PROPAGATION OF CHAOS FOR MODERATELY INTERACTING PARTICLE SYSTEMS RELATED TO SINGULAR KINETIC MCKEAN-VLASOV SDES

ZIMO HAO, JEAN-FRANCOIS JABIR, STÉPHANE MENOZZI, MICHAEL RÖCKNER, XICHENG ZHANG

Abstract. We study the propagation of chaos in a class of moderately interacting particle systems for the approximation of singular kinetic McKean-Vlasov SDEs driven by α-stable processes. Diffusion parts include Brownian ($\alpha = 2$) and pure-jump ($\alpha \in (1,2)$) perturbations and interaction kernels are considered in a non-smooth anisotropic Besov space.Using Duhamel formula, sharp density estimates (recently issued in [17]), and suitable martingale functional inequalities, we obtain direct estimates on the convergence rate between the empirical measure of the particle systems toward the McKean-Vlasov distribution. These estimates further lead to quantitative propagation of chaos results in the weak and strong sense.

Keywords: Weak/strong propagation of chaos; Moderately interacting particle systems; Kinetic McKean-Vlasov SDEs; Distributional interaction kernel.

AMS Classification: **Primary:** 60H10 ; **Secondary**: 60G52, 35Q70.

1. Introduction

1.1. **McKean-Vlasov SDEs.** In this paper, we establish quantitative propagation of chaos results (in the weak and pathwise sense) for a class of moderately interacting particle systems related to the (formal) second order stable-driven McKean-Vlasov SDEs given, up to some (possibly infinite) time horizon T, by

$$\ddot{X}_t = (b_t * \mu_t)(X_t, \dot{X}_t) + \dot{L}_t^\alpha, \quad (X_0, \dot{X}_0) \sim \mu_0, \quad 0 \leqslant t \leqslant T, \tag{1.1}$$

μ_t standing for the joint law of $(X_t, \dot{X}_t)$. The driving noise L^α is defined as an $\mathbb{R}^d$-valued isotropic α-stable Lévy process, $\mu_0 \in \mathscr{P}(\mathbb{R}^{2d})$ corresponds to the fixed initial probability distribution and b is a time-dependent $\mathbb{R}^d$-valued (Schwartz) distribution over the phase space $\mathbb{R}^{2d}$. Finally, the component $b_t * \mu_t$ denotes, whenever it exists, the convolution between b_t and the law of the McKean-Vlasov SDE on the phase space for a.e. t.

The stable noise is assumed to be given with a stability parameter α in $(1,2]$ (the special case $\alpha = 2$ corresponding to the classical Gaussian noise with $L_t^\alpha = \sqrt{2}W_t$ for W a d-dimensional Brownian motion, and the case $\alpha \in (1,2)$ to the pure-jump case, the associated infinitesimal generator being given by the fractional Laplacian $\Delta^{\alpha/2} := -(-\Delta)^{\alpha/2}$).

Weak and strong wellposedness of (1.1) have been established in [17], in the case where $t \mapsto b_t$ lies in a (negatively smooth) mixed Besov space:

$$b \in L^q([0,T]; \mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}(\mathbb{R}^{2d})) = L^q_T \mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}},$$

where L^q denotes the classical Lebesgue space on the time interval $[0,T]$ and $\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}$ denotes an anisotropic Besov space (see Section 1.3 for a precise definition), for a suitable set of integrability and regularity parameters $q, \boldsymbol{p}$ and $\beta_b < 0$ respectively (the index $\boldsymbol{a}$, given in (1.11), will reflect the

This work is partially supported by National Key R&D program of China (No. 2023YFA1010103) and NNSFC grant of China (No. 12131019), and the German Research Foundation (DFG) through the Collaborative Research Centre(CRC) 1283/2 2021 - 317210226 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications". The second author acknowledges the support of the Russian Basic Research program and Z. Hao, J-F. Jabir and X. Zhang ackowledge the support of the Russian Science Foundation project (project No. 24-11-00123).

intrinsic scales of the underlying kinetic system). Notably, the authors in [17] established that μ_t sits in a balanced duality with b, the resulting drift component $b*\mu$ belonging to $L^s([0,T];L^\infty(\mathbb{R}^{2d}))$ for some appropriate $s>2$. (See also [8, 9] for similar results in the non-degenerate setting.)

Empirically, Equation (1.1) describes the momentum of a generic body located at the position X_t at any time t, evolving in a (possibly anomalous for $\alpha\in(1,2)$) medium and subject to the action of the distribution dependent force field $b_t*\mu_t$, where b models a given interaction kernel. Formally, (1.1) arises as the mean field limit of the interacting particle system:

$$\begin{cases}\ddot X^{N,i}_t=\dfrac{1}{N}\displaystyle\sum_{j=1}^N b_t(X^{N,i}_t-X^{N,j}_t,\dot X^{N,i}_t-\dot X^{N,j}_t)+\dot L^{\alpha,i}_t,\\ (X^{N,i}_0,\dot X^{N,i}_0)\sim\mu_0,\quad 1\leqslant i\leqslant N,\quad 0\leqslant t\leqslant T.\end{cases}\tag{1.2}$$

Introducing the velocity component $V_t:=\dot X_t$, (1.1) can be written as the system of first-order degenerate SDEs:

$$\begin{cases}X_t=X_0+\displaystyle\int_0^t V_s\mathrm{d}s,\quad 0\leqslant t\leqslant T,\\ V_t=V_0+\displaystyle\int_0^t(b_s*\mu_s)(X_s,V_s)\mathrm{d}s+L^\alpha_t.\end{cases}\tag{1.3}$$

If b does not depend on the position variable x (i.e. $b(t,x,v)=b(t,v)$) then SDE (1.3) reduces to the following (first-order) non-degenerate (autonomous) McKean-Vlasov SDE:

$$\mathrm{d}V_t=(b_t*\mu_t)(V_t)\mathrm{d}t+L^\alpha_t,\quad 0\leqslant t\leqslant T.\tag{1.4}$$

Formally, the one-time marginal laws $\{\mu_t\}_{t\geqslant0}$ of $Z_t:=(X_t,V_t)$ give a distributional solution to the following kinetic nonlinear Fokker-Planck equation:

$$\text{for a.e. } 0\leqslant t\leqslant T,\quad \partial_t\mu_t=(\Delta_v^{\frac{\alpha}{2}}-v\cdot\nabla_x)\mu_t-\mathrm{div}_v((b_t*\mu_t)\mu_t)\text{ on }\mathbb{R}^{2d}.\tag{1.5}$$

Establishing wellposedness of (1.5), in the class of anisotropic Besov spaces mentioned above and described in Section 1.3, is a key step to derive weak and/or strong wellposedness of the nonlinear SDE (1.1). From a modeling point of view, while Gaussian noise remains central for the representation of diffusion for molecular motions, the interest for considering more general stable noises has been growing in many applicative fields such as Physics, Biology or advances in deep learning. Experiments have notably evidenced that much more general non-Gaussian noises are at play in nature. We can for instance refer to heavy tailed and the "jump and tumble" phenomena in cell motions, [34]; fractal or Lévy flights pattern in turbulence modelling, [39] and the interest to learning processes based on stochastic gradient descent methods which naturally appear in the occurrence of fast and large excursions facilitating the exploration of multiple minima, [40], [46]. Likewise, the interest for stochastic kinetic models naturally appears in the microscopic description of aggregative social and economical population dynamics ([33] and references therein), cell motions (see again [34]), in Computational Fluid Dynamics and the Lagrangian modeling of turbulent flows ([35], [2]) or for the design of under damped stochastic gradient descents ([11]).

Let us point out that for a singular kernel, giving a rigorous meaning to the formal particle system in (1.2) is a rather involved task. For Lebesgue spaces this has been done in a non degenerate (i.e. no dependence in the x variable for the kernel b) Brownian setting in [41] and in [22] through Girsanov type arguments, and in [16] through a Picard linearization approximation. It remains an open problem to extend the result to other stable noises in that setting or to more singular drifts.

A natural approach to circumvent this difficulty consists in introducing a moderately interacting version of the particle systems, where the initial interaction kernel is replaced by its convolution with an appropriate mollifier. This mollification operation provides the advantage that there are no issues about the definition of the corresponding particle system which is intrinsically well-posed. Propagation of chaos properties can be then estimated by attuning the order of the mollification with the particles size N. This procedure has been methodically employed in the literature, from the seminal papers [28, 29, 30] (from which the terminology *moderately interacting particle* used thereafter originates) to the recent works [14] and [31]. Notably, in [29, 30] (see also [25]), the author derives propagation of chaos results from a fluctuation analysis between a (non-degenerate) case of McKean-Vlasov SDE with local interaction and a moderately interacting particle approximation. To briefly illustrate this derivation, taking (1.4) as a toy model, fluctuations are established from a non-asymptotic control of the distance, set in a suitable functional space, between $\mathrm{Law}(V_t)$ and $\phi_N \star \frac{1}{N}\sum_{i=1}^N \delta_{\{V_t^{N,i}\}}$ (ϕ_N denoting the mollifier for the interaction). This way of capturing propagation of chaos naturally quantifies the cost of the mollification of the interactions, and is further natural for numerical applications and the validation of stochastic particle methods for nonlinear PDEs (see [3, 4], for the particular case of the one-dimensional Burgers equation). The more recent paper [31] proposes a further generalisation of Oelschläger's approach and establishes original quantitative propagation of chaos results for cut-off moderately interacting particle systems with initial local L^p- or Riesz-like - interaction kernels (we may also refer to [32] for more particular physical and biological applications). Our results below further systematize the original setting in [31], addressing kinetic dynamics, distributional interaction kernels and enclosing, in a consistent way, driving noises ranging from the classical Brownian case to pure-jump situations. While some current technical issues (that we briefly discuss in Remark 1.5) restrict the spread of our results, characteristic thresholds (e.g. [24, 5]) and practical instances of singular stochastic kinetic dynamics will be addressed in future work.

To analyse quantitatively propagation of chaos for a mollified version of (1.2), the procedure consists in comparing the Duhamel representation of the Fokker-Planck equation (1.5) and the expansion of the convolution of the empirical measure $\frac{1}{N}\sum_{i=1}^N \delta_{\{X_t^{N,i},V_t^{N,i}\}}$ with the underlying mollifier evaluated along the transport associated with the kinetic operator (see Section 2 for details). This comparison somehow amounts to measure a *moderate propagation of chaos* (see Theorem 2.1 below) and can consequently be used to derive pathwise and weak propagation of chaos. The expansion yields an SPDE. There are then two sources of propagation of chaos:

- the first one, which is of *stability type* and roughly corresponds to the difference between the Duhamel expansion and the drift of the SPDE, is handled through a priori regularity controls on the Fokker-Planck equation. Such controls were obtained in the current kinetic setting in [17], and we can also refer to [8, 9] in the non-degenerate case. We can mention that SDEs with non-regular coefficients have been thoroughly studied, i.e. for time inhomogeneous distributional drifts without non-linear dependence on the law. We can cite e.g. [13], [1] or [7] in the Young regime and [12] or [26] in which the authors manage to go beyond the thresholds appearing in the previous references assuming some appropriate underlying rough path structure. In that last setting defining the drift is a rather delicate point. Importantly, the current McKean-Vlasov framework, which tackles singular kernels, also benefits from the regularizing effects of the law through the convolution. In particular, from the smoothness of the law, it was derived in the quoted works [17], [8, 9] that the nonlinear drift could eventually be seen as a *usual* drift in some Lebesgue space from which some strong uniqueness results, that are important for pathwise propagation of chaos (see e.g. Theorem 1.2 below), were derived.
- the second source is associated with a stochastic integral coming from the expansion of the empirical measure. This term needs to be evaluated in a suitable function space, i.e. the one for which the error is investigated. This naturally induces considering martingale type inequalities in Banach spaces of type M or enjoying the UMD (unconditional martingale difference) property. We refer to Section D and [36] for further details.

The paper is organized as follows: In Sections 1.2 and 1.3, we introduce the moderately interacting particle approximation of Equation (1.3) and essential preliminaries on anisotropic Besov spaces, including key properties that will be used throughout the paper. Our main results on weak and strong propagation of chaos properties of the moderately interacting particle systems are stated in

Theorem 1.2 in Section 1.4. The core of the corresponding proof is presented in Section 2, where we introduce and prove the aforementioned *moderate propagation of chaos* in Theorem 2.1. The proof of Theorem 2.1 is given in Section 2.2, along some auxiliary estimates (Lemmas 2.8 and 2.9) effectively proven in Section 3. From Theorem 2.1, the proof of Theorem 1.2 is addressed in Section 2.3.In the final part of the paper, we append extra-technical results essentially related to anisotropic Besov spaces and used throughout the paper. These results are: about weighted anisotropic Besov and related heat kernel estimates (Appendix A); about the scaling properties of mollifiers under the norm of the anisotropic Besov space (Appendix B); sampling error of the initial condition (Appendix C); functional properties relating the space $\mathbf{B}^{\beta}_{\boldsymbol{p};\boldsymbol{a}}$ to UMD and M-type Banach spaces (Appendix D); and Gronwall's inequality of Volterra type (Appendix E).

1.2. **Moderately interacting particle systems.** Consider the following particle system, with *moderate interactions* , as an approximation for the McKean-Vlasov SDE (1.3):

$$\begin{cases} X^{N,i}_t = \xi^1_i + \int_0^t V^{N,i}_s \,\mathrm{d}s, \quad i = 1, 2, \cdots, N, \\ V^{N,i}_t = \xi^2_i + \int_0^t (b^N_s * \mu^N_s)(Z^{N,i}_s)\,\mathrm{d}s + L^{\alpha,i}_t, \end{cases} \tag{1.6}$$

where $Z^{N,i}_t := (X^{N,i}_t, V^{N,i}_t)$, $Z^{N,i}_0 = \xi_i = (\xi^1_i, \xi^2_i)$, $\mu^N_t := \frac{1}{N}\sum_{i=1}^N \delta_{\{Z^{N,i}_t\}}$, and

$$b^N_t(z) := b_t * \Gamma_t\phi_N(z), \quad \Gamma_t f(x, v) := f(x - tv, v). \tag{1.7}$$

Here, the $\{L^{\alpha,i}\}_{i=1}^N$ are independent copies of L^α, and $\{\xi_i\}_{i=1}^N$ is a family of i.i.d. $\mathbb{R}^{2d}$-valued random variables with common law μ_0. Eventually, for some $\zeta \in (0, 1]$ and a smooth compactly supported and symmetric probability density function $\phi : \mathbb{R}^{2d} \to [0, \infty)$, we define the mollifier

$$\phi_N(x, v) := N^{(2+\alpha)\zeta d}\phi(N^{(1+\alpha)\zeta}x, N^\zeta v). \tag{1.8}$$

The scales in the mollifier reflect the homogeneity of the underlying distance, see (1.14), which will be used to define the corresponding anisotropic Besov spaces. In parallel, the parameter ζ somehow quantifies the strength of the mollification which is needed to define the particle system to compensate the irregularity of b. The operator Γ_t represents the characteristic kinetic transport and its application to ϕ_N will become relevant in our computations later on. Note that the presence of the composition $\Gamma_t\phi_N$ does not alter the nature of the mollifier as, in the distributional sense, $\lim_{N\to\infty} \Gamma_t\phi_N = \delta_{\{0\}}$.

1.3. **Anisotropic Besov space and kinetic semi-group.** In this section we recall the definition of anisotropic Besov spaces with mixed integrability indices as well as their basic properties (see [42], [44] and [17]).

For a multi-index $\boldsymbol{p} = (p_x, p_v) \in [1, \infty]^2$, we first define the Bochner-type iterated space $\mathbb{L}^{\boldsymbol{p}} = L^{p_v}(\mathbb{R}^d; L^{p_x}(\mathbb{R}^d)) = \{f : \mathbb{R}^{2d} \to \mathbb{R}$ Borel measurable : $\|f\|_{\mathbb{L}^{\boldsymbol{p}}} < \infty\}$ for $L^p(\mathbb{R}^d)$ denoting the Lebesgue space on $\mathbb{R}^d$ and

$$\|f\|_{\mathbb{L}^{\boldsymbol{p}}} := \|f\|_{\boldsymbol{p}} := \left(\int_{\mathbb{R}^d} \|f(\cdot, v)\|^{p_v}_{p_x} \mathrm{d}v\right)^{1/p_v}. \tag{1.9}$$

We adopt here the usual convention when one integrability index or the two integrability indices are ∞. That is, for $\boldsymbol{p} = (p_x, \infty)$, $\|f\|_{\boldsymbol{p}} := \mathrm{esssup}_v \|f(\cdot, v)\|_{p_x}$ and $\boldsymbol{p} = (\infty, \infty)$, the $\mathbb{L}^{\boldsymbol{p}}$-norm corresponds to $\mathrm{esssup}_{v,x}\|f(x, v)\|$. Note that the above $\mathbb{L}^{\boldsymbol{p}}$-norm is invariant under the translation operator Γ_t. Namely

$$\|\Gamma_t f\|_{\boldsymbol{p}} = \left(\int_{\mathbb{R}^d} \|f(\cdot - tv, v)\|^{p_v}_{p_x}\mathrm{d}v\right)^{1/p_v} = \left(\int_{\mathbb{R}^d} \|f(\cdot, v)\|^{p_v}_{p_x}\mathrm{d}v\right)^{1/p_v} = \|f\|_{\boldsymbol{p}}. \tag{1.10}$$

This property will be extensively used and the operator Γ_t naturally appears in the current degenerate setting, as it encodes the transport associated with the (linear) first order vector field in

(1.5). Plugging it into the mollifier will actually allow to get rid of the degenerate terms in the stability analysis of Section 2.

Let $\boldsymbol{a}$ be the scaling parameter given by

$$\boldsymbol{a} := (1+\alpha, 1). \tag{1.11}$$

When $\alpha = 2$, i.e. in the Brownian case, the vector $\boldsymbol{a}$ encodes the time exponents of the variances of the entries of $(\int_0^t Z_s \mathrm{d}s, Z_t) = \sqrt{2}(\int_0^t W_s \mathrm{d}s, W_t)$, i.e. $\mathbb{E}[(\int_0^t W_s \mathrm{d}s)^2] = \frac{2t^3}{3}, \mathbb{E}[W_t^2] = 2t$. When $\alpha \in (1,2)$, $\boldsymbol{a}$ still appears when considering invariance or scaling properties of the joint density (see (2.13) below).

For notational convenience, we shall write from here on:

$$\tfrac{1}{\boldsymbol{p}} := \left(\tfrac{1}{p_x}, \tfrac{1}{p_v}\right), \qquad \boldsymbol{a}\cdot\tfrac{d}{\boldsymbol{p}} := \tfrac{(1+\alpha)d}{p_x} + \tfrac{d}{p_v},$$

for any $\boldsymbol{p}, \boldsymbol{q} \in [1,\infty]^2$,

$$\boldsymbol{p} \geqslant \boldsymbol{q} \Leftrightarrow p_x \geqslant q_x,\ p_v \geqslant q_v, \qquad \mathscr{A}_{\boldsymbol{p},\boldsymbol{q}} := \boldsymbol{a}\cdot(\tfrac{d}{\boldsymbol{p}} - \tfrac{d}{\boldsymbol{q}}), \tag{1.12}$$

as well as

$$\boldsymbol{p} \vee \boldsymbol{2} = (p_x \vee 2, p_v \vee 2), \quad \boldsymbol{p} \wedge \boldsymbol{2} = (p_x \wedge 2, p_v \wedge 2).$$

Similarly , we use bold symbols to denote constant vectors in $\mathbb{R}^2$, that is,

$$\boldsymbol{1} := (1,1), \quad \boldsymbol{2} = (2,2), \quad \boldsymbol{d} := (d,d), \quad \boldsymbol{\alpha} := (\alpha,\alpha), \ \boldsymbol{\infty} := (\infty,\infty).$$

Finally, we use the symbol $\boldsymbol{p}'$ to denote the conjugate index of $\boldsymbol{p}$, i.e.,

$$\tfrac{1}{\boldsymbol{p}} + \tfrac{1}{\boldsymbol{p}'} = \boldsymbol{1} \Leftrightarrow \boldsymbol{p}' = (p'_x, p'_v) \text{ with } \frac{1}{p_x} + \frac{1}{p'_x} = 1 = \frac{1}{p_v} + \frac{1}{p'_v}. \tag{1.13}$$

These general notations settled, we shall now introduce the class of anisotropic Besov spaces $\mathbf{B}^{\beta,q}_{\boldsymbol{p};\boldsymbol{a}}$ characterizing the singularity of the interaction kernel b. To this aim, let us introduce some preliminaries. For an L^1-integrable function f on $\mathbb{R}^{2d}$, let $\hat{f}$ be the Fourier transform of f defined by

$$\hat{f}(\xi) := \int_{\mathbb{R}^{2d}} \mathrm{e}^{-\mathrm{i}\xi\cdot z} f(z)\mathrm{d}z, \quad \xi \in \mathbb{R}^{2d},$$

and $\check{f}$ the Fourier inverse transform of f defined by

$$\check{f}(z) := (2\pi)^{-2d} \int_{\mathbb{R}^{2d}} \mathrm{e}^{\mathrm{i}\xi\cdot z} f(\xi)\mathrm{d}\xi, \quad z \in \mathbb{R}^{2d}.$$

For $z = (x,v)$ and $z' = (x',v')$ in $\mathbb{R}^{2d}$, we introduce the anisotropic distance

$$|z - z'|_{\boldsymbol{a}} := |x - x'|^{1/(1+\alpha)} + |v - v'|. \tag{1.14}$$

Note that $z \mapsto |z|_{\boldsymbol{a}}$ is not smooth at the origin. This distance appears very naturally in connection with the homogeneity of the underlying linear degenerate operator. In the kinetic setting we can refer to the work by Priola [37], or [10, 18, 23, 19, 17], who employed this distance for establishing Schauder type estimates. In the current work, it will be used in order to define the corresponding anisotropic Besov spaces (see Definition 1.1 below).

For $r > 0$ and $z \in \mathbb{R}^{2d}$, we also introduce the ball centred at z and with radius r with respect to the above distance as follows:

$$B^{\boldsymbol{a}}_r(z) := \{z' \in \mathbb{R}^{2d} : |z' - z|_{\boldsymbol{a}} \leqslant r\}, \ \ B^{\boldsymbol{a}}_r := B^{\boldsymbol{a}}_r(0).$$

Let $\chi^{\boldsymbol{a}}_0$ be a symmetric C^∞-function on $\mathbb{R}^{2d}$ with

$$\chi^{\boldsymbol{a}}_0(\xi) = 1 \text{ for } \xi \in B^{\boldsymbol{a}}_1 \text{ and } \chi^{\boldsymbol{a}}_0(\xi) = 0 \text{ for } \xi \notin B^{\boldsymbol{a}}_2.$$

We define

$$\phi^{\boldsymbol{a}}_j(\xi) := \begin{cases} \chi^{\boldsymbol{a}}_0(2^{-j\boldsymbol{a}}\xi) - \chi^{\boldsymbol{a}}_0(2^{-(j-1)\boldsymbol{a}}\xi), & j \geqslant 1, \\ \chi^{\boldsymbol{a}}_0(\xi), & j = 0, \end{cases}$$

where for $s \in \mathbb{R}$ and $\xi = (\xi_1, \xi_2)$,

$$2^{s\boldsymbol{a}}\xi = (2^{(1+\alpha)s}\xi_1, 2^s\xi_2).$$

Note that

$$\mathrm{supp}(\phi_j^{\boldsymbol{a}}) \subset \big\{\xi : 2^{j-1} \leqslant |\xi|_{\boldsymbol{a}} \leqslant 2^{j+1}\big\},\ j \geqslant 1,\ \mathrm{supp}(\phi_0^{\boldsymbol{a}}) \subset B_2^{\boldsymbol{a}}, \tag{1.15}$$

and

$$\sum_{j\geqslant 0} \phi_j^{\boldsymbol{a}}(\xi) = 1, \quad \forall \xi \in \mathbb{R}^{2d}. \tag{1.16}$$

Let $\mathscr{S}$ be the space of all Schwartz functions on $\mathbb{R}^{2d}$ and $\mathscr{S}'$ be the dual space of $\mathscr{S}$, that is the tempered distribution space. For given $j \geqslant 0$, the dyadic anisotropic block operator $\mathcal{R}_j^{\boldsymbol{a}}$ is defined on $\mathscr{S}'$ by

$$\mathcal{R}_j^{\boldsymbol{a}} f(z) := (\phi_j^{\boldsymbol{a}} \hat{f})\check{}\,(z) = \check{\phi}_j^{\boldsymbol{a}} * f(z), \tag{1.17}$$

where the convolution is understood in the distributional sense and by scaling,

$$\check{\phi}_j^{\boldsymbol{a}}(z) = 2^{(j-1)(2+\alpha)d}\check{\phi}_1^{\boldsymbol{a}}(2^{(j-1)\boldsymbol{a}}z), \quad j \geqslant 1. \tag{1.18}$$

Similarly, we can define the classical isotropic block operator $\mathcal{R}_j f = \check{\phi}_j * f$ in $\mathbb{R}^d$, for $\{\phi_j\}_{j\geqslant 0}$ a smooth partition of unity of $\mathbb{R}^d$ where

$$\mathrm{supp}(\phi_j) \subset \big\{\xi : 2^{j-1} \leqslant |\xi| \leqslant 2^{j+1}\big\},\ j \geqslant 1,\ \mathrm{supp}(\phi_0) \subset B_2(0). \tag{1.19}$$

Now we introduce the following anisotropic Besov spaces and mixed (anisotropic) Besov spaces.

Definition 1.1. *Let $s \in \mathbb{R}$, $q \in [1, \infty]$ and $\boldsymbol{p} \in [1, \infty]^2$. The anisotropic Besov space is defined by*

$$\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}} := \left\{ f \in \mathscr{S}'(\mathbb{R}^{2d}) \ :\ \|f\|_{\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}} := \left(\sum_{j\geqslant 0} \left(2^{js}\|\mathcal{R}_j^{\boldsymbol{a}} f\|_{\boldsymbol{p}}\right)^q \right)^{1/q} < \infty \right\},$$

where $\|\cdot\|_{\boldsymbol{p}}$ is defined in (1.9). Similarly, one defines the usual isotropic Besov spaces $\mathbf{B}^{s,q}_p$ in $\mathbb{R}^d$ in terms of isotropic block operators $\mathcal{R}_j$. If there is no confusion, we shall write

$$\mathbf{B}^{s}_{\boldsymbol{p};\boldsymbol{a}} := \mathbf{B}^{s,\infty}_{\boldsymbol{p};\boldsymbol{a}}.$$

For $s_0, s_1 \in \mathbb{R}$, the mixed Besov space is defined by

$$\mathbf{B}^{s_0,s_1}_{\boldsymbol{p};x,\boldsymbol{a}} := \left\{ f \in \mathscr{S}'(\mathbb{R}^{2d}) \ :\ \|f\|_{\mathbf{B}^{s_0,s_1}_{\boldsymbol{p};x,\boldsymbol{a}}} := \sup_{k,j\geqslant 0} 2^{\frac{ks_0}{1+\alpha}} 2^{js_1} \|\mathcal{R}_k^x \mathcal{R}_j^{\boldsymbol{a}} f\|_{\boldsymbol{p}} < \infty \right\},$$

where for any $f : \mathbb{R}^{2d} \to \mathbb{R}$,

$$\mathcal{R}_k^x f(x, v) := \mathcal{R}_k f(\cdot, v)(x).$$

The space $\mathbf{B}^s_{\boldsymbol{p};\boldsymbol{a}}$ (with $s < 0$) will characterize the class of distributional interaction kernel b and the irregularity-integrability ranges of μ_0. The mixed Besov space $\mathbf{B}^{s_0;s_1}_{\boldsymbol{p};x,\boldsymbol{a}}$ will be essentially used in the study of strong convergence for the propagation of chaos.

1.4. **Main results.** To state our main results, we introduce the following assumptions:

(H) $b \in L^\infty(\mathbb{R}_+; \mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}})$ *and* $\mu_0 \in \mathscr{P}(\mathbb{R}^{2d}) \cap \mathbf{B}^{\beta_0}_{\boldsymbol{p}_0;\boldsymbol{a}}$ *with* $\boldsymbol{p}_0 = (p_{x,0}, p_{v,0}) \in [1, \infty]^2$, $\boldsymbol{p}_b = (p_{x,b}, p_{v,b}) \in [1, \infty)^2$ *satisfying* $\frac{1}{\boldsymbol{p}_0} + \frac{1}{\boldsymbol{p}_b} \geqslant \mathbf{1}$, *and* $\beta_b \leqslant 0$ *and* $\beta_0 \in (-1, 0)$ *are such that*

$$0 < \Lambda := \mathcal{A}_{\boldsymbol{p}_0,\boldsymbol{p}_b'} - \beta_0 - \beta_b = \boldsymbol{a} \cdot \tfrac{d}{\boldsymbol{p}_0} - \beta_0 + \boldsymbol{a} \cdot \tfrac{d}{\boldsymbol{p}_b} - \beta_b - (2 + \alpha)d < \alpha - 1. \tag{1.20}$$

We shall use the following parameter set:

$$\Theta := \big(\alpha, d, \boldsymbol{p}_0, \boldsymbol{p}_b, \beta_0, \beta_b, \|\mu_0\|_{\mathbf{B}^{\beta_0}_{\boldsymbol{p}_0;\boldsymbol{a}}}, \|b\|_{L^\infty(\mathbb{R}_+;\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}})}\big),$$

to highlight, when pertinent, the dependency of estimates on the parameters α, d, etc., e.g. when we say $C = C(\Theta)$, it means that the constant C may depend on all or part of the parameters in Θ. Additional dependency on the order ζ of the mollification or other relevant parameters will be emphasized in the same way.

Under the above assumptions, by [17, Theorem 1.3], there exists a finite time horizon $T_0 = T_0(\Theta) > 0$ for which the McKean-Vlasov SDE (1.3) admits a unique weak solution $Z_t = (X_t, V_t)$ on the time interval $[0, T_0]$. Under the stronger condition that b is in $L^\infty([0, T_0], \mathbf{B}^{s_0,s_1}_{\boldsymbol{p}_0;x,\boldsymbol{a}})$, where $\mathbf{B}^{s_0,s_1}_{\boldsymbol{p}_0;x,\boldsymbol{a}}$ is as in Definition 1.1 with the regularity parameters s_0, s_1 and Λ as in **(H)** satisfying

$$s_0 = 1 + \tfrac{\alpha}{2}, \qquad s_1 = \tfrac{\alpha}{2} - 1 + \beta_b, \qquad \Lambda \in \Big(0, \tfrac{3}{2}\alpha - 2\Big), \tag{1.21}$$

pathwise uniqueness also holds (see again [17, Theorem 1.3]).

In order to give the range of ζ in the definition of the mollifier ϕ_N, we introduce two quantities $\mathfrak{m}_\alpha$ and θ_α for later use:

$$\mathfrak{m}_\alpha := 1/((p_{x,0} \wedge p_{v,0} \wedge 2) \vee \alpha), \quad \theta_\alpha := \begin{cases} \mathscr{A}_{\mathbf{1},\mathbf{2}} = \mathbf{a} \cdot \frac{d}{\mathbf{2}} = 2d, \ \alpha = 2, \\ \big[\mathscr{A}_{\mathbf{1},\boldsymbol{p}_0}\big] \vee \big[\mathscr{A}_{\boldsymbol{p}_b,\boldsymbol{\infty}} - (1+\alpha)\beta_b\big], \ \alpha \in (1,2). \end{cases} \tag{1.22}$$

The main result of this paper is encompassed in the following:

Theorem 1.2. *Suppose that* **(H)** *holds and either* $\boldsymbol{p}_0 > \boldsymbol{\alpha}$, *or* $(\alpha, \boldsymbol{p}_0) = (2, \mathbf{1})$ *and* μ_0 *satisfies:*

$$\mu_0(|\cdot|^m_{\boldsymbol{a}}) := \mathbb{E}[|(X_0, V_0)|^m_{\boldsymbol{a}}] < \infty, \quad \forall m \in \mathbb{N}.$$

Then, for any $\beta, \zeta > 0$ *with the following upper bounds*

$$\beta < \bar{\beta}_\alpha := \begin{cases} 1 - \Lambda \ \textit{if}\ \alpha = 2, \\ (\alpha - 1 - \Lambda) \wedge ((\alpha + \beta_0 - \Lambda)/2) \ \textit{if}\ \alpha \in (1,2), \end{cases} \quad \text{and } \zeta < (1 - \mathfrak{m}_\alpha)/\theta_\alpha,$$

and, for any $\varepsilon > 0$ *there is a constant* $C = C(\Theta, m, \beta, \zeta, \varepsilon) > 0$ *such that for all* $N \geqslant 1$, *and* $t \in [0, T_0]$,

$$\|\mathbb{P} \circ (Z^{N,1}_t)^{-1} - \mathbb{P} \circ Z_t^{-1}\|_{\rm var} \lesssim_C N^{-\beta\zeta} + N^{\mathfrak{m}_\alpha - 1 + \zeta\theta_\alpha + \varepsilon}, \tag{1.23}$$

where $\|\cdot\|_{\rm var}$ *is the total variation distance on* $\mathscr{P}(\mathbb{R}^{2d})$. *Moreover, if in addition* $1 - \alpha/2 < (\alpha + \beta_0 - \Lambda)/2$, b *lies in the mixed Besov space* $L^\infty(\mathbb{R}_+; \mathbf{B}^{s_0,s_1}_{\boldsymbol{p}_b;x,\boldsymbol{a}})$ *and* (1.21) *holds, then for any* ζ *as above and for* β *in the range*

$$\beta \in (1 - \alpha/2, \bar{\beta}_\alpha),$$

we have

$$\Big\|\sup_{t \in [0,T_0]} |Z^{N,1}_t - Z^1_t|\Big\|_{L^2(\Omega)} \lesssim_C N^{-\beta\zeta} + N^{\mathfrak{m}_\alpha - 1 + \zeta\theta_\alpha + \varepsilon}, \tag{1.24}$$

where $Z^1 = (X^1, V^1)$ *is the strong solution of* (1.3) *driven by the noise* $L^{\alpha,1}$ *and starting from* ξ_1.

