

**On partially observed jump diffusions III
regularity of the filtering density**

Germ, Fabian; Gyöngy, István

DOI

[10.1007/s40072-024-00338-9](https://doi.org/10.1007/s40072-024-00338-9)

Publication date

2024

Document Version

Final published version

Published in

Stochastics and Partial Differential Equations: Analysis and Computations

Citation (APA)

Germ, F., & Gyöngy, I. (2024). On partially observed jump diffusions III: regularity of the filtering density. *Stochastics and Partial Differential Equations: Analysis and Computations*, 13(1), 531-583. <https://doi.org/10.1007/s40072-024-00338-9>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



On partially observed jump diffusions III: regularity of the filtering density

Fabian Germ¹ · István Gyöngy²

Received: 15 March 2024 / Revised: 4 September 2024 / Accepted: 16 September 2024
© The Author(s) 2024

Abstract

The filtering equations associated to a partially observed jump diffusion model $(Z_t)_{t \in [0, T]} = (X_t, Y_t)_{t \in [0, T]}$, driven by Wiener processes and Poisson martingale measures are considered. Building on results from two preceding articles on the filtering equations, the regularity of the conditional density of the signal X_t , given observations $(Y_s)_{s \in [0, t]}$, is investigated, when the conditional density of X_0 given Y_0 exists and belongs to a Sobolev space, and the coefficients satisfy appropriate smoothness and growth conditions.

Keywords Nonlinear filtering · Random measures · Lévy processes

Mathematics Subject Classification Primary 60H05 · 60H15; Secondary 35R60

Contents

1	Introduction
2	Formulation of the main results
3	Preliminaries
4	Sobolev estimates
5	Solvability of the filtering equations in Sobolev spaces
6	Proof of Theorem 2.1
	References

✉ Fabian Germ
f.germ@tudelft.nl
István Gyöngy
i.gyongy@ed.ac.uk

¹ Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

² School of Mathematics and Maxwell Institute, University of Edinburgh, Scotland, UK

1 Introduction

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$ be a complete filtered probability space, carrying a $d_1 + d'$ -dimensional \mathcal{F}_t -Wiener process $(W_t, V_t)_{t \geq 0}$ and independent \mathcal{F}_t -Poisson martingale measures $\tilde{N}_i(d\mathfrak{z}, dt) = N_i(d\mathfrak{z}, dt) - \nu_i(d\mathfrak{z})dt$ on $\mathbb{R}_+ \times \mathfrak{Z}_i$, for $i = 0, 1$, with σ -finite characteristic measures ν_0 and ν_1 on a separable measurable space $(\mathfrak{Z}_0, \mathcal{Z}_0)$ and on $(\mathfrak{Z}_1, \mathcal{Z}_1) = (\mathbb{R}^{d'} \setminus \{0\}, \mathcal{B}(\mathbb{R}^{d'} \setminus \{0\}))$, respectively, where $\mathcal{B}(V)$ denotes the Borel σ -algebra on V for topological spaces V .

We consider the signal and observation model

$$\begin{aligned}
 dX_t &= b(t, X_t, Y_t) dt + \sigma(t, X_t, Y_t) dW_t + \rho(t, X_t, Y_t) dV_t \\
 &\quad + \int_{\mathfrak{Z}_0} \eta(t, X_{t-}, Y_{t-}, \mathfrak{z}) \tilde{N}_0(d\mathfrak{z}, dt) + \int_{\mathfrak{Z}_1} \xi(t, X_{t-}, Y_{t-}, \mathfrak{z}) \tilde{N}_1(d\mathfrak{z}, dt) \\
 dY_t &= B(t, X_t, Y_t) dt + dV_t + \int_{\mathfrak{Z}_1} \mathfrak{z} \tilde{N}_1(d\mathfrak{z}, dt),
 \end{aligned}
 \tag{1.1}$$

