\|_p^\alpha + \|\mu_s^2\|_p^\alpha) \, \mathrm{d}s.
\end{aligned}
$$

Setting $Z_t := 1 + |X_t^1|^\alpha + |X_t^2|^\alpha + \|\mu_t^1\|_p^\alpha + \|\mu_t^2\|_p^\alpha$, it follows from the pathwise bound (3.4) that

$$\|Z\|_{C_T} \lesssim_{h,\alpha} 1 + \|Y^1\|_{C_T}^\alpha + \|Y^2\|_{C_T}^\alpha + \|\mu^{Y^1}\|_p^\alpha + \|\mu^{Y^2}\|_p^\alpha \quad \mathbb{P}\text{-a.s.}; \qquad (3.17)$$

moreover by Grönwall's lemma, for any $t \in [0,T]$ we have the $\mathbb{P}$-a.s. pathwise estimate

$$
\begin{aligned}
\sup_{r \leq t} |X_r^1 - X_r^2| &\leq \exp\left(\int_0^t g_s Z_s \, \mathrm{d}s \right) \left(\|Y^1 - Y^2\|_{C_T} + \int_0^t g_s \, d_p(\mu_s^1, \mu_s^2) Z_s \, \mathrm{d}s \right) \\
&\lesssim \exp\left((1 + \|g\|_{L_T^1}) \|Z\|_{C_T} \right) \left(\|Y^1 - Y^2\|_{C_T} + \int_0^t g_s \, d_p(\mu_s^1, \mu_s^2) \, \mathrm{d}s \right)
\end{aligned}
\qquad (3.18)
$$

Now taking the L_Ω^p-norm on both sides, using an Hölder inequality on the r.h.s. (using the fact that $q > p$), we can find another constant $C' = C'(\|g\|_{L_T^1}, p, q)$ such that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{r \leq t} |X_r^1 - X_r^2|^p \right]^{\frac{1}{p}} &\lesssim \mathbb{E}[\exp(C'\|Z\|_{C_T})] \left(\mathbb{E}[\|Y^1 - Y^2\|_{C_T}^q]^{\frac{1}{q}} + \int_0^t g_s \, d_p(\mu_s^1, \mu_s^2) \, \mathrm{d}s \right) \\
&\lesssim \mathbb{E}[\exp(C'\|Z\|_{C_T})] \left(d_q(\mu^{Y^1}, \mu^{Y^2}) + \int_0^t g_s \, \mathbb{E}\left[\sup_{r \leq s} |X_r^1 - X_r^2|^p \right]^{\frac{1}{p}} \mathrm{d}s \right).
\end{aligned}
$$

Applying Grönwall's lemma again, we deduce that

$$d_p(\mu^{X^1}, \mu^{X^2}) \leq \mathbb{E}[\|X^1 - X^2\|_{C_T}^p]^{1/p} \lesssim e^{\|g\|_{L_T^1}} \mathbb{E}[\exp(C'\|Z\|_{C_T})] \, d_q(\mu^{Y^1}, \mu^{Y^2}).$$

Now consider the quantities $\exp(\|\mu^{Y^i}\|_p^\alpha) = \exp(\mathbb{E}[\|Y^i\|_{C_T}^p]^{\alpha/p})$ for $i = 1, 2$. In the case $\alpha > p$ one can directly apply Jensen's inequality to give that $\exp(\mathbb{E}[\|Y^i\|_{C_T}^p]^{\alpha/p}) \leq \mathbb{E}[\exp(\|Y^i\|_{C_T}^\alpha)]$. For $\alpha < p$, we first apply Lemma 3.28 with $\gamma = \alpha/p$, followed by Jensen's inequality to give, for some $C_\gamma > 0$,

$$\exp\left(\mathbb{E}[\|Y^i\|_{C_T}^p]^{\alpha/p}\right) \leq \exp\left((C_\gamma + \mathbb{E}[\|Y^i\|_{C_T}^p])^{\alpha/p}\right) \leq \mathbb{E}\left[\exp((C_\gamma + \|Y^i\|_{C_T}^p)^{\alpha/p})\right]$$
$$\lesssim_{\alpha,p} \mathbb{E}\left[\exp(\|Y^i\|_{C_T}^\alpha)\right].$$

In either case we have $\exp(\|\mu^{Y^i}\|_p^\alpha) \lesssim_{\alpha,p} \mathbb{E}[\exp(\|Y^i\|_{C_T}^\alpha)]$ and so combined with (3.17) we finally obtain the stability estimate (3.16) for some new constant C.

As the estimate holds for any possible pair of solutions X^1, X^2, taking $Y^1 = Y^2$ we deduce pathwise uniqueness and this completes the proof. $\qquad\square$

Remark 3.30. It's clear from the proof above that Assumption 3.27 admits other variants, for instance we could instead require, i., iii. and

ii′. $|B_t(x,\mu) - B_t(y,\nu)| \leq g_t(|x-y| + d_p(\mu,\nu))(1 + |x|^\alpha + |y|^\alpha + \langle e^{c|\cdot|^\alpha}, \mu\rangle + \langle e^{c|\cdot|^\alpha}, \nu\rangle)$

for some $c > 0$. The only difference is that in this case existence of solutions doesn't follow from a straightforward application of Proposition 3.10. However, it's easy to check that, due to the available a priori estimates, the same compactness argument can be readapted in this setting.

The next example provides concrete choices of B satisfying Assumption 3.27.

Example 3.31. *If $Y = \xi + W$, where ξ has Gaussian tails and W is a continuous Gaussian process (e.g. fractional Brownian motion), then condition iii. is satisfied for any $\alpha \in [0, 2)$.*

Suppose $b : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ satisfies for some $\alpha > 0$

$$|b_t(x,y)| \leq g_t(1 + |x| + |y|),$$
$$|b_t(x,x') - b_t(y,y')| \leq g_t(|x-y| + |x'-y'|)(1 + |x|^\alpha + |y|^\alpha + |x'|^\alpha + |y'|^\alpha),$$

for some integrable function g. Then a few elementary computations reveal that

$$B_t(x,\mu) := \int_{\mathbb{R}^d} b_t(x,x')\mu(\mathrm{d}x')$$

satisfies conditions i., ii. for $p = \alpha + 1$. Indeed by Hölder's and Minkowski's inequalities

$$|B_t(x,\mu) - B_t(y,\nu)| \leq \int_{\mathbb{R}^{2d}} |b_t(x,x') - b_t(y,y')| m(\mathrm{d}x', \mathrm{d}y')$$
$$\leq g_t\left(|x-y| + \left(\int_{\mathbb{R}^{2d}} |x'-y'|^p\, m(\mathrm{d}x', \mathrm{d}y')\right)^{1/p}\right) \times$$
$$\times \left(1 + |x|^\alpha + |y|^\alpha + \left(\int_{\mathbb{R}^{2d}} (|x'|^\alpha + |y'|^\alpha)^{p'}\, m(\mathrm{d}x', \mathrm{d}y')\right)^{1/p'}\right)$$
$$\leq g_t(|x-y| + d_p(\mu,\nu))(1 + |x|^\alpha + |y|^\alpha + \|\mu\|_{p'\alpha}^\alpha + \|\nu\|_{p'\alpha}^\alpha)$$

once we take m to be an optimal coupling for $d_p(\mu,\nu)$; choosing $p = \alpha + 1$, so that $p'\alpha = p$, we get the desired estimate for ii. and condition i. can be checked similarly.

The above assumption on b is satisfied for instance if Db satisfies a growth condition of the form $|Db_t(x,x')| \lesssim g_t(1 + |x|^\alpha + |x'|^\alpha)$, for all $(x,x') \in \mathbb{R}^{2d}$.

4 DDSDEs with convolutional structure

In this section we will focus on a particular subcase of (3.1), given by drifts with the convolutional structure $B_t(x, \mu) = (b_t * \mu)(x)$. Contrary to the previous sections, we now impose the noise term Y to be of the form $Y_t = \xi + W_t$ where ξ and W are independent random variables taking respectively values in $\mathbb{R}^d$ and C_T; the reason for this, which will become clearer later on, is to exploit the transport structure to show that some integrability features of $\mathcal{L}(\xi)$ are propagated at positive times. Still, no assumption on W are imposed apart from its continuity.

The DDSDE in consideration therefore becomes

$$X_t = \xi + \int_0^t (b_s * \mathcal{L}(X_s))(X_s) \, \mathrm{d}s + W_t = \xi + \int_0^t \int_{\mathbb{R}^d} b_s(X_s - y) \, \mathcal{L}(X_s)(\mathrm{d}y) \, \mathrm{d}s + W_t \quad (4.1)$$

for some measurable $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$.

The convolutional structure arises naturally in the study of particle systems with pairwise interactions; it also allows to discuss kernels b with poor spatial regularity, possibly unbounded and merely integrable.

Since the drift is not necessarily continuous anymore, we require a new notion of solution, compared to Definition 3.1, that ensures that the integral appearing in (4.1) is meaningful in the Lebesgue sense. As usual, we work for simplicity on a finite time interval $[0, T]$.

Definition 4.1 (Weak Solutions II). Let $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ be a measurable map. We say that a tuple $(\Omega, \mathcal{F}, \mathbb{P}; X, W, \xi)$, consisting of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable map $(X, W, \xi) : \Omega \to C_T \times C_T \times \mathbb{R}^d$ is a weak solution to (4.1) on $[0, T]$ if:

1. For all $t > 0$, $\mathcal{L}(X_t) = \mathbb{P} \circ X_t^{-1} =: \mu_t$ is absolutely continuous w.r.t. Lebesgue.

2. There exists a measurable representative of $(t, x) \mapsto |b_t| * \mathcal{L}(X_t)(x)$ such that

$$\int_0^T |b_s| * \mathcal{L}(X_s)(X_s) \, \mathrm{d}s < \infty, \quad \mathbb{P}\text{-a.s.}.$$

3. The relation (4.1) holds $\mathbb{P}$-a.s., the integral being interpreted in the Lebesgue sense.

Remark 4.2. Definition 4.1 is insensitive to the choice of measurable representative for $b_t * \mathcal{L}(X_s)$ appearing in (2). Consider any two measurable maps $f, \tilde{f} : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ such that $f(t, x) = \tilde{f}(t, x)$ for Lebesgue a.e. (t, x), then by condition (1) and Fubini we have

$$\mathbb{E}\Big[\int_0^T |f(s, X_s) - \tilde{f}(s, X_s)| \mathrm{d}t \Big] = \int_0^T \int_{\mathbb{R}^d} |f(s, x) - \tilde{f}(s, x)| \, \mathcal{L}(X_s)(\mathrm{d}x) \, \mathrm{d}s = 0.$$

As a consequence, if $\mathbb{P}$-a.s. $\int_0^T |f(s, X_s)| \, \mathrm{d}s < \infty$, then the same holds for $\tilde{f}$ and one has that $\int_0^{\cdot} f(s, X_s) \, \mathrm{d}s = \int_0^{\cdot} \tilde{f}(s, X_s) \, \mathrm{d}s$ as C_T-valued random variables. A similar argument shows that the definition is insensitive to modifications of the kernel b on Lebesgue negligible sets; for this reason we can consider kernels belonging to equivalence classes like $L_T^1 L_x^p$.