Remark 1.3 (*About the convergence rates.*)**.** *The convergence rate in the previous Theorem is mainly associated with the characteristic parameters* β *and* $\mathfrak{m}_\alpha$ *(we recall that* ζ *appeared in the definition of the mollifier in the particle system, see* (1.6)-(1.8)*).*

- *For the parameter* β, *two thresholds appear in the upper-bound. The first one,* $\alpha - 1 - \Lambda$, *comes from the smoothness of the underlying Fokker-Planck equation* (1.5) *and corresponds to the gap in the condition* (1.20) *(see e.g.* (2.48)*). The other contribution,* $(\alpha + \beta_0 - \Lambda)/2$, *is related to time integrability issues associated with the norm for which we estimate the error. It only appears in the pure stable case due to the constraint* $\boldsymbol{p}_0 > \boldsymbol{\alpha}$ *whereas considering* $\boldsymbol{p}_0 = \mathbf{1}$ *in the Brownian case* $\alpha = 2$ *spares this contribution. The division by 2 follows from the quadratic nature of the* *nonlinearity in* (1.5) *which also appears for the error analysis when* $\boldsymbol{p}_0 \neq \mathbf{1}$ *(see the proof of Theorem 2.1 p. 14).*

On the other hand, when considering the pathwise approach, a lower bound is also needed in order to guarantee the sufficient smoothness on the underlying PDE to apply a Zvonkin type argument (see (2.49),(2.50)*).*

- *For the parameter* $\mathfrak{m}_\alpha$ *the range mainly corresponds to that of a limit Theorem. It actually appears from the control of a stochastic integral that appears in the SPDE satisfied by the error (see as well Theorem 2.1 and* (2.23), (2.17)*). If* $\alpha = 2$ *then* $\mathfrak{m}_\alpha = \frac{1}{2}$*, which is the usual rate in a Gaussian central limit theorem. For* $\alpha \in (1,2)$*, since* $\boldsymbol{p}_0 > \boldsymbol{\alpha}$*,* $\mathfrak{m}_\alpha = 1/((p_{x,0} \wedge p_{v,0} \wedge 2)$*. If* $\boldsymbol{p}_0$ *is somehow close to* $\boldsymbol{\alpha}$ *then* $\mathfrak{m}_\alpha$ *is close to* $\frac{1}{\alpha}$ *yielding a principal rate which is close to the stable limit theorem, i.e. in that case* $1 - \mathfrak{m}_\alpha \simeq 1 - \frac{1}{\alpha}$*. On the other hand, for large values of* $\boldsymbol{p}_0$*, the Gaussian regime again prevails.*
In any case, the previous rates need to be deflated (for integrability reasons because an underlying singular term (convolution between the mollifier and the gradient of the corresponding stable heat kernel). This naturally again makes the parameter ζ *appear (see Lemma 2.9,* (2.23), (2.17)*). Eventually, the parameter* θ_α *is constrained by the choice of the function space in which we analyze the error that actually allows to exploit suitable controls on the underlying Fokker-Planck equation* (1.5)*.*
- *Balancing the two previous errors terms: in order that the two errors have the same range we have to set*

$$N^{-\beta\zeta} = N^{\mathfrak{m}_\alpha - 1 + \zeta\theta_\alpha + \varepsilon} \iff \zeta(\theta_\alpha + \beta) = 1 - \mathfrak{m}_\alpha - \varepsilon \iff \zeta = \frac{1 - \mathfrak{m}_\alpha - \varepsilon}{(\theta_\alpha + \beta)},$$

yielding a final bound in

$$N^{-\frac{\beta}{\theta_\alpha + \beta}(1 - \mathfrak{m}_\alpha - \varepsilon)}.$$

- *For* $\alpha = 2$*, observe first that since* $\boldsymbol{p}_0 = 1$*,* $\Lambda = -\beta_0 + \boldsymbol{a} \cdot \frac{d}{\boldsymbol{p}_b} - \beta_b$*. In that case* Λ *can be arbitrarily small provided* β_0, β_b *and* $\boldsymbol{p}$ *are respectively small and large enough. In this setting,* β *could be taken arbitrarily close to 1 and the corresponding rate would read* $N^{-\frac{1}{2d+1}(\frac{1}{2} - \varepsilon)} = N^{-\frac{1}{4d+2}(1-2\varepsilon)} = N^{-\frac{1}{\boldsymbol{a}\cdot\boldsymbol{d}+2}(1-2\varepsilon)}$*. Since* β *is close to 1 in that case, the exponent in the previous convergence rate also corresponds to the typical magnitude of the parameter* ζ*. For* N *particles in dimension* $2d$ *this has to be compared to the expected interaction range, which from the multi-scale effect due to the kinetic dynamics would actually read*[1] $(1/d + 3d)^- = (1/4d)^-$.
- *For* $\alpha \in (1,2)$*, if* $\boldsymbol{p}_0$*,* $\boldsymbol{p}_b'$ *are sufficiently close and* β_0, β_b *are small enough, then* Λ *is again small. Since* $(\alpha - 1 - \Lambda) < (\alpha - \Lambda)/2 \iff \alpha - \Lambda < 2$*, for* β_0 *small enough, one can take* β *arbitrarily close to* $\alpha - 1$*. If now* $\boldsymbol{p}_0 > \boldsymbol{\alpha}$ *but close to* $\boldsymbol{\alpha}$ *the rate writes:*

$$N^{-\frac{\alpha-1}{\theta_\alpha + \alpha - 1}(1 - \frac{1}{\alpha} - \varepsilon)}.$$

Eventually for $\boldsymbol{p}_0, \boldsymbol{p}_b'$ *close and* β_b *small* $\theta_\alpha \simeq \mathscr{A}_{\mathbf{1},\boldsymbol{p}_0} = \boldsymbol{a} \cdot \frac{d}{\boldsymbol{p}_0'} \simeq d(2+\alpha)\frac{\alpha^+ - 1}{\alpha^+}$ *and the rate rewrites as*

$$N^{-\frac{\alpha-1}{d(2+\alpha)+\alpha}(1 - \frac{\alpha}{\alpha-1}\varepsilon)} = N^{-\frac{\alpha-1}{\boldsymbol{a}\cdot\boldsymbol{d}+\alpha}(1 - \frac{\alpha}{\alpha-1}\varepsilon)}.$$

There is then a continuity in the stability parameter for the final convergence rate. The previous ones are somehow the best achievable rates. The more singular the parameter the worse the rate.

Let us finally point out that the weak propagation of chaos result (1.23) *shall be understood as a yet-to-be improved threshold, the rate being directly, as for* (1.24)*, derived from the moderate propagation of chaos stated in Theorem 2.1. This leaves open the possibility of extended more refined semigroup techniques (e.g. [6]) to our present singular framework open. This will be addressed in future works.*

[1]From now on for $a \in \mathbb{R}$ we denote be a^- (resp. a^+) any real number of the form $a - \varepsilon$ (resp. $a = \varepsilon$), $\varepsilon > 0$.

Remark 1.4 (About the sampling of the initial condition)**.** *Our results can be extended to the case where the initial data ξ_i are not necessarily i.i.d. variables. Specifically, assuming only exchangeability of the initial states $(X_0^{N,i}, V_0^{N,i})$, the convergence rates in Theorem 1.2 (and Theorem 2.1) still hold if*

$$\lim_{N\to\infty} N^{c_0\zeta} \|\phi_N * (\mu_0^N - \mu_0)\|^n_{L^n(\Omega;\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}})} = 0$$

*with some constants $c_0 > 0$ and $n \in \mathbb{N}$. This assertion can be explicitly drawn from the proof of Theorem 2.1 below (see the estimates (2.42) to (2.44) in **Step 3**) from which Theorem 1.2 is derived. Since the representation involving c_0 and n can be intricate, for simplicity, we have assumed independence and identical distribution of the initial data. The i.i.d. assumption is made for simplicity.*

Remark 1.5 (About the integrability parameters $\boldsymbol{p}_b$ and $\boldsymbol{p}_0$)**.** *We would like to specify some points concerning the assumptions on the integrability parameter $\boldsymbol{p}_b \in [1, +\infty)^2$ of the singular drift b and the particular thresholds set on $\boldsymbol{p}_0$ in Theorem 1.2 (and Theorem 2.1). The restrictions $\boldsymbol{p}_b \neq (\infty, \infty)$ and $\boldsymbol{p}_0 \leqslant \boldsymbol{p}_b'$ (equivalent to $\boldsymbol{p}_b \leqslant \boldsymbol{p}_0'$) stated in* **(H)** *are both inherent to the wellposedness of the McKean-Vlasov (1.3), and might potentially be overcome in future works (see e.g. [9] for the non-degenerate case). As it will be actually seen through the propagation of chaos analysis below, the more particular condition $\boldsymbol{p}_0 > \boldsymbol{\alpha}$ arises from the analysis of the stochastic fluctuations in the mollified empirical measure $\Gamma_t\phi_N * \mu_t^N$ driving (1.6) (see precisely (2.17), (2.18) and (2.19) below). To control these fluctuations we have to rely on martingale type inequalities for stochastic integrals valued in Banach spaces satisfying some appropriate functional properties, notably, as in [31, 32], the UMD property for the Brownian case $\alpha = 2$, and the further M-type Banach space property for the jump case $\alpha \in (1, 2)$. (Precise definitions of these spaces and properties related to our setting are presented in Appendix D.) These functional properties are primarily to be satisfied by the subspace $\mathbb{L}^{\boldsymbol{p}_0} \cap \mathbb{L}^{\boldsymbol{p}_b'}$. Importantly, $\mathbb{L}^{\boldsymbol{p}}$ spaces have the UMD property for $\boldsymbol{p} = (p_x, p_v)$ when neither p_x nor p_v belonged to $\{1, \infty\}$. Whenever $p_{x,b} = 1$ or $p_{v,b} = 1$ (or $\boldsymbol{p} = \mathbf{1}$) we are led to consider $p_{x,b'} = \infty$ or $p_{v,b'} = \infty$ and $\mathbb{L}^{\boldsymbol{p}_b'}$ does not fulfill the aforementioned UMD property. This difficulty can be circumvented using the embedding (2.12) between Besov spaces which ensures b lies in the larger $\mathbf{B}^{\tilde\beta_b}_{\tilde{\boldsymbol{p}}_b;\boldsymbol{a}}$-space with $\boldsymbol{p}_b < \tilde{\boldsymbol{p}}_b$ ($\Rightarrow \boldsymbol{p}_b' > \tilde{\boldsymbol{p}}_b'$) and $\tilde\beta_b = \beta_b - \mathscr{A}_{\boldsymbol{p}_b,\tilde{\boldsymbol{p}}_b}$. One can so decrease the integrability exponent to a finite one, up to a slight decrease of the regularity of b. Since the differential index $\tilde\beta_b - \boldsymbol{a}\cdot\frac{d}{\tilde{\boldsymbol{p}}_b} = \beta_b - \boldsymbol{a}\cdot\frac{d}{\boldsymbol{p}_b}$ remains unchanged, the rates (1.23) and (1.24) in Theorem 1.2 and the rate (2.6) in Theorem 2.1 are preserved. This approach applies to both the diffusive case $\alpha = 2$ and the pure jump case $\alpha \in (1, 2)$.*

The particular case $\boldsymbol{p}_0 = (1, 1)$ features a more significant threshold. In the Brownian case, similarly to e.g. [31], [32], one can proceed with weighted $\mathbb{L}^{\boldsymbol{p}}$ space with some $\boldsymbol{p} > \mathbf{1}$ and the cost of introducing a weight is somehow absorbed by the Gaussian noise. Such procedure naturally seemingly fails in the pure jump case, and for the case $\alpha \in (1, 2)$ the condition $\boldsymbol{p}_0 > \boldsymbol{\alpha}$ (ensuring $\mathbb{L}^{\boldsymbol{p}_0}$ is a M-type Banach) is needed to exhibit a convergence rate.

The situation $\alpha \in (1, 2)$ for large integrability parameter $\boldsymbol{p}_b$ or integrability parameter $\boldsymbol{p}_0$ below $\boldsymbol{\alpha}$ is so a particular open case in our paper and a natural restriction in regard to applications. Extending this present situation is at the moment beyond our proof arguments and would rather require an additional and suitable alteration (e.g. a truncation) of the Lévy noise $L^{\alpha,i}$ in the mollified particle system (1.6). Importantly, observe that the constraint $p_{x,0} \wedge p_{v,0} > \alpha$ actually prevents the situations $p_{x,b} = \infty$ or $p_{v,b} = \infty$ and to consider the case formally embedding (when $\beta_b = 0$) fully bounded interactions, or partially bounded interactions. Namely when the kernel depends only on the velocity ($b_t(x, v) = h_1(t, v)$) or only the position component ($b_t(x, v) = h_2(t, x)$), the integrability index $\boldsymbol{p}_b = (\boldsymbol{p}_{x,b}, \boldsymbol{p}_{v,b})$ including so an ∞ component. When the dependence occurs in the velocity, $\boldsymbol{p}_b = (\infty, \boldsymbol{p}_{v,b}), \boldsymbol{p}_{v,b} < +\infty, \boldsymbol{p}_{v,b}' > \alpha$, this leads to somehow consider an autonomous non-degenerate McKean-Vlasov equation for which our propagation of chaos analysis would apply (for the sake of concision, details concerning this case have been purposely left aside). On the other *hand, when $\boldsymbol{p}_b = (\boldsymbol{p}_{x,b}, \infty), \boldsymbol{p}_{x,b} < +\infty$, the presence of the transport of the velocity component makes things much more tricky. This specific point will be the subject of further research.*

2. Convergence of empirical measure via Duhamel's formula

Throughout this section we assume **(H)** is in force and recall that for the time horizon $T_0 > 0$, the McKean-Vlasov SDE (1.3) admits a unique weak solution on $[0, T_0]$. Moreover, by [17, Theorem 3.11, (i)], the time marginal law $\mu_s(\mathrm{d}z) = u_s(z)\mathrm{d}z$ satisfies the Fokker-Planck equation (1.5) and for any $\beta \geqslant 0$,

$$\sup_{s\in(0,T_0]} \left(s^{\frac{\beta-\beta_0}{\alpha}} \|u_s\|_{\mathbf{B}^{\beta,1}_{\boldsymbol{p}_0;\boldsymbol{a}}} + s^{\frac{\Lambda+\beta}{\alpha}} \|u_s\|_{\mathbf{B}^{\beta-\beta_b,1}_{\boldsymbol{p}'_b;\boldsymbol{a}}} \right) < \infty, \tag{2.1}$$

where $\boldsymbol{p}'_b$ is the conjugate index of $\boldsymbol{p}_b$ in the sense of (1.13).

Let ϕ_N be defined by (1.8). Consider the following mollified approximation of the empirical measure μ^N_t:

$$u^N_t(z) := \mu^N_t * \Gamma_t\phi_N(z) = \frac{1}{N}\sum_{i=1}^{N} (\Gamma_t\phi_N)(Z^{N,i}_t - z), \tag{2.2}$$

where $\Gamma_t\phi_N$ corresponds to the composition of the transport operator Γ_t with ϕ_N namely:

$$\Gamma_t\phi_N(x,v) = N^{(2+\alpha)\zeta d}\phi(N^{(1+\alpha)\zeta}(x - tv), N^{\zeta}v),$$

where ϕ is symmetric. Note that this specific choice for the mollifier involving the transport operator will precisely allow to derive a suitable SPDE for the difference $u_t - u^N_t$ (see (2.21) below).

Let $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$, β_0, $\boldsymbol{p}_0$ and Λ be as in **(H)**. For any $f \in \mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}} \cap \mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}'_b;\boldsymbol{a}}$ and $\beta, t \geqslant 0$, we introduce

$$\|f\|_{\mathbb{S}^\beta_t(b)} := (1\wedge t)^{\frac{\beta-\beta_0}{\alpha}} \|f\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}} + (1\wedge t)^{\frac{\beta+\Lambda}{\alpha}} \|b_t * f\|_{\boldsymbol{\infty}}. \tag{2.3}$$

The main result of this section is the following convergence of the mollified empirical measure (2.2).

Theorem 2.1. *Assume that either $\boldsymbol{p}_0 > \boldsymbol{\alpha}$ or $(\alpha, \boldsymbol{p}_0) = (2, \mathbf{1})$ and*

$$\mu_0(|\cdot|^m_{\boldsymbol{a}}) < \infty, \quad \forall m \in \mathbb{N}. \tag{2.4}$$

Given m_α and θ_α as in (1.22), for any $\beta, \zeta > 0$ such that

$$\beta < (\alpha - 1 - \Lambda) \wedge ((\alpha + \beta_0 - \Lambda)/2), \quad \zeta < (1 - \mathfrak{m}_\alpha)/\theta_\alpha, \tag{2.5}$$

and for any $\varepsilon > 0$, $m > 0$, there exists a constant $C = C(\Theta, \zeta, \beta, \varepsilon, m) > 0$ such that for all $N \geqslant 1$,

$$\sup_{t\in[0,T_0]} \|u^N_t - u_t\|_{L^m(\Omega;\mathbb{S}^\beta_t(b))} \lesssim_C N^{-\beta\zeta} + N^{\mathfrak{m}_\alpha - 1 + \zeta\theta_\alpha + \varepsilon}. \tag{2.6}$$

The rest of the section will be dedicated to the proof of Theorem 2.1.

2.1. Technical preliminaries on the anisotropic Besov spaces.

The following Bernstein inequality is standard and proven in [44].

Lemma 2.2 (Bernstein's inequality). *For any $\boldsymbol{k} := (k_1, k_2) \in (\mathbb{N}\setminus\{0\})^2$ and $\boldsymbol{p}, \boldsymbol{p}_1 \in [1,\infty]^2$ with $\boldsymbol{p}_1 \leqslant \boldsymbol{p}$, there is a constant $C = C(d, \boldsymbol{k}, \boldsymbol{p}, \boldsymbol{p}_1) > 0$ such that, for all $j \geqslant 0$, $f = f(x,v) \in \mathcal{S}'$,*

$$\|\nabla^{k_1}_x \nabla^{k_2}_v \mathcal{R}^{\boldsymbol{a}}_j f\|_{\boldsymbol{p}} \lesssim_C 2^{j\boldsymbol{a}\cdot(\boldsymbol{k}+\frac{d}{\boldsymbol{p}_1}-\frac{d}{\boldsymbol{p}})} \|\mathcal{R}^{\boldsymbol{a}}_j f\|_{\boldsymbol{p}_1} = 2^{j\left((1+\alpha)k_1+k_2+\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}}\right)} \|\mathcal{R}^{\boldsymbol{a}}_j f\|_{\boldsymbol{p}_1}, \tag{2.7}$$

for $\nabla^{k_1}_x$ and $\nabla^{k_2}_v$ denoting respectively the k^{th}_1 and k^{th}_2 differential operators with respect to the variable x and the variable v.

For a function $f:\mathbb{R}^{2d}\to\mathbb{R}$, the first-order difference operator is defined by

$$\delta_h^{(1)}f(z):=\delta_h f(z):=f(z+h)-f(z),\quad z,h\in\mathbb{R}^{2d}. \tag{2.8}$$

For $M\in\mathbb{N}$, the $M^{\rm th}$-order difference operator is defined recursively by

$$\delta_h^{(M+1)}f(z)=\delta_h\circ\delta_h^{(M)}f(z).$$

The following characterization is well-known (see e.g. [44] and [19, Theorem 2.7]).

Proposition 2.3. *For $s>0$, $q\in[1,\infty]$ and $\boldsymbol{p}\in[1,\infty]^2$, an equivalent norm of $\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}$ is given by*

$$\|f\|_{\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}}\asymp\|f\|_{\boldsymbol{p}}+\left(\int_{|h|_{\boldsymbol{a}}\leqslant 1}\left(\frac{\|\delta_h^{([s]+1)}f\|_{\boldsymbol{p}}}{|h|^s_{\boldsymbol{a}}}\right)^q\frac{\mathrm{d}h}{|h|_{\boldsymbol{a}}^{(2+\alpha)d}}\right)^{1/q}, \tag{2.9}$$

where $[s]$ is the integer part of s. In particular, $\mathbf{C}^s_{\boldsymbol{a}}:=\mathbf{B}^{s;\infty}_{\boldsymbol{\infty};\boldsymbol{a}}$ is the anisotropic Hölder-Zygmund space, and for $s\in(0,1)$, there is a constant $C=C(\alpha,d,s)>0$ such that

$$\|f\|_{\mathbf{C}^s_{\boldsymbol{a}}}\asymp_C\|f\|_{\boldsymbol{\infty}}+\sup_{z\neq z'}|f(z)-f(z')|/|z-z'|^s_{\boldsymbol{a}}.$$

The above characterization will be notably used, in the appendix section A, for establishing weighted versions of key estimates in this Section including the Bernstein inequality and Lemma 2.4 below.

We recall the following lemma proved in [17, Lemma 2.6] on embeddings and Young convolution inequalities related to $\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}$.

Lemma 2.4. *(i) For any $\boldsymbol{p}\in[1,\infty]^2$, $s'>s$ and $q\in[1,\infty]$, it holds that*

$$\mathbf{B}^{0,1}_{\boldsymbol{p};\boldsymbol{a}}\hookrightarrow\mathbb{L}^{\boldsymbol{p}}\hookrightarrow\mathbf{B}^{0,\infty}_{\boldsymbol{p};\boldsymbol{a}},\quad\mathbf{B}^{s',\infty}_{\boldsymbol{p};\boldsymbol{a}}\hookrightarrow\mathbf{B}^{s,1}_{\boldsymbol{p};\boldsymbol{a}}\hookrightarrow\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}. \tag{2.10}$$

(ii) For any $\beta,\beta_1,\beta_2\in\mathbb{R}$, $q,q_1,q_2\in[1,\infty]$ and $\boldsymbol{p},\boldsymbol{p}_1,\boldsymbol{p}_2\in[1,\infty]^2$ with

$$\beta=\beta_1+\beta_2,\quad 1+\tfrac{1}{\boldsymbol{p}}=\tfrac{1}{\boldsymbol{p}_1}+\tfrac{1}{\boldsymbol{p}_2},\quad\tfrac{1}{q}=\tfrac{1}{q_1}+\tfrac{1}{q_2},$$

it holds that, for some universal constant $C>0$,

$$\|f*g\|_{\mathbf{B}^{\beta,q}_{\boldsymbol{p};\boldsymbol{a}}}\leqslant C\|f\|_{\mathbf{B}^{\beta_1,q_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}\|g\|_{\mathbf{B}^{\beta_2,q_2}_{\boldsymbol{p}_2;\boldsymbol{a}}}. \tag{2.11}$$

(iii) For $\mathbf{1}\leqslant\boldsymbol{p}_1\leqslant\boldsymbol{p}\leqslant\infty$, $q\in[1,\infty]$ and $s=s_1+\boldsymbol{a}\cdot\big(\frac{d}{\boldsymbol{p}}-\frac{d}{\boldsymbol{p}_1}\big)$, there is a $C>0$ such that

$$\|f\|_{\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}}\lesssim_C\|f\|_{\mathbf{B}^{s_1,q}_{\boldsymbol{p}_1;\boldsymbol{a}}}. \tag{2.12}$$

Now we recall the basic estimates for the kinetic semi-group. Let $p_t(z)$ be the distributional density of $Z_t:=(\int_0^t L^\alpha_s\mathrm{d}s,L^\alpha_t)$. By scaling, it is easy to see that

$$p_t(z)=t^{-(1+\frac{2}{\alpha})d}p_1(t^{-\frac{1+\alpha}{\alpha}}x,t^{-\frac{1}{\alpha}}v),\quad t>0. \tag{2.13}$$

The kinetic semigroup of operator $\Delta_v^{\frac{\alpha}{2}}-v\cdot\nabla_x$ is given by

$$P_tf(z):=\mathbb{E}\big[f(\Gamma_tz+Z_t)\big]=\Gamma_t(p_t*f)(z)=(\Gamma_tp_t*\Gamma_tf)(z)=\int_{\mathbb{R}^{2d}}\Gamma_tp_t(z-z')\Gamma_tf(z')\mathrm{d}z', \tag{2.14}$$

where

$$\Gamma_tf(z):=f(\Gamma_tz):=f(x-tv,v).$$

For $f\in\mathcal{S}'$ and $t>0$, let

$$\mathscr{I}^f_t(x,v):=\int_0^tP_{t-s}f(s,x,v)\mathrm{d}s.$$

By the definition (2.14) of P_t, it is easy to see that in the distributional sense,

$$\partial_t\mathscr{I}^f_t=\Delta_v^{\frac{\alpha}{2}}\mathscr{I}^f_t-v\cdot\nabla_x\mathscr{I}^f_t+f.$$

The following estimates are proven in [17, Lemma 2.12].

Lemma 2.5. *Let $\boldsymbol{p},\boldsymbol{p}_1\in[1,\infty]^2$ with $\boldsymbol{p}_1\leqslant\boldsymbol{p}$. For any $\beta\in\mathbb{R}$ and $\gamma\geqslant 0$, there is a constant $C=C(\alpha,d,\boldsymbol{p},\boldsymbol{p}_1,\beta,\gamma)>0$ such that for any $f\in\mathbf{B}^{\beta}_{\boldsymbol{p};\boldsymbol{a}}$ and all $j\geqslant 0$, and $t>0$,*

$$\|\mathcal{R}^{\boldsymbol{a}}_j P_t f\|_{\boldsymbol{p}}\lesssim_C 2^{j(\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}}-\beta)}((2^{j\alpha}t)^{-\gamma}\wedge 1)\|f\|_{\mathbf{B}^{\beta}_{\boldsymbol{p}_1;\boldsymbol{a}}}. \tag{2.15}$$

The following lemma provides useful estimates about the kinetic semigroup P_t (see [17, Lemma 2.16]).

Lemma 2.6. *Let $\boldsymbol{p},\boldsymbol{p}_1\in[1,\infty]^2$ with $\boldsymbol{p}_1\leqslant\boldsymbol{p}$. For any $\beta,\beta_1\in\mathbb{R}$, there is a $C_0=C_0(\alpha,d,\beta,\beta_1,\boldsymbol{p},\boldsymbol{p}_1)>0$ such that for all $t>0$,*

$$\mathbb{1}_{\{\beta\neq\beta_1-\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}}\}}\|P_tf\|_{\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}}}+\|P_tf\|_{\mathbf{B}^{\beta}_{\boldsymbol{p};\boldsymbol{a}}}\lesssim_{C_0}(1\wedge t)^{-\frac{(\beta-\beta_1+\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}})\vee 0}{\alpha}}\|f\|_{\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}. \tag{2.16}$$

2.2. **Proof of Theorem 2.1.** For notation simplicity, we shall write

$$\mathcal{U}^N_t(z):=u_t(z)-u^N_t(z),$$

and, for any smooth function $f:\mathbb{R}^{2d}\to\mathbb{R}$ and all $z_0=(x_0,v_0)\in\mathbb{R}^d\times\mathbb{R}^d$, we shall denote respectively by $\nabla_xf(z_0)$ and $\nabla_vf(z_0)$ the gradients of f w.r.t. the x and v variable, evaluated at point z_0. To prove (2.6), we shall exploit the Duhamel formula related to the evolution of $\mathcal{U}^N_t$. For a.e. $t\in(0,T_0]$, the density u_t satisfies

$$\partial_tu_t=\big(\triangle^{\frac{\alpha}{2}}_v-v\cdot\nabla_x\big)u_t-\mathrm{div}_v\big((b_t*u_t)u_t\big)$$

for div_v denoting the divergence operator on the variable v. The Duhamel formula for u_t (see e.g. [17, Section 1]) then gives:

$$u_t(z)=P_t\mu(z)-\int_0^tP_{t-s}(\mathrm{div}_v(b_s*u_s)u_s)(z)\,\mathrm{d}s,\quad t\in[0,T_0].$$

We next look at the the evolution equation satisfied by $\mathcal{U}^N_t(z)$. For any fixed $z_0\in\mathbb{R}^{2d}$, applying Itô's formula to $t\to(\Gamma_t\phi_N)(Z^{N,i}_t-z_0)$, we have

$$\begin{aligned}\mathrm{d}(\Gamma_t\phi_N)(Z^{N,i}_t-z_0)&=\big(\partial_t\Gamma_t\phi_N+\Delta^{\frac{\alpha}{2}}_v\Gamma_t\phi_N\big)(Z^{N,i}_t-z_0)\mathrm{d}t+V^{N,i}_t\cdot(\nabla_x\Gamma_t\phi_N)(Z^{N,i}_t-z_0)\mathrm{d}t\\&\quad+(b^N_t*\mu^N_t)(Z^{N,i}_t)\cdot(\nabla_v\Gamma_t\phi_N)(Z^{N,i}_t-z_0)\mathrm{d}t+\mathrm{d}M^{N,i}_t(z_0),\end{aligned}$$

where

$$M^{N,i}_t(z_0):=\begin{cases}\sqrt{2}\displaystyle\int_0^t(\nabla_v\Gamma_s\phi_N)(Z^{N,i}_s-z_0)\mathrm{d}W^i_s, & \alpha=2,\\ \displaystyle\int_0^t\int_{\mathbb{R}^d\setminus\{0\}}\delta^{(1)}_{(0,v)}(\Gamma_s\phi_N)(Z^{N,i}_s-z_0)\tilde{\mathcal{N}}^i(\mathrm{d}s,\mathrm{d}v), & \alpha\in(1,2),\end{cases} \tag{2.17}$$

with $\delta^{(1)}_{(0,v)}$ being the (forward) difference operator on the variable v defined in (2.8), - i.e. $\delta^{(1)}_{(0,v)}f(z)=f(z+(0,v))-f(z)$ - and $\tilde{\mathcal{N}}^i$ standing for the compensated Poisson measure associated with $L^{\alpha,i}$, the driving noise of the i^{th} particle of the system (1.6).

Noting that for any smooth function $f:\mathbb{R}^{2d}\to\mathbb{R}$,

$$\partial_t\Gamma_tf(z_0)=-v_0\cdot\nabla_x\Gamma_tf(z_0),\quad\forall z_0=(x_0,v_0)\in\mathbb{R}^d\times\mathbb{R}^d,$$

we have

$$\begin{aligned}(\partial_t\Gamma_t\phi_N)(Z^{N,i}_t-z_0)+V^{N,i}_t\cdot(\nabla_x\Gamma_t\phi_N)(Z^{N,i}_t-z_0)&=v_0\cdot\nabla_x\Gamma_t\phi_N(Z^{N,i}_t-z_0)\\&=-v_0\cdot\nabla_x\big(\Gamma_t\phi_N(Z^{N,i}_t-\cdot)\big)(z_0),\end{aligned}$$

denoting by $\nabla_x(f(Z^{N,i}_t-\cdot))(z_0)$ the x-derivative of $z\mapsto f(Z^{N,i}_t-\cdot)$ evaluated in z_0, and

$$\Delta^{\frac{\alpha}{2}}_vu^N_t(z_0)=\Delta^{\frac{\alpha}{2}}_v(\Gamma_t\phi_N*\mu^N_t)(z_0)=(\Delta^{\frac{\alpha}{2}}_v\Gamma_t\phi_N*\mu^N_t)(z_0).$$

Thus, by the definition of $u_t^N(z)$ in (2.2) and for b_t^N as in (1.7), we have for all $z_0=(x_0,v_0)\in\mathbb{R}^{2d}$,

$$\begin{aligned}\mathrm{d}u_t^N(z_0) &= \big((\Delta_v^{\frac{\alpha}{2}}-v_0\cdot\nabla_x)u_t^N(z_0)-\langle(b_t^N*\mu_t^N)\cdot(\nabla_v\Gamma_t\phi_N)(\cdot-z_0),\mu_t^N(\cdot)\rangle\big)\mathrm{d}t+\mathrm{d}M_t^N(z_0)\\ &= \Big((\Delta_v^{\frac{\alpha}{2}}-v_0\cdot\nabla_x)u_t^N(z_0)-\mathrm{div}_vG_t^N(z_0)\Big)\,\mathrm{d}t+\mathrm{d}M_t^N(z_0),\end{aligned}\tag{2.18}$$

where

$$M_t^N(z_0):=\frac{1}{N}\sum_{i=1}^N M_t^{N,i}(z_0),\tag{2.19}$$

with $M_t^{N,i}(z_0)$ being defined in (2.17) . As $b_t^N*\mu_t^N=b_t*u_t^N$,

$$G_t^N(z_0):=\langle(b_t^N*\mu_t^N)(\cdot)(\Gamma_t\phi_N)(\cdot-z_0),\mu_t^N\rangle=\langle(b_t*u_t^N)(\cdot)(\Gamma_t\phi_N)(\cdot-z_0),\mu_t^N\rangle.\tag{2.20}$$

Combining (2.18) with (1.5) and changing z_0 to z, we obtain that $\mathcal{U}_t^N$ solves the SPDE

$$\mathrm{d}\mathcal{U}_t^N=\Big((\Delta_v^{\frac{\alpha}{2}}-v\cdot\nabla_x)\mathcal{U}_t^N-\mathrm{div}_vH_t^N\Big)\,\mathrm{d}t-\mathrm{d}M_t^N,\tag{2.21}$$

where

$$H_t^N:=(b_t*u_t)u_t-G_t^N.\tag{2.22}$$

In particular, we derive from (2.21)the Duhamel formula,

$$\mathcal{U}_t^N=P_t(u_0-u_0^N)-\int_0^t P_{t-s}\mathrm{div}_vH_s^N\mathrm{d}s-\int_0^t P_{t-s}\mathrm{d}M_s^N=:\mathcal{I}_t^N-\mathcal{H}_t^N-\mathcal{M}_t^N,\tag{2.23}$$

where $\mathcal{M}_t^N=\frac{1}{N}\sum_{i=1}^N\mathcal{M}_t^{N,i}$ and

$$\mathcal{M}_t^{N,i}(z_0):=\begin{cases}\sqrt{2}\displaystyle\int_0^t P_{t-s}(\nabla_v\Gamma_s\phi_N)(Z_s^{N,i}-z_0)\mathrm{d}W_s^i, & \alpha=2,\\ \displaystyle\int_0^t\int_{\mathbb{R}^d\setminus\{0\}}P_{t-s}\delta_{(0,v)}^{(1)}(\Gamma_s\phi_N)(Z_s^{N,i}-z_0)\tilde{n}^i(\mathrm{d}s,\mathrm{d}v), & \alpha\in(1,2).\end{cases}\tag{2.24}$$

Applying the norm $\|\cdot\|_{\mathbb{S}_t^\beta(b)}$ set in (2.3) to the above, it follows that

$$\|\mathcal{U}_t^N\|_{\mathbb{S}_t^\beta(b)}\leqslant\|\mathcal{I}_t^N\|_{\mathbb{S}_t^\beta(b)}+\|\mathcal{H}_t^N\|_{\mathbb{S}_t^\beta(b)}+\|\mathcal{M}_t^N\|_{\mathbb{S}_t^\beta(b)}.\tag{2.25}$$

Now, we first look at the estimate of $\|\mathcal{I}_t^N\|_{\mathbb{S}_t^\beta(b)}$.