where $b = (b^i)$, $B = (B^i)$, $\sigma = (\sigma^{ij})$ and $\rho = (\rho^{il})$ are Borel functions on $\mathbb{R}_+ \times \mathbb{R}^{d+d'}$, with values in \mathbb{R}^d , $\mathbb{R}^{d'}$, $\mathbb{R}^{d \times d_1}$ and $\mathbb{R}^{d \times d'}$, respectively, $\eta = (\eta^i)$ and $\xi = (\xi^i)$ are \mathbb{R}^d -valued $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}^{d+d'}) \otimes \mathcal{Z}_0$ -measurable and \mathbb{R}^d -valued $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}^{d+d'}) \otimes \mathcal{Z}_1$ -measurable functions on $\mathbb{R}_+ \times \mathbb{R}^{d+d'} \times \mathfrak{Z}_0$ and $\mathbb{R}_+ \times \mathbb{R}^{d+d'} \times \mathfrak{Z}_1$, respectively.

This paper is a continuation of [5, 7]. In [7] we derive the filtering equations, describing the time evolution of the conditional distribution $P_t(dx) = P(X_t \in dx | \mathcal{F}_t^Y)$ and the unnormalised conditional distribution $\mu_t(dx) = P_t(dx)\lambda_t$ of the unobserved component X_t given \mathcal{F}_t^Y , the σ -algebra generated by the observations $(Y_s)_{s \in [0, t]}$, where $(\lambda_t)_{t \in [0, T]}$ is a normalising positive process. The equation for $(\mu_t(dx))_{t \in [0, T]}$, referred to as Zakai equation, is linear, and therefore questions about existence, uniqueness and regularity of its solution for a given initial condition can be studied easier than those for the nonlinear filtering equation satisfied by $(P_t(dx))_{t \in [0, T]}$. For more details on the filtering equations for partially observed (jump) diffusions as well as for a historical account we refer to [7] and the references therein. In [5] it is shown that the conditional density $\pi_t = dP_t/dx$ exists for $t > 0$ and belongs to L_q for $q \in [1, p]$, if $\pi_0 = dP_0/dx \in L_p$ for a $p \geq 2$, the coefficients of (1.1) satisfy appropriate Lipschitz and growth conditions, and the derivatives of ξ and η in x are equicontinuous in x , uniformly in their other variables. The aim of the present paper is to show that if in addition to the Lipschitz and growth conditions in [5, 7], we assume that the coefficients have continuous and bounded derivatives in x up to order $m + 1$, for some integer $m \geq 0$, and $\pi_0 \in W_p^m$ for some $p \geq 2$, then $(\pi_t)_{t \geq 0}$ is a W_p^m -valued weakly cadlag process.

For partially observed diffusion processes, i.e., when $\xi = \eta = 0$ and the observation process Y does not have jumps, the existence and the regularity properties of the conditional density π_t have been extensively studied in the literature. In [14], an early work on the regularity of the filtering density for continuous diffusions, it was shown that if the coefficients are bounded, σ, ρ admit bounded derivatives in $x \in \mathbb{R}^d$ up to order $m + 1$, b, B admit bounded derivatives in x up to order m , $\sigma\sigma^*$ is uniformly non-degenerate and $\pi_0 \in W_p^m \cap W_2^m$, then the filtering density $(\pi_t)_{t \in [0, T]}$ is weakly

continuous as an W_p^m -valued process, where $p \geq 2$ and $m \geq 0$. Later, under the uniform non-degeneracy condition on $\sigma\sigma^*$, stronger results are obtained in [9, 13] by the help of the L_p -theory of SPDEs developed there. Generalisations of the linear filtering theory are presented in [12] by the help of the theory of SPDEs with VMO leading coefficients and growing lower order coefficients, see [10, 11].