In order for the requirement (2) in Definition 4.1 to be met, it is desirable to have some information on the integrability properties of $\mathcal{L}(X_t)$. To derive it, we need to recall known facts from classical ODE and transport PDE theory.

Given $x_0 \in \mathbb{R}^d$, a drift $\bar{b} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ with continuous and bounded first derivative and a continuous path $\omega \in C_T$, there exists a unique solution $x \in C_T$ to the Cauchy problem

$$x_t = x_0 + \int_0^t \bar{b}(s, x_s)\mathrm{d}s + \omega_t \quad \forall\, t \in [0,T]; \tag{4.2}$$

similarly to Section 2, we can define the solution map $x_t = \Phi(t, x_0; \omega)$ associated to the drift $\bar{b}$, which is a continuous map $\Phi : [0,T] \times \mathbb{R}^d \times C_T \to \mathbb{R}^d$. For fixed (t, ω), the map $x_0 \mapsto \Phi(t, x_0; \omega)$ is a diffeomorphism of $\mathbb{R}^d$, with Jacobian given by the formula

$$\det D\Phi(t, x_0; \omega) = \exp\left(\int_0^t \mathrm{div}\,\bar{b}(s, \Phi(s, x_0; \omega))\mathrm{d}s \right).$$

In particular we have the following two-sided estimate independent of ω:

$$0 < \exp\left(-\|\mathrm{div}\,\bar{b}\|_{L_T^1 L_x^\infty} \right) \leq \det D\Phi(t, x_0; \omega) \leq \exp\left(\|\mathrm{div}\,\bar{b}\|_{L_T^1 L_x^\infty} \right) < \infty; \tag{4.3}$$

similarly for $\det D\Psi(t, x_0; \omega)$, where $\Psi(t, \cdot\,; \omega)$ denotes the inverse of $x_0 \mapsto \Phi(t, x_0; \omega)$.

On the other hand, we can consider the solution map as a continuous (thus measurable) map $(x_0, \omega) \mapsto \Phi(\cdot, x_0; \omega)$ from $\mathbb{R}^d \times C_T$ into C_T; in particular if we now consider a random variable (ξ, W) defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider the associated variable $\Phi(\cdot, \xi; W)$, which amounts to the solution of the random ODE problem (4.2).

With this considerations in mind, we can prove the following preliminary a priori estimate.

Proposition 4.3. *Let $b \in L_T^1 C_b^1$, $p \in [1, \infty]$ and (ξ, W) be independent input data s.t. $\mathcal{L}(\xi)(\mathrm{d}x) = \rho(x)\,\mathrm{d}x$ with $\rho \in L_x^p$ for $p \in [1, \infty]$ and let X denote the unique solution to (4.1). Then there exists a constant $C = C(p, \|\mathrm{div}\,b\|_{L_T^1 L_x^\infty})$ such that*

$$\sup_{t \in [0,T]} \|\mathcal{L}(X_t)\|_{L_x^p} \leq C\|\rho\|_{L_x^p} \tag{4.4}$$

Furthermore, only using the assumption that ρ is a probability density, for any $\varepsilon > 0$ there exists a $\delta > 0$, only depending on ρ and $\|\mathrm{div}\,b\|_{L_T^1 L_x^\infty}$, s.t.

$$\sup_{t \in [0,T]} \mathcal{L}(X_t)(A) \leq \varepsilon \quad \forall\, A \subset \mathbb{R}^d,\ |A| \leq \delta. \tag{4.5}$$

Proof. Let X be the aforementioned solution, which exists by the results of Section 2. Then X is also a solution to the random ODE (4.2) for the choice $\bar{b}_t(x) = b_t * \mathcal{L}(X_t)(x)$ and we can represent it as $X_t = \Phi(t, \xi; W)$; without loss of generality we can take (ξ, W) to be the canonical variables on the probability space $(\mathbb{R}^d \times C_T, \mathcal{B}(\mathbb{R}^d \times C_T), (\rho\,\mathrm{d}x) \otimes \mu)$ where $\mu = \mathcal{L}(W)$. Moreover it holds that

$$\|\mathrm{div}\,\bar{b}\|_{L_T^1 L_x^\infty} = \|(\mathrm{div}\,b.) * \mathcal{L}(X.)\|_{L_T^1 L_x^\infty} \leq \|\mathrm{div}\,b\|_{L_T^1 L_x^\infty}.$$

For any measurable and bounded $f : \mathbb{R}^d \to \mathbb{R}$ and any $t \in [0,T]$ we have

$$\langle f, \mathcal{L}(X_t) \rangle = \mathbb{E}[f(\Phi(t, \xi; W))] = \int_{C_T} \int_{\mathbb{R}^d} f(\Phi(t, x; \omega))\rho(x)\mathrm{d}x\, \mu(\mathrm{d}\omega)$$

$$= \int_{C_T} \int_{\mathbb{R}^d} f(y)\rho(\Psi(t, y; \omega)) \det D\Psi(t, y; \omega)\, \mathrm{d}y\, \mu(\mathrm{d}\omega)$$

$$= \int_{\mathbb{R}^d} f(y)\left[\int_{C_T} \rho(\Psi(t, y; \omega)) \det D\Psi(t, y; \omega)\, \mu(\mathrm{d}\omega) \right] \mathrm{d}y.$$

As a consequence, $\mathcal{L}(X_t)$ is absolutely continuous w.r.t. Lebesgue with density given by

$$\rho_t(y) := \int_{C_T} \rho(\Psi(t,y;\omega)) \det D\Psi(t,y;\omega)\,\mu(\mathrm{d}\omega).$$

Given the bounds on the Jacobian of Φ and Ψ as above, for $p \in [1,\infty)$ we can estimate $\|\rho_t\|_{L^p}$ by

$$\begin{aligned}
\int_{\mathbb{R}^d} |\rho_t(y)|^p\,\mathrm{d}y &\leq \int_{\mathbb{R}^d} \int_{C_T} |\rho(\Psi(t,y;\omega))|^p |\det D\Psi(t,y;\omega)|^p\,\mu(\mathrm{d}\omega)\,\mathrm{d}y \\
&\lesssim_{\mathrm{div}\,b} \int_{C_T} \int_{\mathbb{R}^d} |\rho(\Psi(t,y;\omega))|^p\,\mathrm{d}y\,\mu(\mathrm{d}\omega) \\
&= \int_{C_T} \int_{\mathbb{R}^d} |\rho(x)|^p \det D\Phi(t,x;\omega)\,\mathrm{d}x\,\mu(\mathrm{d}\omega) \lesssim_{\mathrm{div}\,b} \int_{\mathbb{R}^d} |\rho(x)|^p\,\mathrm{d}x;
\end{aligned}$$

the estimate for $p = \infty$ is similar. To prove the last statement, applying several times the fact that the incompressibility constants of $\Phi(t,\cdot;\omega)$ are controlled by $\exp(\|\mathrm{div}\,b\|_{L^1_T L^\infty_x})$ uniformly in $\omega \in C_T$ for any $t \in [0,T]$ and $A \subset \mathbb{R}^d$, see (4.3), we have that

$$\mathcal{L}(X_t)(A) \lesssim_h \int_A \int_{C_T} \rho(\Psi(t,y;\omega))\,\mu(\mathrm{d}\omega)\,\mathrm{d}y \lesssim_h \int_{C_T} \int \rho(x)\mathbb{1}_{\psi(t,A;\omega)}(x)\,\mathrm{d}x\,\mu(\mathrm{d}\omega);$$

the conclusion follows by observing that $\rho \in L^1_x$ is uniformly integrable and that $|\psi(t,A;\omega)| \lesssim_h |A|$ uniformly in t,ω. $\qquad\square$

With the a priori bound (4.4) at hand, we are now ready to prove existence of solutions to (4.1) for integrable drifts b, up to the price of requiring additional assumptions on $\mathrm{div}\,b$ and ρ.

Proposition 4.4. *Suppose $b \in L^1_T L^q_x$, $\mathrm{div}\,b \in L^1_T L^\infty_x$ and $\mathcal{L}(\xi)(\mathrm{d}x) := \rho(x)\,\mathrm{d}x$ for a probability density $\rho \in L^p_x$ with $p,q \in [1,\infty]$ such that*

$$\frac{1}{2q} + \frac{1}{p} \leq 1. \tag{4.6}$$

Then there exists at least one weak solution X_t to (4.1) in the sense of Definition 4.1, which moreover satisfies $\sup_{t \in [0,T]} \|\mathcal{L}(X_t)\|_{L^p_x} < \infty$.

Proof. The proof is based on classical compactness arguments; since $\rho \in L^1_x \cap L^p_x$, by interpolation w.l.o.g. we can assume equality holds in (4.6).

Let $\{b^n\}_n$ be a sequence in $L^1_T C^1_b$ satisfying $\|b^n\|_{L^1_T L^q_x} \leq \|b\|_{L^1_T L^q_x}$, $\|\mathrm{div}\,b^n\|_{L^1_T L^\infty_x} \leq \|\mathrm{div}\,b\|_{L^1_T L^\infty_x}$ (take for instance $b^n_s := b_s * \varphi_{1/n}$ with $\{\varphi_\varepsilon\}_{\varepsilon > 0}$ standard mollifiers); by Section 2, for each $n \geq 1$ there exists a unique strong solution $(X^n, \mathcal{L}(X^n)) \in C_T \times \mathcal{P}(C_T)$ to the DDSDE associated to b^n, with input data (ξ, W). By Proposition 4.3, we can find some $C = C(\|\mathrm{div}\,b\|_{L^1_T L^\infty_x})$ such that

$$\sup_{t \in [0,T]} \|\mathcal{L}(X^n_t)\|_{L^p} \leq C\|\rho\|_{L^p_x} \quad \forall\, n \in \mathbb{N}.$$

Set $Z^n_t := \int_0^t b^n_s * \mathcal{L}(X^n_s)(X^n_s)\,\mathrm{d}s = X^n_t - \xi - W_t$. Our first goal is to establish uniform bounds on $\{Z_n\}_n$. It holds that

$$\begin{aligned}
\mathbb{E}[\|Z^n\|_{W^{1,1}_T}] &\leq \mathbb{E}\left[\int_0^T |b^n_s| * \mathcal{L}(X^n_s)(X^n_s)\,\mathrm{d}s\right] \\
&\leq \int_0^T \langle |b^n_s| * \mathcal{L}(X^n_s), \mathcal{L}(X^n_s)\rangle\,\mathrm{d}s \\
&\leq \|b^n\|_{L^1_T L^q_x} \sup_{t \in [0,T]} \|\mathcal{L}(X^n_t)\|^2_{L^p_x} \lesssim \|b\|_{L^1_T L^q_x} \|\rho\|^2_{L^p_x}
\end{aligned}$$

where in the third passage we applied Hölder's and Young's inequalities.