Lemma 2.7. *For any $\beta>0$, there is a constant $C=C(\Theta,\beta)>0$ such that, for all $N\geqslant 1$,*

$$\sup_{t>0}\|\mathcal{I}_t^N\|_{\mathbb{S}_t^\beta(b)}\lesssim N^{-\beta\zeta}\|u_0\|_{\mathbf{B}^{\beta_0}_{\boldsymbol{p}_0;\boldsymbol{a}}}+\|\phi_N*(\mu_0^N-u_0)\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}},\tag{2.26}$$

where $\zeta>0$ is from the definition (1.8) of ϕ_N.

Proof. Recalling that $\beta-\beta_0>0$, by the heat kernel estimate (2.16) in Lemma 2.6, we have

$$\|\mathcal{I}_t^N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}}\lesssim(1\wedge t)^{-\frac{\beta-\beta_0}{\alpha}}\|\phi_N*\mu_0^N-u_0\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}},$$

and recalling the condition $\boldsymbol{p}_0\leqslant\boldsymbol{p}_b'$ in (**H**), the definition (1.20) of Λ, and that $\beta>0,\Lambda>0$, (2.16) also yields

$$\|\mathcal{I}_t^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}}\lesssim(1\wedge t)^{-\frac{\beta+\Lambda}{\alpha}}\|\phi_N*\mu_0^N-u_0\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}}.$$

Hence, as the convolution inequality (2.11) ensures that $\|b_t*\mathcal{I}_t^N\|_\infty\lesssim\|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}}\|\mathcal{I}_t^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}}$,

$$\begin{aligned}\|\mathcal{I}_t^N\|_{\mathbb{S}_t^\beta(b)} &\lesssim(1\wedge t)^{\frac{\beta-\beta_0}{\alpha}}\|\mathcal{I}_t^N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}}+(1\wedge t)^{\frac{\beta+\Lambda}{\alpha}}\|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}}\|\mathcal{I}_t^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}}\\ &\lesssim\|\phi_N*\mu_0^N-u_0\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}}\leqslant\|\phi_N*u_0-u_0\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}}+\|\phi_N*(\mu_0^N-u_0)\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}}.\end{aligned}\tag{2.27}$$

Note that a duality argument and the Besov equivalence (2.9) yields, for all $z \in \mathbb{R}^d$,

$$\|u_0(\cdot - z) - u_0(\cdot)\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}} \lesssim |z|^\beta_{\boldsymbol{a}} \|u_0\|_{\mathbf{B}^{\beta_0}_{\boldsymbol{p}_0;\boldsymbol{a}}}.$$

We thus have, according to the scaling estimate (B.1) for ϕ_N stated in Section B of the appendix,

$$\|\phi_N * u_0 - u_0\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}} \lesssim \||\cdot|^\beta_{\boldsymbol{a}} \phi_N\|_{\mathbf{1}} \|u_0\|_{\mathbf{B}^{\beta_0}_{\boldsymbol{p}_0;\boldsymbol{a}}} = N^{-\beta\zeta} \||\cdot|^\beta_{\boldsymbol{a}} \phi_1\|_{\mathbf{1}} \|u_0\|_{\mathbf{B}^{\beta_0}_{\boldsymbol{p}_0;\boldsymbol{a}}}.$$

Plugging this inequality into (2.27) we obtain (2.26). □

The estimates for handling the components $\|\mathcal{H}^N_t\|_{\mathbb{S}^\beta_t(b)}$ and the moments of $\|\mathcal{M}^N_t\|_{\mathbb{S}^\beta_t(b)}$ in (2.25) are stated in the two next lemmas. The first one exhibits a term which goes to zero with a rate depending on the smoothness of the underlying Fokker-Planck equation and an integral contribution writing essentially in terms of $\mathcal{U}^N_t$ and $\|u^N_t\|_{\boldsymbol{p}_0}$, i.e. these terms will enter in a Gronwall or circular type argument. The second lemma, associated with the stochastic integral, essentially produces a convergence rate close to the one of the corresponding stable limit theorem. Proofs are postponed to the next Section 3.

Lemma 2.8. *(Estimate for $\mathcal{H}^N_t$) For any $\beta \in \big(0, (\alpha + \beta_0 - \Lambda) \wedge 1)\big)$, there is a constant $C = C(\Theta, \beta) > 0$ such that, for all $t \in (0, T_0]$ and $N \geqslant 1$,*

$$\|\mathcal{H}^N_t\|_{\mathbb{S}^\beta_t(b)} \lesssim N^{-\beta\zeta} + \int_0^t G_\beta(t,s)\Big(\|b_s * \mathcal{U}^N_s\|_{\boldsymbol{\infty}} \|u^N_s\|_{\boldsymbol{p}_0} + s^{-\frac{\Lambda}{\alpha}} \|\mathcal{U}^N_s\|_{\boldsymbol{p}_0}\Big) \mathrm{d}s, \tag{2.28}$$

where

$$G_\beta(t,s) := t^{\frac{\beta-\beta_0}{\alpha}} (t-s)^{-\frac{1}{\alpha}} + t^{\frac{\beta+\Lambda}{\alpha}} (t-s)^{-\frac{\Lambda+\beta_0+1}{\alpha}}. \tag{2.29}$$

Lemma 2.9 (Estimate for $\mathcal{M}^N$)**.** *Under the assumptions of Theorem 2.1, for any $\varepsilon > 0$, $m \in \mathbb{N}$ and $\beta, \zeta > 0$ with the following upper bounds*

$$\zeta < (1 - \mathfrak{m}_\alpha)/\theta_\alpha, \tag{2.30}$$

there is a constant $C = C(\varepsilon, m, \beta, \zeta) > 0$ such that for all $N \in \mathbb{N}$,

$$\sup_{t\in[0,T_0]} \|\mathcal{M}^N_t\|_{L^m(\Omega;\mathbb{S}^\beta_t(b))} \lesssim_C N^{\mathfrak{m}_\alpha - 1 + \zeta\theta_\alpha + \varepsilon}. \tag{2.31}$$

With the above preparatory results we are now able to demonstrate Theorem 2.1.

Proof of Theorem 2.1. Applying Lemmas 2.7 and 2.8 to (2.25), we have

$$\|\mathcal{U}^N_t\|_{\mathbb{S}^\beta_t(b)} \lesssim I^N_0 + \|\mathcal{M}^N_t\|_{\mathbb{S}^\beta_t(b)} + \int_0^t G_\beta(t,s)\Big(\|b_s * \mathcal{U}^N_s\|_\infty \|u^N_s\|_{\boldsymbol{p}_0} + s^{-\frac{\Lambda}{\alpha}} \|\mathcal{U}^N_s\|_{\boldsymbol{p}_0}\Big) \mathrm{d}s, \tag{2.32}$$

where $\beta \in \big(0, (\alpha + \beta_0 - \Lambda) \wedge 1\big)$, $G_\beta(t,s)$ is defined in (2.29) and

$$I^N_0 := N^{-\beta\zeta} + \|\phi_N * (\mu^N_0 - u_0)\|_{\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}}}. \tag{2.33}$$

It now remains to handle through a suitable circular argument the terms involving $\mathcal{U}^N_s$ in the right hand side of (2.32), for which we need to make the $\|\cdot\|_{\mathbb{S}^\beta_s(b)}$ norm appear. In particular, the technical difficulty arises from the factor $\|u^N_s\|_{\boldsymbol{p}_0}$, which requires to distinguish the cases $\boldsymbol{p}_0 = (1,1)$ and $\boldsymbol{p}_0 \neq \mathbf{1}$. For the former, only the case $\alpha = 2$ is of interest, the pure-jump case requiring the property $\alpha < p_{x,0} \wedge p_{v,0}$. We divide the proofs into three steps.

(Step 1: Case $\alpha = 2$ and $\boldsymbol{p}_0 = \mathbf{1}$). For $\beta \in \big(0, (\alpha + \beta_0 - \Lambda) \wedge 1)\big)$, noting that

$$\|u^N_s\|_{\boldsymbol{p}_0} = \|u^N_s\|_{\mathbf{1}} = \|\mu^N_s * \Gamma_s \phi_N\|_{\mathbf{1}} = 1,$$

we have, by the embeddings (2.10) and the convolution inequality (2.11)

$$\begin{aligned}\int_0^t G_\beta(t,s)\Big(\|b_s*\mathcal{U}_s^N\|_{\boldsymbol{\infty}}\|u_s^N\|_{\boldsymbol{p}_0}+s^{-\frac{\Lambda}{\alpha}}\|\mathcal{U}_s^N\|_{\boldsymbol{p}_0}\Big)\mathrm{d}s\\ \lesssim \int_0^t G_\beta(t,s)\Big(\|b_s*\mathcal{U}_s^N\|_{\boldsymbol{\infty}}+s^{-\frac{\Lambda}{\alpha}}\|\mathcal{U}_s^N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}}\Big)\mathrm{d}s\\ \lesssim \int_0^t G_\beta(t,s)\Big((1\wedge s)^{-\frac{\beta+\Lambda}{\alpha}}+(1\wedge s)^{\frac{\beta_0-\beta-\Lambda}{\alpha}}\Big)\|\mathcal{U}_s^N\|_{\mathbb{S}_s^\beta(b)}\mathrm{d}s,\end{aligned}\tag{2.34}$$

recalling (2.3) for the last inequality.

Hence, for any $t\in[0,T_0]$ and $m\in\mathbb{N}$, by (2.32) and $\beta_0<0$,

$$\|\mathcal{U}_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))}\lesssim\|I_0^N\|_{L^m(\Omega)}+\|\mathcal{M}_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))}+\int_0^t G_\beta(t,s)s^{\frac{\beta_0-\beta-\Lambda}{\alpha}}\|\mathcal{U}_s^N\|_{L^m(\Omega;\mathbb{S}_s^\beta(b))}\mathrm{d}s.$$

Since $\beta<\alpha-\Lambda+\beta_0$ (recalling (1.20)), by Gronwall's inequality of Volterra type stated in Lemma E.3, we get, for any $T\in(0,T_0]$,

$$\sup_{t\in(0,T]}\|\mathcal{U}_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))}\lesssim\|I_0^N\|_{L^m(\Omega)}+\sup_{t\in(0,T]}\|\mathcal{M}_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))}.\tag{2.35}$$

(Step 2: Case $\alpha\in(1,2]$ and $\boldsymbol{p}_0\neq\mathbf{1}$). In this situation, we cannot simply rely on the bound $\|\mu_s^N\|_{\mathbf{B}^0_{\mathbf{1};\boldsymbol{a}}}<\infty$ in Equation (2.34). Indeed, owing to (2.11) and the estimate (B.8), this would yield

$$\|u_s^N\|_{\boldsymbol{p}_0}\lesssim\|\Gamma_s\phi_N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}}\|\mu_s^N\|_{\mathbf{B}^0_{\mathbf{1};\boldsymbol{a}}}\lesssim N^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}_0}}.$$

To avoid this cost, we again make the difference $\mathcal{U}_s^N$ appear, for $s\in(0,t]$. Using the control $\|u_s\|_{\boldsymbol{p}_0}\lesssim\|u_s\|_{B^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}}\lesssim s^{\frac{\beta_0}{\alpha}}$ deduced from (2.10) and (2.1) and recalling the definition (2.3), we derive

$$\begin{aligned}\|b_s*\mathcal{U}_s^N\|_{\boldsymbol{\infty}}\|u_s^N\|_{\boldsymbol{p}_0}&\leqslant\|b_s*\mathcal{U}_s^N\|_{\boldsymbol{\infty}}\|\mathcal{U}_s^N\|_{\boldsymbol{p}_0}+\|b_s*\mathcal{U}_s^N\|_{\boldsymbol{\infty}}\|u_s\|_{\boldsymbol{p}_0}\\ &\leqslant s^{\frac{\beta_0-2\beta-\Lambda}{\alpha}}\|\mathcal{U}_s^N\|^2_{\mathbb{S}_s^\beta(b)}+s^{\frac{\beta_0-\beta-\Lambda}{\alpha}}\|\mathcal{U}_s^N\|_{\mathbb{S}_s^\beta(b)}.\end{aligned}$$

Substituting this estimate into (2.32), we get, for any $0<t\leqslant T_0$,

$$\|\mathcal{U}_t^N\|_{\mathbb{S}_t^\beta(b)}\lesssim I_0^N+\|\mathcal{M}_t^N\|_{\mathbb{S}_t^\beta(b)}+\sum_{i=1}^2\int_0^t G_\beta(t,s)s^{\frac{\beta_0-i\beta-\Lambda}{\alpha}}\|\mathcal{U}_s^N\|^i_{\mathbb{S}_s^\beta(b)}\mathrm{d}s.$$

Note that by (2.5),

$$\beta_0-2\beta-\Lambda>-\alpha,\quad \beta_0-2\beta-\Lambda>-(\alpha-1-\beta_0+\beta).$$

Note also that, as $\frac{\alpha}{2}>\alpha-1$, $\beta<\frac{\alpha}{2}-\Lambda$ and, as $\beta_0<0$, according to **(H)**, $\Lambda<\alpha-1-\beta_0$. These bounds ensure:

$$\sup_{t\in(0,T_0]}\sum_{i=1}\int_0^t G_\beta(t,s)s^{\frac{\beta_0-i\beta-\Lambda}{\alpha}}\mathrm{d}s\asymp\sup_{t\in(0,T_0]}\sum_{i=1}\int_0^t\Big((t-s)^{-\frac{1}{\alpha}}+(t-s)^{-\frac{\Lambda+\beta_0+1}{\alpha}}\Big)s^{\frac{\beta_0-i\beta-\Lambda}{\alpha}}\mathrm{d}s<\infty.$$

Thus one can use the Gronwall inequality (E.9) stated in Lemma E.3-(iii) to derive, for all $s\in(0,T_0]$,

$$\begin{aligned}\|\mathcal{U}_s^N\|_{\mathbb{S}_s^\beta(b)}\lesssim I_0^N+\|\mathcal{M}_s^N\|_{\mathbb{S}_s^\beta(b)}+\int_0^s G_\beta(s,r)r^{\frac{\beta_0-\beta-\Lambda}{\alpha}}\|\mathcal{M}_r^N\|_{\mathbb{S}_r^\beta(b)}\mathrm{d}r\\ +\int_0^s G_\beta(s,r)r^{\frac{\beta_0-2\beta-\Lambda}{\alpha}}\|\mathcal{U}_r^N\|^2_{\mathbb{S}_r^\beta(b)}\mathrm{d}r.\end{aligned}\tag{2.36}$$

Define

$$F_t^N:=\sup_{t'\in(0,t]}\int_0^{t'}G_\beta(t',s)s^{\frac{\beta_0-2\beta-\Lambda}{\alpha}}\|\mathcal{U}_s^N\|^2_{\mathbb{S}_s^\beta(b)}\mathrm{d}s$$

and

$$\mathscr{T}_N := I_0^N + \left(\sup_{t\in(0,T_0]} \int_0^t G_\beta(t,s) s^{\frac{\beta_0-\beta-\Lambda}{\alpha}} \|\mathcal{M}_s^N\|^2_{\mathbb{S}_s^\beta(b)} \mathrm{d}s \right)^{\frac12}. \tag{2.37}$$

Taking the square and multiplying $G_\beta(t',s)s^{\frac{\beta_0-2\beta-\Lambda}{\alpha}}$ for both sides of (2.36), integrating the resulting expression for $s\in(0,t')$, by Fubini's theorem and (E.8) (in Lemma E.3-(ii)), we obtain that for some $c_0,c_1>0$,

$$F_t^N \leqslant (c_0\mathscr{T}_N)^2 + (c_1 F_t^N)^2, \quad \forall 0\leqslant t\leqslant T_0.$$

In particular,

$$\big(c_1 F_t^N - (2c_1)^{-1}\big)^2 \geqslant (2c_1)^{-2} - (c_0\mathscr{T}_N)^2.$$

Thus, on the event $\Omega_0 := \big\{\omega\in\Omega : c_0\mathscr{T}_N < (2c_1)^{-1}\big\}$,

$$|c_1F_t^N - (2c_1)^{-1}| \geqslant \sqrt{(2c_1)^{-2} - c_0(\mathscr{T}_N)^2}. \tag{2.38}$$

Observe now that, using successively the control (2.1) on u, the control $\sup_N \|\mu_t^N\|_{\mathbf{B}^0_{\mathbf{1};\boldsymbol{a}}} < \infty$, (2.10) and finally applying the estimate (B.8) on $\|\Gamma_t\phi_N\|_{\mathbf{B}^\beta_{\boldsymbol{p};\boldsymbol{a}}}$ (taking $\ell=0$, $\lambda=N^\zeta$, $\ell=0$ and successively $\boldsymbol{p}=\boldsymbol{p}_0$ and $\boldsymbol{p}=\boldsymbol{p}_b'$), we have

$$\begin{aligned}
\|\mathcal{U}_t^N\|_{\mathbb{S}_t^\beta(b)} &= t^{\frac{\beta-\beta_0}{\alpha}} \|u_t - \Gamma_t\phi_N * \mu_t^N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}} + t^{\frac{\Lambda+\beta}{\alpha}} \|b_t * \big(u_t - \Gamma_t\phi_N*\mu_t^N\big)\|_\infty \\
&\overset{(2.11)}{\lesssim} t^{\frac{\beta-\beta_0}{\alpha}} \|u_t - \Gamma_t\phi_N * \mu_t^N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}} + t^{\frac{\Lambda+\beta}{\alpha}} \|u_t - \Gamma_t\phi_N*\mu_t^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}} \\
&\lesssim 1 + \|\Gamma_t\phi_N * \mu_t^N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}} + \|\Gamma_t\phi_N * \mu_t^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}} \\
&\leqslant 1 + \|\Gamma_t\phi_N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}} + \|\Gamma_t\phi_N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}} \\
&\overset{(2.10)}{\lesssim} 1 + \|\Gamma_t\phi_N\|_{\mathbf{B}^{\varepsilon}_{\boldsymbol{p}_0;\boldsymbol{a}}} + \|\Gamma_t\phi_N\|_{\mathbf{B}^{\varepsilon-\beta_b}_{\boldsymbol{p}_b';\boldsymbol{a}}} \overset{(B.8)}{\lesssim} N^{\theta_0\zeta},
\end{aligned} \tag{2.39}$$

where $\varepsilon>0$ and

$$\theta_0 = [\mathscr{A}_{\mathbf{1},\boldsymbol{p}_0} + (1+\alpha)\varepsilon] \vee [\mathscr{A}_{\boldsymbol{p}_b,\infty} + (1+\alpha)(\varepsilon-\beta_b)].$$

Hence, it is easy to see that $t\mapsto F_t^N$ is continuous and $\lim_{t\downarrow 0} F_t^N = 0$. Thus, by (2.38),

$$F_t^N \leqslant \Big((2c_1)^{-1} - \sqrt{(2c_1)^{-2} - (c_0\mathscr{T}_N)^2}\Big)/c_1 \leqslant c_2\mathscr{T}_N, \quad \forall t\in[0,T_0], \text{ on } \Omega_0,$$

for $c_2 = c_0/c_1$. Plugging this into (2.36) and recalling (2.37), the above yields

$$\|\mathcal{U}_t^N\|_{\mathbb{S}_t^\beta(b)} \mathbb{1}_{\Omega_0} \lesssim \Big(\mathscr{T}_N + \|\mathcal{M}_t^N\|_{\mathbb{S}_t^\beta(b)}\Big) \mathbb{1}_{\Omega_0}.$$

Therefore, for all $m,n\geqslant 1$, by (2.39) and Chebyshev's inequality, we get

$$\begin{aligned}
\sup_{t\in(0,T_0]} \mathbb{E}[\|\mathcal{U}_t^N\|^m_{\mathbb{S}_t^\beta(b)}] &\leqslant \sup_{t\in(0,T_0]} \mathbb{E}[\|\mathcal{U}_t^N\|^m_{\mathbb{S}_t^\beta(b)} \mathbb{1}_{\Omega_0^c}] + \sup_{t\in(0,T_0]} \mathbb{E}[\|\mathcal{U}_t^N\|^m_{\mathbb{S}_t^\beta(b)} \mathbb{1}_{\Omega_0}] \\
&\lesssim N^{m\theta_0\zeta}\mathbb{P}(\Omega_0^c) + \mathbb{E}[(\mathscr{T}_N)^m] + \sup_{t\in(0,T_0]} \mathbb{E}[\|\mathcal{M}_t^N\|^m_{\mathbb{S}_t^\beta(b)}] \\
&\leqslant N^{m\theta_0\zeta}(2c_1c_0)^n \mathbb{E}[(\mathscr{T}_N)^n] + \mathbb{E}[(\mathscr{T}_N)^m] + \sup_{t\in(0,T_0]} \mathbb{E}[\|\mathcal{M}_t^N\|^m_{\mathbb{S}_t^\beta(b)}].
\end{aligned} \tag{2.40}$$

We now turn to the estimation of $\mathbb{E}[(\mathscr{T}_N)^n]$. By Hölder's inequality, there is a p large enough so that

$$\int_0^t G_\beta(t,r) r^{\frac{\beta_0-\beta-\Lambda}{\alpha}} \|\mathcal{M}_r^N\|^2_{\mathbb{S}_r^\beta(b)} \mathrm{d}r \lesssim \left(\int_0^t \|\mathcal{M}_r^N\|^{2p}_{\mathbb{S}_r^\beta(b)} \mathrm{d}r\right)^{1/p}.$$

Thus, if we take $n \geqslant 2p$, then

$$\mathbb{E}[(\mathcal{J}_N)^n] \lesssim \mathbb{E}\big[(I_0^N)^n\big] + \mathbb{E}\left[\sup_{t\in(0,T_0]}\left(\int_0^t G_\beta(t,r) r^{\frac{\beta_0-\beta-\Lambda}{\alpha}} \|\mathcal{M}_r^N\|^2_{\mathbb{S}_r^\beta(b)} \mathrm{d}r\right)^{\frac{n}{2}}\right]$$

$$\lesssim \mathbb{E}\big[(I_0^N)^n\big] + \mathbb{E}\left(\int_0^{T_0} \|\mathcal{M}_r^N\|^n_{\mathbb{S}_r^\beta(b)} \mathrm{d}r\right) \lesssim \mathbb{E}\big[(I_0^N)^n\big] + \sup_{t\in(0,T_0]} \mathbb{E}\left[\|\mathcal{M}_t^N\|^n_{\mathbb{S}_t^\beta(b)}\right].$$

Substituting this into (2.40), we obtain that for any $n, m \geqslant 2p$,

$$\sup_{t\in(0,T_0]} \|\mathcal{U}_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))} \lesssim N^{\theta_0\zeta}\left(\|I_0^N\|_{L^n(\Omega)} + \sup_{t\in(0,T_0]} \|\mathcal{M}_t^N\|_{L^n(\Omega;\mathbb{S}_t^\beta(b))}\right)^{n/m} + \|I_0^N\|_{L^m(\Omega)} + \sup_{t\in(0,T_0]} \|\mathcal{M}_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))}. \tag{2.41}$$

(Step 3). In view of (2.41) and, as the control of $\|\mathcal{M}_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))}$ is already set by Lemma 2.9, it remains to estimate $\|I_0^N\|_{L^n(\Omega)}$. Recalling the definition (2.33) of I_0^N, we have, for any $n \in \mathbb{N}$,

$$\|I_0^N\|_{L^n(\Omega)} \lesssim N^{-\beta\zeta} + \|\phi_N * (\mu_0^N - u_0)\|_{L^n(\Omega;\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}})}. \tag{2.42}$$

Now let us again consider separately the cases $\boldsymbol{p}_0 > \boldsymbol{\alpha}$ and $\boldsymbol{p}_0 = \mathbf{1}$, $\alpha = 2$. For the former, let $q = p_{x,0} \wedge p_{v,0} \wedge 2$. By (C.2) in Lemma C.3 , with $\boldsymbol{p} = \boldsymbol{p}_0$ and $\beta_0 - \beta < 0$, we have

$$\|\phi_N * (\mu_0^N - u_0)\|_{L^n(\Omega;\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}})} \lesssim N^{\frac{1}{q}-1+\zeta\mathcal{A}_{\mathbf{1},\boldsymbol{p}_0}} \leqslant N^{\mathfrak{m}_\alpha-1+\zeta\theta_\alpha}, \tag{2.43}$$

where $\mathfrak{m}_\alpha$ and θ_α are defined in (1.22). For the case $(\alpha, \boldsymbol{p}_0) = (2, \mathbf{1})$, by (C.2) with $\boldsymbol{p} = \mathbf{2}$ and $\beta_0 - \beta < 0$ and by (2.4), we have

$$\|\phi_N * (\mu_0^N - u_0)\|_{L^n(\Omega;\mathbf{B}^{\beta_0-\beta}_{\boldsymbol{p}_0;\boldsymbol{a}})} \lesssim N^{-\frac{1}{2}+\zeta\mathcal{A}_{\mathbf{1},\mathbf{2}}} \leqslant N^{\mathfrak{m}_2-1+\zeta\theta_2}. \tag{2.44}$$

Therefore, combining (2.35), (2.41)-(2.44) and applying Lemma 2.9, we obtain that for any $n, m \geqslant 2p$,

$$\sup_{t\in(0,T_0]} \|\mathcal{U}_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))} \lesssim N^{\theta_0\zeta}(N^{-\beta\zeta} + N^{\mathfrak{m}_\alpha-1+\zeta\theta_\alpha})^{n/m} + N^{-\beta\zeta} + N^{\mathfrak{m}_\alpha-1+\zeta\theta_\alpha}.$$

As n can be chosen arbitrarily, given N fixed, we can choose n large enough, so that the $(N^{-\beta\zeta} + N^{\mathfrak{m}_\alpha-1+\zeta\theta_\alpha})^{n/m-1}$ dominates $N^{\theta_0\zeta}$. This enables to derive the desired estimate. □

2.3. **Proof of Theorem 1.2.** By (2.11) and the essential control (2.1) on u_s, for any $\beta \geqslant 0$, we have

$$\|b_s * \mu_s\|_{\mathbf{B}^\beta_{\boldsymbol{\infty};\boldsymbol{a}}} = \|b_s * u_s\|_{\mathbf{B}^\beta_{\boldsymbol{\infty};\boldsymbol{a}}} \lesssim \|b_s\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}} \|u_s\|_{\mathbf{B}^{\beta-\beta_b}_{\boldsymbol{p}_b';\boldsymbol{a}}} \lesssim s^{-\frac{\Lambda+\beta}{\alpha}} \|b\|_{L^\infty_{T_0}\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}}, \tag{2.45}$$

and, according to (2.10),

$$\|b_s * \mu_s\|_{\mathbb{L}^\infty} \lesssim \|b_s * u_s\|_{\mathbf{B}^{0,1}_{\boldsymbol{\infty};\boldsymbol{a}}} \lesssim s^{-\frac{\Lambda}{\alpha}} \|b\|_{L^\infty_{T_0}\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}}. \tag{2.46}$$

In particular, (2.45) yields: for any $\beta \in [0, \alpha - 1 - \Lambda)$, there exists $q > \frac{\alpha}{\alpha-1}$ such that the drift of (1.3)

$$B_s(x,v) := b_s * \mu_s(x,v) = b_s * u_s(x,v)$$

is in $L^q_{T_0}\mathbf{B}^\beta_{\boldsymbol{\infty};\boldsymbol{a}}$. We split the proof of Theorem 1.2 by establishing successively the estimates (1.23) and (1.24).

Proof of (1.23). Fix $T \in (0, T_0]$ and $\varphi \in C_b^\infty(\mathbb{R}^{2d})$, and set $B_t^T := B_{T-t}(= b_{T-t} * \mu_{T-t})$. By [17, Theorem 4.2-(i)], there is a unique smooth solution to the following PDE:

$$\partial_t u = (\Delta_v^{\frac{\alpha}{2}} - v \cdot \nabla_x)u + B^T \cdot \nabla_v u \text{ on } [0,T] \times \mathbb{R}^{2d}, \qquad u(0) = \varphi.$$

Its Duhamel formulation is further given by

$$u(t)=P_t\varphi+\int_0^t P_{t-s}(B^T_s\cdot\nabla_v u(s))\mathrm{d}s,\qquad t\in[0,T].$$

Noting that by (2.16),

$$\|\nabla_v P_t f\|_{\mathbb{L}^\infty}\lesssim\|\nabla_v P_t f\|_{\mathbf{B}^{0,1}_{\infty;\boldsymbol{a}}}\lesssim t^{-\frac{1}{\alpha}}\|f\|_{\mathbf{B}^{0}_{\infty;\boldsymbol{a}}}\lesssim t^{-\frac{1}{\alpha}}\|f\|_{\mathbb{L}^\infty},$$

by (2.46), we have

$$\begin{aligned}\|\nabla_v u(t)\|_{\mathbb{L}^\infty}&\lesssim t^{-\frac{1}{\alpha}}\|\varphi\|_{\mathbb{L}^\infty}+\int_0^t(t-s)^{-\frac{1}{\alpha}}\|B^T_s\|_{\mathbb{L}^\infty}\|\nabla_v u(s)\|_{\mathbb{L}^\infty}\mathrm{d}s\\&\lesssim t^{-\frac{1}{\alpha}}\|\varphi\|_{\mathbb{L}^\infty}+\int_0^t(t-s)^{-\frac{1}{\alpha}}s^{-\frac{\Lambda}{\alpha}}\|\nabla_v u(s)\|_{\mathbb{L}^\infty}\mathrm{d}s.\end{aligned}$$

Since $\Lambda<1$, by the classical Gronwall inequality of Volterra type (or applying (E.5) in Lemma E.2), we have

$$\|\nabla_v u(t)\|_{\mathbb{L}^\infty}\lesssim t^{-\frac{1}{\alpha}}\|\varphi\|_{\mathbb{L}^\infty}.\tag{2.47}$$

By applying the generalized version of Itô's formula to $(t,z)\mapsto u(T-t,z)$ stated in [17, Lemma 4.3], we have

$$\mathbb{E}\big[\varphi(Z_T)\big]=\mathbb{E}\big[u(0,Z_T)\big]=\mathbb{E}\big[u(T,Z_0)\big],$$

and

$$\mathbb{E}\big[\varphi(Z^{N,1}_T)\big]=\mathbb{E}\big[u(0,Z^{N,1}_T)\big]=\mathbb{E}\big[u(T,\xi_1)\big]+\mathbb{E}\Big[\int_0^T\big(b_s*u^N_s-B_s\big)\cdot\nabla_v u(T-s,Z^{N,1}_s)\mathrm{d}s\Big].$$

Thus, by (2.11) and (2.47),

$$\begin{aligned}|\mathbb{E}\big[\varphi(Z_T)\big]-\mathbb{E}\big[\varphi(Z^{N,1}_T)\big]|&\leqslant\mathbb{E}\Big[\int_0^T\|b_s*u^N_s-b_s*u_s\|_{\mathbb{L}^\infty}\|\nabla_v u(T-s)\|_{\mathbb{L}^\infty}\mathrm{d}s\Big]\\&\lesssim\|\varphi\|_{\mathbb{L}^\infty}\int_0^T(T-s)^{-\frac{1}{\alpha}}\|b_s*u^N_s-b_s*u_s\|_{\mathbb{L}^\infty}\mathrm{d}s.\end{aligned}$$

Estimate (1.23) now follows from (2.6) which gives

$$\begin{aligned}&\mathbb{E}\Big[\int_0^T(T-s)^{-\frac{1}{\alpha}}\|b_s*u^N_s-b_s*u_s\|_{\mathbb{L}^\infty}\mathrm{d}s\Big]\\&\lesssim\int_0^T(T-s)^{-\frac{1}{\alpha}}(1\wedge s)^{-\frac{\beta+\Lambda}{\alpha}}\mathbb{E}\big[\|u^N_s-u_s\|_{\mathbb{S}^\beta_s(b)}\big]\mathrm{d}s\lesssim N^{-\beta\zeta}+N^{\mathtt{m}_\alpha-1+\zeta\theta_\alpha+\varepsilon},\end{aligned}\tag{2.48}$$

the limit $\beta<\alpha-1-\Lambda\Leftrightarrow\frac{\beta+\Lambda}{\alpha}<1-\frac{1}{\alpha}$ ensuring the finiteness of the integral $\int_0^T(T-s)^{-\frac{1}{\alpha}}(1\wedge s)^{-\frac{\beta+\Lambda}{\alpha}}\,\mathrm{d}s$. □

Proof of (1.24). Since $b\in L^\infty_{T_0}\mathbf{B}^{1+\frac{\alpha}{2},\frac{\alpha}{2}-1+\beta_b}_{\boldsymbol{p}_b;x,\boldsymbol{a}}$, according to Lemma 4.7-(*ii*) in [17], for any $\beta\in\big(1-\frac{\alpha}{2},\bar\beta_\alpha\big)$, B belongs to the space $L^q_{T_0}(\mathbf{C}^\beta_{\boldsymbol{a}}\cap\mathbf{C}^{\frac{\alpha+\beta}{\alpha+1}}_x)$, for $C^{\frac{\alpha+\beta}{\alpha+1}}_x$ denoting the (isotropic) Hölder-Zygmund space on the x-variable.