In [20] it was proven that the non-degeneracy condition can be dropped if one imposes $m + 2$ bounded derivatives on σ, ρ in x , as well as $m + 1$ derivatives on b, B in x , to get that π is a W_p^m -valued weakly continuous process if $\pi_0 \in W_p^m \cap W_2^m$ for a $p \geq 2$ and $m \geq 1$. In [15] it was shown that the conditional density dP_t/dx exists for any $t > 0$ and it is in L_2 if $\pi_0 = dP_0/dx$ exists, it belongs to L_2 , and the coefficients are bounded and Lipschitz continuous. To achieve this, a nice calculation is presented in [15] to show that the $L_2(\Omega \times \mathbb{R}^d, \mathbb{R})$ -norm of $P_t^{(\varepsilon)}$, the conditional distribution P_t mollified by Gaussian kernels, can be estimated by the $L_2(\Omega \times \mathbb{R}^d, \mathbb{R})$ -norm of π_0 , independently of the mollification parameter $\varepsilon > 0$. For classical results in filtering theory we refer to [1] and the references therein.

More recently filtering densities associated to signal and observation models $(X_t, Y_t)_{t \geq 0}$ with jumps have been investigated in a growing number of publications. The above mentioned method from [15] was used in [2, 3, 17, 19], under various conditions on the filtering models, to prove that the conditional density $\pi_t = P(X_t \in dx | (Y_s)_{s \in [0, t]})/dx$ exists for $t > 0$ and it is in L_2 , when the initial density exists and belongs to L_2 . In [2] only the observation process has jumps. In [3] only the signal process has jumps due to an additive noise component which is a cadlag process of bounded variation, adapted to the observation process. In [17] a fairly general jump diffusion model is considered, but, as in [2, 3], the driving noises in the signal are independent of those in the observation. In [19] a general jump diffusion model with correlated signal and observation noises is considered, but as in the articles [2, 3, 17], the coefficients in the signal do not depend on the observation process. In the above publications, [2, 3, 17, 19], the coefficients of the SDE describing the signal and observation models are bounded and satisfy appropriate Lipschitz conditions. An approach from [16] is adapted to study the uniqueness of measure-valued solutions to the filtering equations for a model with jumps in [18], and the existence of the filtering density in L_2 is obtained when the initial filtering density is in L_2 and the jump component in the signal is a symmetric α -stable Lévy process. In this publication the signal and observation noises are independent of each other, the drift and diffusion coefficients in the signal process depend only on x , the variable for the signal, they are bounded and their derivatives up to first order and up second order, respectively, exist and are bounded functions.

As the present paper is a direct continuation of [5], it builds on the results of the latter. In [5] the method from [15], combined with methods from the theory of SPDEs, is applied to partially observed jump diffusions to show that the filtering density π_t exists for $t > 0$ and belongs to L_q for a $q \in [1, p]$, provided $\pi_0 \in L_p$ for a $p \geq 2$, the coefficients in the SDE satisfy appropriate Lipschitz conditions, the drift coefficient in the observation process is bounded, and the other coefficients satisfy a linear growth condition. In the present paper we investigate the regularity of the filtering density for the same filtering model as in [5]. In addition to these assumptions from [5], in the

present paper we assume that for an integer $m \geq 0$ the initial conditional density π_0 is in W_p^m for some $p \geq 2$ and the coefficients of the SDE admit bounded derivatives in x up to order $m + 1$, excluding the zeroth order. Under these conditions we prove that $(\pi_t)_{t \geq 0}$ is a W_p^m -valued weakly cadlag process. Moreover, we show that if the coefficients are also bounded, then $(\pi_t)_{t \geq 0}$ is a W_p^s -valued strongly cadlag process for any $s < m$.

We would like to mention that our method of proving the smoothness of the conditional density is different from those used in the case of partially observed diffusion processes, and we require the existence and boundedness of the derivatives of the coefficients only up to order $m + 1$ in order to get the conditional density in W_p^m .