Then Helly's selection theorem, together with the fact that $W_T^{1,1}$ embeds compactly in L_T^r for any $r < \infty$ as well as in the topology of pointwise convergence, implies that $\{Z^n\}_n$ is tight in L_T^r and $\{(\xi, W, Z^n)\}_n$ is tight in $\mathbb{R}^d \times C_T \times L_T^r$.

We can therefore apply Prokhorov's and Skorohod's theorems to construct another probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, carrying a family of random variables $(\tilde{\xi}^n, \tilde{W}^n, \tilde{Z}^n)$ converging $\tilde{\mathbb{P}}$-a.s. to $(\tilde{\xi}, \tilde{W}, \tilde{Z})$ in $\mathbb{R}^d \times C_T \times L_T^r$, such that $\mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{\xi}^n, \tilde{W}^n, \tilde{Z}^n) = \mathcal{L}_{\mathbb{P}}(\xi, W, Z^n)$. From this we deduce as before that $\tilde{\mathbb{P}}$-a.s. $\tilde{Z}^n \to \tilde{Z}$ also in the sense of pointwise convergence, similarly for $\tilde{X}^n := \tilde{\xi}^n + \tilde{W}^n + \tilde{Z}^n$; moreover $\tilde{X}^n$ are still solutions to the DDSDE associated to b^n and $\tilde{Z}_t^n = \int_0^t b_s^n * \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{X}_s^n)(\tilde{X}_s^n)\, \mathrm{d}s$.

In order to conclude, it remains to show that $(\tilde{X}, \tilde{W}, \tilde{\xi})$ is a weak solution to the DDSDE associated to b, on the probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$. For notational simplicity we will drop the tildes from now on.

We first treat the case $p > 1$. Since $\mathbb{P}$-a.s. $X_t^n \to X^n$, $\mathcal{L}(X_t^n) \rightharpoonup \mathcal{L}(X_t)$ weakly and they are uniformly bounded in L_x^p with $p > 1$; thus $\mathcal{L}(X_t) \in L_x^p$ and by lower semicontinuity of weak convergence (weak-$*$ if $p = \infty$) it holds that

$$\sup_{t \in [0,T]} \|\mathcal{L}(X_t)\|_{L_x^p} \leq \sup_{t \in [0,T]} \liminf_{n \to \infty} \|\mathcal{L}(X_t^n)\|_{L_x^p} \leq C\|\rho\|_{L_x^p},$$

which checks requirement (1) of Definition 4.1. The verification of (2) is similar and we are left with showing that $Z^n \to Z$ for $Z_t = \int_0^t b_s * \mathcal{L}(X_s)(X_s)\, \mathrm{d}s$.

In the case $p > 1$, $q < \infty$, we may further assume that $b^n \to b$ in $L_T^1 L_x^q$. We fix $\varepsilon > 0$ and choose $\bar{b} \in C_b^1([0,T] \times \mathbb{R}^d; \mathbb{R}^d)$ such that $\|b - \bar{b}\|_{L_T^1 L_x^q} < \varepsilon$ and split the estimate as

$$\|Z^n - Z\|_{C_T} \leq \int_0^T |b_s^n - \bar{b}_s| * \mathcal{L}(X_s^n)(X_s^n)\, \mathrm{d}s + \int_0^T |\bar{b}_s * \mathcal{L}(X_s^n)(X_s^n) - \bar{b}_s * \mathcal{L}(X_s)(X_s)|\, \mathrm{d}s$$

$$+ \int_0^T |\bar{b}_s - b_s| * \mathcal{L}(X_s)(X_s)\, \mathrm{d}s$$

$$:= I_1^n + I_2^n + I_3.$$

Regarding I_1^n, we can apply the uniform estimates on $\|\mathcal{L}(X_t^n)\|_{L_x^p}$ to obtain

$$\mathbb{E}[I_1^n] \lesssim \sup_{s \in [0,T]} \|\mathcal{L}(X_s^n)\|_{L_x^p}^2 \|b^n - \bar{b}\|_{L_T^1 L_x^q} \lesssim \|b^n - b\|_{L_T^1 L_x^q} + \|b - \bar{b}\|_{L_T^1 L_x^q};$$

similarly for I_3. On the other hand, since $\bar{b} \in C_b^1$ and $\mathcal{L}(X^n) \rightharpoonup \mathcal{L}(X)$, $\bar{b} * \mathcal{L}(X^n)$ converges uniformly to $\bar{b} * \mathcal{L}(X)$; combined with $X^n \to X$ $\mathbb{P}$-a.s. and the uniform bounds, by dominated convergence we have $\mathbb{E}[I_2^n] \to 0$. Thus overall we obtain

$$\limsup_{n \to \infty} \mathbb{E}[\|Z^n - Z\|_{C_T}] \leq \limsup_{n \to \infty} (\mathbb{E}[I_1^n] + \mathbb{E}[I_2^n] + \mathbb{E}[I_3])$$

$$\lesssim \limsup_{n \to \infty} \|b^n - b\|_{L_T^1 L_x^q} + 2\|b - \bar{b}\|_{L_T^1 L_x^q} \lesssim 2\varepsilon;$$

by the arbitrariness of $\varepsilon > 0$ we deduce that $\mathbb{E}[\|Z^n - Z\|_{C^T}] \to 0$ and that X is a weak solution.

In the case $p = 1, q = \infty$ we cannot assume $\|b^n - b\|_{L_T^1 L_x^\infty} \to 0$, but only that $b_t^n(x) \to b_t(x)$ for Lebesgue a.e. $(t,x) \in [0,T] \times \mathbb{R}^d$. Since $X_t^n \rightharpoonup X_t$, the sequence $\{\mathcal{L}(X_t^n)\}_n$ is tight, which together with (4.5) implies its equiintegrability. But then by Dunford–Pettis theorem $\mathcal{L}(X_t) \in L_x^1$ and $\mathcal{L}(X_t^n) \rightharpoonup \mathcal{L}(X_t)$ weakly in L_x^1, checking requirement (1) of Definition 4.1; (2) follows trivially since $b \in L_T^1 L_x^\infty$. Finally, since $\mathcal{L}(X_t^n) \rightharpoonup \mathcal{L}(X_t)$ weakly in L_x^1, we can apply Lemma B.1 from Appendix B at any fixed $t \in [0,T]$ to conclude that $\mathbb{E}[\|Z^n - Z\|_{C_T}] \to 0$ still holds. $\qquad\square$

Having established existence under suitable integrability assumptions on b, it is natural to establish uniqueness under similar requirements on the derivative Db. We start with a conditional result of uniqueness in a class of sufficiently regular solutions.

Proposition 4.5. *Let q, $p \in (1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} < 1$ and assume $b \in L_T^1 W_x^{1,q}$, $\rho \in L_x^p$. Then both uniqueness in law and pathwise uniqueness holds for* (4.1) *in the class of solutions satisfying*

$$\sup_{t \in [0,T]} \|\mathcal{L}(X_t)\|_{L_x^p} < \infty \tag{4.7}$$

in the following sense: if X^1 and X^2 are both solutions satisfying (4.7)*, then their laws as probabilities on the path space C_T coincide; if X^i are defined on the same probability space and solve the DDSDE w.r.t. same input data (ξ, W), then $X^1 = X^2$ $\mathbb{P}$-a.s.*

Proof. Given two solutions X^i as above, let us set $\mu_t^i = \mathcal{L}(X_t^i)$ and $b_t^i(x) = (b_t * \mu_t^i)(x)$; from Proposition 4.4, for all $t \in [0,T]$, $\mu_t^i \in L_x^1 \cap L_x^p$ for $p > q'$, by interpolation $\mu^i \in L_t^\infty L_x^{q'}$ and so by Young's inequality

$$\|b^i\|_{L_T^1 W_x^{1,\infty}} \lesssim \|b\|_{L_T^1 W^{1,q}} \|\mu^i\|_{L_T^\infty L_x^{q'}} \lesssim 1.$$

In particular, both drifts b^i are spatially Lipschitz continuous; moreover by arguing as in Proposition 3.5, we only need to check pathwise uniqueness for solutions X^i to (4.1) satisfying (4.7).

First observe that the difference $Y := X^1 - X^2$ satisfies

$$
\begin{aligned}
|Y_t| &\leq \int_0^t |b_s^1(X_s^1) - b_s^1(X_s^2)| \, \mathrm{d}s + \int_0^t |(b^1 - b^2)_s(X_s^2)| \, \mathrm{d}s \\
&\leq \int_0^t \|b_s * \mu_s^1\|_{W_x^{1,\infty}} |Y_s| \, \mathrm{d}s + \int_0^t \|b_s * (\mu_s^1 - \mu_s^2)\|_{L_x^\infty} \, \mathrm{d}s \\
&\lesssim_{\mu^i} \int_0^t \|b_s\|_{W_x^{1,q}} (1 + |Y_s|) \, \mathrm{d}s;
\end{aligned}
$$

applying Grönwall's lemma, we find a deterministic constant $C = C(b, \mu^i) > 0$ such that $\|X^1 - X^2\|_{C_T} = \|Y\|_{C_T} \leq C$ $\mathbb{P}$-a.s.

As a consequence, we also have that $d_{r'}(\mu_t^1, \mu_t^2) \leq C < \infty$ for any $r' \in (1, \infty)$; in particular, the assumptions of Lemma B.2 from Appendix B are met, as there exists $r > 1$ such that $r/q + 1/p \leq 1$. It follows that

$$\|b_s * (\mu_s^1 - \mu_s^2)\|_{L_x^\infty} \lesssim \|b_s\|_{W^{1,q}} \big(\|\mu_s^1\|_{L_x^p}^{1/r} + \|\mu_s^2\|_{L_x^p}^{1/r}\big) d_{r'}(\mu_s^1, \mu_s^2) \lesssim_{\mu^i} \|b_s\|_{W^{1,q}} \|Y_s\|_{L_\Omega^{r'}}.$$

We use this new bound to improve the estimates on Y, giving:

$$|Y_t| \lesssim_{\mu^i} \int_0^t \|b_s\|_{W_x^{1,q}} (|Y_s| + \|Y_s\|_{L_\Omega^{r'}}) \, \mathrm{d}s$$

and now applying Minkowski's integral inequality (with $r' > 1$) we obtain

$$\|Y_t\|_{L_\Omega^{r'}} \lesssim_{\mu^i} \int_0^t \|b_s\|_{W_x^{1,q}} \|Y_s\|_{L_\Omega^{r'}} \, \mathrm{d}s.$$

However, by the hypothesis $b \in L_T^1 W_x^{1,q}$ and applying Grönwall's lemma, we deduce $\|Y_t\|_{L_\Omega^{r'}} = 0$ for all $t \in [0,T]$ and so $X^1 = X^2$ $\mathbb{P}$-a.s. $\qquad \square$

Combining Propositions 4.4 and 4.5 we get the following existence and uniqueness result.