Then, for any fixed $\lambda>0$, by [17, Theorem 4.2-(i)], there is a unique solution u to the following (Zvonkin type) backward PDE for $T\in(0,T_0)$:

$$\partial_t u+(\Delta_v^{\frac{\alpha}{2}}-v\cdot\nabla_x-\lambda)u+B\cdot\nabla_v u=B,\quad u(T)=0,\tag{2.49}$$

such that, setting $\nabla u=(\nabla_x u,\nabla_v u)$, by [17, Theorem 4.2-(iii)], for λ large enough :

$$\|\nabla u\|_{\mathbb{L}^\infty_T}:=\|\nabla u\|_{L^\infty((0,T);\mathbb{L}^\infty)}\leqslant\tfrac{1}{2},\qquad g(s):=\|\nabla u(s)\|_{C^\delta_v}\in L^q([0,T_0]),\tag{2.50}$$

where, if $\alpha = 2$, $\delta = 1$ and $\|\nabla u(s)\|_{C_v^1} := \|\nabla_x \nabla_v u(s)\|_{\mathbb{L}^\infty} + \|\nabla_v^2 u(s)\|_{\mathbb{L}^\infty}$, and, if $\alpha \in (1,2)$, $\delta \geqslant \frac{\alpha}{2} + \varepsilon_0$ for some $\varepsilon_0 > 0$, and

$$\|\nabla u(s)\|_{C_v^\delta} := \|\nabla u(s)\|_{\mathbb{L}^\infty} + \sup_{v \in \mathbb{R}^d \setminus \{0\}} \frac{\|\delta_{(0,v)}^{(1)} \nabla u(s)\|_{\mathbb{L}^\infty}}{|v|^\delta},$$

$\delta_{(0,v)}^{(1)}$ being the difference operator defined in (2.8). In particular, for each $t \in [0,T]$, $z \mapsto \Phi_t(z) := z + (0, u(t,z))$ forms a C^1-diffeomorphism on $\mathbb{R}^{2d}$. By Itô's formula, we have, for any $t \in [0,T]$,

$$\Phi_t(Z_t^{N,1}) = \Phi_0(\xi_1) + \int_0^t (V_s^{N,1}, \lambda u(s, Z_s^{N,1})) \mathrm{d}s + M_t^{N,1} + \int_0^t [b_s * (u_s^N - u_s) \cdot \nabla_v \Phi_s](Z_s^{N,1}) \mathrm{d}s,$$

where

$$M_t^{N,1} = \begin{cases} \sqrt{2} \int_0^t \nabla_v \Phi_s(Z_s^{N,1}) \mathrm{d}W_s^1, & \text{if } \alpha = 2, \\ \int_0^t \int_{\mathbb{R}^d} \delta_{(0,v)}^{(1)} \Phi_s(Z_s^{N,1}) \tilde{\mathcal{N}}^1(\mathrm{d}s, \mathrm{d}v), & \text{if } \alpha \in (1,2). \end{cases}$$

Similarly, for $Z_t^1 = (X_t^1, V_t^1)$, the solution to (1.3) driven by $L^{\alpha,1}$ and starting at the initial $Z_0^1 = \xi_1$, we have

$$\Phi_t(Z_t^1) = \Phi_0(\xi_1) + \int_0^t (V_s^1, \lambda u(s, Z_s^1)) \mathrm{d}s + M_t,$$

where

$$M_t = \begin{cases} \sqrt{2} \int_0^t \nabla_v \Phi_s(Z_s^1) \mathrm{d}W_s^1, & \text{if } \alpha = 2, \\ \int_0^t \int_{\mathbb{R}^d} \delta_{(0,v)}^{(1)} \Phi_s(Z_s^1) \tilde{\mathcal{N}}^1(\mathrm{d}s, \mathrm{d}v), & \text{if } \alpha \in (1,2). \end{cases}$$

Thus, according to (2.50),

$$\begin{aligned} |\Phi_t(Z_t^{N,1}) - \Phi_t(Z_t^1)| \lesssim & (1 + \lambda \|\nabla u\|_{\mathbb{L}_T^\infty}) \int_0^t |Z_s^{N,1} - Z_s^1| \mathrm{d}s + |M_t^{N,1} - M_t| \\ & + \|\nabla_v \Phi\|_{\mathbb{L}_T^\infty} \int_0^t |b_s * (u_s^N - u_s)(Z_s^{N,1})| \mathrm{d}s, \end{aligned}$$

where

$$M_t^{N,1} - M_t = \begin{cases} \sqrt{2} \int_0^t \left(\nabla_v u(s, Z_s^{N,1}) - \nabla_v u(s, Z_s^1) \right) \mathrm{d}W_s^1, & \text{if } \alpha = 2, \\ \int_0^t \int_{\mathbb{R}^d} \left\{ \delta_{(0,v)}^{(1)} \Phi_s(Z_s^{N,1}) - \delta_{(0,v)}^{(1)} \Phi_s(Z_s^1) \right\} \tilde{\mathcal{N}}^1(\mathrm{d}s, \mathrm{d}v), & \text{if } \alpha \in (1,2). \end{cases}$$

Observe in this latter case that:

$$|\delta_{(0,v)}^{(1)} \Phi_s(Z_s^{N,1}) - \delta_{(0,v)}^{(1)} \Phi_s(Z_s^1)| \lesssim |Z_s^{N,1} - Z_s^1| \Big(\|\nabla u(s,\cdot)\|_{\mathbb{L}^\infty} \mathbb{1}_{\{|v|>1\}} + \|\nabla u(s,\cdot)\|_{C_v^\delta} |v|^\delta \mathbb{1}_{\{|v| \leqslant 1\}} \Big).$$

Combining BDG's inequality and (2.50), we have

$$\begin{aligned} \mathbb{E}\big[|M_t^{N,1} - M_t|^2\big] & \lesssim \begin{cases} \sqrt{2} \int_0^t \mathbb{E}\big[|\nabla_v u(s, Z_s^{N,1}) - \nabla_v u(s, Z_s^1)|^2\big] \mathrm{d}s, & \text{if } \alpha = 2, \\ \int_0^t \int_{\mathbb{R}^d} \mathbb{E}\big[\big|\delta_{(0,v)}^{(1)} \Phi_s(Z_s^{N,1}) - \delta_{(0,v)}^{(1)} \Phi_s(Z_s^1)\big|^2\big] \nu(\mathrm{d}v) \mathrm{d}s, & \text{if } \alpha \in (1,2). \end{cases} \\ & \lesssim \int_0^t |g(s)|^2 \mathbb{E}|Z_s^{N,1} - Z_s^1|^2 \mathrm{d}s, \end{aligned}$$

observing in particular that, in the case $\alpha \in (1,2)$, since $2\delta > \alpha$,

$$\int |v|^{2\delta} \mathbb{1}_{\{|v|\leqslant 1\}}\nu(\mathrm{d}v) < \infty.$$

We note that $|g|^2 \in L^{q/2}([0,T_0])$ with $q > \alpha/(\alpha-1) \geqslant 2$. Therefore

$$\mathbb{E}\left[\sup_{s\in[0,t]} |Z_s^{N,1} - Z_s^1|^2\right] \lesssim \mathbb{E}\left[\sup_{s\in[0,t]} |\Phi_s(Z_s^{N,1}) - \Phi_s(Z_s^1)|^2\right]$$
$$\lesssim \int_0^t |g(s)|^2 \mathbb{E}|Z_s^{N,1} - Z_s^1|^2 \mathrm{d}s + \mathbb{E}\left[\left(\int_0^T \|b_s * (u_s^N - u_s)\|_{\mathbb{L}^\infty} \mathrm{d}s\right)^2\right],$$

which implies, by Gronwall's inequality, that

$$\mathbb{E}\left[\sup_{s\in[0,T]} |Z_s^{N,1} - Z_s^1|^2\right] \lesssim \mathbb{E}\left[\int_0^T \|b_s * (u_s^N - u_s)\|_{\mathbb{L}^\infty}^2 \mathrm{d}s\right]$$
$$\lesssim \left(\int_0^T (1\wedge s)^{-2(\frac{\beta+\Lambda}{\alpha})} \mathbb{E}\Big[\|u_s^N - u_s\|^2_{\mathbb{S}_s^\beta(b)}\Big] \mathrm{d}s\right).$$

Observe that $\beta < \alpha - 1 - \Lambda$ implies $\beta < \frac{\alpha}{2} - \Lambda$. Recall indeed that since $\alpha \in (1,2]$ then $\alpha - 1 \leqslant \frac{\alpha}{2}$. Hence, $\int_0^T (1\wedge s)^{-2(\frac{\beta+\Lambda}{\alpha})}\mathrm{d}s$ is finite. Estimate (1.24) again follows by (2.6). □

3. Proof of the main technical Lemmas

3.1. **Proof of Lemma 2.8.** First of all, applying (2.16) (for $\beta = 0$, $\beta_1 = -1$, $\boldsymbol{p}_1 = \boldsymbol{p} = \boldsymbol{p}_0$), we have

$$\|\mathscr{H}_t^N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}} \lesssim \int_0^t \|P_{t-s}\mathrm{div}_v H_s^N\|_{\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}}}\mathrm{d}s \lesssim \int_0^t (t-s)^{-\frac{1}{\alpha}} \|\mathrm{div}_v H_s^N\|_{\mathbf{B}^{-1}_{\boldsymbol{p}_0;\boldsymbol{a}}}\mathrm{d}s$$
$$\lesssim \int_0^t (t-s)^{-\frac{1}{\alpha}} \|H_s^N\|_{\mathbf{B}^{0}_{\boldsymbol{p}_0;\boldsymbol{a}}}\mathrm{d}s \lesssim \int_0^t (t-s)^{-\frac{1}{\alpha}} \|H_s^N\|_{\boldsymbol{p}_0}\mathrm{d}s, \tag{3.1}$$

using as well (2.10) for the two last inequalities. In the same way, using again (2.16) (for $\beta = -\beta_b$, $\beta_1 = -1$, $\boldsymbol{p} = \boldsymbol{p}_b'$, $\boldsymbol{p}_1 = \boldsymbol{p}_0$), recalling, from **(H)** that, since $\frac{1}{\boldsymbol{p}_0} + \frac{1}{\boldsymbol{p}_b} \geqslant 1$, $\boldsymbol{p}_0 \leqslant \boldsymbol{p}_b'$ and, from (1.20), that $\Lambda + \beta_0 + 1 = \mathscr{A}_{\boldsymbol{p}_0,\boldsymbol{p}_b'} - \beta_b + 1 > 0$, we get

$$\|\mathscr{H}_t^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}} \lesssim \int_0^t \|P_{t-s}\mathrm{div}_v H_s^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}}\mathrm{d}s \lesssim \int_0^t (t-s)^{-\frac{\Lambda+\beta_0+1}{\alpha}} \|H_s^N\|_{\boldsymbol{p}_0}\mathrm{d}s. \tag{3.2}$$

In order to estimate $\|H_s^N\|_{\boldsymbol{p}_0}$, by adding and subtracting the elements $(b_s * u_s)u_s^N$ and $\langle (b_s * u_s)(\Gamma_s\phi_N)(\cdot - z), \mu_s^N\rangle$, we obtain from (2.22) and (2.20) the following decomposition for $H_s^N(z)$:

$$\begin{aligned} H_s^N(z) &= \langle (b_s * \mathcal{U}_s^N)(\Gamma_s\phi_N)(\cdot - z), \mu_s^N\rangle + ((b_s * u_s)\mathcal{U}_s^N)(z) \\ &\quad - \Big(\langle (b_s * u_s)(\Gamma_s\phi_N)(\cdot - z), \mu_s^N\rangle - ((b_s * u_s)u_s^N)(z)\Big) \\ &=: H_s^{1,N}(z) + H_s^{2,N}(z) - H_s^{3,N}(z). \end{aligned} \tag{3.3}$$

For $H_s^{1,N}(z)$, noting that for any $z \in \mathbb{R}^{2d}$,

$$|H_s^{1,N}(z)| = \Big|\langle (b_s * \mathcal{U}_s^N)(\Gamma_s\phi_N)(\cdot - z), \mu_s^N\rangle\Big| \leqslant \|b_s * \mathcal{U}_s^N\|_{\boldsymbol{\infty}} \langle (\Gamma_s\phi_N)(\cdot - z), \mu_s^N\rangle = \|b_s * \mathcal{U}_s^N\|_{\boldsymbol{\infty}} u_s^N(z),$$

we have

$$\|H_s^{1,N}\|_{\boldsymbol{p}_0} \leqslant \|b_s * \mathcal{U}_s^N\|_{\boldsymbol{\infty}} \|u_s^N\|_{\boldsymbol{p}_0}.$$

For $H_s^{2,N}(z)$, we directly have

$$\|H_s^{2,N}\|_{\boldsymbol{p}_0} \leqslant \|b_s * u_s\|_{\boldsymbol{\infty}} \|\mathcal{U}_s^N\|_{\boldsymbol{p}_0} \overset{(2.11)}{\lesssim} \|b_s\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}} \|u_s\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}} \|\mathcal{U}_s^N\|_{\boldsymbol{p}_0} \overset{(2.1)}{\lesssim} s^{-\frac{\Lambda}{\alpha}} \|\mathcal{U}_s^N\|_{\boldsymbol{p}_0}.$$

For $H^{3,N}_s(z)$, by definition, we can write

$$H^{3,N}_s(z) = -\int_{\mathbb{R}^{2d}} (b_s * u_s(z) - b_s * u_s(\bar z))(\Gamma_s\phi_N)(z-\bar z)\mu^N_s(\mathrm{d}\bar z). \tag{3.4}$$

Note that, for any $\kappa\in(0,1)$, and, for $\mathbf{C}^\kappa_{\boldsymbol a} = \mathbf{B}^{\kappa,\infty}_{\boldsymbol\infty;\boldsymbol a}$ as in Proposition 2.3,

$$\begin{aligned}|b_s * u_s(z) - b_s * u_s(\bar z)| &\leqslant |(b_s*u_s)(x,v) - (b_s*u_s)(x-s(v-\bar v),v)| \\ &\quad + |(b_s*u_s)(x-s(v-\bar v),v) - (b_s*u_s)(\bar x,\bar v)| \\ &\leqslant s^\kappa|v-\bar v|^\kappa \|b_s*u_s\|_{\mathbf{C}^{(1+\alpha)\kappa}_{\boldsymbol a}} + |\Gamma_s(z-\bar z)|^\kappa_{\boldsymbol a}\|b_s*u_s\|_{\mathbf{C}^\kappa_{\boldsymbol a}}.\end{aligned}$$

By (2.11) and (2.1), we have, for any $\beta\geqslant 0$,

$$\|b_s*u_s\|_{\mathbf{C}^\beta_{\boldsymbol a}} \lesssim \|b_s\|_{\mathbf{B}^{\beta_b}_{\boldsymbol p_b;\boldsymbol a}}\|u_s\|_{\mathbf{B}^{\beta-\beta_b,1}_{\boldsymbol p_b';\boldsymbol a}} \lesssim s^{-\frac{\Lambda+\beta}{\alpha}}.$$

Hence,

$$\begin{aligned}|b_s * u_s(z) - b_s * u_s(\bar z)| &\lesssim \left(s^{-\frac{\Lambda+(1+\alpha)\kappa}{\alpha}+\kappa}|v-\bar v|^\kappa + s^{-\frac{\Lambda+\kappa}{\alpha}}|\Gamma_s(z-\bar z)|^\kappa_{\boldsymbol a}\right) \\ &\lesssim s^{-\frac{\Lambda+\kappa}{\alpha}}\left(|v-\bar v|^\kappa + |\Gamma_s(z-\bar z)|^\kappa_{\boldsymbol a}\right).\end{aligned}$$

Note that for any $z=(x,v)\in\mathbb{R}^{2d}$, thanks to $\mathrm{supp}\phi_N\subset\{(x,v): |N^{(1+\alpha)\zeta}x|^{\frac{1}{1+\alpha}} + |N^\zeta v|\leqslant C\} = \{z: |z|_{\boldsymbol a}\leqslant CN^{-\zeta}\}$

$$(|v| + |\Gamma_s z|_{\boldsymbol a})^\kappa \Gamma_s\phi_N(z) \lesssim N^{-\zeta\kappa}\Gamma_s\phi_N(z).$$

Substituting these into (3.4), we get for any $\kappa\in[0,1)$,

$$\begin{aligned}\|H^{3,N}_s\|_{\boldsymbol p_0} &\lesssim s^{-\frac{\kappa+\Lambda}{\alpha}}\left\|\int_{\mathbb{R}^{2d}}(|\cdot-\bar v|^\kappa + |\Gamma_s(\cdot-\bar z)|^\kappa_{\boldsymbol a})(\Gamma_s\phi_N)(\cdot-\bar z)\mu^N_s(\mathrm{d}\bar z)\right\|_{\boldsymbol p_0} \\ &\lesssim s^{-\frac{\kappa+\Lambda}{\alpha}}N^{-\zeta\kappa}\left\|\int_{\mathbb{R}^{2d}}\Gamma_s\phi_N(\cdot-\bar z)\mu^N_s(\mathrm{d}\bar z)\right\|_{\boldsymbol p_0} = s^{-\frac{\kappa+\Lambda}{\alpha}}N^{-\kappa\zeta}\|u^N_s\|_{\boldsymbol p_0}.\end{aligned}$$

This implies that

$$\begin{aligned}\|H^{3,N}_s\|_{\boldsymbol p_0} &\lesssim \left[(s^{-\frac{\kappa+\Lambda}{\alpha}}N^{-\kappa\zeta})\wedge s^{-\frac{\Lambda}{\alpha}}\right]\|u^N_s\|_{\boldsymbol p_0} \leqslant s^{-\frac{\Lambda}{\alpha}}\|\mathcal{U}^N_s\|_{\boldsymbol p_0} + s^{-\frac{\kappa+\Lambda}{\alpha}}N^{-\kappa\zeta}\|u_s\|_{\boldsymbol p_0} \\ &\overset{(2.1)}{\lesssim} s^{-\frac{\Lambda}{\alpha}}\|\mathcal{U}^N_s\|_{\boldsymbol p_0} + s^{\frac{\beta_0-\kappa-\Lambda}{\alpha}}N^{-\kappa\zeta}.\end{aligned}$$

Combining the above calculations, we obtain that: for any $\kappa\in(0,1)$,

$$\|H^N_s\|_{\boldsymbol p_0} \lesssim \|b_s*\mathcal{U}^N_s\|_{\boldsymbol\infty}\|u^N_s\|_{\boldsymbol p_0} + s^{-\frac{\Lambda}{\alpha}}\|\mathcal{U}^N_s\|_{\boldsymbol p_0} + s^{\frac{\beta_0-\kappa-\Lambda}{\alpha}}N^{-\kappa\zeta}. \tag{3.5}$$

By (3.1), (3.2) and (3.5), we get, for $\kappa<(\alpha+\beta_0-\Lambda)\wedge 1$,

$$\|\mathcal{H}^N_t\|_{\mathbf{B}^{0,1}_{\boldsymbol p_0;\boldsymbol a}} \lesssim \int_0^t (t-s)^{-\frac{1}{\alpha}}\Big(\|b_s*\mathcal{U}^N_s\|_{\boldsymbol\infty}\|u^N_s\|_{p_0} + s^{-\frac{\Lambda}{\alpha}}\|\mathcal{U}^N_s\|_{p_0}\Big)\mathrm{d}s + t^{\frac{\alpha-1+\beta_0-\kappa-\Lambda}{\alpha}}N^{-\kappa\zeta},$$

and (remembering that, from (1.20) and $\beta_0\leqslant 0$, $\frac{\Lambda+\beta_0+1}{\alpha}<1$)

$$\|\mathcal{H}^N_t\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol p_b';\boldsymbol a}} \lesssim \int_0^t (t-s)^{-\frac{\Lambda+\beta_0+1}{\alpha}}\Big(\|b_s*\mathcal{U}^N_s\|_{\boldsymbol\infty}\|u^N_s\|_{\boldsymbol p_0} + s^{-\frac{\Lambda}{\alpha}}\|\mathcal{U}^N_s\|_{p_0}\Big)\mathrm{d}s + t^{\frac{\alpha-1-\kappa-2\Lambda}{\alpha}}N^{-\kappa\zeta}.$$

In particular, if we take $\kappa=\beta$ with $\beta\in(0,(\alpha+\beta_0-\Lambda)\wedge 1)$, then it is easy to see that

$$\begin{aligned}\|\mathcal{H}^N_t\|_{\mathbb{S}^\beta_t(b)} &\overset{(2.11)}{\lesssim} t^{\frac{\beta-\beta_0}{\alpha}}\|\mathcal{H}^N_t\|_{\mathbf{B}^{0,1}_{\boldsymbol p_0;\boldsymbol a}} + t^{\frac{\beta+\Lambda}{\alpha}}\|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol p_b;\boldsymbol a}}\|\mathcal{H}^N_t\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol p_b';\boldsymbol a}} \\ &\lesssim N^{-\beta\zeta} + \int_0^t G_\beta(t,s)\Big(\|b_s*\mathcal{U}^N_s\|_{\boldsymbol\infty}\|u^N_s\|_{\boldsymbol p_0} + s^{-\frac{\Lambda}{\alpha}}\|\mathcal{U}^N_s\|_{\boldsymbol p_0}\Big)\mathrm{d}s,\end{aligned}$$

where $G_\beta(t,s)$ is given as in (2.29).

3.2. **Proof of Lemma 2.9.** Our proof arguments are first focused on establishing general estimates on $\mathcal{M}_t^N$ successively in the Brownian case $\alpha = 2$ (Theorem 3.2 and Lemma 3.3 below) and next the pure-jump case $\alpha \in (1,2)$ (Theorem 3.6). Combining these results, the proof of Lemma 2.9 is achieved at the end of the section.

We first show the following estimate.

Lemma 3.1. *Let $\alpha = 2$, $\boldsymbol{p} \in [2,\infty]^2$ and $\boldsymbol{p}_1 \in [1,\infty)^2$ with $\boldsymbol{p}_1 \leqslant \boldsymbol{p}$ and $\beta, \beta_1 \in \mathbb{R}$. For any $m \geqslant 1$ and $\beta_2 > (\beta - \beta_1) \vee 0$, there is a constant $C = C(m, \boldsymbol{p}, \boldsymbol{p}_1, \beta, \beta_1, \beta_2, d) > 0$ such that, for any $N \geqslant 1$ and $f \in \mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}$,*

$$\sup_{t>0} \|f * \mathcal{M}_t^N\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}})} \lesssim_C N^{-\frac{1}{2}+\zeta(3\beta_2+\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}})} \|f\|_{\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}. \tag{3.6}$$

In particular, for any $\boldsymbol{p} \in [2,\infty]^2$, $\beta \in \mathbb{R}$, $\beta_2 > \beta \vee 0$ and $m \geqslant 1$,

$$\|\mathcal{M}_t^N\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}})} \lesssim_C N^{-\frac{1}{2}+\zeta(3\beta_2+\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}})}. \tag{3.7}$$

Proof. By the embedding (2.12), without loss of generality we may assume $\boldsymbol{p} \in [2,\infty)^2$. By the definition of the anisotropic Besov norm and Minkowski's inequality, we have

$$\|f * \mathcal{M}_t^N\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}})} \leqslant \sum_{j\geqslant -1} 2^{\beta j} \|f * \mathcal{R}_j^{\boldsymbol{a}} \mathcal{M}_t^N\|_{L^m(\Omega;\mathbb{L}^{\boldsymbol{p}})}.$$

Recalling the definition (2.24) of $\mathcal{M}_t^N$, we have

$$\begin{aligned} f * \mathcal{R}_j^{\boldsymbol{a}} \mathcal{M}_t^N(z) &= \frac{1}{N}\sum_{i=1}^N \int_0^t f * \mathcal{R}_j^{\boldsymbol{a}} P_{t-s}((\nabla_v \Gamma_s \phi_N)(Z_s^{N,i} - \cdot))(z) \mathrm{d}W_s^i \\ &= \frac{1}{N}\sum_{i=1}^N \int_0^t (f * \mathcal{R}_j^{\boldsymbol{a}} P_{t-s} \nabla_v \Gamma_s \phi_N)(\Gamma_{t-s} Z_s^{N,i} - z) \mathrm{d}W_s^i, \end{aligned}$$

where we have used that the symmetry property $\phi_N(z) = \phi_N(-z)$ and, for a function $g : \mathbb{R}^{2d} \to \mathbb{R}$ and $z' \in \mathbb{R}^{2d}$,

$$P_{t-s}(g(z' - \cdot))(z) = (P_{t-s} g)(\Gamma_{t-s} z' - z). \tag{3.8}$$

Next let us define, for t fixed, the stopped process

$$f * \mathcal{R}_j^{\boldsymbol{a}} \mathcal{M}_{u,t}^N(z) = \frac{1}{N}\sum_{i=1}^N \int_0^u (f * \mathcal{R}_j^{\boldsymbol{a}} P_{t-s} \nabla_v \Gamma_s \phi_N)(\Gamma_{t-s} Z_s^{N,i} - z) \mathrm{d}W_s^i, \quad u \in [0,t],$$

which defines an $\mathbb{L}^{\boldsymbol{p}}$-valued martingale. With this, we are in position (as in [31]) to apply a functional BDG (Burkhölder-Davis-Gundy) inequality to derive the control (3.6). Following e.g. [43] (see Theorem 16.1.1 therein), this BDG inequality states that, given $(E, |\cdot|_E)$ a UMD (unconditional martingale difference) Banach space, for any E-valued martingale $\{m_t\}_{t\geqslant 0}$ with quadratic variation $\{[m]_t\}_{t\geqslant 0}$ and any $p \in (1,\infty)$, we have, for all $t \geqslant 0$,

$$\|\sup_{0\leqslant u\leqslant t} |m_u|_E\|_{L^p(\Omega)} \lesssim_{c_{E,p}} \sup_{0\leqslant u\leqslant t} \||m_u|_E\|_{L^p(\Omega)} \lesssim_{C_{E,p}} \||[m]_t^{1/2}|_E\|_{L^p(\Omega)}, \tag{3.9}$$

for some constants $C_{E,p} = C(E,p)$ and $c_{E,p} = c(E,p)$. The property that the mixed space $\mathbb{L}^{\boldsymbol{p}}$ is indeed a UMD space for $\boldsymbol{p} \in (1,\infty)^2$ - along a recall of the notion of UMD spaces - is established in the appendix section (see Corollary D.2). Consequently, applying (3.9), we have

$$\begin{aligned}\mathbb{E}\big[\|f*\mathcal{R}^{\boldsymbol{a}}_j\mathcal{M}^N_t\|^m_{\boldsymbol{p}}\big]&\lesssim\sup_{0\leqslant u\leqslant t}\mathbb{E}\big[\|f*\mathcal{R}^{\boldsymbol{a}}_j\mathcal{M}^N_{u,t}\|^m_{\boldsymbol{p}}\big]\lesssim\mathbb{E}\big[\|[f*\mathcal{R}^{\boldsymbol{a}}_j\mathcal{M}^N_{u,t}]_{u=t}\|^m_{\boldsymbol{p}}\big]\\&\lesssim\mathbb{E}\left\|\left(\frac{1}{N^2}\sum_{i=1}^N\int_0^t|(f*\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N)(\Gamma_{t-s}Z^{N,i}_s-\cdot)|^2\mathrm{d}s\right)^{1/2}\right\|^m_{\boldsymbol{p}}\\&\lesssim N^{-m}\mathbb{E}\left\|\left(\sum_{i=1}^N\int_0^t|(f*\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N)(\Gamma_{t-s}Z^{N,i}_s-\cdot)|^2\mathrm{d}s\right)^{1/2}\right\|^m_{\boldsymbol{p}}\\&=N^{-m}\mathbb{E}\left\|\sum_{i=1}^N\int_0^t|(f*\mathcal{R}^{\boldsymbol{a}}_jP^\kappa_{t-s}\nabla_v\Gamma_s\phi_N)(\Gamma_{t-s}Z^{N,i}_s-\cdot)|^2\mathrm{d}s\right\|^{m/2}_{\boldsymbol{p}/2}\\&\leqslant N^{-m}\mathbb{E}\left(\sum_{i=1}^N\int_0^t\|(f*\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N)(\Gamma_{t-s}Z^{N,i}_s-\cdot)\|^2_{\boldsymbol{p}}\mathrm{d}s\right)^{m/2}\\&=N^{-m/2}\left(\int_0^t\|f*\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N\|^2_{\boldsymbol{p}}\mathrm{d}s\right)^{m/2}.\end{aligned}$$

Let $\boldsymbol{p}_2\in[1,\infty]^2$ be defined by $\mathbf{1}+\frac{1}{\boldsymbol{p}}=\frac{1}{\boldsymbol{p}_1}+\frac{1}{\boldsymbol{p}_2}$. By Young's inequality, the Definition 1.1 of the anisotropic Besov norm, together with (1.17), and the heat kernel estimate (2.15), we have

$$\begin{aligned}\|f*\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N\|_{\boldsymbol{p}}&\lesssim 2^{-\beta_1 j}\|f\|_{\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}\|\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N\|_{\boldsymbol{p}_2}\\&\lesssim 2^{j(1-\beta_2-\beta_1)}\|f\|_{\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}((2^{2j}(t-s))^{-1}\wedge 1)\|\nabla_v\Gamma_s\phi_N\|_{\mathbf{B}^{\beta_2-1}_{\boldsymbol{p}_2;\boldsymbol{a}}}.\end{aligned}$$

Observing that

$$\int_0^t((2^{2j}s)^{-1}\wedge 1)^2\mathrm{d}s\lesssim 2^{-2j},\tag{3.10}$$

we further have, by (B.8) in Appendix B,

$$\left(\int_0^t\|f*\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N\|^2_{\boldsymbol{p}}\mathrm{d}s\right)^{1/2}\lesssim 2^{-j(\beta_2+\beta_1)}\|f\|_{\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}N^{\zeta(3\beta_2+\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}})}.$$

Combining the above calculations, we get

$$\|f*\mathcal{M}^N_t\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}})}\lesssim\sum_{j\geqslant 0}2^{\beta j}2^{-j(\beta_2+\beta_1)}\|f\|_{\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}N^{\zeta(3\beta_2+\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}})}.$$

The result (3.7) now follows by $\beta<\beta_2+\beta_1$. The estimate (3.7) is deduced next by taking f the Dirac measure on 0, which lies in $\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}$ for $\beta_1=0$ and $\boldsymbol{p}_1=\mathbf{1}$. □

The previous lemma is naturally not enough for the proof of Lemma 2.9 since it requires $\boldsymbol{p}\geqslant 2$. To drop this restriction, we use, as in [31], weight function techniques to show the following stronger estimate. The price we have to pay for this procedure is that we need uniform moment estimates on the process $Z^{N,1}_t$.