This article is structured as follows. Section 2 contains the main results along with the required assumptions. In Section 3 we state some important results from [7] and [5] which we build on. Section 4 contains Sobolev estimates necessary to obtain a priori estimates for the smoothed filtering measures. In section 5 we investigate some solvability properties of the Zakai equation. Section 6 finally contains the proof of our main theorem, as well as some auxiliary results.

In conclusion we present important notions and notations used in this paper. For an integer $n \geq 0$ the notation $C_b^n(\mathbb{R}^d)$ means the space of real-valued bounded continuous functions on \mathbb{R}^d , which have bounded and continuous derivatives up to order n . (If $n = 0$, then $C_b^0(\mathbb{R}^d) = C_b(\mathbb{R}^d)$ denotes the space of real-valued bounded continuous functions on \mathbb{R}^d). We denote by $\mathbb{M} = \mathbb{M}(\mathbb{R}^d)$ the set of finite Borel measures on \mathbb{R}^d , and by $\mathfrak{M} = \mathfrak{M}(\mathbb{R}^d)$ the set of finite signed Borel measures on \mathbb{R}^d . For $\mu \in \mathfrak{M}$ we use the notation

$$\mu(\varphi) = \int_{\mathbb{R}^d} \varphi(x) \mu(dx)$$

for Borel functions φ on \mathbb{R}^d . We say that a function $v : \Omega \rightarrow \mathbb{M}$ is \mathcal{G} -measurable for a σ -algebra $\mathcal{G} \subset \mathcal{F}$, if $v(\varphi)$ is a \mathcal{G} -measurable random variable for every bounded Borel function φ on \mathbb{R}^d . An \mathbb{M} -valued stochastic process $v = (v_t)_{t \in [0, T]}$ is said to be weakly cadlag if almost surely $v_t(\varphi)$ is a cadlag function of t for all $\varphi \in C_b(\mathbb{R}^d)$. An \mathfrak{M} -valued process $(v_t)_{t \in [0, T]}$ is weakly cadlag, if it is the difference of two \mathbb{M} -valued weakly cadlag processes. For processes $U = (U_t)_{t \in [0, T]}$ we use the notation \mathcal{F}_t^U for the P -completion of the σ -algebra generated by $\{U_s : s \leq t\}$. By an abuse of notation, we often write \mathcal{F}_t^U when referring to the filtration $(\mathcal{F}_t^U)_{t \in [0, T]}$, whenever this is clear from the context. For a measure space $(\mathfrak{Z}, \mathcal{Z}, \nu)$ and $p \geq 1$ we use the notation $L_p(\mathfrak{Z}) = L_p(\mathfrak{Z}, \mathbb{R}^d)$ for the L_p -space of \mathbb{R}^d -valued \mathcal{Z} -measurable functions defined on \mathfrak{Z} . However, if not otherwise specified, the function spaces are considered to be over \mathbb{R}^d . We always use without mention the summation convention, by which repeated integer valued indices imply a summation. For multi-indices $\alpha = (\alpha_1, \dots, \alpha_d)$ of nonnegative integers $\alpha_i, i = 1, \dots, d$, functions φ of $x = (x^1, \dots, x^d) \in \mathbb{R}^d$ and nonnegative integers k we use the notations $|\alpha| = \sum_i \alpha_i$,

$$D^\alpha \varphi(x) = D_1^{\alpha_1} D_2^{\alpha_2} \dots D_d^{\alpha_d} \varphi(x), \quad \text{as well as} \quad |D^k \varphi|^2 = \sum_{|\gamma|=k} |D^\gamma \varphi|^2,$$

where $D_i = \frac{\partial}{\partial x^i}$ and $|\cdot|$ denotes an appropriate norm. We also use the notation $D_{ij} = D_i D_j$. If we want to stress that the derivative is taken in a variable x , we write D_x^α . If the norm $|\cdot|$ is not clear from the context, we sometimes use appropriate subscripts, as in $|\varphi|_{L_p}$ for the $L_p(\mathbb{R}^d)$ -norm of φ . For $p \geq 1$ and integers $m \geq 0$ we use the notation W_p^m for Borel functions $f = f(x)$ on \mathbb{R}^d such that