Theorem 4.6. *Let q, $p \in (1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} < 1$ and $b \in L^1_T W^{1,q}_x$ with $\operatorname{div} b \in L^1_T L^\infty_x$. Then for any $\rho \in L^p_x$ there exists a strong solution X_t satisfying $\sup_{t \in [0,T]} \|\mathcal{L}(X_t)\|_{L^p_x} < \infty$, which is unique in this class (in the sense of Proposition 4.5).*

Proof. The assumption on (p, q) implies that (4.6) holds, thus by Proposition 4.4 weak existence holds; as observed in the proof of Proposition 4.5, the drift $\bar{b}_t = b_t * \mathcal{L}(X_t) \in L^1_T W^{1,\infty}_x$ and so arguing as in Proposition 3.7 we deduce strong existence. Uniqueness follows again from Proposition 4.5. $\qquad\square$

Theorem 4.6, together with the a priori estimates from Proposition 4.3, guarantees that such solution X is the only possible accumulation point of solutions X^n constructed from mollified drifts $b^n = b * \varphi_{1/n}$ as in the proof of Proposition 4.4. It doesn't however exclude the existence of different solutions $\tilde{X}$, associated to the same input data (ξ, W), which fail to meet the requirement $\sup_{t \in [0,T]} \|\mathcal{L}(\tilde{X}_t)\|_{L^p_x} < \infty$.

In order to strengthen the result and truly establish uniqueness in law, without any regularity assumption on $\mathcal{L}(\tilde{X})_t$, we need more restrictive conditions on b; the advantage is that we deduce some strong stability estimates, in the form of equation (4.8) below. This result relies on the results of [11] and in particular Lemma B.3.

Proposition 4.7. *Let $(p, q, r) \in (1, \infty)^3$, $b \in L^1_T W^{1,q}_x$ with $\operatorname{div} b \in L^1_T L^\infty_x$ and*

$$q > d, \quad \frac{r}{p} + \frac{1}{q} \le 1, \quad \frac{1}{r} + \frac{1}{r'} = 1.$$

Also assume we are given input data $\mathcal{L}(\xi^1, W^1) = \mu^{\xi^1} \otimes \mu^{W^1} \in \mathcal{P}_{r'}(\mathbb{R}^d \times C_T)$, such that $\mu^{\xi^1}(\mathrm{d}x) = \rho(x) \, \mathrm{d}x$ for some $\rho \in L^q_x$. Denote by μ^{X^1} the unique solution associated to $\mu^{\xi^1 + W^1}$, in the sense of Theorem 4.6. Then there exists a constant C, depending on $d, T, p, q, r, \|\operatorname{div} b\|_{L^1_T L^\infty_x}, \|b\|_{L^1_T W^{1,p}_x}$ and $\|\rho\|_{L^q}$, with the following property: for any other solution μ^{X^2} to (4.1) with input μ^{Y^2},

$$d_{r'}(\mu^{X^1}, \mu^{X^2}) \le C d_{r'}(\mu^{\xi^1 + W^1}, \mu^{Y^2}). \tag{4.8}$$

Proof. Since $q > d$, by Sobolev embeddings $b \in L^1_T C^0_b$, so that the DDSDE (4.1) is meaningful in the sense of Definition 3.1, without the need of any regularity assumption on μ^{X^2}. By Remark 3.6, given any coupling of $\mu^{\xi^1 + W^1}, \mu^{Y^2}$, we can construct a coupling of the associated solutions μ^{X^1}, μ^{X^2} as well, so from now on we will work with X^i, ξ^1, W^1, Y^2 all defined on the same probability space; also for simplicity we will adopt the compact notation $Y^1 = \xi^1 + W^1$. Since X^i are solutions to the DDSDE and $b \in L^1_T C^0_b$, we have the trivial estimate

$$\mathbb{E}\big[\|X^i\|^{r'}_{C_T}\big] \lesssim \|b\|^{r'}_{L^1_T W^{1,q}_x} + \mathbb{E}\big[\|Y^i\|^{r'}_{C_T}\big]$$

which shows that $\mu^{X^i} \in \mathcal{P}_{r'}(C_T)$.

Having made these preparations, we can now pass to the proof of estimate (4.8). Set $\mu^i_t = \mathcal{L}(X^i_t)$, then the process $X^1 - X^2$ satisfies

$$|X^1_t - X^2_t| \le \int_0^t |b_s * \mu^1_s(X^1_s) - b_s * \mu^2_s(X^2_s)| \, \mathrm{d}s + |Y^1_t - Y^2_t|;$$

since $\sup_{t \in [0,T]} \|\mu^1_t\|_{L^p_x} < \infty$, we can apply Lemma B.3 from Appendix B to obtain

$$|X^1_t - X^2_t| \lesssim \int_0^t \|b_s\|_{W^{1,q}_x} (1 + \|\mu^1_s\|_{L^p_x}) \, (|X^1_s - X^2_s| + d_{r'}(\mu^1_s, \mu^2_s)) \, \mathrm{d}s + |Y^1_t - Y^2_t|$$

$$\lesssim_{\operatorname{div} b, \rho} \int_0^t \|b_s\|_{W^{1,q}_x} \, (|X^1_s - X^2_s| + d_{r'}(\mu^1_s, \mu^2_s)) \, \mathrm{d}s + \|Y^1 - Y^2\|_{C_T};$$

taking the $L_\Omega^{r'}$-norm on both sides and applying Minkowski's inequality we find

$$\|X_t^1 - X_t^2\|_{L_\Omega^{r'}} \lesssim \int_0^t \|b_s\|_{W_x^{1,q}} \|X_s^1 - X_s^2\|_{L_\Omega^{r'}} \, ds + \|\|Y^1 - Y^2\|_{C_T}\|_{L_\Omega^{r'}}.$$

By Grönwall's lemma we can find some constant κ such that

$$\sup_{t \in [0,T]} d_{r'}(\mu_t^1, \mu_t^2) \lesssim e^{\kappa \|b\|_{L_T^1 W_x^{1,q}}} \|\|Y^1 - Y^2\|_{C_T}\|_{L_\Omega^{r'}}.$$

In order to obtain a bound in the path space C_T it suffices to observe that the previous inequalities also imply, again by Grönwall's lemma,

$$\|X^1 - X^2\|_{C_T} \lesssim_b \|b\|_{L_T^1 W_x^{1,q}} \sup_{t \in [0,T]} d_{r'}(\mu_t^1, \mu_t^2) + \|Y^1 - Y^2\|_{C_T};$$

inserting the estimate for $d_{r'}(\mu_t^1, \mu_t^2)$ and taking expectation, we find a constant C as above s.t.

$$d_{r'}(\mu^{X^1}, \mu^{X^2}) \leq \mathbb{E}[\|X^1 - X^2\|_{C_T}^{r'}]^{1/r'} \leq C \, \mathbb{E}[\|Y^1 - Y^2\|_{C_T}^{r'}]^{1/r'}.$$

Minimizing over all possible couplings (Y^1, Y^2) gives the conclusion. $\qquad\square$

5 Mean field convergence

In this section we prove mean field limits for the interacting particle systems associated to the drifts considered in Sections 3-4. We start by presenting Tanaka's idea in a rather abstract fashion and then apply it to our cases of interest.

5.1 Abstract criteria

Let us shortly recall the setup. Given a family of C_T-valued random variables $\{Y^i, i \in \mathbb{N}\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable drift $B : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ (possibly with $\mathcal{P}(\mathbb{R}^d)$ replaced by $\mathcal{P}_p(\mathbb{R}^d)$ for some $p \in [1, \infty)$), we say that a family $X^{(N)} := \{X^{i,N}\}_{i=1}^N$, seen as a $(C_T)^N$-valued random variable, is a solution to the N-particle system if for $\mathbb{P}$-a.e. $\omega \in \Omega$ it holds

$$X_t^{i,N}(\omega) = \int_0^t B_s(X_s^{i,N}(\omega), L^N(X_s^{(N)}(\omega))) \, ds + Y_t^i \quad \forall t \in [0,T], \, i = 1, \ldots, N \qquad (5.1)$$

where the integral is imposed to be meaningful in the Lebesgue sense for any such ω and

$$L^N(X_t^{(N)}(\omega)) := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}(\omega)}$$

denotes the empirical measure of the system at time t. For notational simplicity we will assume different particle systems to be defined on the same probability space, although we could allow $(\Omega, \mathcal{F}, \mathbb{P})$ to vary as a function of N. In this case we would adopt the notation $Y^{(N)} := \{Y^{i,N}\}_{i=1}^N$ for the sequence of input data below.

We will adopt the following notion of mean field convergence.

Definition 5.1 (Mean Field Convergence). Let $\{Y^i\}_{i \in \mathbb{N}}$ be a sequence of input data such that, defining $Y^{(N)} = \{Y^i\}_{i=1}^N$, one has $\mathbb{P}$-a.s.

$$L^N(Y^{(N)}(\omega)) = \frac{1}{N} \sum_{i=1}^N \delta_{Y^i(\omega)} \rightharpoonup \mu^Y \text{ in } \mathcal{P}(C_T) \text{ as } N \to \infty,$$

with Y such that there exists a unique solution, μ^X to the DDSDE (3.1) with input μ^Y.

Then we say that mean field convergence of the particle system (1.2) to the DDSDE (1.1) holds along $\{Y^i\}_{i\in\mathbb{N}}$ if for any sequence of solutions $\{X^{i,N}\}$ to the associated N-particle system, $\mathbb{P}$-a.s. we have

$$L^N(X^{(N)}(\omega)) \rightharpoonup \mu^X \text{ in } \mathcal{P}(C_T) \text{ as } N \to \infty.$$

Let us remark that although our definition is non-standard, it is well suited to our purposes below: it doesn't require the solution $\{X^{i,N}\}$ to be unique, only the candidate limit μ^X; similarly we don't impose the property to hold for all possible families $\{Y^i\}$, allowing us some freedom in their choice. As the applications in the next section will show, it covers the i.i.d. case for a large class of $\mathcal{L}(Y) \in \mathcal{P}(C_T)$.

We are now ready to present Tanaka's argument; as already mentioned in the introduction, it can be regarded as a "transfer principle" from the DDSDE to the particle system.