Theorem 3.2. *Let $\alpha=2$, $\beta\geqslant 0$, $\boldsymbol{p}\in[1,\infty]^2$ with $p_x\wedge p_v<2$ and $\ell>\mathcal{A}_{\boldsymbol{p},\boldsymbol{p}\vee\mathbf{2}}$. For any $\beta_1>\beta$ and $m\geqslant 2$, there is a constant $C=C(\Theta,\boldsymbol{p},\ell,\beta,\beta_1,m)>0$ such that, for any $N\geqslant 1$,*

$$\sup_{t\in[0,T]}\|\mathcal{M}^N_t\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}})}\lesssim_C\left[1+\sup_{s\in[0,T]}\big(\mathbb{E}\big[|Z^{N,1}_s|^{\ell m}_{\boldsymbol{a}}\big]\big)^{1/m}\right]N^{\zeta(3\beta_1+\mathcal{A}_{\mathbf{1},\boldsymbol{p}\vee\mathbf{2}})-\frac{1}{2}}.\tag{3.11}$$

Proof. By the embedding (2.12), without loss of generality, we may (again) assume $\boldsymbol{p} \in [1,\infty)^2$. Note again from the Minkowski inequality that

$$\|\mathcal{M}^N_t\|_{L^m(\Omega;\mathbf{B}^{\beta;1}_{\boldsymbol{p};\boldsymbol{a}})} \leqslant \sum_{j\geqslant 0} 2^{\beta j}\|\mathcal{R}^{\boldsymbol{a}}_j\mathcal{M}^N_t\|_{L^m(\Omega;\mathbb{L}^{\boldsymbol{p}})}. \tag{3.12}$$

Whenever $\boldsymbol{p} < \mathbf{2}$, one cannot directly make an estimate by BDG's inequality since $\mathbb{L}^{\boldsymbol{p}/2}$ is not a Banach space. For simplicity of notation, we write $\boldsymbol{p_1} := \boldsymbol{p} \vee \mathbf{2}$ and let $\boldsymbol{p_2} \in [1,\infty]^2$ be defined by $\frac{1}{\boldsymbol{p}} = \frac{1}{\boldsymbol{p}_1} + \frac{1}{\boldsymbol{p}_2}$. To overcome the difficulty, we now specifically use a weight function:

$$\omega_\ell(z) := \big(1 + (1+|x|^2)^{\frac{1}{1+\alpha}} + |v|^2\big)^{\ell/2}, \quad z = (x,v) \in \mathbb{R}^{2d}. \tag{3.13}$$

for $\ell > \boldsymbol{a}\cdot\frac{d}{\boldsymbol{p}_2} = \mathcal{A}_{\boldsymbol{p},\boldsymbol{p}_1}$. The main properties and related key estimates on ω_ℓ are stated in Appendix A. Clearly, we have

$$\omega_\ell(z+z') \lesssim \omega_\ell(z) + \omega_\ell(z'), \quad z,z' \in \mathbb{R}^{2d},$$

and the choice of ℓ also precisely guarantees the following integrability of the inverse weight,

$$\|\omega_\ell^{-1}\|_{\boldsymbol{p_2}} < \infty.$$

Since $\mathbb{L}^{\boldsymbol{p}_1}$ is (now) a UMD space, by Hölder's inequality and BDG's inequality (3.9) applied for $\mathbb{L}^{\boldsymbol{p}_1}$-valued martingale (and introducing, for the application of the inequality, an appropriate stopped version of $\mathcal{R}^{\boldsymbol{a}}_j\mathcal{M}^N_t$ as in the proof of Lemma 3.1), we have

$$\begin{aligned}
\mathbb{E}\big[\|\mathcal{R}^{\boldsymbol{a}}_j\mathcal{M}^N_t\|^m_{\boldsymbol{p}}\big] &\leqslant \mathbb{E}\big[\big\|\mathcal{R}^{\boldsymbol{a}}_j\mathcal{M}^N_t\omega_\ell\big\|^m_{\boldsymbol{p}_1}\|\omega_\ell^{-1}\|^m_{\boldsymbol{p}_2}\big]\\
&\lesssim N^{-m}\mathbb{E}\left\|\left(\sum_{i=1}^N\int_0^t|(\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N)(\Gamma_{t-s}Z^{N,i}_s-\cdot)\omega_\ell|^2\mathrm{d}s\right)^{1/2}\right\|^m_{\boldsymbol{p}_1}\\
&= N^{-m}\mathbb{E}\left\|\sum_{i=1}^N\int_0^t|(\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N)(\Gamma_{t-s}Z^{N,i}_s-\cdot)\omega_\ell|^2\mathrm{d}s\right\|^{m/2}_{\boldsymbol{p}_1/2}\\
&\leqslant N^{-m}\mathbb{E}\left(\sum_{i=1}^N\int_0^t\|(\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N)(\Gamma_{t-s}Z^{N,i}_s-\cdot)\omega_\ell\|^2_{\boldsymbol{p}_1}\mathrm{d}s\right)^{m/2}.
\end{aligned} \tag{3.14}$$

Noting that for a nonnegative function $g : \mathbb{R}^{2d} \to [0,\infty)$,

$$\begin{aligned}
g(\Gamma_{t-s}Z^{N,i}_s-\cdot)\omega_\ell &\lesssim (g\omega_\ell)(\Gamma_{t-s}Z^{N,i}_s-\cdot) + g(\Gamma_{t-s}Z^{N,i}_s-\cdot)\omega_\ell(\Gamma_{t-s}Z^{N,i}_s)\\
&\lesssim (g\omega_\ell)(\Gamma_{t-s}Z^{N,i}_s-\cdot)(1+|Z^{N,i}_s|^\ell),
\end{aligned}$$

by the translation invariance of $\mathbb{L}^{\boldsymbol{p}}$-norm, we have

$$\|(\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N)(\Gamma_{t-s}Z^{N,i}_s-\cdot)\omega_\ell\|_{\boldsymbol{p}_1} \lesssim \|\big(\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N\big)\omega_\ell\|_{\boldsymbol{p}_1}(1+|Z^{N,i}_s|^\ell_{\boldsymbol{a}}).$$

Substituting this into (3.14) and by $Z^{N,i}_s \overset{(\text{law})}{=} Z^{N,1}_s$, we further have

$$\mathbb{E}\big[\|\mathcal{R}^{\boldsymbol{a}}_j\mathcal{M}^N_t\|^m_{\boldsymbol{p}}\big] \lesssim N^{-m/2}\left(1+\sup_{s\in[0,t]}\mathbb{E}\big[|Z^{N,1}_s|^{\ell m}_{\boldsymbol{a}}\big]\right)\left(\int_0^t\big\|\big(\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N\big)\omega_\ell\big\|^2_{\boldsymbol{p}_1}\mathrm{d}s\right)^{m/2}. \tag{3.15}$$

At this stage, we can apply Lemma A.3 from Appendix A - with $\beta = \beta_1$, $\boldsymbol{p} = \boldsymbol{p_1}$ and $\gamma = 1$ - to get, for any $j \geqslant 0$,

$$\|\big(\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\nabla_v\Gamma_s\phi_N\big)\omega_\ell\|^2_{\boldsymbol{p}_1} \lesssim 2^{2j(1-\beta_1)}\big((2^{2j}(t-s))^{-1}\wedge 1\big)^2\|(\Gamma_s\phi_N)\omega_\ell\|^2_{\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}.$$

Hence, applying Lemma B.2, it follows that

$$\begin{aligned}\int_0^t &\|(\mathcal{R}_j^{\boldsymbol{a}} P_{t-s}\nabla_v\Gamma_s\phi_N)\omega_\ell\|_{\boldsymbol{p}_1}^2\mathrm{d}s\\ &\lesssim 2^{2j(1-\beta_1)}\int_0^t((2^{2j}(t-s))^{-1}\wedge 1)^2\|\omega_\ell\Gamma_s\phi_N\|^2_{\mathbf{B}^{\beta_1}_{\boldsymbol{p}_1;\boldsymbol{a}}}\mathrm{d}s\\ &\lesssim (1+(tN^{2\zeta})^{2\beta_1})N^{2\zeta(\beta_1+\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_1})}2^{2j(1-\beta_1)}\int_0^t((2^{2j}s)^{-1}\wedge 1)^2\mathrm{d}s\\ &\lesssim N^{2\zeta(3\beta_1+\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_1})}2^{-2j\beta_1},\end{aligned}$$

using again (3.10) for the last inequality. Substituting this into (3.15) and by (3.12), we obtain

$$\|\mathcal{M}_t^N\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}})}\lesssim N^{-1/2}\left(1+\sup_{s\in[0,t]}(\mathbb{E}\big[|Z_s^{N,1}|_{\boldsymbol{a}}\big]^{\ell m})^{1/m}\right)\sum_j 2^{\beta j}N^{\zeta(3\beta_1+\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_1})}2^{-j\beta_1},$$

which immediately implies the estimate (3.11) by $\beta_1>\beta$. □

To conclude the proof of Lemma 2.9 for $\alpha=2$, it remains to establish the uniform moment estimates for the solution to the interacting particle system (1.6) when $\boldsymbol{p}_0=\boldsymbol{1}$. These estimates are given by the following lemma.

Lemma 3.3. *Let $\alpha=2$ and $\boldsymbol{p}_0=\boldsymbol{1}$ and assume the condition* **(H)** *holds. Then, for any $m\in\mathbb{N}$, there is a constant $C=C(\Theta,m)>0$ such that for all $t\in(0,T_0]$ and $N\geqslant 1$,*

$$\|b_t*u_t^N\|_{L^m(\Omega;\mathbb{L}^\infty)}\lesssim_C t^{-\frac{\Lambda}{2}}+\sup_{t\in[0,T_0]}\|b_t*\mathcal{M}_t^N\|_{L^m(\Omega;\mathbb{L}^\infty)}. \tag{3.16}$$

In particular,

$$\sup_{t\in[0,T_0]}\mathbb{E}\big[|Z_t^{N,1}|_{\boldsymbol{a}}^m\big]\lesssim_C 1+\mathbb{E}\big[|Z_0^{N,1}|_{\boldsymbol{a}}^m\big]+\sup_{t\in[0,T_0]}\|b_t*\mathcal{M}_t^N\|^m_{L^m(\Omega;\mathbb{L}^\infty)}. \tag{3.17}$$

Proof. By (2.18) and Duhamel's formula (with the notations of (2.23)), we have

$$u_t^N(z)=P_tu_0^N(z)+\int_0^t P_{t-s}\mathrm{div}_vG_s^N(z)\mathrm{d}s+\mathcal{M}_t^N(z).$$

By (2.11) and (2.16), we thus get

$$\begin{aligned}\|b_t*u_t^N\|_{\mathbb{L}^\infty}&\leqslant\|b_t*P_tu_0^N\|_{\mathbb{L}^\infty}+\int_0^t\|b_t*P_{t-s}\mathrm{div}_vG_s^N\|_{\mathbb{L}^\infty}\mathrm{d}s+\|b_t*\mathcal{M}_t^N\|_{\mathbb{L}^\infty}\\ &\lesssim\|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}}\|P_tu_0^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}}+\int_0^t\|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}}\|P_{t-s}\mathrm{div}_vG_s^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}}\mathrm{d}s+\|b_t*\mathcal{M}_t^N\|_{\mathbb{L}^\infty}\\ &\lesssim t^{(\beta_b-\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_b'})/2}\|u_0^N\|_{\mathbf{B}^0_{\boldsymbol{1};\boldsymbol{a}}}+\int_0^t(t-s)^{(\beta_b-\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_b'}-1)/2}\|G_s^N\|_{\mathbf{B}^0_{\boldsymbol{1};\boldsymbol{a}}}\mathrm{d}s+\|b_t*\mathcal{M}_t^N\|_{\mathbb{L}^\infty}.\end{aligned}$$

Observing as well that

$$\|u_0^N\|_{\mathbf{B}^0_{\boldsymbol{1};\boldsymbol{a}}}\lesssim\|u_0^N\|_{\boldsymbol{1}}=\|\mu_t^N*\Gamma_t\phi_N\|_{\boldsymbol{1}}=1,$$

and, from (2.20),

$$|G_s^N(z)|\leqslant\|b_s*u_s^N\|_{\mathbb{L}^\infty}u_s^N(z)\Rightarrow\|G_s^N\|_{\boldsymbol{1}}\leqslant\|b_s*u_s^N\|_{\mathbb{L}^\infty}.$$

From the above computations, using again (2.10) and from (1.20),

$$\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_b'}-\beta_b\leqslant\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_b'}-\beta_b-\beta_0=\Lambda<1,$$

we have, for all $t\in(0,T_0]$,

$$\|b_t*u_t^N\|_{\mathbb{L}^\infty}\lesssim t^{-\frac{\Lambda}{2}}+\int_0^t(t-s)^{-\frac{1+\Lambda}{2}}\|b_s*u_s^N\|_{\mathbb{L}^\infty}\mathrm{d}s+\|b_t*\mathcal{M}_t^N\|_{\mathbb{L}^\infty}.$$

Therefore,

$$\|b_t * u_t^N\|_{L^m(\Omega;\mathbb{L}^\infty)} \lesssim t^{-\frac{\Lambda}{2}} + \|b_t * \mathcal{M}_t^N\|_{L^m(\Omega;\mathbb{L}^\infty)} + \int_0^t (t-s)^{-\frac{1+\Lambda}{2}} \|b_s * u_s^N\|_{L^m(\Omega;\mathbb{L}^\infty)} \mathrm{d}s.$$

Since $\Lambda < 1$, by Gronwall's inequality of Volterra-type (see Lemma E.2), we obtain (3.16).

In particular, by (1.6), we have, for any $m \geqslant 1$,

$$\sup_{t\in[0,T_0]} \mathbb{E}\big[|Z_t^{N,1}|_{\boldsymbol{a}}^m\big] \lesssim 1 + \mathbb{E}\big[|Z_0^{N,1}|_{\boldsymbol{a}}^m\big] + \int_0^{T_0} \mathbb{E}\big[\|b_s * u_s^N\|_{\mathbb{L}^\infty}^m\big] \mathrm{d}s.$$

Estimate (3.17) now follows by (3.16). □

Remark 3.4. *For $\boldsymbol{p}_0 \neq \mathbf{1}$, let us point out that we have a direct treatment for $\|b_t * u_t^N\|_{\mathbb{L}^\infty}$, but with a worse convergence rate. Indeed, by (2.11) and (2.2),*

$$\begin{aligned}\|b_t * u_t^N\|_{\mathbb{L}^\infty} &\lesssim \|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}} \|u_t^N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}} = \|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}} \|\mu_t^N * \Gamma_t\phi_N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}} \\ &\leqslant \|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}} \|\mu_t^N\|_{\mathbf{B}^0_{\mathbf{1};\boldsymbol{a}}} \|\Gamma_t\phi_N\|_{\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}}} \lesssim \|\Gamma_t\phi_N\|_{\mathbf{B}^{-\beta_b+\frac{\varepsilon}{3}}_{\boldsymbol{p}_b';\boldsymbol{a}}} \lesssim N^{\zeta(-3\beta_b+\varepsilon+\boldsymbol{a}\cdot\frac{d}{\boldsymbol{p}_b})}\end{aligned}$$

using (2.10) for the last but one inequality and Lemma B.2 for the last one.

Next we turn to the treatment of $\alpha \in (1,2)$. In this case, we need to use a specific martingale inequality established in [20]. We first state the following estimate.

Lemma 3.5. *For any $\beta \in \mathbb{R}$, $\boldsymbol{p} \in [1,\infty]^2$ and $0 \leqslant \varepsilon \leqslant \theta < 1$, there is a constant $C > 0$ such that for all $t > 0$, $v \in \mathbb{R}^d$ and $f \in \mathbf{B}^{\beta+\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}$,*

$$\|P_t \delta^{(1)}_{(0,v)} f\|_{\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}}} \lesssim_C \big[(|v|^\theta t^{-\frac{\theta-\varepsilon}{\alpha}}) \wedge 1\big] \|f\|_{\mathbf{B}^{\beta+2\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}.$$

Proof. Let $\boldsymbol{p} \in [1,\infty]^2$. By (2.10) and the heat kernel estimate (2.16), we have, for any $\gamma \geqslant 0$,

$$\big\|P_t \delta^{(1)}_{(0,v)} f\big\|_{\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}}} \lesssim \big\|P_t \delta^{(1)}_{(0,v)} f\big\|_{\mathbf{B}^{\beta+\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}} \lesssim t^{-\frac{\gamma}{\alpha}} \|\delta^{(1)}_{(0,v)} f\|_{\mathbf{B}^{\beta-\gamma+\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}.$$

On the other hand, by the equivalent form of (2.9), we have, for any $\beta \in \mathbb{R}$ and $\gamma \in [0,1)$,

$$\begin{aligned}\|\delta^{(1)}_{(0,v)} f\|_{\mathbf{B}^{\beta}_{\boldsymbol{p};\boldsymbol{a}}} &= \sup_{j\geqslant 0} 2^{j\beta} \|\mathcal{R}^{\boldsymbol{a}}_j \delta^{(1)}_{(0,v)} f\|_{\boldsymbol{p}} = \sup_{j\geqslant 0} 2^{j\beta} \|\delta^{(1)}_{(0,v)} \mathcal{R}^{\boldsymbol{a}}_j f\|_{\boldsymbol{p}} \\ &\lesssim |v|^\gamma \sup_{j\geqslant 0} 2^{j\beta} \|\mathcal{R}^{\boldsymbol{a}}_j f\|_{\mathbf{B}^{\gamma}_{\boldsymbol{p};\boldsymbol{a}}} \lesssim |v|^\gamma \|f\|_{\mathbf{B}^{\beta+\gamma}_{\boldsymbol{p};\boldsymbol{a}}}.\end{aligned}$$

Combining the above two estimates, we obtain

$$\big\|P_t \delta^{(1)}_{(0,v)} f\big\|_{\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}}} \lesssim \|f\|_{\mathbf{B}^{\beta+\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}$$

and for any $0 \leqslant \varepsilon \leqslant \theta < 1$,

$$\big\|P_t \delta^{(1)}_{(0,v)} f\big\|_{\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}}} \lesssim t^{-\frac{\theta-\varepsilon}{\alpha}} \|\delta^{(1)}_{(0,v)} f\|_{\mathbf{B}^{\beta-\theta+2\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}} \lesssim |v|^\theta t^{-\frac{\theta-\varepsilon}{\alpha}} \|f\|_{\mathbf{B}^{\beta+2\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}.$$

The proof is complete. □

Now we can show the following estimate.

Theorem 3.6. *Let $T > 0$, $1 < \alpha < 2$ and $\boldsymbol{p} = (p_x, p_v) \in (1,\infty]^2$ with*

$$\alpha < q := 2 \wedge p_x \wedge p_v.$$

Then, for any $\beta \geqslant 0$, $m \geqslant 1$ and $\epsilon > 0$, there exists $C = C(T,\alpha,\boldsymbol{p},\beta,m,\varepsilon) > 0$ such that for all $N \geqslant 1$,

$$\sup_{t\in[0,T]} \|\mathcal{M}_t^N\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}})} \lesssim_C N^{\frac{1}{q}-1+\zeta((\alpha+1)\beta+\mathcal{A}_{\mathbf{1},\boldsymbol{p}}+\varepsilon)}. \tag{3.18}$$

As a preliminary to the proof of Theorem 3.6, let us highlight that applying the functional BDG inequality (3.9) previously used for the case Brownian case $\alpha = 2$ would naturally bring rather intricate moment issues. It is more natural to rely on a specific martingale inequality which fits the case of integrals with compensated Poisson measures. As we refer the interested reader to [27, Section 5] for a historical account of maximal inequalities for such integrals, we will make use below of the particular estimate established in [20].

Theorem 3.7 ([20], Theorem 2.13 and Corollary 2.14)**.** *Let $(E, |\cdot|_E)$ be a separable Banach space of martingale type p with $1 < p \leqslant 2$, $(Z, \mathcal{Z})$ be a measurable space and ν be a positive σ-finite measure on Z. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ be a complete filtered probability space with right continuous filtration. Assume that $\tilde{\eta}$ is a compensated time homogeneous Poisson random measure on Z over $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ with intensity ν. Then, for any progressively measurable process $\xi : \mathbb{R}^+ \times \Omega \to E$ such that, for some $n \geqslant 1$,*

$$\mathbb{E}\Big[\int_0^\infty \int_Z |\xi(t,z)|_E^p \nu(dz)\,\mathrm{d}t + \int_0^\infty \int_Z |\xi(t,z)|_E^{p^n} \nu(dz)\,\mathrm{d}t\Big] < \infty$$

we have

$$\mathbb{E}\Big[\sup_{0\leqslant t\leqslant T} \Big|\int_0^t \int_Z \xi(s,z)\widetilde{\eta}(ds,dz)\Big|_E^{p^n}\Big)\Big] \lesssim_C \sum_{k=1}^n \mathbb{E}\Bigg[\Big(\int_0^T \int_Z |\xi(s,z)|_E^{p^k} \nu(dz)\,\mathrm{d}s\Big)^{p^{n-k}}\Bigg], \tag{3.19}$$

for some positive constant $C = C(E, p, n)$.

As for the notion of UMD spaces, the definition of Banach space of martingale type p is recalled in Appendix D. Therein, we also establish that, under the assumption $\alpha < 1 \wedge p_x \wedge p_v$, the inequality (3.19) applies to the case $E = \mathbb{L}^{\boldsymbol{p}}$ (see Corollary D.4).

Proof of Theorem 3.6. As in the previous proof of this section, according to the embedding (2.12), we may assume $\boldsymbol{p} \in (1, \infty)^2$ without loss of generality. As $\alpha \in (1, 2)$, by (2.17), (2.19) and (2.23), we can write

$$\mathcal{M}_t^N(z) = \frac{1}{N} \sum_{i=1}^N \int_0^t \int_{\mathbb{R}^d\setminus\{0\}} \xi_t^i(s, v_i)(z)\tilde{\mathcal{N}}^i(\mathrm{d}s, \mathrm{d}v_i),$$

where (using again the difference operator (2.8)),

$$\xi_t^i(s, v)(z) := P_{t-s}\Big(\delta_{(0,v)}^{(1)} \Gamma_s \Phi_N\Big)(Z_{s-}^{N,i} - \cdot)(z).$$

To use appropriately the inequality (3.19), we first need to lift $\mathcal{M}^N$ on the product space $\mathbb{R}_0^{Nd} := \mathbb{R}^{Nd} \setminus \{0\}$ as follows: For

$$\boldsymbol{L}_t^N := (L_t^{\alpha,1}, \cdots, L_t^{\alpha,N}),$$

the overall noise driving the particle system (1.6), let $\boldsymbol{\mathcal{N}}^N((0,t], U)$ denote the jump measure of $\boldsymbol{L}^N$ and $\widetilde{\boldsymbol{\mathcal{N}}}^N$ the related compensated measure, respectively defined as: for all $t \in (0, T]$ and $U \in \mathscr{B}(\mathbb{R}_0^{Nd})$,

$$\boldsymbol{\mathcal{N}}^N((0,t], U) := \sum_{0<s\leqslant t} \mathbb{1}_U(\Delta \boldsymbol{L}_s^N), \quad \widetilde{\boldsymbol{\mathcal{N}}}^N(\mathrm{d}s, \mathrm{d}\boldsymbol{v}) := \boldsymbol{\mathcal{N}}^N(\mathrm{d}s, \mathrm{d}\boldsymbol{v}) - \boldsymbol{\nu}(\mathrm{d}\boldsymbol{v})\mathrm{d}s,$$

where $\boldsymbol{\nu}(\mathrm{d}\boldsymbol{v})$ is the Lévy measure of $\boldsymbol{L}^N$. Since the $L^{\alpha,i}$ are independent, their jumps $\Delta L^{\alpha,i} \neq 0$ never occur at the same time and $\boldsymbol{\nu}$ and $\boldsymbol{\mathcal{N}}^N$ admit the following representations:

$$\boldsymbol{\nu}(\mathrm{d}\boldsymbol{v}) = \sum_{i=1}^N \delta_{\{0\}}(\mathrm{d}v_1) \cdots \delta_{\{0\}}(\mathrm{d}v_{i-1})\nu(\mathrm{d}v_i)\delta_{\{0\}}(\mathrm{d}v_{i+1}) \cdots \delta_{\{0\}}(\mathrm{d}v_N), \tag{3.20}$$

$$\boldsymbol{\mathcal{N}}^N(ds, \mathrm{d}\boldsymbol{v}) = \sum_{i=1}^N \delta_{\{0\}}(\mathrm{d}v_1) \cdots \delta_{\{0\}}(\mathrm{d}v_{i-1})\mathcal{N}^i(ds, \mathrm{d}v_i)\delta_{\{0\}}(\mathrm{d}v_{i+1}) \cdots \delta_{\{0\}}(\mathrm{d}v_N),$$

for $\delta_{\{0\}}$ the Dirac measure in 0. In particular, for any i, since $\xi^i_t(s,0)(z)=0$ and since the measure $\widetilde{\mathcal{n}}^N$ only supports one jump at a given time, we have:

$$\begin{aligned}&\int_{\mathbb{R}^{Nd}_0}\xi^i_t(s,v_i)(z)\widetilde{\mathcal{n}}^N(\mathrm{d}s,\mathrm{d}\boldsymbol{v})\\&=\sum_{j=1,j\neq i}^N\int_{\mathbb{R}^d_0}\xi^i_t(s,0)(z)\,\mathbb{1}_{\{v_j\neq 0\}}\widetilde{\mathcal{n}}^j(\mathrm{d}s,\mathrm{d}v_j)+\int_{\mathbb{R}^d_0}\xi^i_t(s,v_i)(z)\widetilde{\mathcal{n}}^i(\mathrm{d}s,\mathrm{d}v_i)\\&=\int_{\mathbb{R}^d_0}\xi^i_t(s,v_i)(z)\widetilde{\mathcal{n}}^i(\mathrm{d}s,\mathrm{d}v_i).\end{aligned}$$

As such, if we next introduce the predictable process

$$\boldsymbol{\xi}^N_t(s,\boldsymbol{v})(z):=\frac{1}{N}\sum_{i=1}^N\xi^i_t(s,v_i)(z),\quad 0\leqslant s\leqslant t\leqslant T, z\in\mathbb{R}^d,$$

then $\mathcal{m}^N_t(z)$ can be written as

$$\mathcal{m}^N_t(z)=\int_0^t\int_{\mathbb{R}^{Nd}_0}\boldsymbol{\xi}^N_t(s,\boldsymbol{v})(z)\widetilde{\mathcal{n}}^N(\mathrm{d}s,\mathrm{d}\boldsymbol{v}).$$

Since $\mathbb{L}^{\boldsymbol{p}}$ is a space of martingale q-type (recall that $q\in(\alpha,2\wedge p_x\wedge p_v])$, applying Theorem 3.7 to the stopped martingale

$$\mathcal{m}^N_{u,t}(z)=\int_0^u\int_{\mathbb{R}^{Nd}_0}\boldsymbol{\xi}^N_t(s,\boldsymbol{v})(z)\widetilde{\mathcal{n}}^N(\mathrm{d}s,\mathrm{d}\boldsymbol{v}),\quad u\in[0,t],$$

we have: for any $n\in\mathbb{N}$,

$$\begin{aligned}\Big(\mathbb{E}\Big[\|\mathcal{m}^N_t\|^{q^n}_{\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}}}\Big]\Big)^{1/q^n}&\leqslant\sum_{j\geqslant 0}2^{\beta j}\Big(\mathbb{E}\Big[\|\mathcal{R}^{\boldsymbol{a}}_j\mathcal{m}^N_t\|^{q^n}_{\mathbb{L}^{\boldsymbol{p}}}\Big]\Big)^{1/q^n}\leqslant\sum_{j\geqslant 0}2^{\beta j}\Big(\mathbb{E}\Big[\sup_{0\leqslant u\leqslant t}\|\mathcal{R}^{\boldsymbol{a}}_j\mathcal{m}^N_{u,t}\|^{q^n}_{\mathbb{L}^{\boldsymbol{p}}}\Big]\Big)^{1/q^n}\\&=\sum_{j\geqslant 0}2^{\beta j}\Big(\mathbb{E}\Big[\sup_{0\leqslant u\leqslant t}\Big\|\int_0^u\int_{\mathbb{R}^{Nd}_0}\mathcal{R}^{\boldsymbol{a}}_j\boldsymbol{\xi}^N_t(s,\boldsymbol{v})(z)\widetilde{\mathcal{n}}^N(\mathrm{d}s,\mathrm{d}\boldsymbol{v})\Big\|^{q^n}_{\mathbb{L}^{\boldsymbol{p}}}\Big]\Big)^{1/q^n}\\&\lesssim\sum_{j\geqslant 0}2^{\beta j}\Bigg(\sum_{k=1}^n\mathbb{E}\Bigg[\Big(\int_0^t\int_{\mathbb{R}^{Nd}_0}\|\mathcal{R}^{\boldsymbol{a}}_j\boldsymbol{\xi}^N_t(s,\boldsymbol{v})\|^{q^k}_{\mathbb{L}^{\boldsymbol{p}}}\boldsymbol{\nu}(\mathrm{d}\boldsymbol{v})\,\mathrm{d}s\Big)^{q^{n-k}}\Bigg]\Bigg)^{1/q^n}.\end{aligned}\tag{3.21}$$

According to (3.20),

$$\int_{\mathbb{R}^{Nd}_0}\|\mathcal{R}^{\boldsymbol{a}}_j\boldsymbol{\xi}^N_t(s,\boldsymbol{v})\|^{q^k}_{\mathbb{L}^{\boldsymbol{p}}}\boldsymbol{\nu}(\mathrm{d}\boldsymbol{v})=\frac{1}{N^{q^k}}\sum_{i=1}^N\int_{\mathbb{R}^d_0}\|\mathcal{R}^{\boldsymbol{a}}_j\xi^i_t(s,v_i)\|^{q^k}_{\mathbb{L}^{\boldsymbol{p}}}\nu(\mathrm{d}v_i),$$

and so

$$\mathbb{E}\Bigg[\Big(\int_0^t\int_{\mathbb{R}^{Nd}_0}\|\mathcal{R}^{\boldsymbol{a}}_j\boldsymbol{\xi}^N_t(s,\boldsymbol{v})\|^{q^k}_{\mathbb{L}^{\boldsymbol{p}}}\boldsymbol{\nu}(\mathrm{d}\boldsymbol{v})\,\mathrm{d}s\Big)^{q^{n-k}}\Bigg]=\frac{1}{N^{q^n}}\mathbb{E}\Bigg[\Big(\sum_{i=1}^N\int_0^t\int_{\mathbb{R}^d_0}\|\mathcal{R}^{\boldsymbol{a}}_j\xi^i_t(s,v)\|^{q^k}_{\mathbb{L}^{\boldsymbol{p}}}\nu(\mathrm{d}v)\,\mathrm{d}s\Big)^{q^{n-k}}\Bigg].$$

Using successively the invariance by translation of $\|\cdot\|_{\mathbb{L}^{\boldsymbol{p}}}$ and Lemmas 2.5 and 3.5 further gives

$$\begin{aligned}\|\mathcal{R}^{\boldsymbol{a}}_j\xi^i_t(s,v)\|_{\mathbb{L}^{\boldsymbol{p}}}&=\Big\|\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\Big(\delta^{(1)}_{(0,v)}\Gamma_s\Phi_N\Big)(Z^{N,i}_{s-}-\cdot)\Big\|_{\mathbb{L}^{\boldsymbol{p}}}=\Big\|\mathcal{R}^{\boldsymbol{a}}_jP_{t-s}\Big(\delta^{(1)}_{(0,v)}\Gamma_s\Phi_N\Big)\Big\|_{\mathbb{L}^{\boldsymbol{p}}}\\&\lesssim 2^{-(\beta+\epsilon)j}\|P_{t-s}\delta^{(1)}_{(0,v)}\Gamma_s\Phi_N\|_{\mathbf{B}^{\beta+\varepsilon,1}_{\boldsymbol{p};\boldsymbol{a}}}\lesssim 2^{-(\beta+\epsilon)j}\Big[\big(|v|^\theta(t-s)^{-\frac{\theta}{\alpha}}\big)\wedge 1\Big]\|\Gamma_s\phi_N\|_{\mathbf{B}^{\beta+3\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}.\end{aligned}$$

Hence, plugging the above into (3.21) yields

$$\begin{aligned}&\left(\mathbb{E}\left[\|m_t^N\|^{q^n}_{\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}}}\right]\right)^{1/q^n}\\&\lesssim \frac{1}{N}\sup_{s\in[0,t]}\|\Gamma_s\phi_N\|_{\mathbf{B}^{\beta+3\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}\sum_{j\geqslant 0}2^{-\varepsilon j}\left(\sum_{k=1}^n\left(N\int_0^t\int_{\mathbb{R}^d_0}\frac{(|v|^\theta(t-s)^{-\frac{\theta-\varepsilon}{\alpha}})^{q^k}\wedge 1}{|v|^{d+\alpha}}\mathrm{d}v\mathrm{d}s\right)^{q^{n-k}}\right)^{1/q^n}\\&\lesssim \frac{N^{1/q}}{N}\sup_{s\in[0,t]}\|\Gamma_s\phi_N\|_{\mathbf{B}^{\beta+3\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}\left(\sum_{k=1}^n\left(\int_0^t(t-s)^{\frac{\varepsilon}{\theta}-1}\int_{\mathbb{R}^d_0}\frac{|v|^{\theta q^k}\wedge 1}{|v|^{d+\alpha}}\mathrm{d}v\mathrm{d}s\right)^{q^{n-k}}\right)^{1/q^n}.\end{aligned}$$

Since $q\in(\alpha,2]$, one can choose $\theta\in(\frac{\alpha}{q},1)$ so that the remaining time-integrals on the right-hand side are finite for each $k=1,\cdots,n$. Thus, we obtain

$$\left(\mathbb{E}\left[\|m_t^N\|^{q^n}_{\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}}}\right]\right)^{1/q^n}\lesssim N^{1/q-1}\sup_{s\in[0,t]}\|\Gamma_s\phi_N\|_{\mathbf{B}^{\beta+3\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}.$$

Now for any $m\in\mathbb{N}$, one can choose n large enough so that $m\leqslant q^n$. Thus by Hölder's inequality and (B.8),

$$\|m_t^N\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p};\boldsymbol{a}})}\lesssim N^{1/q-1}\sup_{s\in[0,t]}\|\Gamma_s\phi_N\|_{\mathbf{B}^{\beta+3\varepsilon}_{\boldsymbol{p};\boldsymbol{a}}}\lesssim N^{1/q-1+\zeta((\alpha+1)(\beta+3\varepsilon)+\mathscr{A}_{\mathbf{1},\boldsymbol{p}})}.$$

The proof is complete by choosing n large enough. □

Proof of Lemma 2.9. We separately consider the cases $\alpha=2$ and $\alpha\in(1,2)$.