$$|f|_{W_p^m}^p := \sum_{k=0}^m \int_{\mathbb{R}^d} |D^k f(x)|^p dx < \infty.$$

Throughout the paper we work on the finite time interval $[0, T]$, where $T > 0$ is fixed but arbitrary, as well as on a given complete probability space (Ω, \mathcal{F}, P) equipped with a filtration $(\mathcal{F}_t)_{t \geq 0}$ such that \mathcal{F}_0 contains all the P -null sets. For $p, q \geq 1$ and integers $m \geq 1$ we denote by $\mathbb{W}_p^m = L_p((\Omega, \mathcal{F}_0, P), W_p^m(\mathbb{R}^d))$ and $\mathbb{W}_{p,q}^m \subset L_p(\Omega, L_q([0, T], W_p^m(\mathbb{R}^d)))$ the set of $\mathcal{F}_0 \otimes \mathcal{B}(\mathbb{R}^d)$ -measurable real-valued functions $f = f(\omega, x)$ and \mathcal{F}_t -optional W_p^m -valued processes $g = g_t(\omega, x)$ such that

$$|f|_{\mathbb{W}_p^m}^p := \mathbb{E}|f|_{W_p^m}^p < \infty \quad \text{and} \quad |g|_{\mathbb{L}_{p,q}}^p := \mathbb{E}\left(\int_0^T |g_t|_{W_p^m}^q dt\right)^{p/q} < \infty$$

respectively. If $m = 0$ we set $\mathbb{L}_p = \mathbb{W}_p^0$ and $\mathbb{L}_{p,q} = \mathbb{W}_{p,q}^0$. In case a different σ -algebra \mathcal{G} than \mathcal{F}_0 is considered above, we denote this explicitly by $\mathbb{L}_p(\mathcal{G})$ and $\mathbb{W}_p^m(\mathcal{G})$. If $m \geq 0$ is not an integer and $p > 1$ then W_p^m denotes the space of real-valued generalised functions h on \mathbb{R}^d such that

$$|h|_{W_p^m} := |(1 - \Delta)^{m/2} h|_{L_p} < \infty.$$

A W_p^m -valued process $(\pi_t)_{t \in [0, T]}$ is said to be weakly cadlag if it is a weakly cadlag process in the weak topology in W_p^m , and it is said to be strongly cadlag if it is a cadlag process in the strong topology in W_p^m . Finally, for real-valued functions f and g on \mathbb{R}^d , we often denote by (f, g) the integral of $f \cdot g$ over \mathbb{R}^d .

2 Formulation of the main results

We fix nonnegative constants K_0, K_1, L, K, T and functions $\bar{\xi} \in L_2(\mathfrak{Z}_1) = L_2(\mathfrak{Z}_1, \mathcal{Z}_1, \nu_1), \bar{\eta} \in L_2(\mathfrak{Z}_0) = L_2(\mathfrak{Z}_0, \mathcal{Z}_0, \nu_0)$, used throughout the paper, and make the following assumptions.

Assumption 2.1 (i) For $z_j = (x_j, y_j) \in \mathbb{R}^{d+d'}$ ($j = 1, 2$), $t \geq 0$ and $\mathfrak{z}_i \in \mathfrak{Z}_i$ ($i = 0, 1$),

$$\begin{aligned} &|b(t, z_1) - b(t, z_2)| + |B(t, z_1) - B(t, z_2)| + |\sigma(t, z_1) - \sigma(t, z_2)| \\ &\quad + |\rho(t, z_1) - \rho(t, z_2)| \leq L|z_1 - z_2|, \\ &|\eta(t, z_1, \mathfrak{z}_0) - \eta(t, z_2, \mathfrak{z}_0)| \leq \bar{\eta}(\mathfrak{z}_0)|z_1 - z_2|, \\ &|\xi(t, z_1, \mathfrak{z}_1) - \xi(t, z_2, \mathfrak{z}_1)| \leq \bar{\xi}(\mathfrak{z}_1)|z_1 - z_2|. \end{aligned}$$