Proposition 5.2. *Consider the DDSDE* (3.1) *for a given drift B. Assume there exists $\mu^{Y^1} \in \mathcal{P}(C_T)$ with the following properties:*

i. *there exists a unique solution μ^{X^1} associated to μ^{Y^1};*

ii. *there exist $p \in [1,\infty)$, a measurable function $F : \mathcal{P}_p(C_T) \times \mathcal{P}(C_T) \to [0,+\infty]$ and an increasing continuous function $M : [0,+\infty) \to [0,+\infty)$ with $M(0) = 0$ such that, for any other input μ^{Y^2} and any solution μ^{X^2} associated to μ^{Y^2}, it holds that*

$$d_p(\mu^{X^1}, \mu^{X^2}) \leq F(\mu^{Y^1}, \mu^{Y^2})\, M(d_p(\mu^{Y^1}, \mu^{Y^2})). \tag{5.2}$$

Then for any solution $\{X^{i,N}\}_i$ to the N-particle system associated to $\{Y^i\}_i$, $\mathbb{P}$-a.s.

$$d_p(\mu^{X^1}, L^N(X^{(N)}(\omega)) \leq F(\mu^{Y^1}, L^N(Y^{(N)}(\omega)))\, M\left(d_p\left(\mu^{Y^1}, L^N(Y^{(N)}(\omega))\right)\right). \tag{5.3}$$

Before giving the proof, let us explain the meaning of condition (5.2). In the simplest cases like Section 2, it can be shown that the DDSDE (3.1) is well-posed for all $\mu^Y \in \mathcal{P}(C_T)$ and defines a continuous solution map $S^B : \mathcal{P}(C_T) \to \mathcal{P}(C_T)$ by $\mu^Y \mapsto \mu^X$. In this case condition (5.2) is satisfied for $F \equiv 1$ and M the modulus of continuity of S^B around μ^{Y^1}; if S^B is globally Lipschitz, then $M(r) = Cr$ for some constant $C > 0$. The assumptions of Proposition 5.2 are just a convenient generalization to the case in which S^B is not well-defined anymore, or maybe not on an open neighbourhood of μ^{Y^1} in $\mathcal{P}_p(C_T)$ but only on a strict subset of it (encoded by the presence of F which can take value $+\infty$).

Proof. Let the family $Y^{(N)} = \{Y^i\}_{i=1}^N$ be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, together with a solution $X^{(N)} = \{X^{i,N}\}_{i=1}^N$ to the N-particle system; we can regard $X^{(N)}$, $Y^{(N)}$ as $(C_T)^N$-valued random variables. Define a new probability space $(\Omega_N, \mathcal{F}_N, \mathbb{P}_N)$ by

$$\Omega_N := \{1, \ldots, N\}, \quad \mathcal{F}_N := 2^{\Omega_N}, \quad \mathbb{P}_N := \frac{1}{N}\sum_{i=1}^N \delta_i.$$

We can identify any element $y^{(N)} = \{y^i\}_{i=1}^N \in (C_T)^N$ as a random variable $\Omega_N \ni i \mapsto y^i \in C_T$. With this identification in place, for any $\varphi \in C_b(C_T)$ we have

$$\mathbb{E}_{\mathbb{P}_N}\left[\varphi(y^{(N)})\right] = \frac{1}{N}\sum_{i=1}^N \varphi(y^i) = \left\langle \varphi, \frac{1}{N}\sum_{i=1}^N \delta_{y^i} \right\rangle = \langle \varphi, L^N(y^{(N)}) \rangle;$$

that is, $L^N(y^{(N)}) = \mathcal{L}_{\mathbb{P}_N}(y^{(N)})$. Applying the above to $Y^{(N)}(\omega)$, $X^{(N)}(\omega)$ we deduce that for $\mathbb{P}$-a.e. ω,

$$X_t^{i,N}(\omega) = \int_0^t B_s(X_s^{i,N}(\omega), \mathcal{L}_{\mathbb{P}_N}(X_s^{(N)}(\omega))) \, \mathrm{d}s + Y_t^i(\omega) \quad \forall i \in \Omega_N.$$

Namely for any such ω the solution $X^{(N)}(\omega)$ to the particle system can be regarded as a solution to the DDSDE (3.1) with input $Y^{(N)}(\omega)$ with respect to the probability space $(\Omega_N, \mathcal{F}_N, \mathbb{P}_N)$. Applying assumption (5.2) for the choice $\mu^{Y^2} = \mathcal{L}_{\mathbb{P}_N}(Y^{(N)}(\omega)) = L^N(Y^N(\omega))$, similarly μ^{X^2}, readily gives the conclusion. $\qquad \square$

Remark 5.3. The proof also reveals the following simple fact: if the DDSDE (3.1) is well-posed with a globally defined solution map $S^B : \mathcal{P}_p(C_T) \to \mathcal{P}_p(C_T)$, then any solution to the particle system (5.1) satisfies $X^{(N)}(\omega) = S^B(Y^{(N)}(\omega))$. In particular strong existence, pathwise uniqueness and uniqueness in law hold for the particle system as well.

We immediately deduce the following.

Corollary 5.4. *Suppose μ^{Y^1} satisfies the assumptions of Proposition 5.2. Then the mean field convergence holds along any sequence $\{Y^i\}_i$ such that $\mathbb{P}$-a.s.,*

$$\limsup_{N \to \infty} F\left(\mu^{Y^1}, L^N(Y^{(N)}(\omega))\right) < \infty, \quad \lim_{N \to \infty} d_p\left(\mu^{Y^1}, L^N(Y^{(N)}(\omega))\right) = 0.$$

The above sufficient conditions still need to be checked case-by-case; however if $F \equiv 1$, the next general lemma shows that they hold in the classical case of i.i.d. data.

Lemma 5.5. *Let E be a separable Banach space, $p \in [1, \infty)$ and $\{Z^i\}_{i \geq 1}$ be a family of i.i.d. random variables with common law $\mu \in \mathcal{P}_p(E)$. Then*

$$\lim_{N \to \infty} d_p(L^N(Z^{(N)}(\omega)), \mu) = 0 \quad \text{for } \mathbb{P}\text{-a.e. } \omega. \tag{5.4}$$

Proof. The statement is very close to [17, Lem. 54]; the fact that $L^N(X^N(\omega)) \rightharpoonup \mu$ for $\mathbb{P}$-a.e. ω is a classical result, sometimes referred to as Glivenko–Cantelli theorem. Moreover by the law of large numbers we have

$$\|L^N(X^{(N)})(\omega)\|_p^p = \frac{1}{N} \sum_{i=1}^N \|X^i(\omega)\|_E^p \to \int_E \|x\|_E^p \, \mu(\mathrm{d}x) = \|\mu\|_p^p \quad \mathbb{P}\text{-a.s.}$$

and combining these facts gives the conclusion. $\qquad \square$

Remark 5.6. If the data $\{Y^i\}_{i=1}^N$ are exchangeable (which is true in the i.i.d. case), then so are $\{X^{i,N}\}$. It then follows from [50, Proposition 2.2] that the mean field convergence is equivalent to the property of propagation of chaos; the proof therein is based on purely combinatorial arguments and does not rely on any assumption on Y, thus can be applied directly to our setting.

Remark 5.7. Often one is not just interested in establishing the mean field limit property but also to derive rates of convergence for $d_p(L^N(X^{(N)}), \mathcal{L}(X))$ as a function of N; thanks to Proposition 5.2, once the functions F and M are explicit, it suffices to find similar rates for $d_p(L^N(Y^{(N)}), \mathcal{L}(Y))$. In the case of i.i.d. data, this problem has been studied extensively, see for instance [25, 21]; the most advanced result in the case of measures supported on $\mathbb{R}^d$ we are aware of are the ones presented in [26]. Unfortunately, the results therein cannot be applied to our setting, since they don't deal with convergence in $\mathcal{P}_p(C_T)$ (although they can still be combined with Sznitman's argument to provide explicit estimates, see Section 7.3 from [26]). Indeed, rates of convergence of empirical measures in $\mathcal{P}_p(C_T)$ are much more difficult to obtain and only partial results (mostly relying on Sanov's theorem and Talagrand inequalities) are available, see [5, 6].

5.2 Applications to particular drifts

We are now ready to apply the previous results to the drifts considered in Sections 3-4.

5.2.1 Osgood drifts

We assume that B satisfies Assumption 3.8, with $p = 1$ and Assumption 3.13; it then follows from Theorem 3.18 that the DDSDE has an associated solution map $S^B : \mathcal{P}_1(C_T) \to \mathcal{P}_1(C_T)$ with modulus of continuity M of the form

$$M(r) = G^{-1}\big(G(r) + \|h\|_{L^1_T}\big)$$

for G given by (3.9). By Remark 5.3, the particle system (5.1) is well-posed and $X^{(N)}(\omega) = S^B(Y^{(N)}(\omega))$.

Corollary 5.8. *Under the above assumptions on B, for any $\mu^Y \in \mathcal{P}_1(C_T)$ we have*

$$d_1(L^N(X^{(N)}(\omega)), S^B(\mu^Y)) \leq M(d_1(L^N(Y^{(N)}(\omega)), \mu^Y)) \quad \mathbb{P}\text{-a.s.};$$

as a consequence, mean field convergence holds along any sequence $\{Y^i\}$ such that

$$d_1(L^N(Y^{(N)}(\omega)), \mu^Y) \to 0 \quad \mathbb{P}\text{-a.s.}$$

which is true in particular if $\{Y^i\}$ are taken as i.i.d. variables distributed as μ^Y.

Proof. The first statement is an immediate consequence of Proposition 5.2 for the choice $F \equiv 1$ on $\mathcal{P}_1(C_T)$ and M as above; the second one comes from Corollary 5.4. The claim for i.i.d. variables follows from Lemma 5.5. $\square$

Let us finally mention that the function M can be explicitly computed from the modulus of continuity f, so one can also obtain more precise information on the rate of convergence.

5.2.2 Monotone drifts

We now assume that B satisfies Assumption 3.8 and one between Assumptions 3.20, (3.13) or (3.15) (for the same p); thus Lemma 3.26 is applicable and we have a continuous solution map $S^B : \mathcal{P}_p(C_T) \to \mathcal{P}_p(C_T)$ associated to the DDSDE. As before, by Remark 5.3 the particle system (5.1) is also well-posed.

Corollary 5.9. *Under the above assumptions on B, for any $\mu^Y \in \mathcal{P}_p(C_T)$ we have*

$$d_p(L^N(X^{(N)}(\omega)), S^B(\mu^Y)) \leq M^Y(d_p(L^N(Y^{(N)}(\omega)), \mu^Y)) \quad \mathbb{P}\text{-a.s.}$$

once we take M^Y to be the modulus of continuity of S^B around μ^Y, namely

$$M(r) := \sup\Big\{d_p(S^B(\mu^Y), S^B(\nu)) \mid \nu \in \mathcal{P}_p(C_T),\, d_p(\nu, \mu^{Y^1}) \leq r\Big\}.$$

As a consequence, mean field convergence holds along any sequence $\{Y^i\}$ such that

$$d_p(L^N(Y^{(N)}(\omega)), \mu^Y) \to 0 \quad \mathbb{P}\text{-a.s.}$$

which is true in particular if $\{Y^i\}$ are taken as i.i.d. variables distributed as $\mu^Y \in \mathcal{P}_p(C_T)$.

The proof is identical to that of Corollary 5.8 and thus omitted; let us point out however that in this case only qualitative continuity of S^B is available, thus we can't obtain quantitative rates of convergence.