(Case $\alpha=2$). Fix $\varepsilon>0$. By Lemma 3.1 (with $\beta=0$, $\beta_1=\beta_b$, $\boldsymbol{p}=\boldsymbol{\infty}$ and $\boldsymbol{p_2}=\boldsymbol{p}_b$), we have,

$$\|b_t*m_t^N\|_{L^m(\Omega;\mathbb{L}^{\boldsymbol{\infty}})}\lesssim N^{-\frac{1}{2}+\zeta(\mathscr{A}_{\boldsymbol{p}_b,\infty}-3\beta_b+\varepsilon)}\|b_t\|_{\mathbf{B}^{\beta_b}_{\boldsymbol{p}_b;\boldsymbol{a}}}. \tag{3.22}$$

In particular, if $\zeta\leqslant 1/(2(\mathscr{A}_{\boldsymbol{p}_b,\infty}-3\beta_b+\varepsilon))$, then

$$\sup_{t\in[0,T_0]}\sup_N\|b_t*m_t^N\|_{L^m(\Omega;\mathbb{L}^{\boldsymbol{\infty}})}\lesssim 1. \tag{3.23}$$

If $\boldsymbol{p}_0>\mathbf{2}$, then by (3.7) with $\beta=0$, $\beta_2=\varepsilon/3$ and $\boldsymbol{p}=\boldsymbol{p}_0$, we have

$$\|m_t^N\|_{L^m(\Omega;\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}})}\lesssim N^{-\frac{1}{2}+\zeta(\varepsilon+\mathscr{A}_{\mathbf{1},\boldsymbol{p}_0})}.$$

Hence, by definition of the norm $\|\cdot\|_{\mathbb{S}^\beta_t(b)}$ and (3.22), we have

$$\begin{aligned}\|m_t^N\|_{L^m(\Omega;\mathbb{S}^\beta_t(b))}&=t^{\frac{\beta-\beta_0}{\alpha}}\|m_t^N\|_{L^m(\Omega;\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}})}+t^{\frac{\beta+\Lambda}{\alpha}}\|b_t*m_t^N\|_{L^m(\Omega;\mathbb{L}^{\boldsymbol{\infty}})}\\&\lesssim N^{-\frac{1}{2}+\zeta(\varepsilon+\mathscr{A}_{\mathbf{1},\boldsymbol{p}_0})}+N^{-\frac{1}{2}+\zeta(\mathscr{A}_{\boldsymbol{p}_b,\boldsymbol{\infty}}-3\beta_b+\varepsilon)}\\&\lesssim N^{-\frac{1}{2}+\zeta((\mathscr{A}_{\boldsymbol{p}_b,\boldsymbol{\infty}}-3\beta_b)\vee\mathscr{A}_{\mathbf{1},\boldsymbol{p}_0}+\varepsilon)}.\end{aligned} \tag{3.24}$$

If $\boldsymbol{p}_0=\mathbf{1}$, then, by (3.11) (for $\beta=0$, $\beta_1=\varepsilon/3$, $\boldsymbol{p}=\mathbf{1}$),

$$\begin{aligned}\|m_t^N\|_{L^m(\Omega;\mathbf{B}^{0,1}_{\mathbf{1};\boldsymbol{a}})}&\lesssim\left[1+\sup_{s\in[0,T]}(\mathbb{E}|Z_s^{N,1}|^{\ell m}_{\boldsymbol{a}})^{1/m}\right]N^{\zeta(\varepsilon+\mathscr{A}_{\mathbf{1},\mathbf{2}})-\frac{1}{2}}\\&\overset{(3.17),(3.23)}{\lesssim}\left[1+(\mathbb{E}|Z_0^{N,1}|^{\ell m}_{\boldsymbol{a}})^{1/m}\right]N^{\zeta(\varepsilon+\mathscr{A}_{\mathbf{1},\mathbf{2}})-\frac{1}{2}}.\end{aligned}$$

Substituting this into (3.24) and by (3.22), we obtain

$$\begin{aligned}\|m_t^N\|_{L^m(\Omega;\mathbf{B}^{0,1}_{\mathbf{1},;\boldsymbol{a}})}&\lesssim\left[1+(\mathbb{E}|Z_0^{N,1}|^{\ell m}_{\boldsymbol{a}})^{1/m}\right]N^{\zeta(\varepsilon+\mathscr{A}_{\mathbf{1},\mathbf{2}})-\frac{1}{2}}+N^{-\frac{1}{2}+\zeta(\mathscr{A}_{\boldsymbol{p}_b,\infty}-3\beta_b+\varepsilon)}\\&\lesssim\left[1+(\mathbb{E}|Z_0^{N,1}|^{\ell m}_{\boldsymbol{a}})^{1/m}\right]N^{\zeta((\mathscr{A}_{\boldsymbol{p}_b,\infty}-3\beta_b)\vee\mathscr{A}_{\mathbf{1},\mathbf{2}}+\varepsilon)-\frac{1}{2}}.\end{aligned}$$

(Case $\alpha \in (1,2)$). By (2.11), we have

$$\begin{aligned}\|m_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))} &= t^{\frac{\beta-\beta_0}{\alpha}}\|m_t^N\|_{L^m(\Omega;\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}})} + t^{\frac{\beta+\Lambda}{\alpha}}\|b_t * m_t^N\|_{L^m(\Omega;\mathbb{L}^\infty)}\\ &\lesssim t^{\frac{\beta-\beta_0}{\alpha}}\|m_t^N\|_{L^m(\Omega;\mathbf{B}^{0,1}_{\boldsymbol{p}_0;\boldsymbol{a}})} + t^{\frac{\beta+\Lambda}{\alpha}}\|m_t^N\|_{L^m(\Omega;\mathbf{B}^{-\beta_b,1}_{\boldsymbol{p}_b';\boldsymbol{a}})}.\end{aligned}$$

Since $p_{x,0} \wedge p_{v,0} > \alpha$ and $\boldsymbol{p}_0 \leqslant \boldsymbol{p}_b'$, one can choose $(\beta, \boldsymbol{p}) = (0, \boldsymbol{p}_0)$ and $(-\beta_b, \boldsymbol{p}_b')$ separately in Estimate (3.18) so that the exponent q dominating the rate in N is given by $q = p_{x,0} \wedge p_{v,0} \wedge 2$, and

$$\begin{aligned}\|m_t^N\|_{L^m(\Omega;\mathbb{S}_t^\beta(b))} &\lesssim N^{\frac{1}{q}-1+\zeta(\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_0}+\varepsilon)} + N^{\frac{1}{q}-1+\zeta(\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_b'}-(1+\alpha)\beta_b+\varepsilon)}\\ &\lesssim N^{\frac{1}{q}-1+\zeta(\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_0}\vee(\mathcal{A}_{\boldsymbol{p}_b,\boldsymbol{\infty}}-(1+\alpha)\beta_b)+\varepsilon)}.\end{aligned}$$

This completes the proof. □

Appendix A. Weighted estimates

In this section, we state and prove some essential properties on the weighted $\mathbb{L}^{\boldsymbol{p}}$-controls relative to the block operators $\mathcal{R}_j^{\boldsymbol{a}}$ defined in (1.17), for a particular family of weight functions ω_ℓ, specified in (A.1) below. We also establish controls for the kinetic semigroup P_t defined in (2.14). These estimates are notably meaningful to handle the particular setting $\boldsymbol{p}_0 = \boldsymbol{1}$ in the main Theorems 1.2 and 2.1 (notably for Theorem 3.2), and some intermediate results in the subsequent appendices B and C.

The results presented in this section consist of slight extensions of previous statements from [19] and [17] (and also references therein), but, for the sake of completeness, we present here the full proofs.

Consider the class of weight functions, given by: For $\ell \geqslant 0$,

$$\omega_\ell(z) := \big(1 + (1+|x|^2)^{\frac{1}{1+\alpha}} + |v|^2\big)^{\ell/2}, \quad z = (x,v) \in \mathbb{R}^{2d}. \tag{A.1}$$

This weight function is C^∞ and satisfies the following three properties relative to derivatives, growth and integrability control and to finite difference:

$$|\nabla^k \omega_\ell(z)| = |(\nabla_x^k \omega_\ell(z), \nabla_v^k \omega_\ell(z))| \lesssim \omega_\ell(z), \quad k \in \mathbb{N}, z \in \mathbb{R}^{2d}, \tag{A.2}$$

$$\omega_\ell(z+z') \leqslant (2^{\ell-1}\vee 1)(\omega_\ell(z)+\omega_\ell(z')) \leqslant 2(2^{\ell-1}\vee 1)\omega_\ell(z)\omega_\ell(z'), \ z,z' \in \mathbb{R}^{2d}, \tag{A.3}$$

and, for all $\boldsymbol{p} \in [1,\infty]^2$, such that $\ell > \boldsymbol{a}\cdot\frac{d}{\boldsymbol{p}}$,

$$\|(\omega_\ell)^{-1}\|_{\boldsymbol{p}} < \infty. \tag{A.4}$$

Given ω_ℓ as above, we shall consider the quantities

$$\|(\mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}}, \quad j \geqslant 0, \boldsymbol{p} \in [1,\infty]^2, f \in \mathcal{S}'(\mathbb{R}^{2d}).$$

These are naturally related to the weighted anisotropic Besov space:

$$B^{s,q}_{\boldsymbol{p};\boldsymbol{a}}(\omega_\ell) := \left\{ f \in \mathcal{S}' : \|f\|_{\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}(\omega_\ell)} := \left(\sum_{j\geqslant 0}\big(2^{js}\|(\mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}}\big)^q\right)^{1/q} < \infty \right\},$$

for $s \geqslant 0$, $\boldsymbol{p} \in [1,\infty]^2$ and $q \in [1,\infty]$. Theorem 2.7 in [19] previously established, for any $f \in \mathcal{S}'$, the equivalence between the norms $\|f\|_{\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}(\omega_\ell)}$ and $\|\omega_\ell f\|_{\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}}$ in the special case where $\boldsymbol{p} = (p,p) \in [1,\infty]^2$. This equivalence is stated for a broad class of weight functions, ω_ℓ defined in (A.1) being a particular instance of this class. The original proof of Theorem 2.7 in [19] relies on establishing, whenever $s > 0$, that

$$\|f\|_{\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}(\omega_\ell)} \asymp \|f\|_{\widetilde{\mathbf{B}}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}(\omega_\ell)} \asymp \|\omega_\ell f\|_{\mathbf{B}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}}, \tag{A.5}$$

where

$$\|f\|_{\widetilde{\mathbf{B}}^{s,q}_{\boldsymbol{p};\boldsymbol{a}}(\omega_\ell)} := \|\omega_\ell f\|_{\boldsymbol{p}} + \left(\int_{|h|_{\boldsymbol{a}}\leqslant 1}\left(\frac{\|\omega_\ell\delta_h^{([\beta]+1)}f\|_{\boldsymbol{p}}}{|h|_{\boldsymbol{a}}^s}\right)^q \frac{\mathrm{d}h}{|h|^{(2+\alpha)d}}\right)^{1/q},$$

for $\delta_h^{(M)}$ the M^{th} difference operator defined in Section 2.1. Essentially, the equivalence (A.5) is obtained from the composition properties of $\delta_h^{(M)}$, suitable commutations with the block operators $\mathcal{R}_j^{\boldsymbol{a}}$ (see [19], pp. 647), the properties (A.2) and (A.3), and finally the weighted Young inequality:

$$\|(g*h)\omega_\ell\|_{\boldsymbol{p}} \lesssim \|g\omega_\ell\|_{\boldsymbol{p}_1}\|h\omega_\ell\|_{\boldsymbol{p}_2},\quad \frac{1}{\boldsymbol{p}'} = \frac{1}{\boldsymbol{p}_1}+\frac{1}{\boldsymbol{p}_2}, \boldsymbol{p},\boldsymbol{p}_1,\boldsymbol{p}_2\in[1,\infty]^2 \tag{A.6}$$

(for $\boldsymbol{p}'$ the conjugate (1.13) of $\boldsymbol{p}$). The latter can be simply derived from Young's inequality, noticing that

$$g*h(z)\omega_\ell(z) = \int \frac{\omega_\ell(z)}{\omega_\ell(z-\overline{z})\omega_\ell(\overline{z})}(g\omega_\ell)(z-\overline{z})(h\omega_\ell)(\overline{z})\,\mathrm{d}\overline{z},$$

and observing that, by (A.3), $\frac{\omega_\ell(z)}{\omega_\ell(z-\overline{z})\omega_\ell(\overline{z})}\lesssim 1$, for all $z,\overline{z}$.

The restriction $\boldsymbol{p}=(p,p)$ in [19] was only made for simplicity and replicating the initial proof steps in [19], one may extend (A.5) from a common integrability index to a mixed one. For the present paper, we shall restrict this extension on establishing the following estimate.

Lemma A.1. *For any $f\in\mathcal{S}'(\mathbb{R}^{2d})$ such that $\omega_\ell f\in\mathbf{B}^{\beta}_{\boldsymbol{p};\boldsymbol{a}}$ with $\beta\geqslant 0$, $\boldsymbol{p}\in[1,\infty]^2$, we have: for any $j\geqslant 0$,*

$$\|(\mathcal{R}_j^{\boldsymbol{a}}f)\omega_\ell\|_{\boldsymbol{p}} \lesssim 2^{-\beta j}\Big(\|\omega_\ell f\|_{\boldsymbol{p}} + \|\omega_\ell f\|_{\mathbf{B}^{\beta}_{\boldsymbol{p};\boldsymbol{a}}}\Big). \tag{A.7}$$

In particular, for $\beta=0$, $\|(\mathcal{R}_j^{\boldsymbol{a}}f)\omega_\ell\|_{\boldsymbol{p}}\lesssim\|\omega_\ell f\|_{\boldsymbol{p}}$.

Proof. As in [19], the proof essentially consists in establishing that

$$\|(\mathcal{R}_j^{\boldsymbol{a}}f)\omega_\ell\|_{\boldsymbol{p}} \lesssim 2^{-\beta j}\left(\|\omega_\ell f\|_{\boldsymbol{p}} + \sup_{|h|_{\boldsymbol{a}}\leqslant 1}\frac{\|(\delta_h^{([\beta]+1)}f)\omega_\ell\|_{\boldsymbol{p}}}{|h|_{\boldsymbol{a}}^{\beta}}\right). \tag{A.8}$$

Indeed, this estimate gives naturally the claim when $\beta=0$. For $\beta>0$, recalling the Leibniz rule: for any integer $m\geqslant 1$,

$$\delta_h^{(m)}(f\omega_\ell) = \sum_{n=0}^{m}\frac{m!}{n!(m-n)!}\big(\delta_h^{(n)}f\big)\big(\delta_h^{(m-n)}\omega_\ell(\cdot+nh)\big)$$

the quantity $\|(\delta_h^{[\beta]+1}f)\omega_\ell\|_{\boldsymbol{p}}$ is readily bounded by $\|\delta_h^{([\beta]+1)}(f\omega_\ell)\|_{\boldsymbol{p}}$ and $\|(\delta_h^{(k)}f)(\delta_h^{(n)}\omega_\ell(\cdot+nh))\|_{\boldsymbol{p}}$ for $0\leqslant k<m$ and $0\leqslant n\leqslant m$. Further, due to the polynomial form (A.1), by Taylor expansions, one can check that: for all $h\in\mathbb{R}^d$ such that $|h|_{\boldsymbol{a}}\leqslant 1$,

$$|\delta_h^{(n)}\omega_\ell(z+nh)|\lesssim|h|_{\boldsymbol{a}}^n\omega_\ell(z)\omega_\ell(nh),$$

and further

$$\|(\delta_h^{(k)}f)(\delta_h^{(n)}\omega_\ell(\cdot+nh))\|_{\boldsymbol{p}}\lesssim|h|_{\boldsymbol{a}}^n\|(\delta_h^{(k)}f)\omega_\ell\|_{\boldsymbol{p}}.$$

By iteration, it thus follows that

$$\sup_{|h|_{\boldsymbol{a}}\leqslant 1}\frac{\|(\delta_h^{([\beta]+1)}f)\omega_\ell\|_{\boldsymbol{p}}}{|h|_{\boldsymbol{a}}^{\beta}}\lesssim\|\omega_\ell f\|_{\boldsymbol{p}} + \sup_{|h|_{\boldsymbol{a}}\leqslant 1}\frac{\|\delta_h^{([\beta]+1)}(f\omega_\ell)\|_{\boldsymbol{p}}}{|h|_{\boldsymbol{a}}^{\beta}}.$$

The claim (A.7) is finally derived by the equivalent norm (2.9) with $q=\infty$ and $s=\beta>0$.
For the proof of (A.8), recall that, for any integer $m\geqslant 1$ and $h\in\mathbb{R}^{2d}$, the m^{th} difference operator writes as

$$\delta_h^{(m)}f(z) = \sum_{n=0}^{m}(-1)^{m-n}C_m^n f(z+nh).$$

For fixed $m \geqslant 1$, we shall so define the class $\{\phi_k^{m,\boldsymbol{a}}\}_{k\geqslant 0}$ as

$$\phi_k^{m,\boldsymbol{a}}(\xi) = (-1)^m \sum_{n=1}^{m} (-1)^{m-n} C_m^n \phi_k^{\boldsymbol{a}}(n\xi), \quad \xi \in \mathbb{R}^{2d}.$$

Given that, for any $n \in \mathbb{N} \setminus \{0\}$, $\{\phi_j^{\boldsymbol{a}}(n\cdot)\}_{j\geqslant 0}$ is a partition of unity, $\{\phi_k^{m,\boldsymbol{a}}\}_{k\geqslant 0}$ defines itself a partition of unity with

$$\sum_{k\geqslant 0} \phi_k^{m,\boldsymbol{a}}(\xi) = (-1)^m \sum_{n=1}^{m} (-1)^{m-n} C_m^n = 1.$$

Hence, for all $j \geqslant 0$, $\mathcal{R}_j^{\boldsymbol{a}} f = \sum_{k\geqslant 0} \check{\phi}_j^{\boldsymbol{a}} * \check{\phi}_k^{m,\boldsymbol{a}} * f$. Next, since $(\phi_k^{\boldsymbol{a}}(n\cdot))\check{}(z) = n^{-2d}\check{\phi}_k^{\boldsymbol{a}}(n^{-1}z)$ and $n^{-2d}\check{\phi}_j^{\boldsymbol{a}} * \check{\phi}_k^{\boldsymbol{a}}(n^{-1}\cdot) = (\phi_j^{\boldsymbol{a}}\phi_k^{\boldsymbol{a}}(n\cdot))\check{} = 0$ for all $|k-j| > n+2$, from the above, we get

$$\mathcal{R}_j^{\boldsymbol{a}} f(z) = \sum_{k\geqslant 0} \check{\phi}_j^{\boldsymbol{a}} * \check{\phi}_k^{m,\boldsymbol{a}} * f(z) = \sum_{k\geqslant 0, |k-j|\leqslant m+2} \check{\phi}_j^{\boldsymbol{a}} * \Big((-1)^m \sum_{n=1}^{m} (-1)^{m-n} C_m^n (\phi_k^{\boldsymbol{a}}(n\cdot))\check{} * f\Big)(z).$$

Moreover by change of variables and as $\int \check{\phi}_k^{\boldsymbol{a}}(\bar{z})\mathrm{d}\bar{z} = 0$,

$$(-1)^m \sum_{n=1}^{m} (-1)^{m-n} C_m^n n^{-2d} \check{\phi}_k^{\boldsymbol{a}}(n^{-1}\cdot) * f(z) = (-1)^m \sum_{n=1}^{m} (-1)^{m-n} C_m^n \int \check{\phi}_k^{\boldsymbol{a}}(-\bar{z}) f(z + n\bar{z})\mathrm{d}\bar{z}$$
$$= (-1)^m \sum_{n=0}^{m} (-1)^{m-n} C_m^n \int \check{\phi}_k^{\boldsymbol{a}}(-\bar{z}) f(z + n\bar{z})\mathrm{d}\bar{z} = (-1)^m \int \check{\phi}_k^{\boldsymbol{a}}(-\bar{z}) \delta_{\bar{z}}^{(m)} f(z) d\bar{z}.$$

Combining all the above,

$$\mathcal{R}_j^{\boldsymbol{a}} f(z) = (-1)^m \sum_{k\geqslant 0, |k-j|\leqslant m+2} \check{\phi}_j^{\boldsymbol{a}} * \int \check{\phi}_k^{\boldsymbol{a}}(-\bar{z}) \delta_{\bar{z}}^{(m)} f(z) d\bar{z} =: \sum_{k\geqslant 0, |k-j|\leqslant m+2} \check{\phi}_j^{\boldsymbol{a}} * \tilde{\mathcal{R}}_k^{\boldsymbol{a}} f(z).$$

Applying the weighted Young inequality (A.6), it follows that

$$\|(\mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}} \lesssim \sum_{k\geqslant 0, |k-j|\leqslant m+2} \|\omega_\ell \check{\phi}_j^{\boldsymbol{a}}\|_{\boldsymbol{1}} \|(\tilde{\mathcal{R}}_k^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}}.$$

As $\phi_j^{\boldsymbol{a}}$ is a Schwartz function, $\check{\phi}_j^{\boldsymbol{a}}$ is itself a Schwartz function, and according to (1.18), we have $\sup_j \|\omega_\ell \check{\phi}_j^{\boldsymbol{a}}\|_{\boldsymbol{1}} < \infty$. Thus,

$$\|(\tilde{\mathcal{R}}_k^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}} = \|\int \check{\phi}_k^{\boldsymbol{a}}(-\bar{z})\big(\omega_\ell \delta_{\bar{z}}^{(m)} f\big)(\cdot)\,\mathrm{d}\bar{z}\|_{\boldsymbol{p}} \leqslant \int |\check{\phi}_k^{\boldsymbol{a}}(-\bar{z})| \|\omega_\ell \delta_{\bar{z}}^{(m)} f\|_{\boldsymbol{p}}\,\mathrm{d}\bar{z}$$
$$\leqslant \int_{|\bar{z}|_{\boldsymbol{a}}\leqslant 1} |\check{\phi}_k^{\boldsymbol{a}}(-\bar{z})| \|\omega_\ell \delta_{\bar{z}}^{(m)} f\|_{\boldsymbol{p}}\,\mathrm{d}\bar{z} + \int_{|\bar{z}|_{\boldsymbol{a}}>1} |\check{\phi}_k^{\boldsymbol{a}}(-\bar{z})| \|\omega_\ell \delta_{\bar{z}}^{(m)} f\|_{\boldsymbol{p}}\,\mathrm{d}\bar{z}.$$

Taking now $m = [\beta]+1$, and as $\||\check{\phi}_k^{\boldsymbol{a}}| \cdot |_{\boldsymbol{a}}^{\beta}\|_{\boldsymbol{1}} \lesssim 2^{-\beta j}$

$$\int_{|z|_{\boldsymbol{a}}\leqslant 1} |\check{\phi}_k^{\boldsymbol{a}}(-\bar{z})| \|\omega_\ell \delta_{\bar{z}}^{(m)} f\|_{\boldsymbol{p}}\,\mathrm{d}\bar{z} \lesssim 2^{-\beta j} \sup_{|\bar{z}|_{\boldsymbol{a}}\leqslant 1} \frac{\|(\delta_{\bar{z}}^{([\beta]+1)} f)\omega_\ell\|_{\boldsymbol{p}}}{|\bar{z}|_{\boldsymbol{a}}^{\beta}}.$$

Meanwhile, since, for all $z, h \in \mathbb{R}^{2d}$, (A.2) and (A.3) ensure that $|\delta_h^{(1)}\omega(z)| \lesssim |h|_{\boldsymbol{a}}\omega_\ell(h)\omega_\ell(z)$ for all $h \in \mathbb{R}^{2d}$, by change of variables,

$$\|(\delta_{\bar{z}}^{(1)} f)\omega_\ell\|_{\boldsymbol{p}} = \|(\delta_{-\bar{z}}^{(1)}\omega_\ell) f\|_{\boldsymbol{p}} \lesssim |\bar{z}|_{\boldsymbol{a}}\omega_\ell(\bar{z})\|\omega_\ell f\|_{\boldsymbol{p}}.$$

Iterating this estimate, it follows that

$$\int_{|z|_{\boldsymbol{a}}>1} |\check{\phi}_k^{\boldsymbol{a}}(-\bar{z})| \|\omega_\ell \delta_{\bar{z}}^{(m)} f\|_{\boldsymbol{p}}\,\mathrm{d}\bar{z} \lesssim \Big(\int_{|z|_{\boldsymbol{a}}>1} |\check{\phi}_k^{\boldsymbol{a}}(-\bar{z})| |\bar{z}|_{\boldsymbol{a}}^m \omega_\ell(\bar{z})^m\,\mathrm{d}\bar{z}\Big)\|\omega_\ell f\|_{\boldsymbol{p}} \lesssim 2^{-j\beta}\|\omega_\ell f\|_{\boldsymbol{p}}.$$

Hence, we can conclude that

$$\|(\mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}} \lesssim 2^{-\beta j}\Big(\sup_{|\bar{z}|_{\boldsymbol{a}}\leqslant 1} \frac{\|(\delta_{\bar{z}}^{([\beta]+1)} f)\omega_\ell\|_{\boldsymbol{p}}}{|\bar{z}|_{\boldsymbol{a}}} + \|\omega_\ell f\|_{\boldsymbol{p}}\Big),$$

and so conclude (A.8). □

We next establish a weighted version of the estimate (2.15) in Lemma 2.5 for the kinetic semigroup P_t in the case $\alpha = 2$. To this aim, let us first establish a weighted version of the Bernstein inequality in Lemma 2.2 (we refer to [19, Lemma 2.4] for the special case $\boldsymbol{p} = (p,p)$).

Lemma A.2. *For any positive integers $\boldsymbol{k} = (k_1, k_2)$, $\boldsymbol{p}, \boldsymbol{p}_1 \in [1,\infty]$ with $\boldsymbol{p}_1 \leqslant \boldsymbol{p}$, and for $\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}}$ defined as in (1.12), there is a constant $C = C(\boldsymbol{k}, \boldsymbol{p}, \boldsymbol{p}_1, d, \ell) > 0$ such that for all $j \geqslant 0$, $f \in \mathscr{S}'$,*

$$\|(\nabla_x^{k_1}\nabla_v^{k_2}\mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}} \lesssim 2^{j\left((1+\alpha)k_1+k_2+\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}}\right)}\|(\mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}_1}. \tag{A.9}$$

Proof. Using (again) the partition of unity defined by $\{\phi_k^{\boldsymbol{a}}\}_{k\geqslant 0}$ and since, according to (1.15), $\mathrm{supp}(\phi_j^{\boldsymbol{a}}\phi_k^{\boldsymbol{a}}) = 0$ whenever $|k-j| > 2$, we have

$$\nabla_x^{k_1}\nabla_v^{k_2}\mathcal{R}_j^{\boldsymbol{a}} f = \sum_{k\geqslant 0, |k-j|\leqslant 2} \check{\phi}_k^{\boldsymbol{a}} * \nabla_x^{k_1}\nabla_v^{k_2}\mathcal{R}_j^{\boldsymbol{a}} f = \sum_{k\geqslant 0, |k-j|\leqslant 2} \nabla_x^{k_1}\nabla_v^{k_2}\check{\phi}_k^{\boldsymbol{a}} * \mathcal{R}_j^{\boldsymbol{a}} f.$$

Applying the weighted Young inequality (A.6), it follows that

$$\begin{aligned}\|(\nabla_x^{k_1}\nabla_v^{k_2}\mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}} &\leqslant \sum_{k\geqslant 0, |k-j|\leqslant 2} \|(\nabla_x^{k_1}\nabla_v^{k_2}\check{\phi}_k^{\boldsymbol{a}} * \mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}} \\ &\lesssim \sum_{k\geqslant 0, |k-j|\leqslant 2} \|(\mathcal{R}_j^{\boldsymbol{a}} f)\omega_\ell\|_{\boldsymbol{p}_1}\|(\nabla_x^{k_1}\nabla_v^{k_2}\check{\phi}_k^{\boldsymbol{a}})\omega_\ell\|_{\boldsymbol{p}_2},\end{aligned}$$

where $\frac{1}{\boldsymbol{p}'} = \frac{1}{\boldsymbol{p}_1} + \frac{1}{\boldsymbol{p}_2}$ Since each $\check{\phi}_k^{\boldsymbol{a}}$ is a Schwartz function, $\|(\nabla_x^{k_1}\nabla_v^{k_2}\check{\phi}_0^{\boldsymbol{a}})\omega_\ell\|_{\boldsymbol{p}_2} < \infty$. For $k \geqslant 1$, owing to the scaling property (1.18) and by change of variables, we get

$$\begin{aligned}\|\check{\phi}_k^{\boldsymbol{a}}\omega_\ell\|_{\boldsymbol{p}_2} &= 2^{(k-1)\left((1+\alpha)k_1+k_2+(2+\alpha)d\right)}\|(\nabla_x^{k_1}\nabla_v^{k_2}\check{\phi}_1^{\boldsymbol{a}})(2^{(k-1)\boldsymbol{a}}\cdot)\omega_\ell\|_{\boldsymbol{p}_2} \\ &= 2^{(k-1)\left((1+\alpha)k_1+k_2+\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_2}\right)}\|(\nabla_x^{k_1}\nabla_v^{k_2}\check{\phi}_1^{\boldsymbol{a}})\omega_\ell(2^{-(k-1)\boldsymbol{a}}\cdot)\|_{\boldsymbol{p}_2}.\end{aligned}$$

Since $\omega_\ell(2^{-(k-1)\boldsymbol{a}}\cdot) \lesssim_{C(\ell)} \omega_\ell$ and $\mathcal{A}_{\boldsymbol{1},\boldsymbol{p}_2} = \mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}}$, this gives the claim. □

Owing to this preliminary, we have the following heat kernel estimate:

Lemma A.3. *Let P_t be defined as in (2.14) for $\alpha = 2$. For any $\boldsymbol{p}, \boldsymbol{p}_1 \in [1,\infty]^2$ with $\boldsymbol{p}_1 \leqslant \boldsymbol{p}$, $\beta > 0$, $\gamma > 0$ and $0 \leqslant T < \infty$, there is a constant $C = C(d, \beta, \gamma, \boldsymbol{p}, \boldsymbol{p}_1, \ell, T) > 0$ such that for any $f \in \mathbf{B}^{\beta}_{\boldsymbol{p}_1;\boldsymbol{a}}(\omega_\ell)$, $0 < t \leqslant T$, and $j \geqslant 0$,*

$$\|(\mathcal{R}_j^{\boldsymbol{a}} P_t\nabla_v f)\omega_\ell\|_{\boldsymbol{p}} \lesssim_C 2^{j(\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}}+1-\beta)}\big((2^{2j}t)^{-\gamma}\wedge 1\big)\|\omega_\ell f\|_{\mathbf{B}^{\beta}_{\boldsymbol{p}_1;\boldsymbol{a}}}, \tag{A.10}$$

where $\mathcal{A}_{\boldsymbol{p}_1,\boldsymbol{p}}$ is defined as in (1.12), with $\alpha = 2$ and $\boldsymbol{a} = (3,1)$.

Proof. The proof essentially consists in replicating the original arguments of the items (*i*) and (*ii*) in [17, Lemma 2.12], the auxiliary Lemma 2.11 therein and Lemma 3.1 in [44].