5.2.3 Locally Lipschitz drifts

Here we impose B to satisfy Assumption 3.27; by Proposition 3.29, we can talk of stability of solutions, but technically we don't have a well defined solution map on all of $\mathcal{P}_p(C_T)$ since the result only applies to μ^Y belonging to the class

$$\mathcal{E}^\alpha := \left\{ \nu \in \mathcal{P}(C_T) : \int_{C_T} e^{\lambda \|\omega\|_{C_T}^\alpha} \, \nu(\mathrm{d}\omega) < \infty \text{ for all } \lambda \in \mathbb{R} \right\}.$$

Nevertheless, this class always includes convex combinations of Diracs and so by Remark 5.3 the result transfers to well-posedness of the particle system (5.1).

Corollary 5.10. *Let B satisfy the above assumptions and fix $p, q \in [1, \infty)$ with $p < q$. Then there exists a constant C such that for any $\mu^Y \in \mathcal{E}^\alpha$ we have*

$$d_p(L^N(X^{(N)}(\omega)), \mu^X)) \leq C \, F(L^N(Y^{(N)}(\omega)), \mu^Y) \, d_q(L^N(Y^{(N)}(\omega)), \mu^Y) \quad \mathbb{P}\text{-a.s.}$$

for F given by

$$F(\nu^1, \nu^2) := \int_{C_T} e^{C\|\omega\|_{C_T}^\alpha} \, \nu^1(\mathrm{d}\omega) + \int_{C_T} e^{C\|\omega\|_{C_T}^\alpha} \, \nu^2(\mathrm{d}\omega).$$

As a consequence, mean field convergence holds along any sequence $\{Y^i\}$ such that

$$\limsup_{N \to \infty} F(L^N(Y^{(N)}(\omega)), \mu^Y) < \infty, \quad d_q(L^N(Y^{(N)}(\omega)), \mu^Y) \to 0 \quad \mathbb{P}\text{-a.s.}$$

which is true in particular if $\{Y^i\}$ are taken as i.i.d. variables distributed as $\mu^Y \in \mathcal{E}^\alpha$.

Proof. The first statement is a consequence of Propositions 3.29 and 5.2; technically here contrary to (5.2) we have $p < q$, but the proof can be readapted to cover this case. The second statement is analogous to Corollary 5.4 and so it only remains to cover the i.i.d. claim. Convergence of $L^N(Y^{(N)})$ to μ^Y in $\mathcal{P}_q$ follows from Lemma 5.5; we have

$$F(L^N(Y^{(N)}(\omega)), \mu^{Y^1}) = \frac{1}{N} \sum_{i=1}^N e^{C\|Y^i\|_{C_T}^\alpha} + \int_{C_T} e^{C\|\omega\|_{C_T}^\alpha} \, \mu^{Y^1}(\mathrm{d}\omega)$$

and by SLLN it holds

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N e^{C\|Y^i\|_{C_T}^\alpha} = \int_{C_T} e^{C\|\omega\|_{C_T}^\alpha} \, \mu^{Y^1}(\mathrm{d}\omega) \quad \mathbb{P}\text{-a.s.}$$

implying the conclusion. □

5.2.4 Convolutional drifts

Let us now assume $B = b * \mu$ for b satisfying the assumptions of Proposition 4.7; recall in particular that $q > d$ and $L_T^1 W_x^{1,q} \hookrightarrow L_T^1 C_x^\alpha$ for some $\alpha > 0$. As a consequence, the equation to the particle system is always meaningful and existence of weak solutions to (5.1) holds by standard arguments; uniqueness however does not need to hold.

Recall that here the input is of the form $Y = \xi + W$, for independent ξ, W with $\mathcal{L}(\xi) \in L_x^p$; if we also considered $Y^{(N)}$ to be of the form $Y^i = \xi^i + W^i$, one would then invoke the more refined results from [11] to show that there exists a unique solution to the particle system (5.1) for $\mathbb{P}$-a.e. realisation of $W^{(N)}(\omega)$ and Lebesgue-a.e. choice of initial data $\xi^{(N)} \in \mathbb{R}^{Nd}$. However we will not need this fact here and we will therefore not impose such an assumption on $Y^{(N)}$.

Corollary 5.11. *Let b, ρ satisfy the assumptions of Proposition 4.7 and denote by μ^X the unique solution to the DDSDE (4.1) associated to $\mu^Y = \mu^{\xi+W}$. Then for any input $Y^{(N)}$ and any solution $X^{(N)}$ to the associated particle system (5.1) it holds that*

$$d_{r'}(L^N(X^{(N)}(\omega)), \mu^{X^1}) \leq C d_{r'}(L^N(Y^{(N)}(\omega)), \mu^Y)) \quad \mathbb{P}\text{-a.s.}$$

for the parameter $r' \in (1, \infty)$ as defined in Proposition 4.7 and given by the relation

$$\frac{1}{r'} = 1 - \frac{q'}{p}, \quad \frac{1}{q} + \frac{1}{q'} = 1.$$

We deduce that mean field convergence holds along any sequence $\{Y^i\}$ such that $\mathbb{P}$-a.s.

$$d_{r'}(L^N(Y^{(N)}(\omega)), \mu^{Y^1}) \to 0$$

which includes the case of $\{Y^i\}$ i.i.d. distributed as $\mu^{\xi+W} \in \mathcal{P}_{r'}(C_T)$.

Proof. The proof follows the same line of arguments as the previous corollaries, this time based on Propositions 5.2 and 4.7. $\quad\square$

In particular Corollary 5.11 holds in the case $\mu^{\xi+W} \in \cap_{n\in\mathbb{N}} \mathcal{P}_n(C_T)$, completing the proof of Theorem 1.1.

A Approximation in metric spaces

Our main result from Section 3.1 relies on approximating suitable drifts $B_t(\mu, x)$ by more regular ones. For this reason, we provide here an abstract result on approximation by Lipschitz functions in metric spaces (E, d); although our main focus is given by $E = \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d)$, the result is true in general.

Standards results of this kind can be found in [34, Theorem 6.8] and [16, Section 6]; however, as they are not directly suited for our purpose, we present the necessary modifications below.

Assumption A.1. *Given a metric space (E, d), we say that a measurable map $f : \mathbb{R}_+ \times E \to \mathbb{R}^d$, has time inhomogeneous, linear growth if there exists a locally integrable function $h : \mathbb{R}_+ \to \mathbb{R}_+$ and some (w.l.o.g. any) $z_0 \in E$ such that*

$$|f_t(z)| \leq h_t(1 + d(z, z_0)) \quad \forall t \geq 0, z \in E. \tag{A.1}$$

In order to prove Proposition A.3 below, we first need the following lemma.

Lemma A.2. *Let $g : E \to \mathbb{R}$ be a bounded map, uniformly continuous on bounded sets. Define*
$$g^n(z) = \inf_{w \in E} \{g(w) + n d_E(z, w)\}; \tag{A.2}$$

the sequence $\{g^n\}_n$ has the following properties:

  i. g^n is n-Lipschitz;

  ii. $g^n(z) \leq g^{n+1}(z) \leq g(z)$ and $|g^n(z)| \leq \|g\|_{C^0}$ for all $z \in E$ and $n \in \mathbb{N}$;

  iii. $g^n \to g$ uniformly on bounded sets.

Proof. Let us set $K = \|g\|_{C^0}$; we can assume $K \neq 0$. Point i. and the first part of ii. follow from [16, Theorem 6.4.1]; by the fact that $g(w) + n d_E(z, w) \geq g(w) \geq -K$ we deduce $g^n(z) \geq -K$ and thus $|g^n(z)| \leq K$ for all $z \in E$ and $n \in \mathbb{N}$. The proof of iii. is mostly

the same as [16, Theorem 6.4.1], which however only deals with globally uniformly continuous functions; for this reason we give it explicitly.

Let us take n big enough so that $2K/n < 1$ and fix $z_0 \in E$, $R > 0$; let m_R denote the modulus of continuity of g on $B_{R+1}(z_0)$, which exists by assumption. For any $w \in E$ such that $d(z, w) \geq 2K/n$ it holds that

$$g(z) + n d_E(z, z) = g(z) \leq K = 2K - K \leq g(w) + n d_E(z, w)$$

which implies that

$$g^n(z) = \inf \big\{ g(w) + n d_E(z, w) : w \in E, d_E(w, z) \leq 2K/n \big\}.$$

In particular if $z \in B_R(z_0)$, since $2K/n < 1$, the infimum is taken over a subset of $B_{R+1}(z_0)$. Fix $\varepsilon > 0$, $z \in B_R(z_0)$; we can find w_n s.t. $d_E(w_n, z) \leq 2K/n$ and $g(w_n) + n d_E(w_n, z) \leq g^n(z) + \varepsilon$. Thus

$$g(w_n) - g^n(z) \leq g(w_n) - g^n(z) + n d_E(w_n, z) \leq \varepsilon$$

and

$$0 \leq g(z) - g^n(z) = g(z) \mp g(w_n) - g^n(z) \leq m_R\Big(\frac{2K}{n}\Big) + \varepsilon \to \varepsilon \quad \text{as } n \to \infty.$$

It follows that for all n large enough,

$$\sup_{z \in B_R(z_0)} |g(z) - g^n(z)| \leq 2\varepsilon;$$

by the arbitrariness of $\varepsilon > 0$, uniform convergence on bounded sets follows. □

Applying Lemma A.2 to time inhomogeneous functions gives the following result.

Proposition A.3. *Let $f : [0, T] \times E \to \mathbb{R}^d$ be as in Assumption A.1 and such that, for every $t \in [0, T]$, $f_t(\,\cdot\,)$ is uniformly continuous on bounded sets. Then there exists a sequence of functions $\{f^n\}_n : [0, T] \times E \to \mathbb{R}^d$ such that:*

  i. *For every $n \in \mathbb{N}$ and $t \in [0, T]$, $f_t^n(\,\cdot\,)$ is globally Lipschitz with constant $C = C(n) > 0$ independent of t.*

  ii. *For every $n \in \mathbb{N}$, $f_t^n(\,\cdot\,)$ has time inhomogeneous linear growth, in the sense of Assumption A.1, with growth function h_t.*

  iii. *For every $t \in [0, T]$, $f_t^n(\,\cdot\,) \to f_t(\,\cdot\,)$ uniformly on bounded sets.*

Proof. Define the map
$$g_t(z) := \frac{f_t(z)}{(1 + d_E(z, z_0))};$$
for every $t \in [0, T]$, by assumption $g_t(\,\cdot\,)$ is bounded and uniformly continuous on bounded sets, so we can apply Lemma A.2; this way we obtain a sequence $g_t^n(x)$, defined for fixed t by equation (A.2). Next, consider $\rho \in C_c^\infty(\mathbb{R}; [0, 1])$ such that $\rho(x) \equiv 1$ for $|x| \leq 1$ and $\rho(x) \equiv 0$ for $|x| \geq 2$, and set

$$f_t^n(z) := g_t^n(z) \left(1 + d_E(z, z_0)\right) \rho\left(\frac{d_E(z, z_0)}{n}\right).$$

We claim that f^n has the desired properties. Indeed, since $|g_t^n(z)| \leq |g_t(z)| \leq h_t$, we also have $|f_t^n(z)| \leq |g_t^n(z)|(1 + d_E(z, z_0)) \leq h_t(1 + d_E(z, z_0))$, so that ii. is satisfied. Item i. is clear from the construction of f^n, while the convergence of item iii. follows from iii. of Lemma A.2. □

B Analytic lemmas and inequalities involving maximal functions

We start by presenting a useful weak-strong convergence lemma.

Lemma B.1. *Let X^n be a sequence of $\mathbb{R}^d$-valued r.v. defined on the same probability space such that $X^n \to X$ in probability; assume that $\mu^n := \mathcal{L}(X^n)$, $\mu := \mathcal{L}(X)$ all belong to L^1_x and $\mu^n \rightharpoonup \mu$ weakly in L^1_x. Finally, consider a sequence $b^n \in L^\infty_x$ such that $\|b^n\|_{L^\infty_x} \leq C$ and $b^n(x) \to b(x)$ for Lebesgue-a.e. x. Then it holds that*

$$\lim_{n \to \infty} \mathbb{E}[|(b^n * \mu^n)(X^n) - (b * \mu)(X)|] = 0.$$

Proof. We split the estimate as follows:

$$
\begin{aligned}
\mathbb{E}[|(b^n * \mu^n)(X^n) - (b * \mu)(X)|] &\leq \mathbb{E}[(|b^n - b| * \mu^n)(X^n)] + \mathbb{E}[|b * (\mu^n - \mu)|(X^n)] \\
&\quad + \mathbb{E}[|(b * \mu)(X^n) - (b * \mu)(X)|] \\
&=: I_1^n + I_2^n + I_3^n.
\end{aligned}
$$

To handle I_1^n, we recall the following classical fact, which can be checked using Egorov's theorem: since $|b^n - b|$ is a uniformly bounded sequence converging Lebesgue a.e. to 0 and μ^n is equiintegrable, $|b^n - b| * \mu^n$ converges pointwise to 0; but then iterating the procedure to $f^n := |b^n - b| * \mu^n$ in place of $|b^n - b|$, we can conclude that $I_1^n = \int f^n \, \mathrm{d}\mu^n \to 0$ as well.

Convergence of I_2^n follows from similar arguments.

Finally, since $b \in L^\infty_x$ and $\mu \in L^1_x$, $b * \mu \in C^0_b$ and so convergence of I_3^n follows from the uniform bounds and the convergence $X^n \to X$ in probability. $\qquad\square$

The remainder of this appendix is devoted to recalling some classical and more recent properties of maximal functions and their use in estimates involving Wasserstein distances.

Maximal functions play a fundamental role in ODEs driven by Sobolev drifts, see for instance [19] for their application; let us recall some fundamental facts, which can be found in [49].

Given $b \in L^p(\mathbb{R}^d)$, $p \in [1, \infty]$, its maximal function Mb is defined by

$$Mb(x) := \sup_{r > 0} \frac{1}{\lambda_d \, r^d} \int_{B(x,r)} |b(y)| \, \mathrm{d}y$$

where λ_d stands for the Lebesgue measure of $B(0,1)$ in $\mathbb{R}^d$. It is well known that if $p \in (1, \infty]$, then $Mf \in L^p(\mathbb{R}^d)$ and

$$\|Mb\|_p \leq c_{d,p}\|b\|_p$$

for some constant $c_{d,p} > 0$; similar definition and properties hold in the case of vector-valued drifts $b \in L^p(\mathbb{R}^d; \mathbb{R}^m)$ (in which case $c = c_{d,p,m}$).

If $b \in W^{1,p}(\mathbb{R}^d; \mathbb{R}^d)$, then there exists a Lebesgue-negligible set $N \subset \mathbb{R}^d$ and a constant $c_d > 0$ such that the Hajlasz inequality holds:

$$|b(x) - b(y)| \leq c_d \, |x - y| \, (MDb(x) + MDb(y)) \quad \forall \, x, y \in \mathbb{R}^d \setminus N. \tag{B.1}$$

The above and similar inequalities allows one to control the map $\mu \mapsto b * \mu$ in Wasserstein spaces; these results are relevant for applications in Section 4.

Lemma B.2. *Let $(p, q, r) \in (1, \infty)^3$ be such that $b \in W^{1,p}_x$, $\mu, \nu \in L^q_x$ and*

$$\frac{r}{p} + \frac{1}{q} \leq 1.$$

Then there exists a constant $C = C(d, p, q, r)$ such that

$$\|b * (\mu - \nu)\|_\infty \leq C\|b\|_{W^{1,p}_x} \left(\|\mu\|_{L^q_x}^{1/r} + \|\nu\|_{L^q_x}^{1/r}\right) d_{r'}(\mu, \nu)$$

Proof. If $d_{r'}(\mu, \nu) = +\infty$ the inequality is trivially true (recall our definition of $d_{r'}(\mu, \nu)$, which is valid for any $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$), so we can assume $d_{r'}(\mu, \nu) < \infty$. Moreover by hypothesis $\mu, \nu \in L_x^{\tilde{q}}$ for any $\tilde{q} \in [1, q]$, therefore w.l.o.g. we can assume $r/p + 1/q = 1$.

Let $m \in \Pi(\mu, \nu)$ be any coupling of (μ, ν) and let $N \subset \mathbb{R}^d$ be as in (B.1); since μ, ν are absolutely continuous w.r.t. Lebesgue, it holds that $m((N \times \mathbb{R}^d) \cup (\mathbb{R}^d \times N)) = 0$. Therefore applying (B.1), for any fixed $x \in \mathbb{R}^d$, gives that

$$
\begin{aligned}
|b * (\mu - \nu)(x)| &= \left| \int_{\mathbb{R}^{2d}} [b(x-y) - b(x-z)] \, m(\mathrm{d}y, \mathrm{d}z) \right| \\
&\lesssim \int_{((\mathbb{R}^d \times N) \cup (N \times \mathbb{R}^d))^c} |y - z| \, (MDb(x-y) + MDb(x-z)) \, m(\mathrm{d}y, \mathrm{d}z) \\
&\leq \left(\int_{\mathbb{R}^{2d}} |y - z|^{r'} \, m(\mathrm{d}y, \mathrm{d}z) \right)^{1/r'} \\
&\quad \times \left(\int_{\mathbb{R}^{2d}} |MDb(x-y) + MDb(x-z)|^r \, m(\mathrm{d}y, \mathrm{d}z) \right)^{1/r}.
\end{aligned}
$$

Moreover we have

$$
\int_{\mathbb{R}^d} |MDb(x-y)|^r \, m(\mathrm{d}y, \mathrm{d}z) = \int_{\mathbb{R}^d} |MDb(x-y)|^r \mu(\mathrm{d}y) \leq \|MDb\|_{L^{rq'}}^r \|\mu\|_{L^q}
$$

$$
\lesssim \|Db\|_{L^{rq'}}^r \|\mu\|_{L^q};
$$

since $rq' = p$, applying a similar estimate to $MDb(x-z)$ we obtain that

$$
\sup_{x \in \mathbb{R}^d} |b * (\mu - \nu)(x)| \lesssim \|b\|_{W^{1,p}} \left(\|\mu\|_{L^q}^{1/r} + \|\nu\|_{L^q}^{1/r} \right) \left(\int_{\mathbb{R}^{2d}} |y - z|^{r'} \, m(\mathrm{d}y, \mathrm{d}z) \right)^{1/r'}.
$$

Minimizing over all possible $m \in \Pi(\mu, \nu)$ we obtain the conclusion. $\qquad\square$

Inequality (B.1) is an example of a family of possibly asymmetric inequalities involving Mb, see [10] for a general discussion. It was shown in [11, Lemma 5.1] that, under the stricter requirement that $b \in W^{1,p}(\mathbb{R}^d; \mathbb{R}^d)$ for $p > d$, there exists a function $g \in L^p(\mathbb{R}^d)$ (which can be taken as $g = c_{d,p} (M|Db|^{\tilde{p}})^{1/\tilde{p}}$ for any $\tilde{p} \in (d, p)$) and a Lebesgue-negligible set $N \subset \mathbb{R}^d$ such that $\|g\|_{L^p} \lesssim_{d,p} \|b\|_{W^{1,p}}$ and

$$
|b(x) - b(y)| \leq g(x) |x - y| \quad \forall x, y \in \mathbb{R}^d \setminus N. \tag{B.2}
$$

Contrary to Lemma B.2, which requires both $\mu, \nu \in L_x^q$, the one-sided inequality (B.2) provides the following estimate.

Lemma B.3. *Let $(p, q, r) \in (1, \infty)^3$ be parameters satisfying*

$$
p > d, \quad \frac{r}{p} + \frac{1}{q} \leq 1. \tag{B.3}
$$

Then there exists a constant $C = C(d, p, q, r)$ such that, for any $b \in W_x^{1,p}$ and $\mu, \nu \in \mathcal{P}$ with $\mu \in L_x^q$, it holds that

$$
|(b * \mu)(x) - (b * \nu)(y)| \leq C \|b\|_{W^{1,p}} (1 + \|\mu\|_{L^q}) \left(|x - y| + d_{r'}(\mu, \nu) \right).
$$

Proof. As in the proof of Lemma B.2, we only need to consider the case $d_{r'}(\mu, \nu) < \infty$ and may take b to be smooth and $r/p + 1/q = 1$ if needed.

Since $p > d$, by the Sobolev embedding $W^{1,p} \hookrightarrow C_b^0$, so $b * \mu$ and $b * \nu$ are well defined for any $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$; by Young's inequality

$$
|(b * \mu)(x) - (b * \mu)(y)| \leq \|b * \mu\|_{W^{1,\infty}} |x - y| \lesssim \|b\|_{W^{1,p}} (1 + \|\mu\|_{L^q}) |x - y|.
$$

Let $m \in \Pi(\mu, \nu)$, then by virtue of (B.2) (as before the negligible set N doesn't play any role), going through similar calculations to Lemma B.2 we have

$$
\begin{aligned}
|(b * \mu)(y) - (b * \nu)(y)| &\leq \left(\int_{\mathbb{R}^d} |g|^r(y - z_1) \, \mathrm{d}\mu(z_1) \right)^{1/r} \left(\int_{\mathbb{R}^{2d}} |z_1 - z_2|^{r'} \, \mathrm{d}m(z_1, z_2) \right)^{1/r'} \\
&\leq \|g\|_{L^{rq'}} \|\mu\|_{L^q}^{1/r} \left(\int_{\mathbb{R}^{2d}} |z_1 - z_2|^{r'} \, \mathrm{d}m(z_1, z_2) \right)^{1/r'} \\
&\lesssim \|b\|_{W^{1,p}} (1 + \|\mu\|_{L^q}) \left(\int_{\mathbb{R}^{2d}} |z_1 - z_2|^{r'} \, \mathrm{d}m(z_1, z_2) \right)^{1/r'}
\end{aligned}
$$

where in the last passage we used the fact that $rq' = p$ and $\|g\|_{L^p} \lesssim \|b\|_{W^{1,p}}$. Combining the two estimates above and taking the infimum over all possible couplings $m \in \Pi(\mu, \nu)$ yields the conclusion. $\qquad \square$

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR2401600

[2] Ismaël Bailleul, Rémi Catellier, and François Delarue, *Propagation of chaos for mean field rough differential equations*, Ann. Probab. **49** (2021), no. 2, 944–996. MR4255135

[3] Martin Bauer, Thilo Meyer-Brandis, and Frank Proske, *Strong solutions of mean-field stochastic differential equations with irregular drift*, Electron. J. Probab. **23** (2018), Paper No. 132, 35. MR3896869

[4] Paul R. Beesack, *On integral inequalities of Bihari type*, Acta Math. Acad. Sci. Hungar. **28** (1976), no. 1-2, 81–88. MR430377

[5] Emmanuel Boissard, *Simple bounds for convergence of empirical and occupation measures in 1-Wasserstein distance*, Electron. J. Probab. **16** (2011), no. 83, 2296–2333. MR2861675

[6] François Bolley, *Quantitative concentration inequalities on sample path space for mean field interaction*, ESAIM Probab. Stat. **14** (2010), 192–209. MR2741965

[7] François Bolley, José A. Cañizo, and José A. Carrillo, *Stochastic mean-field limit: non-Lipschitz forces and swarming*, Math. Models Methods Appl. Sci. **21** (2011), no. 11, 2179–2210. MR2860672

[8] Timothee Bonnefoi, Ajay Chandra, Augustin Moinat, and Hendrik Weber, *A priori bounds for rough differential equations with a non-linear damping term*, arXiv preprint arXiv:2011.06645 (2020).