For the case $j = 0$, starting from (2.14) - with p_t the density function as in (2.13) -, by direct calculations, applying twice the weighted Young inequality (A.6) and using the invariance of $\mathbb{L}^{\boldsymbol{p}}$ by Γ_t, it follows that

$$\begin{aligned}\|(\mathcal{R}_0^{\boldsymbol{a}} P_t\nabla_v f)\omega_\ell\|_{\boldsymbol{p}} &= \|\big(\check{\phi}_0^{\boldsymbol{a}} * \Gamma_t(p_t * \nabla_v f)\big)\omega_\ell\|_{\boldsymbol{p}} = \|\big(\Gamma_{-t}\check{\phi}_0^{\boldsymbol{a}} * (p_t * \nabla_v f)\big)\Gamma_{-t}\omega_\ell\|_{\boldsymbol{p}_1} \\ &\lesssim \|\big(\Gamma_{-t}\check{\phi}_0^{\boldsymbol{a}} * (p_t * \nabla_v f)\big)\omega_\ell\|_{\boldsymbol{p}} = \|\big(\nabla_v\Gamma_{-t}\check{\phi}_0^{\boldsymbol{a}} * (p_t * f)\big)\omega_\ell\|_{\boldsymbol{p}} \\ &\lesssim \|(\nabla_v\Gamma_{-t}\check{\phi}_0^{\boldsymbol{a}})\omega_\ell\|_{\boldsymbol{p}_2}\|(p_t * f)\omega_\ell\|_{\boldsymbol{p}_1} \\ &\lesssim \|(\nabla_v\Gamma_{-t}\check{\phi}_0^{\boldsymbol{a}})\omega_\ell\|_{\boldsymbol{p}_2}\|p_t\omega_\ell\|_{\boldsymbol{1}}\|f\omega_\ell\|_{\boldsymbol{p}_1},\end{aligned}$$

for $\boldsymbol{p}_2 \in [1,\infty]^2$ such that $1/\boldsymbol{p}_2 = 1/\boldsymbol{p}_1 + 1/\boldsymbol{p}'$. Owing to the scaling properties of p_t (with $\alpha = 2$) and the polynomial form of ω_ℓ, $\sup_{t>0} \|p_t\omega_\ell\|_{\boldsymbol{1}} < \infty$. Since $\nabla_v \Gamma_{-t}\check{\phi}^{\boldsymbol{a}}_j = \Gamma_{-t}\nabla_v\check{\phi}^{\boldsymbol{a}}_j + t\Gamma_{-t}\nabla_x\check{\phi}^{\boldsymbol{a}}_j$, $|\Gamma_t\nabla_v\omega_\ell| \lesssim_{C(T)} \omega_\ell$, and since $\check{\phi}^{\boldsymbol{a}}_0$ is a Schwartz function, we have

$$\|\nabla_v\Gamma_{-t}\check{\phi}^{\boldsymbol{a}}_0\omega_\ell\|_{\boldsymbol{p}_2} \lesssim_{C(T)} \|\nabla_v\check{\phi}^{\boldsymbol{a}}_0\omega_\ell\|_{\boldsymbol{p}_2} + \|\nabla_x\check{\phi}^{\boldsymbol{a}}_0\omega_\ell\|_{\boldsymbol{p}_2} < \infty.$$

Therefore, as $\beta > 0$, the embedding (2.10) yields $\mathbf{B}^\beta_{\boldsymbol{p}_1;\boldsymbol{a}} \hookrightarrow \mathbb{L}^{\boldsymbol{p}_1}$ and, for all $\gamma > 0$,

$$\|\big(\mathcal{R}^{\boldsymbol{a}}_0 P_t f\big)\omega_\ell\|_{\boldsymbol{p}} \lesssim_C \|\omega_\ell f\|_{\boldsymbol{p}_1} \lesssim_{C(T)} \big(t^{-\gamma}\wedge 1\big)\|\omega_\ell f\|_{\mathbf{B}^\beta_{\boldsymbol{p}_1;\boldsymbol{a}}}.$$

For the general case $j \geqslant 1$, as the $\phi^{\boldsymbol{a}}_j$'s define a partition of unity, $\sum_{k\geqslant 0}\mathcal{R}^{\boldsymbol{a}}_k f = f$ and thus

$$\begin{aligned}\mathcal{R}^{\boldsymbol{a}}_j P_t\nabla_v f &= \sum_{k\geqslant 0}\mathcal{R}^{\boldsymbol{a}}_j\Gamma_t p_t * \Gamma_t\mathcal{R}^{\boldsymbol{a}}_k\nabla_v f = \sum_{k\geqslant 0}\check{\phi}^{\boldsymbol{a}}_j * \Gamma_t p_t * \Gamma_t\check{\phi}^{\boldsymbol{a}}_k * \Gamma_t\nabla_v f \\ &= \sum_{k\geqslant 0}\check{\phi}^{\boldsymbol{a}}_j * \Gamma_t\check{\phi}^{\boldsymbol{a}}_k * \Gamma_t p_t * \Gamma_t\nabla_v f.\end{aligned} \tag{A.11}$$

Since $\Gamma_t\check{\phi}^{\boldsymbol{a}}_k = (\phi^{\boldsymbol{a}}(\xi_1,\xi_2 - t\xi_1))\check{} = (\tilde{\Gamma}_t\phi^{\boldsymbol{a}}_k)\check{}$, we can write $\check{\phi}^{\boldsymbol{a}}_j * \Gamma_t\check{\phi}^{\boldsymbol{a}}_k =: (\phi^{\boldsymbol{a}}_j\tilde{\Gamma}_t\phi^{\boldsymbol{a}}_k)\check{}$. Observe next that (with the convention $2^{k-1} = 0$ if $k = 0$) the support of $\phi^{\boldsymbol{a}}_j\tilde{\Gamma}_t\phi^{\boldsymbol{a}}_k$ is given by

$$\mathrm{supp}\big(\phi^{\boldsymbol{a}}_j\tilde{\Gamma}_t\phi^{\boldsymbol{a}}_k\big) = \Big\{\xi = (\xi_1,\xi_2) : 2^{j-1} \leqslant |\xi_1|^{1/3} + |\xi_2| \leqslant 2^{j+1},\ 2^{k-1} \leqslant |\xi_1|^{1/3} + |\xi_2 - t\xi_1| \leqslant 2^{k+1}\Big\}.$$

By triangular inequality, each element of this set has to satisfy: $|\xi_1|^{1/3} \leqslant 2^{j+1}\wedge 2^{k+1} = 2^{j\wedge k+1}$,

$$\begin{aligned}&(2^{k-1} - t|\xi_1|)\vee 2^{j-1} \leqslant |\xi_1|^{1/3} + |\xi_2| \leqslant 2^{j+1}\wedge(2^{k+1} + t|\xi_1|) \leqslant 2^{j+1}\wedge(2^{k+1} + t2^{3(k+1)}),\\ &(2^{j-1} - t|\xi_1|)\vee 2^{k-1} \leqslant |\xi_1|^{1/3} + |\xi_2 - t\xi_1| \leqslant 2^{k+1}\wedge(2^{j+1} + t|\xi_1|) \leqslant 2^{k+1}\wedge(2^{j+1} + t2^{3(j+1)}).\end{aligned}$$

In view of these delimiters, given $j \geqslant 0$ and $t \in [0,T]$, for all k outside the family of indices

$$\Theta^t_j := \{k \in \mathbb{N} : 2^{k-1} \leqslant 2^{j+1} + t2^{3(j+1)} \text{ and } 2^{j-1} \leqslant 2^{k+1} + t2^{3(k+1)}\},$$

the support set $\mathrm{supp}\big(\phi^{\boldsymbol{a}}_j\tilde{\Gamma}_t\phi^{\boldsymbol{a}}_k\big)$ is empty and $\phi^{\boldsymbol{a}}_j\tilde{\Gamma}_t\phi^{\boldsymbol{a}}_k = 0$. Additionally, for all $k \in \Theta^t_j$,

$$2^{k-1} \leqslant \big(2^{j+1} + t2^{3(j+1)}\big) \leqslant 2^{j+1}\big(1 + t2^{2(j+1)}\big),$$

and so

$$k \leqslant j + 2 + (\ln(2))^{-1}\ln\big(1 + t2^{2(j+1)}\big) =: j + 2 + \Theta_+.$$

From this upper-bound, we further deduce a lower-bound on Θ^t_j, observing that

$$2^{k+1} \geqslant 2^{j-1}\big(1 + t2^{2(k+1)}\big)^{-1} \geqslant 2^{j-1}\big(1 + t2^{2(j+3+\Theta_+)}\big)^{-1}$$

and so

$$k \geqslant \Big(j - 2 - (\ln(2))^{-1}\ln\big(1 + t2^{2(j+2+\Theta_+)}\big)\Big)\vee 0 =: \big(j - 2 - \Theta_-\big)\vee 0.$$

Therefore, as $\beta > 0$, $2^{\beta\Theta_-} = \big(1 + t2^{j+1+\Theta_+}\big)^\beta$ and $2^{\Theta_+} = (1 + t2^{2(j+1)})^2$,

$$\sum_{k\in\Theta^t_j} 2^{-k\beta} = \frac{2^{-\beta(j-2-\Theta_-)\vee 0} - 2^{-\beta(j+2+\Theta_+)}}{1 - 2^{-\beta}} \lesssim_{C(\beta)} 2^{-\beta j}\big(1 + t2^{3(j+1)}\big)^\beta. \tag{A.12}$$

Coming back to (A.11), it follows that

$$\mathcal{R}^{\boldsymbol{a}}_j P_t f = \sum_{k\in\Theta^t_j}\check{\phi}^{\boldsymbol{a}}_j * \Gamma_t\check{\phi}^{\boldsymbol{a}}_k * \Gamma_t p_t * \Gamma_t\nabla_v f = \sum_{k\in\Theta^t_j}\mathcal{R}^{\boldsymbol{a}}_j\Gamma_t p_t * \Gamma_t\mathcal{R}^{\boldsymbol{a}}_k\nabla_v f.$$

Now, using again the weighted Young inequality (A.6), we get

$$\|\big(\mathcal{R}^{\boldsymbol{a}}_j P_t\nabla_v f\big)\omega_\ell\|_{\boldsymbol{p}_1} \lesssim \sum_{k\in\Theta^t_j}\|\big((\mathcal{R}^{\boldsymbol{a}}_j\Gamma_t p_t) * (\Gamma_t\mathcal{R}^{\boldsymbol{a}}_k f)\big)\omega_\ell\|_{\boldsymbol{p}_1} \lesssim \sum_{k\in\Theta^t_j}\|(\mathcal{R}^{\boldsymbol{a}}_j\Gamma_t p_t)\omega_\ell\|_{\boldsymbol{1}}\|(\Gamma_t\mathcal{R}^{\boldsymbol{a}}_k\nabla_v f)\omega_\ell\|_{\boldsymbol{p}_1}.$$

Since, for any $z \in \mathbb{R}^{2d}$, $|\Gamma_{-t}\omega_\ell(z)| \lesssim_{C(T,\ell)} \omega_\ell(z)$, applying successively the weighted Bernstein inequality (A.9) (with $\boldsymbol{k} = (0,1)$, $\boldsymbol{p} = \boldsymbol{p}_1$) and the estimate (A.7) in Lemma A.1 yields

$$\|(\Gamma_t \mathcal{R}^{\boldsymbol{a}}_k \nabla_v f)\omega_\ell\|_{\boldsymbol{p}_1} \lesssim 2^k \|(\mathcal{R}^{\boldsymbol{a}}_k f)\omega_\ell\|_{\boldsymbol{p}_1} \lesssim 2^{k(1-\beta)} \|f\omega_\ell\|_{\mathbf{B}^\beta_{\boldsymbol{p}_1;\boldsymbol{a}}}.$$

Finally, extending the proof arguments of Lemma 3.1 in [44] to a weighted setting, we may claim that

$$\|(\mathcal{R}^{\boldsymbol{a}}_j \Gamma_t p_t)\omega_\ell\|_{\mathbf{1}} \lesssim (2^{2j}t)^{-\gamma} \wedge 1. \tag{A.13}$$

From this estimate, and using (A.12) with $\gamma + 3\beta$ in place of γ, we obtain that

$$\sum_{k\in\Theta^t_j} \|(\mathcal{R}^{\boldsymbol{a}}_j \Gamma_t p_t)\omega_\ell\|_{\mathbf{1}} \|(\Gamma_t \mathcal{R}^{\boldsymbol{a}}_k f)\omega_\ell\|_{\boldsymbol{p}_1} \lesssim \Big((2^{2j}t)^{-\gamma-\beta} \wedge 1\Big)\|f\omega_\ell\|_{\mathbf{B}^\beta_{\boldsymbol{p}_1;\boldsymbol{a}}} \sum_{k\in\Theta^t_j} 2^{k(1-\beta)},$$

and so

$$\sum_{k\in\Theta^t_j} \|(\mathcal{R}^{\boldsymbol{a}}_j P_t \nabla_v f)\omega_\ell\|_{\mathbf{1}} \lesssim 2^{j(1-\beta)}\Big((2^{2j}t)^{-\gamma-3\beta} \wedge 1\Big)(1 + t2^{3j})^\beta \|f\omega_\ell\|_{\mathbf{B}^\beta_{\boldsymbol{p}_1;\boldsymbol{a}}}.$$

This gives (A.10).

The proof of the estimate (A.13) proceeds as follows: recalling (1.18) with $\alpha = 2$, $\phi^{\boldsymbol{a}}_j(\xi) = 2^{4(j-1)d}\phi^{\boldsymbol{a}}_1(2^{(j-1)\boldsymbol{a}}\xi)$ (for $2^{s\boldsymbol{a}}\xi = (2^3\xi_1, 2\xi_2)$). Using (again) the scaling properties of p_t, by change of variables, we have, for $\hbar := (2^j\sqrt{t})^{-1}$, and $2^{-\boldsymbol{a}}\overline{z} = (2^{-3}\overline{z}_1, 2^{-1}\overline{z_2})$,

$$\begin{aligned}\|(\mathcal{R}^{\boldsymbol{a}}_j \Gamma_t p_t)\omega_\ell\|_{\mathbf{1}} &= 2^{-4d}\int \omega_\ell(z)\Big|\int \check{\phi}^{\boldsymbol{a}}_1(2^{-\boldsymbol{a}}\overline{z})p_1(x - \hbar^3\overline{x} + v - \hbar\overline{v}, v - \hbar\overline{v})\mathrm{d}\overline{z}\Big|\mathrm{d}z\\ &=: \int \omega_\ell(z)\Big|\int \check{\phi}^{\boldsymbol{a}}_1(2^{-\boldsymbol{a}}\overline{z})H(z - \Gamma_\hbar \hbar^{\boldsymbol{a}}\overline{z})\mathrm{d}\overline{z}\Big|\mathrm{d}z.\end{aligned}$$

For $m \in \mathbb{N}$, defining next the operator $\triangle^m_z$ and its inverse $\triangle^{-m}_\xi$, as

$$\triangle^m_z = \big(\triangle_x + \triangle_v\big)^m, \quad (\triangle^{-m}_\xi f)\hat{}(\xi_1,\xi_2) = \big(|\xi_1|^2 + |\xi_2|^2\big)^{-m}\hat{f}(\xi_1,\xi_2),$$

we get

$$\int \check{\phi}^{\boldsymbol{a}}_1(2^{-\boldsymbol{a}}\overline{z})H(z - \Gamma_\hbar \hbar^{\boldsymbol{a}})\mathrm{d}\overline{z} = \int \check{(}\triangle^{-m}_{\overline{z}}\phi^{\boldsymbol{a}}_1)(2^{\boldsymbol{a}}\overline{z})\triangle^m_{\overline{z}}H(z - \Gamma_\hbar \hbar^{\boldsymbol{a}})\mathrm{d}\overline{z}.$$

Using the growth property (A.3) of ω_ℓ, it follows that

$$\|(\mathcal{R}^{\boldsymbol{a}}_j \Gamma_t p_t)\omega_\ell(z)\|_{\mathbf{1}} \lesssim \|\big(\triangle^{-m}_{\overline{z}}\check{\phi}^{\boldsymbol{a}}_1(2^{\boldsymbol{a}}\overline{z})\big)\omega_\ell(\overline{z})\|_{\mathbf{1}} \sup_{\overline{z}} \|\big(\triangle^m_{\overline{z}}H(\cdot - \hbar\Gamma_\hbar\overline{z})\big)\omega_\ell(\cdot - \hbar\Gamma_\hbar\overline{z})\|_{\mathbf{1}}.$$

Owing to the regularity of p_1,

$$\|\triangle^m_{\overline{z}}H(\cdot - \hbar\Gamma_\hbar\overline{z})\omega_\ell(\cdot - \hbar\Gamma_\hbar\overline{z})\|_{\mathbf{1}} \lesssim \hbar^{2m},$$

and since $\check{\phi}^{\boldsymbol{a}}_1$ is a Schwartz function, $\|\triangle^{-m}_{\overline{z}}\check{\phi}^{\boldsymbol{a}}_1(2^{-\boldsymbol{a}}\overline{z})\omega_\ell(\overline{z})\|_{\mathbf{1}}$ is finite. Hence

$$\|\mathcal{R}^{\boldsymbol{a}}_j \Gamma_t p_t \omega_\ell\| \lesssim \hbar^{2m} = (2^{2j}t)^{-m}.$$

As m is arbitrary, (A.13) follows. □

Appendix B. Besov norm of scale functions

For $\phi \in \mathcal{S}(\mathbb{R}^{2d})$, we define a dilation operator as: for $\lambda > 0$,

$$\phi_\lambda(z) := \mathcal{D}_\lambda\phi(x,v) := \lambda^{(2+\alpha)d}\phi(\lambda^{1+\alpha}x, \lambda v).$$

Let $\boldsymbol{p}, \boldsymbol{p}' \in [1,\infty]^2$ with $\frac{1}{\boldsymbol{p}'} + \frac{1}{\boldsymbol{p}} = \mathbf{1}$. For any $\beta \geqslant 0$, by the scaling properties of the mixed norm $\|\cdot\|_{\boldsymbol{p}}$ defined in (1.9) and of the anisotropic distance (1.14), it is easy to see that, for all $\lambda \geqslant 0$,

$$\||\cdot|^\beta_{\boldsymbol{a}}\phi_\lambda(\cdot)\|_{\boldsymbol{p}} = \lambda^{\boldsymbol{a}\cdot d/\boldsymbol{p}'-\beta}\||\cdot|^\beta_{\boldsymbol{a}}\phi(\cdot)\|_{\boldsymbol{p}} = \lambda^{\mathcal{A}_{1,\boldsymbol{p}}-\beta}\||\cdot|^\beta_{\boldsymbol{a}}\phi(\cdot)\|_{\boldsymbol{p}}. \tag{B.1}$$

We first establish the following crucial lemma for later use.

Lemma B.1. *Let ψ be a smooth function so that $\hat{\psi} \in C_c^\infty(\mathbb{R}^{2d} \setminus \{0\})$ and set $\psi_\ell = \mathscr{D}_\ell \psi$. For any $\boldsymbol{p} \in [1, \infty]^2$ and $\beta \geqslant 0$, there is a constant $C = C(\beta, d, \boldsymbol{p}, \psi, \phi) > 0$ such that for all $\lambda, \ell \geqslant 0$ and $t \in \mathbb{R}$,*

$$\|\psi_\ell * \Gamma_t \phi_\lambda\|_{\boldsymbol{p}} \lesssim \lambda^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} (1 + |t|\lambda^\alpha)^\beta (\lambda/\ell)^\beta. \tag{B.2}$$

Proof. Set $\hbar := \lambda/\ell$. First of all, by the very definition of $\mathscr{D}_\lambda$ and changes of variables, it is easy to see that

$$\psi_\ell * (\Gamma_t \phi_\lambda) = \mathscr{D}_\ell\big(\mathscr{D}_{\ell^{-1}}(\psi_\ell * (\Gamma_t \mathscr{D}_\lambda \phi))\big) = \mathscr{D}_\ell(\psi * \mathscr{D}_{\ell^{-1}}(\Gamma_t \mathscr{D}_\lambda \phi)) = \mathscr{D}_\ell(\psi * \mathscr{D}_\hbar(\mathscr{D}_{\lambda^{-1}} \Gamma_{t\lambda^\alpha} \phi)),$$

the last equality following from the property $\Gamma_t \mathscr{D}_\lambda \phi(x, v) = \mathscr{D}_\lambda \Gamma_{\lambda^\alpha t} \phi(x, v)$. By the embedding (2.10), the Young inequality (2.11), (B.1) and since $\|\Gamma_t f\|_{\boldsymbol{p}} = \|f\|_{\boldsymbol{p}}$, we have

$$\|\psi_\ell * \Gamma_t \phi_\lambda\|_{\boldsymbol{p}} = \ell^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} \|\psi * \mathscr{D}_\hbar(\mathscr{D}_{\lambda^{-1}} \Gamma_{t\lambda^\alpha} \phi)\|_{\boldsymbol{p}} \lesssim \ell^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} \|\psi\|_{\mathbf{B}^{0,1}_{\mathbf{1};\boldsymbol{a}}} \|\mathscr{D}_\hbar \Gamma_{t\lambda^\alpha} \phi\|_{\boldsymbol{p}} = \lambda^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} \|\psi\|_1 \|\phi\|_{\boldsymbol{p}}, \tag{B.3}$$

which gives (B.2) for $\beta = 0$. Next we consider the case $\beta > 0$. Introducing the Laplacian operator, $\triangle := \triangle_x + \triangle_v$, for any $m \in \mathbb{N}$, since $\mathrm{supp}(\hat{\psi}) \subset \mathbb{R}^{2d} \setminus \{0\}$, $\Delta^{-m}\psi$ is a Schwartz function, we further have

$$\|\psi_\ell * \Gamma_t \phi_\lambda\|_{\boldsymbol{p}} = \ell^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} \|\Delta^{-m}\psi * \Delta^m \mathscr{D}_\hbar \Gamma_{t\lambda^\alpha} \phi\|_{\boldsymbol{p}} \leqslant \ell^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} \|\Delta^{-m}\psi\|_{\mathbf{B}^{0,1}_{\mathbf{1};\boldsymbol{a}}} \|\nabla^{2m} \mathscr{D}_\hbar \Gamma_{t\lambda^\alpha} \phi\|_{\boldsymbol{p}}, \tag{B.4}$$

where $\nabla = (\nabla_x, \nabla_v)$. Noting that by the chain rule,

$$|\nabla^{2m} \mathscr{D}_\hbar \Gamma_{t\lambda^\alpha} \phi| \leqslant \sum_{k=0}^{2m} |\nabla_x^{2m-k} \nabla_v^k \mathscr{D}_\hbar \Gamma_{t\lambda^\alpha} \phi| = \sum_{k=0}^{2m} \hbar^{(2m-k)(1+\alpha)+k} |\mathscr{D}_\hbar \nabla_x^{2m-k} \nabla_v^k \Gamma_{t\lambda^\alpha} \phi|,$$

by (B.1), we have

$$\|\nabla^{2m} \mathscr{D}_\hbar \Gamma_{t\lambda^\alpha} \phi\|_{\boldsymbol{p}} \leqslant \sum_{k=0}^{2m} \hbar^{2(1+\alpha)m - \alpha k} \hbar^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} \|\nabla_x^{2m-k} \nabla_v^k \Gamma_{t\lambda^\alpha} \phi\|_{\boldsymbol{p}}. \tag{B.5}$$

On the other hand, noting that since $\nabla_x \Gamma_t = \Gamma_t \nabla_x$ and $\nabla_v \Gamma_t = \Gamma_t \circ (-t\nabla_x + \nabla_v)$,

$$\nabla_x^{2m-k} \nabla_v^k \Gamma_{t\lambda^\alpha} \phi = \Gamma_{t\lambda^\alpha} \nabla_x^{2m-k} (-t\lambda^\alpha \nabla_x + \nabla_v)^k \phi = \Gamma_{t\lambda^\alpha} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} (-t\lambda^\alpha)^i \nabla_x^{2m-k+i} \nabla_v^{k-i} \phi.$$

As $\|\Gamma_t f\|_{\boldsymbol{p}} = \|f\|_{\boldsymbol{p}}$, we have, for any $k = 0, \cdots, 2m$,

$$\|\nabla_x^{2m-k} \nabla_v^k \Gamma_{t\lambda^\alpha} \phi\|_{\boldsymbol{p}} \leqslant \sum_{i=0}^{k} (|t|\lambda^\alpha)^i \|\nabla_x^{2m-k+i} \nabla_v^{k-i} \phi\|_{\boldsymbol{p}} \lesssim (1 + |t|\lambda^\alpha)^{2m}. \tag{B.6}$$

Hence, by (B.4)-(B.6), we have, for any $m \in \mathbb{N}$,

$$\begin{aligned} \|\psi_\ell * \Gamma_t \phi_\lambda\|_{\boldsymbol{p}} &\lesssim \ell^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} \sum_{k=0}^{2m} \hbar^{2(1+\alpha)m - \alpha k} \hbar^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} (1 + |t|\lambda^\alpha)^{2m} \\ &\lesssim \lambda^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} (1 + |t|\lambda^\alpha)^{2m} \sum_{k=0}^{2m} \hbar^{2(1+\alpha)m - \alpha k}. \end{aligned}$$

If $\hbar > 1$, then

$$\|\psi_\ell * \Gamma_t \phi_\lambda\|_{\boldsymbol{p}} \lesssim \lambda^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} (1 + |t|\lambda^\alpha)^{2m} \hbar^{2(1+\alpha)m} \leqslant \lambda^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} (1 + |t|\lambda^\alpha)^{2(1+\alpha)m} \hbar^{2(1+\alpha)m}.$$

If $\hbar \leqslant 1$, then

$$\|\psi_\ell * \Gamma_t \phi_\lambda\|_{\boldsymbol{p}} \lesssim \lambda^{\mathscr{A}_{\mathbf{1},\boldsymbol{p}}} (1 + |t|\lambda^\alpha)^{2m} \hbar^{2m}.$$

Thus by (B.3) and interpolation, we obtain (B.2). $\square$

Let us now recall the weight function previously defined in (A.1),

$$\omega_\ell(z) := \big(1 + (1+|x|^2)^{\frac{1}{1+\alpha}} + |v|^2\big)^{\ell/2}, \quad z=(x,v)\in\mathbb{R}^{2d}.$$

Observing that

$$|\omega_\ell \phi_\lambda| \leqslant \mathscr{D}_\lambda |\omega_\ell \phi|, \tag{B.7}$$

we have the following useful corollary.

Lemma B.2. *Let $\phi \in C_c^\infty(\mathbb{R}^{2d})$ and $\phi_\lambda := \mathscr{D}_\lambda \phi$. For any $T,\ell,\beta \geqslant 0$ and $\boldsymbol{p}\in[1,\infty]^2$, there is a constant $C = C(T,\ell,d,\boldsymbol{p},\beta,\phi) > 0$ such that for all $t\in[0,T]$ and $\lambda \geqslant 1$,*

$$\|\omega_\ell \Gamma_t \phi_\lambda\|_{\mathbf{B}^\beta_{\boldsymbol{p};\boldsymbol{a}}} \lesssim_C \lambda^{(1+\alpha)\beta + \mathscr{A}_{1,\boldsymbol{p}}}. \tag{B.8}$$

In particular, for $\ell = 0$ (namely $\omega_\ell = 1$), we have

$$\|\Gamma_t \phi_\lambda\|_{\mathbf{B}^\beta_{\boldsymbol{p};\boldsymbol{a}}} \lesssim_C \lambda^{(1+\alpha)\beta + \mathscr{A}_{1,\boldsymbol{p}}}.$$

Proof. First of all we show (B.8) for $\ell = 0$. Let $\beta \geqslant 0$ and $\boldsymbol{p}\in[1,\infty]^2$. For $j\geqslant 1$, noting that

$$\mathscr{R}^{\boldsymbol{a}}_j \Gamma_t \phi_\lambda = \check{\phi}^{\boldsymbol{a}}_j * \Gamma_t \phi_\lambda.$$

According to (B.2) - in the preceding lemma - applied to $\psi_\ell = \mathscr{D}_\ell \psi = \check{\phi}^{\boldsymbol{a}}_j$ with, in view of (1.18), $\ell = 2^j$ and $\psi(z) = 2^{-1}\check{\phi}^{\boldsymbol{a}}_2(2^{-\boldsymbol{a}}z))$, there is a constant $C>0$ such that, for all $t\geqslant 0, \lambda>0$ and $j\geqslant 1$,

$$\|\mathscr{R}^{\boldsymbol{a}}_j \Gamma_t \phi_\lambda\|_{\boldsymbol{p}} \lesssim_C (1+t\lambda^\alpha)^\beta \lambda^{\boldsymbol{a}\cdot\frac{d}{\boldsymbol{p}'}} (\lambda/2^j)^\beta.$$

Hence, for any $t\geqslant 0$ and $\lambda > 0$, for the high frequency part of $2^{j\beta}\|\mathscr{R}^{\boldsymbol{a}}_j \Gamma_t\phi_\lambda\|_{\boldsymbol{p}}$, $j\geqslant 1$, we have

$$\sup_{j\geqslant 1}\Big(2^{j\beta}\|\mathscr{R}^{\boldsymbol{a}}_j \Gamma_t \phi_\lambda\|_{\boldsymbol{p}}\Big) \lesssim_C (1+t\lambda^\alpha)^\beta \lambda^{\boldsymbol{a}\cdot\frac{d}{\boldsymbol{p}'}+\beta}.$$

On the other hand, for the lower frequency part $j=0$, by (B.1), we have, for $\lambda\geqslant 1$,

$$\|\mathscr{R}^{\boldsymbol{a}}_0 \Gamma_t \phi_\lambda\|_{\boldsymbol{p}} \lesssim \|\Gamma_t\phi_\lambda\|_{\boldsymbol{p}} = \|\phi_\lambda\|_{\boldsymbol{p}} = \lambda^{\boldsymbol{a}\cdot\frac{d}{\boldsymbol{p}'}}\|\phi\|_{\boldsymbol{p}} \lesssim (1+t\lambda^\alpha)^\beta \lambda^{\beta+\boldsymbol{a}\cdot\frac{d}{\boldsymbol{p}'}}.$$

Combining the above calculations, we obtain

$$\|\Gamma_t \phi_\lambda\|_{\mathbf{B}^\beta_{\boldsymbol{p};\boldsymbol{a}}} \lesssim_{C_T} \lambda^{(1+\alpha)\beta + \mathscr{A}_{1,\boldsymbol{p}}}. \tag{B.9}$$

Next we consider the general case $\ell > 0$. Let $\psi\in C_c^\infty(\mathbb{R}^{2d})$ such that for all $\lambda\geqslant 1$, $\psi\phi_\lambda = \phi_\lambda$ (since the supports of ϕ_λ shrink as λ increases, ψ can be simply taken equal to one on $\mathrm{supp}(\phi_1)$). By definition of Γ_t, we clearly have

$$\|\omega_\ell\Gamma_t\phi_\lambda\|_{\mathbf{B}^\beta_{\boldsymbol{p};\boldsymbol{a}}} = \|\omega_\ell\Gamma_t(\psi\phi_\lambda)\|_{\mathbf{B}^\beta_{\boldsymbol{p};\boldsymbol{a}}} = \|\big(\omega_\ell\Gamma_t(\psi)\big)\Gamma_t(\phi_\lambda)\|_{\mathbf{B}^\beta_{\boldsymbol{p};\boldsymbol{a}}} \lesssim_C \|\omega_\ell\Gamma_t\psi\|_{\mathbf{B}^{\beta,1}_{\boldsymbol{\infty},\boldsymbol{a}}}\|\Gamma_t\phi_\lambda\|_{\mathbf{B}^\beta_{\boldsymbol{p};\boldsymbol{a}}},$$

using Theorem 2 p. 177 in [38], which readily extends to the current anisotropic setting, for the control of the Besov norm of the product. Since ψ has compact support, it is easy to see that for any $T>0$, there is a bounded domain $D_T \subset \mathbb{R}^{2d}$ so that

$$\mathrm{supp}(\Gamma_t\psi)\subset D_T, \quad \forall t\in[0,T],\ \lambda\geqslant 1.$$

Hence,

$$\sup_{t\in[0,T]}\sup_{\lambda\geqslant 1}\|\omega_\ell\Gamma_t\psi\|_{\mathbf{B}^{\beta,1}_{\boldsymbol{\infty};\boldsymbol{a}}} < \infty.$$

The result now follows. $\square$

Appendix C. Convergence rate of empirical measure

We first state the following elementary lemma.

Lemma C.1. *Let $f_i, i = 1, \cdots, N$ be a sequence of Borel functions on $\mathbb{R}^{2d}$. Let $\boldsymbol{p} = (p_x, p_v) \in [1, \infty]^2$ and $q := p_x \wedge p_v \wedge 2$. Then*

$$\Big\|\sum_{i=1}^N |f_i|^2\Big\|_{\boldsymbol{p}/2} \leqslant \left(\sum_{i=1}^N \|f_i\|^q_{\boldsymbol{p}}\right)^{2/q}.$$

Proof. The estimate follows by $(\sum_{i=1}^N |a_i|)^\gamma \leqslant \sum_{i=1}^N |a_i|^\gamma$ for $\gamma \in (0, 1]$ and Minkowski's inequality. □

Now we show the following convergence rate estimate for the empirical measure of i.i.d random variables.

Lemma C.2. *Let $(\xi_1, \cdots, \xi_N)$ be i.i.d random variables in $\mathbb{R}^{2d}$ with common distribution μ and let $\bar\mu_N := \frac{1}{N}\sum_{i=1}^N \delta_{\{\xi_i\}}$ be the related empirical distribution measure. Let $\boldsymbol{p}_1 \in [1, \infty)^2$ and $\boldsymbol{p}_1 \leqslant \boldsymbol{p} \in (1, \infty)^2$. If $\boldsymbol{p}_1 = \boldsymbol{p}$, we let $\ell = 0$; otherwise, we choose $\ell > \mathscr{A}_{\boldsymbol{p}_1,\boldsymbol{p}}$. For any continuous $\phi : \mathbb{R}^{2d} \to \mathbb{R}$ and for all $m \geqslant 1$ for which $\mathbb{E}[|\xi_1|^{\ell m}] < \infty$, there is a constant $C = C(d, \boldsymbol{p}_1, \boldsymbol{p}, m, \ell) > 1$ such that for all $N \geqslant 1$,*

$$\|\phi * (\bar\mu_N - \mu)\|_{L^m(\Omega;\mathbb{L}^{\boldsymbol{p}_1})} \lesssim_C \Big(1 + (\mathbb{E}\big[|\xi_1|^{\ell m}_{\boldsymbol{a}}\big])^{1/m}\Big) N^{\frac{1}{q}-1}\|\omega_\ell\phi\|_{\boldsymbol{p}}, \tag{C.1}$$

where $q = p_x \wedge p_v \wedge 2$ and where ω_ℓ is the weight function defined in (A.1).

Proof. For each $i = 1, \cdots, N$, we define a family of centered i.i.d. random fields on $\mathbb{R}^{2d}$ with zero mean by

$$Y_i(z) := \phi * (\delta_{\{\xi_i\}} - \mu)(z), \quad z \in \mathbb{R}^{2d}.$$

Let ω_ℓ be defined as in (A.1) with ℓ as in the lemma and let $\boldsymbol{p}_2 \in [1, \infty]^2$ be defined by $\frac{1}{\boldsymbol{p}_1} = \frac{1}{\boldsymbol{p}} + \frac{1}{\boldsymbol{p}_2}$. By Hölder's inequality and applying (A.4) with $\ell > \boldsymbol{a} \cdot \frac{d}{\boldsymbol{p}_2} = \mathscr{A}_{\boldsymbol{p}_1,\boldsymbol{p}}$, we have

$$\Big\|\sum_{i=1}^N Y_i\Big\|_{\boldsymbol{p}_1} \leqslant \Big\|\sum_{i=1}^N \omega_\ell Y_i\Big\|_{\boldsymbol{p}}\|\omega_\ell^{-1}\|_{\boldsymbol{p}_2} \lesssim \Big\|\sum_{i=1}^N \omega_\ell Y_i\Big\|_{\boldsymbol{p}}.$$

Since $\mathbb{L}^{\boldsymbol{p}}$ is a UMD space for $\boldsymbol{p} \in (1, \infty)^2$ (see Corollary D.2 in the Appendix D below), by the functional BDG's inequality(see [43] and references therein):

$$\mathbb{E}\Big[\|\mathcal{M}_N\|^m_{\boldsymbol{p}}\Big] \leqslant C_{m,\boldsymbol{p}}\mathbb{E}\Big[\|[\mathcal{M}]^{1/2}_N\|^m_{\boldsymbol{p}}\Big],$$

for discrete-time $\mathbb{L}^{\boldsymbol{p}}$-valued martingale M, we have

$$\mathbb{E}\Big\|\sum_{i=1}^N \omega_\ell Y_i\Big\|^m_{\boldsymbol{p}} \leqslant (C_{m,\boldsymbol{p}})^m\mathbb{E}\Big\|\Big(\sum_{i=1}^N |\omega_\ell Y_i|^2\Big)^{1/2}\Big\|^m_{\boldsymbol{p}} = (C_{m,\boldsymbol{p}})^m\mathbb{E}\Big\|\sum_{i=1}^N |\omega_\ell Y_i|^2\Big\|^{m/2}_{\boldsymbol{p}/2}.$$

By Lemma C.1, we get for $q := p_x \wedge p_v \wedge 2$,

$$\mathbb{E}\Big\|\sum_{i=1}^N \omega_\ell Y_i\Big\|^m_{\boldsymbol{p}} \lesssim (C_{m,\boldsymbol{p}})^m\mathbb{E}\left(\sum_{i=1}^N \|\omega_\ell Y_i\|^q_{\boldsymbol{p}}\right)^{m/q}.$$

Noting that, by the growth property (A.3) of ω_ℓ, and for $\mu(\omega_\ell) := \int \omega_\ell(z)\mu(dz)$,

$$\begin{aligned}|\omega_\ell Y_i|(z) &\leqslant |\omega_\ell(z)\phi * (\delta_{\{\xi_i\}} + \mu)(z)|\\ &\lesssim |(\omega_\ell\phi) * (\delta_{\xi_i} + \mu)|(z) + |\phi * (\delta_{\{\xi_i\}} + \mu)|(z)(\omega_\ell(\xi_i) + \mu(\omega_\ell)).\end{aligned}$$

By Young's inequality and as $1 \leqslant \omega_\ell$, it follows that

$$\|\omega_\ell Y_i\|_{\boldsymbol{p}} \lesssim \|\omega_\ell\phi\|_{\boldsymbol{p}} + \|\phi\|_{\boldsymbol{p}}(\omega_\ell(\xi_i) + \omega_\ell(\mu) + 1) \leqslant \|\omega_\ell\phi\|_{\boldsymbol{p}}(\omega_\ell(\xi_i) + \omega_\ell(\mu) + 1).$$

Combining the above calculations, we get

$$\mathbb{E}\big[\|\phi*(\bar\mu_N-\mu)\|^m_{\boldsymbol{p}_1}\big]=\frac{1}{N^m}\mathbb{E}\Big\|\sum_{i=1}^N Y_i\Big\|^m_{\boldsymbol{p}_1}\lesssim\frac{(C_{m,\boldsymbol{p}})^m}{N^m}N^{\frac{m}{q}}\|\omega_\ell\phi\|_{\boldsymbol{p}}\mathbb{E}\Big[\big(1+\omega_\ell(\xi_1)\big)^m\Big].$$

The proof is complete. □

Now, recalling the dilatation operator $\mathscr{D}_\lambda$ from Appendix B and the notation:

$$\phi_\lambda(z)=\mathscr{D}_\lambda\phi(x,v)=\lambda^{(2+\alpha)d}\phi(\lambda^{1+\alpha}x,\lambda v),\ \ \lambda>0,$$

we have the following useful result.

Theorem C.3. *Let $(\xi_1,\cdots,\xi_N)$ be i.i.d random variables in $\mathbb{R}^{2d}$ with common distribution μ. Let $\bar\mu_N:=\frac{1}{N}\sum_{i=1}^N\delta_{\{\xi_i\}}$ be the empirical distribution measure. Let $\boldsymbol{p_1}\in[1,\infty)^2$ and $\boldsymbol{p_1}\leqslant\boldsymbol{p}\in(1,\infty)^2$. If $\boldsymbol{p_1}=\boldsymbol{p}$, we let $\ell=0$; otherwise, we choose $\ell>\mathscr{A}_{\boldsymbol{p_1},\boldsymbol{p}}$. Then, given*

$$q=p_x\wedge p_v\wedge 2,$$

for any $\beta<0$ and any integer m satisfying

$$\mu\big(|\cdot|^{\ell m}\big):=\mathbb{E}\big[|\xi|^{\ell m}_{\boldsymbol{a}}\big]<\infty,$$

there exists a constant $C>0$ such that, for all $N,\lambda\geqslant1$,

$$\|\phi_\lambda*(\bar\mu_N-\mu)\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p_1};\boldsymbol{a}})}\lesssim_C\Big(1+\mathbb{E}\big[|\xi_1|^{\ell m}_{\boldsymbol{a}}\big]\Big)^{1/m}N^{\frac{1}{q}-1}\lambda^{\mathscr{A}_{\boldsymbol{1},\boldsymbol{p}}}.\tag{C.2}$$

Proof. By Minkowski's inequality, we have

$$\begin{aligned}\|\phi_\lambda*(\bar\mu_N-\mu)\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p_1};\boldsymbol{a}})}&=\Big\|\sum_{j\geqslant0}2^{\beta j}\|\mathcal{R}^{\boldsymbol{a}}_j(\phi_\lambda*(\bar\mu_N-\mu))\|_{\boldsymbol{p_1}}\Big\|_{L^m(\Omega)}\\&\leqslant\sum_{j\geqslant0}2^{\beta j}\|\mathcal{R}^{\boldsymbol{a}}_j(\phi_\lambda*(\bar\mu_N-\mu))\|_{L^m(\Omega;\mathbb{L}^{\boldsymbol{p_1}})}.\end{aligned}$$

Noting that

$$\mathcal{R}^{\boldsymbol{a}}_j(\phi_\lambda*(\bar\mu_N-\mu))=(\mathcal{R}^{\boldsymbol{a}}_j\phi_\lambda)*(\bar\mu_N-\mu)=(\check\phi^{\boldsymbol{a}}_j*\phi_\lambda)*(\bar\mu_N-\mu),$$

by (C.1), we have, for ω_ℓ defined as in (A.1),

$$\|\mathcal{R}^{\boldsymbol{a}}_j(\phi_\lambda*(\bar\mu_N-\mu))\|_{L^m(\Omega;\mathbb{L}^{\boldsymbol{p_1}})}\lesssim_C\Big(1+(\mathbb{E}\big[|\xi_1|^{\ell m}_{\boldsymbol{a}}\big])^{1/m}\Big)N^{\frac{1}{q}-1}\|\omega_\ell\mathcal{R}^{\boldsymbol{a}}_j\phi_\lambda\|_{\boldsymbol{p}}.$$

According to (A.7), multiplying the above by $2^{\beta j}$ (recall that $\beta<0$) and summing the resulting expression over $j\geqslant0$, we get

$$\begin{aligned}\|\phi_\lambda*(\bar\mu_N-\mu)\|_{L^m(\Omega;\mathbf{B}^{\beta,1}_{\boldsymbol{p_1};\boldsymbol{a}})}&\lesssim\Big(1+(\mathbb{E}\big[|\xi_1|^{\ell m}_{\boldsymbol{a}}\big])^{1/m}\Big)N^{\frac{1}{q}-1}\sum_{j\geqslant0}2^{\beta j}\|(\mathcal{R}^{\boldsymbol{a}}_j\phi_\lambda)\omega_\ell\|_{\boldsymbol{p}}\\&\lesssim\Big(1+(\mathbb{E}\big[|\xi_1|^{\ell m}_{\boldsymbol{a}}\big])^{1/m}\Big)N^{\frac{1}{q}-1}\sup_{j\geqslant0}\|(\mathcal{R}^{\boldsymbol{a}}_j\phi_\lambda)\omega_\ell\|_{\boldsymbol{p}}\end{aligned}$$

Applying (A.7) in Lemma A.1, we eventually get

$$\sup_{j\geqslant0}\|(\mathcal{R}^{\boldsymbol{a}}_j\phi_\lambda)\omega_\ell\|_{\boldsymbol{p}}\leqslant\|\omega_\ell\phi_\lambda\|_{\boldsymbol{p}}=\lambda^{\mathscr{A}_{\boldsymbol{1},\boldsymbol{p}}},$$

which ends the proof. □

Appendix D. Characterization of $\mathbb{L}^{\boldsymbol{p}}$ as a UMD space and as a Banach space of martingale type.

Let us first recall the definition of UMD spaces.

Definition D.1 ([36], Chapter 5). *A Banach space $(E, \|\cdot\|_E)$ is said to be a p-unconditional martingale difference (UMD$_p$ in short) space if there exists $p \in (1, \infty)$ and $c = c(E, p) > 0$ such that for every $\{-1, 1\}$-valued sequence $\{\epsilon_n\}_n$ and for every martingale $\{M_n\}_{n\geqslant 1}$ defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geqslant 0}, \mathbb{P})$ and converging as $n \to \infty$ in $L^p(\Omega; E)$, we have*

$$\sup_n \| \sum_{j=0}^{n} \epsilon_j dM_j \|_{L^p(\Omega;E)} \lesssim_c \sup_n \|M_n\|_{L^p(\Omega;E)}, \tag{D.1}$$

for $dM_j := M_j - M_{j-1}$, $dM_0 = 0$.

Simply instances of UMD spaces are the Euclidean space, Hilbert spaces and L^p-spaces for $p \in (1, \infty)$. Further, for any UMD space E, and any measure space $(A, \mathcal{A}, \nu)$, every $L^p(A; E)$, with $p \in (1, \infty)$ is a UMD space, [36], Corollary 5.22. As such, for any $p_x \in (1, \infty)$, $L^{p_x}(\mathbb{R}^d)$ is a UMD space, and by iteration, we can immediately derive the corollary:

Corollary D.2. *For all $\boldsymbol{p} \in (1, \infty)^2$, $\mathbb{L}^{\boldsymbol{p}}$ is a UMD-space.*

For the characterization of $\mathbb{L}^{\boldsymbol{p}}$ as a Banach space of M-type or martingale type p - characterization which is used to apply the martingale inequality in Theorem 3.7 for the proof of Theorem 3.6 -, and for the sake of completeness, let us briefly recall the definition of these spaces.

Definition D.3. *A Banach space $(E, \|\cdot\|_E)$ is said to be of martingale type p, for $p \in [1, \infty]$, if there exists a constant $C = C(E, p)$ such that for all finite E-valued martingale $\{M_k\}_{k=1}^n$, it holds that, for $dM_k = M_k - M_k - 1$ (with the convention $M_0 = 0$)*

$$\sup_{1\leqslant k\leqslant n} \|M_k\|_{L^p(\Omega;E)} \lesssim_C \Big(\sum_{k=1}^{n} \mathbb{E}[\|dM_k\|_E^p] \Big)^{1/p}. \tag{D.2}$$

Characteristically, every Banach space is of martingale type 1, Hilbert spaces are of martingale type 2 and we may refer to [36] for a detailed account of the geometric properties related to (D.2) and to [21] for related martingale transforms.

Similarly to the UMD property, the martingale property (D.2) holds for Lebesgue spaces: for E a Banach space of martingale type $p \in [1, 2]$ and $(A, \mathcal{A}, \nu)$ a measure space, any $L^r(A; E)$ space with $r \in (1, \infty)$ is of martingale type $r \wedge p$ (see [21], Proposition 3.5.30). We have the following result:

Corollary D.4. *For all $\boldsymbol{p} \in (1, \infty)^2$, $\mathbb{L}^{\boldsymbol{p}}$ is a Banach space of martingale type p for all $1 \leqslant p \leqslant p_x \wedge p_v \wedge 2$.*

Proof. Given $\boldsymbol{p} = (p_x, p_v)$, since $\mathbb{R}^{2d}$ is of martingale type 2, $L^{p_x}(\mathbb{R}^d)$ is of martingale type $p_x \wedge 2$. As such, the iterated space $L^{p_v}(\mathbb{R}^d; L^{p_x}(\mathbb{R}^d))$ is of martingale type $p_v \wedge (p_x \wedge 2)$. Since once (D.2) is satisfied for some p, the property also holds true for all integrability index in $[1, p]$ (see Corollary 3.5.28, [21]). This gives the claim. □

Appendix E. Gronwall inequality of Volterra type

In this appendix, we show a Gronwall inequality of Volterra type that is crucial for the proof of Theorem 2.1. First of all, we give the following elementary estimate.

Lemma E.1. *Let $T > 0$ and $a_i, b_i \in (-1, \infty)$ with $a_i + b_i > -1$, $i = 1, 2$. Then there is a constant $C = C(T, a_1, b_1, a_2, b_2) > 0$ such that for all $t \in [0, T]$ and all Borel function $f : [0, T] \to \mathbb{R}_+$,*

$$\int_0^t (t-s)^{a_1} s^{b_1} \left(\int_0^s (s-r)^{a_2} r^{b_2} f(r) \mathrm{d}r \right) \mathrm{d}s \lesssim_C \sum_{i=1}^{2} \int_0^t (t-s)^{a_i} s^{b_i} f(s) \mathrm{d}s. \tag{E.1}$$

In particular, if $b_1 \geqslant b_2$*, then*

$$\int_0^t (t-s)^{a_1} s^{b_1} \left(\int_0^s (s-r)^{a_2} r^{b_2} f(r) \mathrm{d}r \right) \mathrm{d}s \lesssim_C \int_0^t (t-s)^{a_2} s^{b_2} f(s) \mathrm{d}s. \tag{E.2}$$

Proof. By Fubini-Tonelli's theorem we have

$$\int_0^t (t-s)^{a_1} s^{b_1} \left(\int_0^s (s-r)^{a_2} r^{b_2} f(r) \mathrm{d}r \right) \mathrm{d}s = \int_0^t K(t,r) r^{b_2} f(r) \mathrm{d}r,$$

where

$$K(t,r) := \int_r^t (t-s)^{a_1} s^{b_1} (s-r)^{a_2} \mathrm{d}s.$$

(Case $b_1 \geqslant 0$**).** Since $a_1 + 1 > 0$, we have

$$K(t,r) \leqslant T^{b_1} \int_r^t (t-s)^{a_1} (s-r)^{a_2} \mathrm{d}s \lesssim (t-r)^{a_1+a_2+1} \leqslant T^{a_1+1} (t-r)^{a_2}. \tag{E.3}$$

(Case $b_2 \leqslant b_1 < 0$**).** In view of $a_1 + b_1 > -1$, we have

$$K(t,r) \leqslant \int_r^t (t-s)^{a_1} (s-r)^{a_2+b_1} \mathrm{d}s \lesssim (t-r)^{a_1+b_1+a_2+1} \leqslant T^{a_1+b_1+1} (t-r)^{a_2}. \tag{E.4}$$

(Case $b_1 < b_2 \leqslant 0$**).** In this case, for all $s \in [r,t]$, $s^{b_1} = s^{b_1-b_2} s^{b_2} \leqslant (s-r)^{b_1-b_2} r^{b_2}$, and we have

$$K(t,r) \leqslant r^{b_1-b_2} \int_r^t (t-s)^{a_1} (s-r)^{a_2+b_2} \mathrm{d}s \lesssim (t-r)^{a_1+a_2+b_2+1} r^{b_1-b_2} \leqslant T^{a_2+b_2+1} (t-r)^{a_1} r^{b_1-b_2}.$$

(Case $b_1 < 0 < b_2$**).** In this case, we have

$$K(t,r) \leqslant T^{b_2} r^{b_1-b_2} \int_r^t (t-s)^{a_1} (s-r)^{a_2} \mathrm{d}s \lesssim (t-r)^{a_1+a_2+1} r^{b_1-b_2} \leqslant T^{a_2+1} (t-r)^{a_1} r^{b_1-b_2}.$$

Combining the above calculations, we obtain (E.1). (E.2) is from (E.3) and (E.4). □

Now we can show the following Gronwall inequality.

Lemma E.2 (Gronwall's inequality of Volterra type)**.** *Let* $T > 0$*,* $n \in \mathbb{N}$ *and* $a_i, b_i \in (-1, \infty)$ *with* $a_i + b_i > -1$ *for all* $i = 1, .., n$*. Assume that* $f, g : [0,T] \to \mathbb{R}_+$ *are two Borel measurable functions and satisfy that for some* $c_0 > 0$ *and almost all* $t \in (0,T]$*,*

$$f(t) \leqslant g(t) + c_0 \sum_{i=1}^n \int_0^t (t-s)^{a_i} s^{b_i} f(s) \mathrm{d}s. \tag{E.5}$$

Then there is a constant $C = C(T, n, (a_i, b_i)_{i=1}^n, c_0) > 0$ *such that for almost all* $t \in (0,T]$*,*

$$f(t) \lesssim_C g(t) + \sum_{i=1}^n \int_0^t (t-s)^{a_i} s^{b_i} g(s) \mathrm{d}s.$$

Proof. The case $n = 1$ yielding

$$f(t) \leqslant g(t) + c_0 \int_0^t (t-s)^{a_1} s^{b_1} f(s) \mathrm{d}s \Rightarrow f(t) \lesssim_C g(t) + \int_0^t (t-s)^{a_1} s^{b_1} g(s) \mathrm{d}s \tag{E.6}$$

has been established in [15, Lemma A.4]. Now, let $n > 1$ and $\{a_i\}_{i=1}^n, \{b_i\}_{i=1}^n \subset (-1, \infty)$ satisfying $a_i + b_i > -1$. Without loss of generality, we assume that $b_1 \geqslant b_2 \geqslant \cdot\cdot \geqslant b_n$. Let

$$F_1(t) := g(t) + c_0 \sum_{k=2}^n \int_0^t (t-s)^{a_k} s^{b_k} f(s) \mathrm{d}s,$$

and, for any $i = 2, .., n-1$,

$$F_i(t) := g(t) + c_0 \sum_{k=1}^{i-1} \int_0^t (t-s)^{a_k} s^{b_k} g(s) \mathrm{d}s + c_0 \sum_{k=i+1}^n \int_0^t (t-s)^{a_k} s^{b_k} f(s) \mathrm{d}s.$$

Then (E.5) reads as

$$f(t)\leqslant F_1(t)+c_0\int_0^t (t-s)^{a_1}s^{b_1}f(s)\mathrm{d}s.$$

Thus applying (E.6) and next (E.2) (with the ordering $b_1\geqslant b_k$ for $k\geqslant 2$), we have

$$\begin{aligned}
f(t) &\lesssim_C F_1(t)+\int_0^t (t-s)^{a_1}s^{b_1}F_1(s)\mathrm{d}s\\
&= F_1(t)+\int_0^t (t-s)^{a_1}s^{b_1}g(s)\,\mathrm{d}s+c_0\sum_{k=2}^n\int_0^t (t-s)^{a_1}s^{b_1}\Big(\int_0^s (s-r)^{a_k}r^{b_k}f(r)\,\mathrm{d}r\Big)\,\mathrm{d}s\\
&\overset{(E.2)}{\lesssim} F_1(t)+\int_0^t (t-s)^{a_1}s^{b_1}g(s)\,\mathrm{d}s+c_0\sum_{k=2}^n\int_0^t (t-s)^{a_k}s^{b_k}f(r)\,\mathrm{d}r\\
&=F_1(t)+\int_0^t (t-s)^{a_1}s^{b_1}g(s)\mathrm{d}s+c_0\sum_{k=3}^n\int_0^t (t-s)^{a_k}s^{b_k}f(r)\,\mathrm{d}r+c_0\int_0^t (t-s)^{a_2}s^{b_2}f(r)\,\mathrm{d}r\\
&\lesssim F_2(t)+\int_0^t (t-s)^{a_2}s^{b_2}f(s)\mathrm{d}s.
\end{aligned}$$

Iterating the preceding estimate, successively using (E.6) and (E.2)), we obtain

$$f(t)\lesssim F_{n-1}(t)+\int_0^t (t-s)^{a_n}s^{b_n}f(s)\mathrm{d}s,$$

which in turn, applying a last time (E.6), implies the desired estimate. □

Now we apply the previous lemma to the function given in (2.29):

$$G_\beta(t,s)=t^{\frac{\beta-\beta_0}{\alpha}}(t-s)^{-\frac{1}{\alpha}}+t^{\frac{\beta+\Lambda}{\alpha}}(t-s)^{-\frac{\Lambda+\beta_0+1}{\alpha}},\ 0\leqslant s<t<\infty,$$

where $\alpha\in(1,2]$, $\beta\geqslant 0$ and β_0,Λ are as in **(H)**.

Lemma E.3. *Let $T>0$, $\beta\in\big[0,(\beta_0+\alpha)\wedge(\alpha-\Lambda)\big)$ and $\gamma\in[0,\alpha\wedge(\alpha-1+\beta-\beta_0))$. Then the following properties hold.*

(i) There is a $p\geqslant 1$ large enough and $C_T>0$ so that for all $f\in L^p([0,t])$,

$$\int_0^t G_\beta(t,s)s^{-\frac{\gamma}{\alpha}}f(s)\mathrm{d}s\leqslant C_T\|f\|_{L^p([0,t])}. \tag{E.7}$$

(ii) There is a constant $C_T>0$ such that for all $t\in[0,T]$ and measurable $f:[0,T]\to\mathbb{R}_+$,

$$\int_0^t G_\beta(t,s)s^{-\frac{\gamma}{\alpha}}\left(\int_0^s G_\beta(s,r)f(r)\mathrm{d}r\right)\mathrm{d}s\leqslant C_T\int_0^t G_\beta(t,s)f(s)\mathrm{d}s. \tag{E.8}$$

(iii) Assume that $f,g,h:[0,T]\to\mathbb{R}_+$ are three Borel measurable functions and satisfying, for some constant $c_0>0$ and almost all $t\in(0,T]$,

$$f(t)\leqslant g(t)+c_0\int_0^t G_\beta(t,s)s^{-\frac{\gamma}{\alpha}}f(s)\mathrm{d}s+\int_0^t G_\beta(t,s)h(s)\mathrm{d}s. \tag{E.9}$$

Then there is a constant $C=C(T,\alpha,\beta,\beta_0,\Lambda,c_0)>0$ such that for almost all $t\in(0,T]$,

$$f(t)\lesssim_C g(t)+\int_0^t G_\beta(t,s)s^{-\frac{\gamma}{\alpha}}(g(s)+h(s))\mathrm{d}s.$$

Proof. (i) As the exponents $\frac{-1}{\alpha}$, $\frac{-\gamma}{\alpha}$ and $-\frac{\Lambda+\beta_0+1}{\alpha}$ are all strictly above -1, one can take $p'\in(1,\infty)$ close enough to 1 so that $\|G_\beta(t,\cdot)\cdot^{-\frac{\gamma}{\alpha}}\|_{L^{p'}((0,T))}<\infty$. The statement then follows by Hölder's inequality.

(ii) Since $\beta-\beta_0, \beta+\Lambda \in [0,\alpha)$, we have $t^\ell - s^\ell \leqslant (t-s)^\ell$ for $\ell \in \{(\beta-\beta_0)/\alpha, (\beta+\Lambda)/\alpha\}$ and so

$$\begin{aligned}
G_\beta(t,s) &\leqslant \Big(s^{\frac{\beta-\beta_0}{\alpha}} + (t-s)^{\frac{\beta-\beta_0}{\alpha}}\Big)(t-s)^{-\frac{1}{\alpha}} + \Big(s^{\frac{\beta+\Lambda}{\alpha}} + (t-s)^{\frac{\beta+\Lambda}{\alpha}}\Big)(t-s)^{-\frac{\Lambda+\beta_0+1}{\alpha}} \\
&=: \sum_{i=1}^4 (t-s)^{a_i} s^{b_i} \overset{s^\ell \leqslant t^\ell}{\leqslant} G_\beta(t,s) + \sum_{i\in\{2,4\}} (t-s)^{a_i} \leqslant G_\beta(t,s) + 2t^{\frac{\beta-\beta_0}{\alpha}}(t-s)^{a_2-\frac{\beta-\beta_0}{\alpha}} \\
&\leqslant 2G_\beta(t,s), \qquad \text{(E.10)}
\end{aligned}$$

taking

$$a_1 = -\tfrac{1}{\alpha}, b_1 = \tfrac{\beta-\beta_0}{\alpha}, \quad a_2 = \tfrac{\beta-\beta_0-1}{\alpha}, b_2 = 0, \quad a_3 = -\tfrac{\Lambda+\beta_0+1}{\alpha}, b_3 = \tfrac{\beta+\Lambda}{\alpha}, \quad a_4 = \tfrac{\beta-\beta_0-1}{\alpha}, b_4 = 0.$$

Note that

$$\alpha\big(a_i + b_i\big) - \gamma = -1 + \beta - \beta_0 - \gamma > -\alpha, \quad i = 1,2,3,4.$$

Then (E.8) is directly from (E.1). Indeed, (E.1) and (E.10) imply that

$$\begin{aligned}
\int_0^t G_\beta(t,s)s^{-\frac{\gamma}{\alpha}}\left(\int_0^s G_\beta(s,r)f(r)\mathrm{d}r\right)\mathrm{d}s &= \int_0^t G_\beta(t,s)s^{-\frac{\gamma}{\alpha}}\left(\int_0^s G_\beta(s,r)r^{-\frac{\gamma}{\alpha}}[r^{\frac{\gamma}{\alpha}}f(r)]\mathrm{d}r\right)\mathrm{d}s \\
&\leqslant \int_0^t \Big(\sum_{i=1}^4 (t-s)^{a_i} s^{b_i-\frac{\gamma}{\alpha}}\Big)\Big(\sum_{j=1}^4 \int_0^s (s-r)^{a_j} r^{b_j-\frac{\gamma}{\alpha}}[r^{\frac{\gamma}{\alpha}}f(r)]\,\mathrm{d}r\Big)\,\mathrm{d}s \\
&\overset{\text{(E.1)}}{\lesssim} 2\sum_{i=1}^4 \int_0^t (t-r)^{a_i} r^{b_i-\frac{\gamma}{\alpha}}[r^{\frac{\gamma}{\alpha}}f(r)]\,\mathrm{d}r \\
&\lesssim \int_0^t G_\beta(t,r)r^{-\frac{\gamma}{\alpha}}[r^{\frac{\gamma}{\alpha}}f(r)]\mathrm{d}r = \int_0^t G_\beta(t,r)f(r)\mathrm{d}r.
\end{aligned}$$

(iii) Starting from (E.9), by Lemma E.2 (with the source term $g(t) + \int_0^t G_\beta(t,s)h(s)\mathrm{d}s$) and (ii), we have

$$\begin{aligned}
f(t) &\lesssim g(t) + \int_0^t G_\beta(t,s)s^{-\frac{\gamma}{\alpha}}\left(g(s) + \int_0^s G_\beta(s,r)h(r)\mathrm{d}r\right)\mathrm{d}s \\
&\lesssim g(t) + \int_0^t G_\beta(t,s)s^{-\frac{\gamma}{\alpha}}\left(g(s) + h(s)\right)\mathrm{d}s.
\end{aligned}$$

This completes the proof. □

References

[1] Athreya S., Butkovsky O. and Mytnik, L., Strong existence and uniqueness for stable stochastic differential equations with distributional drift, *Ann. Probab.*, 48(1):178–210, 2020.

[2] Bernardin F., Bossy M., Chauvin C., Jabir, J.-F. and Rousseau, A., Stochastic Lagrangian method for downscaling problems in computational fluid dynamics, *M2AN Math. Model. Numer. Anal.*, 44(5): 885-920, 2010.

[3] Bossy M. and Talay D., A stochastic particle method for the McKean-Vlasov and the Burgers equation, *Mathematics of Computation*, 66 (217):157-192, 1997.

[4] Bossy M., Fezoui, L. and Piperno S., Comparison of a stochastic particle method and a finite volume deterministic method applied to Burgers equation, *Monte Carlo Methods and Appl.*, 3 (2):113-140, 1997.

[5] Bresch D., Jabin P.-E. and Soler J., A new approach to the mean-field limit of Vlasov-Fokker-Planck equations. Preprint arXiv: 2203.15747, 2023.

[6] Carmona R. and Delarue F.: Probabilistic theory of mean field games with applications. I. Mean field FBSDEs, control, and games. Probability Theory and Stochastic Modeling, 83. *Springer*, 2018.

[7] Chaudru de Raynal, P.-E and S. Menozzi, S., On multidimensional stable-driven stochastic differential equations with Besov drift, *Electronic Journal of Proba.*, 27, Paper No. 163, 52 pp, 2022.

[8] Chaudru de Raynal P.-E., Jabir J.-F. and Menozzi S., Multidimensional stable driven McKean-Vlasov SDEs with distributional interaction kernel: A regularization by noise perspective. Preprint arXiv:2205.11866. Submitted, 2022.

[9] Chaudru de Raynal P.-E., Jabir J.-F. and Menozzi S., Multidimensional stable driven McKean-Vlasov SDEs with distributional interaction kernel: Critical Thresholds and Related models. Preprint arXiv:2302.09900. Submitted, 2023.
[10] Chen Z.-Q. and Zhang X., Propagation of regularity in L^p-spaces for Kolmogorov-type hypoelliptic operators, *J. Evol. Equ.*, 19(4): 1041-1069, 2019.
[11] Chen X., Chatterji N. S., Bartlett, P. L. and Jordan M. L., Underdamped Langevin MCMC: A non-asymptotic analysis, *Proceedings of Machine Learning Research*, 75: 1-24, 2018.
[12] Delarue F. and Diel R., Rough paths and 1d SDE with a time dependent distributional drift: application to polymers, *Probab. Theory Related Fields*, 165(1-2):1–63, 2016.
[13] Flandoli F., Issoglio E. and Russo F., Multidimensional stochastic differential equations with distributional drift, *Trans. Amer. Math. Soc.*, 369 (3):1665-1688, 2017.
[14] Flandoli F, Olivera C. and Simon M., Uniform approximation of 2 dimensional Navier-Stokes equation by stochastic interacting particle systems, *SIAM J. Math. Anal.*, 52: 5339-5362, 2020.
[15] Hao Z., McKean-Vlasov SDEs with singular drifts. Thesis (Ph.D.)- Bielefeld: Universität Bielefeld; 2023. 179 pp. Online version: Link
[16] Hao Z., Röckner M. and Zhang X., Strong convergence of propagation of chaos for McKean-Vlasov SDEs with singular interactions, *SIAM J. Math. Anal.*, 56, 2024 (Online).
[17] Hao Z., Röckner, M. and Zhang X., Second order fractional mean-field SDEs with singular kernels and measure initial data. Preprint arXiv:2302.04392, 2023.
[18] Hao Z., Wu M. and Zhang X., Schauder estimates for nonlocal kinetic equations and applications, *J. Math. Pures Appl.* (9) 140: 139–184, 2020.
[19] Hao Z., Zhang X., Zhu R. and Zhu X., Singular kinetic equations and applications, *Ann. Probab.* 52 (2): 576-657, 2024.
[20] Hausenblas E., Maximal inequalities of the Itô integral with respect to Poisson random measures or Lévy processes on Banach spaces, *Potential Analysis*, 35: 223-251, 2011.
[21] Hytönen, T, Neerven, J. V., Veraar, V. and Weis, L., Analysis in Banach Spaces Vol. I: Martingales and LittleWood-Paley Theory, Springer 2016.
[22] Hoeksema J., Holding T., Maurelli M. and Tse O., Large deviations for singularly interacting diffusions, *Ann. Inst. H. Poincaré Probab. Statist.*, 60(1): 492–548, 2024.
[23] Imbert C. and Silvestre L., The Schauder estimate for kinetic integral equations, *Anal. PDE*, 14(1): 171–204, 2021.
[24] Jabin P.-E. and Wang Z., Mean field limit and propagation of chaos for Vlasov system with bounded forces, *J. Func. Anal.*, 271 (12): 3588-3627, 2016.
[25] Jourdain B. and Méléard S., Propagation of chaos and fluctuations for a moderate model with smooth initial data, *Ann. Inst. H. Poincaré Probab. Statist.* 34(6): 727-766, 1998.
[26] Kremp, H. and Perkowski, N., Multidimensional SDE with distributional drift and Lévy noise, *Bernoulli*, 28 (3): 1757–1783, 2022.
[27] Marinelli C. and Röckner M., On maximal inequalities for purely discontinuous martingales in infinite dimensions, *Séminaire de Probabilités XLVI, Lecture Notes in Mathematics*, Vol. 2123: 293-315, 2014.
[28] Oelschläger K., A martingale approach to the law of large numbers for weakly interacting stochastic processes, *Ann. Probab.*, 12 (2): 458-479, 1984.
[29] Oelschläger K., A law of large numbers for moderately interacting diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 69, no. 2: 279-322, 1985.
[30] Oelschläger K., A fluctuation theorem for moderately interacting diffusion processes, *Probab. Theory Related Fields*, 74: 591-616, 1987.
[31] Olivera C., Richard A. and Tomašević M., Quantitative particle approximation of nonlinear Fokker-Planck equations with singular kernel, *Ann. Sc. Norm. Super. Pisa Cl. Sci*, (5) Vol. XXIV: 691-749, 2023.
[32] Olivera C., Richard A. and Tomašević M., Quantitative approximation of the Keller-Segel and Burgers equations by moderately interacting particles. Preprint arXiv:2004.03177, 2022.
[33] Pareshi P. and Toscani G., Interacting Multiagent Systems, *Oxford University Press*, 2014.
[34] Perthame, B., PDE models for chemotactic movements: parabolic, hyperbolic and kinetic, *Applications of Mathematics*, 49(6): 539-564, 2004.
[35] Pope S. B., Turbulent Flows, *Cambridge University Press*, 2000.
[36] Pisier G., Martingales in Banach Spaces, *Cambridge University Press*, 2016.
[37] Priola E., Global Schauder estimates for a class of degenerate Kolmogorov equations. *Studia Math.*, 194(2):117–153, 2009.
[38] Runst T. and Sickel W., Sobolev Spaces of Fractional Order, Nemytskij Operators, and Nonlinear Partial Differential Equations, *Walter de Gruyter & Co*, 1997.
[39] Shlesinger, M. F., Zaslavsky, G. M. and Frisch, U., Lévy Flights and Related Topics in Physics, *Springer*, 1995.

[40] Simsekli U., Sagun L. and Gürbüzbalaban M., A tail-index analysis of stochastic gradient noise in deep neural networks, *Proceedings of the 36th International Conference on Machine Learning*, PMLR 97:5827-5837, 2019.
[41] Tomašević M., Propagation of chaos for stochastic particle systems with singular mean-field interaction of Lp - Lq type, *Electron. Commun. Probab.*, 28: 1–13, 2023.
[42] Triebel H., Gagliardo Nirenberg Inequalities, *Proceedings of the Steklov Institute of Mathematics*, 284: 263-279, 2014.
[43] Veraar M. and Yaroslavtsev I., Pointwise properties of Martingales with Values in Banach Function Spaces, *High Dimensional Probability VIII, Birkhäuser*, Chapter 16: 321-340, 2019.
[44] Zhang X. and Zhang X., Cauchy problem of stochastic kinetic equations. *Ann. Appl. Probab.*, 34(1A): 148-202, 2024.
[45] Zhu J., Brzeźniak Z and Hausenblas E., Maximal inequalities for stochastic convolutions driven by compensated Poisson random measures in Banach spaces, *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(2): 937-956, 2017.
[46] Zhu W., Lou Z. and Wu W. B., Beyond sub-gaussian noises: Sharp concentration analysis and stochastic gradient descent, *Journal of Machine Learning Research*, 23, no. 46:1-22, 2022.

Zimo Hao: Fakultät für Mathematik, Universität Bielefeld, 33615, Bielefeld, Germany, Email: zhao@math.uni-bielefeld.de

Jean-Francois Mehdi Jabir: Department of Statistics and Data Analysis, and Laboratory of Stochastic Analysis and its Application, Pokrovski Blvd. 11, office S431, Moscow, Russia Federation., Email address: jjabir@hse.ru

Stéphane Menozzi: LaMME, Université d'Evry Val d'Essonne, Université Paris-Saclay, CNRS (UMR 8071), 23 Boulevard de France 91037 Evry, France., Email address: stephane.menozzi@univ-evry.fr

Michael Röckner: Fakultät für Mathematik, Universität Bielefeld, 33615, Bielefeld, Germany, and Academy of Mathematics and Systems Science, CAS, Beijing, China, Email: roeckner@math.uni-bielefeld.de

Xicheng Zhang: School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China, Email: XichengZhang@gmail.com