[9] Didier Bresch, Pierre-Emmanuel Jabin, and Zhenfu Wang, *On mean-field limits and quantitative estimates with a large class of singular kernels: application to the Patlak-Keller-Segel model*, C. R. Math. Acad. Sci. Paris **357** (2019), no. 9, 708–720. MR4018082

[10] Elia Brué, Maria Colombo, and Camillo De Lellis, *Positive solutions of transport equations and classical nonuniqueness of characteristic curves*, Arch. Ration. Mech. Anal. **240** (2021), no. 2, 1055–1090. MR4244826

[11] Laura Caravenna and Gianluca Crippa, *A directional Lipschitz extension lemma, with applications to uniqueness and Lagrangianity for the continuity equation*, Comm. Partial Differential Equations **46** (2021), no. 8, 1488–1520. MR4286465

[12] José A. Carrillo, Jingwei Hu, Li Wang, and Jeremy Wu, *A particle method for the homogeneous Landau equation*, J. Comput. Phys. X **7** (2020), 100066, 24. MR4121055

[13] Thomas Cass and Terry Lyons, *Evolving communities with individual preferences*, Proc. Lond. Math. Soc. (3) **110** (2015), no. 1, 83–107. MR3299600

[14] Paul-Eric Chaudru de Raynal, *Strong well posedness of McKean-Vlasov stochastic differential equations with Hölder drift*, Stochastic Process. Appl. **130** (2020), no. 1, 79–107. MR4035024

[15] Paul-Eric Chaudru de Raynal and Noufel Frikha, *Well-posedness for some non-linear SDEs and related PDE on the Wasserstein space*, J. Math. Pures Appl. (9) **159** (2022), 1–167. MR4377993

[16] Ştefan Cobzaş, Radu Miculescu, and Adriana Nicolae, *Lipschitz functions*, Lecture Notes in Mathematics, vol. 2241, Springer, Cham, 2019. MR3931701

[17] Michele Coghi, Jean-Dominique Deuschel, Peter K. Friz, and Mario Maurelli, *Pathwise McKean-Vlasov theory with additive noise*, Ann. Appl. Probab. **30** (2020), no. 5, 2355–2392. MR4149531

[18] Sonja Cox, Martin Hutzenthaler, and Arnulf Jentzen, *Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations*, arXiv preprint arXiv:1309.5595 (2013).

[19] Gianluca Crippa and Camillo De Lellis, *Estimates and regularity results for the DiPerna-Lions flow*, J. Reine Angew. Math. **616** (2008), 15–46. MR2369485

[20] Sylvain Delattre, Giambattista Giacomin, and Eric Luçon, *A note on dynamical models on random graphs and Fokker-Planck equations*, J. Stat. Phys. **165** (2016), no. 4, 785–798. MR3568168

[21] Steffen Dereich, Michael Scheutzow, and Reik Schottstedt, *Constructive quantization: approximation by empirical measures*, Ann. Inst. Henri Poincaré Probab. Stat. **49** (2013), no. 4, 1183–1203. MR3127919

[22] Roland L. Dobrušin, *Vlasov equations*, Funktsional. Anal. i Prilozhen. **13** (1979), no. 2, 48–58, 96. MR541637

[23] Gonçalo dos Reis, William Salkeld, and Julian Tugaut, *Freidlin-Wentzell LDP in path space for McKean-Vlasov equations and the functional iterated logarithm law*, Ann. Appl. Probab. **29** (2019), no. 3, 1487–1540. MR3914550

[24] Xavier Erny, *Well-posedness and propagation of chaos for McKean–Vlasov equations with jumps and locally Lipschitz coefficients*, arXiv preprint arXiv:2102.06472 (2021).

[25] Nicolas Fournier and Arnaud Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Related Fields **162** (2015), no. 3-4, 707–738. MR3383341

[26] Nicolas Fournier and Arnaud Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probability Theory and Related Fields **162** (2015), no. 3, 707–738. MR3383341

[27] Nicolas Fournier and Arnaud Guillin, *From a Kac-like particle system to the Landau equation for hard potentials and Maxwell molecules*, Ann. Sci. Éc. Norm. Supér. (4) **50** (2017), no. 1, 157–199. MR3621429

[28] Nicolas Fournier, Maxime Hauray, and Stéphane Mischler, *Propagation of chaos for the 2D viscous vortex model*, J. Eur. Math. Soc. (JEMS) **16** (2014), no. 7, 1423–1466. MR3254330

[29] Lucio Galeati, Fabian Harang, and Avi Mayorcas, *Distribution dependent SDEs driven by additive fractional Brownian motion*, arXiv preprint arXiv:2105.14063 (2021).

[30] Jürgen Gärtner, *On the McKean–Vlasov limit for interacting diffusions*, Math. Nachr. **137** (1988), 197–248. MR968996

[31] William R. P. Hammersley, David Šiška, and Ł ukasz Szpruch, *McKean-Vlasov SDEs under measure dependent Lyapunov conditions*, Ann. Inst. Henri Poincaré Probab. Stat. **57** (2021), no. 2, 1032–1057. MR4260494

[32] Maxime Hauray and Pierre-Emmanuel Jabin, *N-particles approximation of the Vlasov equations with singular potential*, Arch. Ration. Mech. Anal. **183** (2007), no. 3, 489–524. MR2278413

[33] Maxime Hauray and Pierre-Emmanuel Jabin, *Particle approximation of Vlasov equations with singular forces: propagation of chaos*, Ann. Sci. Éc. Norm. Supér. (4) **48** (2015), no. 4, 891–940. MR3377068

[34] Juha Heinonen, *Lectures on analysis on metric spaces*, Universitext, Springer-Verlag, New York, 2001. MR1800917

[35] Pierre-Emmanuel Jabin and Zhenfu Wang, *Mean field limit for stochastic particle systems*, Active particles. Vol. 1. Advances in theory, models, and applications, Model. Simul. Sci. Eng. Technol., Birkhäuser/Springer, Cham, 2017, pp. 379–402. MR3644596

[36] Pierre-Emmanuel Jabin and Zhenfu Wang, *Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels*, Invent. Math. **214** (2018), no. 1, 523–591. MR3858403

[37] Daniel Lacker, *On a strong form of propagation of chaos for McKean-Vlasov equations*, Electron. Commun. Probab. **23** (2018), Paper No. 45, 11. MR3841406

[38] Jean-Michel Lasry and Pierre-Louis Lions, *Mean field games*, Jpn. J. Math. **2** (2007), no. 1, 229–260. MR2295621

[39] Henry P. McKean, Jr., *A class of Markov processes associated with nonlinear parabolic equations*, Proc. Nat. Acad. Sci. U.S.A. **56** (1966), 1907–1911. MR221595

[40] Andrej V. Medvedev, *On a concave differentiable majorant of a modulus of continuity*, Real Anal. Exchange **27** (2001/02), no. 1, 123–129. MR1887686

[41] Sima Mehri and Wilhelm Stannat, *Weak solutions to Vlasov-McKean equations under Lyapunov-type conditions*, Stoch. Dyn. **19** (2019), no. 6, 1950042, 23. MR4033001

[42] Yuliya S. Mishura and Alexander Yu Veretennikov, *Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations*, arXiv preprint arXiv:1603.02212 (2016).

[43] William Fogg Osgood, *Beweis der Existenz einer Lösung der Differentialgleichung $\frac{dy}{dx} = f(x,y)$ ohne Hinzunahme der Cauchy-Lipschitz'schen Bedingung*, Monatsh. Math. Phys. **9** (1898), no. 1, 331–345. MR1546565

[44] Sebastian Riedel and Michael Scheutzow, *Rough differential equations with unbounded drift term*, J. Differential Equations **262** (2017), no. 1, 283–312. MR3567487

[45] Michael Röckner and Xicheng Zhang, *Well-posedness of distribution dependent SDEs with singular drifts*, arXiv preprint arXiv:1809.02216 (2018). MR4255229

[46] Michael Scheutzow, *Uniqueness and nonuniqueness of solutions of Vlasov-McKean equations*, J. Austral. Math. Soc. Ser. A **43** (1987), no. 2, 246–256. MR896631

[47] Sylvia Serfaty, *Mean field limits of the Gross-Pitaevskii and parabolic Ginzburg-Landau equations*, J. Amer. Math. Soc. **30** (2017), no. 3, 713–768. MR3630086

[48] Sylvia Serfaty, *Mean field limit for Coulomb-type flows*, Duke Math. J. **169** (2020), no. 15, 2887–2935, With an appendix by Mitia Duerinckx and Serfaty. MR4158670

[49] Elias M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series, No. 30, vol. 2, Princeton University Press, Princeton, N.J., 1970. MR0290095

[50] Alain-Sol Sznitman, *Topics in propagation of chaos*, École d'Été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, pp. 165–251. MR1108185

[51] Hiroshi Tanaka, *Limit theorems for certain diffusion processes with interaction*, Stochastic analysis (Katata/Kyoto, 1982), North-Holland Math. Library, vol. 32, North-Holland, Amsterdam, 1984, pp. 469–488. MR780770

[52] Cédric Villani, *Optimal transport: old and new*, vol. 338, Springer Science & Business Media, 2008. MR2459454

[53] Feng-Yu Wang, *Distribution dependent SDEs for Landau type equations*, Stochastic Process. Appl. **128** (2018), no. 2, 595–621. MR3739509

[54] Toshio Yamada and Shinzo Watanabe, *On the uniqueness of solutions of stochastic differential equations*, J. Math. Kyoto Univ. **11** (1971), 155–167. MR278420