(ii) For all $z = (x, y) \in \mathbb{R}^{d+d'}$, $t \geq 0$ and $\mathfrak{z}_i \in \mathfrak{Z}_i$ for $i = 0, 1$ we have

$$|b(t, z)| + |\sigma(t, z)| + |\rho(t, z)| \leq K_0 + K_1|z|, \quad |B(t, z)| \leq K, \quad \int_{\mathfrak{Z}_1} |\mathfrak{z}|^2 \nu_1(d\mathfrak{z}) \leq K_0^2,$$

$$|\eta(t, z, \mathfrak{z}_0)| \leq \bar{\eta}(\mathfrak{z}_0)(K_0 + K_1|z|), \quad |\xi(t, z, \mathfrak{z}_1)| \leq \bar{\xi}(\mathfrak{z}_1)(K_0 + K_1|z|).$$

(iii) The initial condition $Z_0 = (X_0, Y_0)$ is an \mathcal{F}_0 -measurable random variable with values in $\mathbb{R}^{d+d'}$.

Assumption 2.2 The functions $\bar{\eta} \in L_2(\mathfrak{Z}_0, \mathcal{Z}_0, \nu_0)$ and $\bar{\xi} \in L_2(\mathfrak{Z}_1, \mathcal{Z}_1, \nu_1)$ are such that for constants K_η and K_ξ we have $\bar{\eta}(\mathfrak{z}_0) \leq K_\eta$ and $\bar{\xi}(\mathfrak{z}_1) \leq K_\xi$ for all $\mathfrak{z}_0 \in \mathfrak{Z}_0$, $\mathfrak{z}_1 \in \mathfrak{Z}_1$.

Assumption 2.3 For some $r > 2$ let $\mathbb{E}|X_0|^r < \infty$ and the measure ν_1 satisfy

$$K_r := \int_{\mathfrak{Z}_1} |\mathfrak{z}|^r \nu_1(d\mathfrak{z}) < \infty.$$

By a well-known theorem of Itô one knows that Assumption 2.1 ensures the existence and uniqueness of a solution $(X_t, Y_t)_{t \geq 0}$ to (1.1) for any given \mathcal{F}_0 -measurable initial value $Z_0 = (X_0, Y_0)$, and for every $T > 0$,

$$\mathbb{E} \sup_{t \leq T} (|X_t|^q + |Y_t|^q) \leq N(1 + \mathbb{E}|X_0|^q + \mathbb{E}|Y_0|^q) \tag{2.1}$$

holds for $q = 2$ with a constant N depending only on $T, K_0, K_1, K_2, |\bar{\eta}|_{L_2}, |\bar{\xi}|_{L_2}$ and $d + d'$. If in addition to Assumption 2.1 we assume Assumptions 2.2 and 2.3, then it is known, see e.g. [4], that the moment estimate (2.1) holds with $q := r$ for every $T > 0$, where now the constant N depends also on r, K_r, K_ξ and K_η .

Assumption 2.4 (i) For a constant $\lambda > 0$ we have

$$\lambda|x - \bar{x}| \leq |x - \bar{x} + \theta(f_i(t, x, y, \mathfrak{z}_i) - f_i(t, \bar{x}, y, \mathfrak{z}_i))|$$

for all $\theta \in [0, 1], t \in [0, T], y \in \mathbb{R}^{d'}, x, \bar{x} \in \mathbb{R}^d, \mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$, and $f_0(t, x, y, \mathfrak{z}_0) = \eta(t, x, y, \mathfrak{z}_0), f_1(t, x, y, \mathfrak{z}_1) = \xi(t, x, y, \mathfrak{z}_1)$.

(ii) For all $(t, y) \in \mathbb{R}_+ \times \mathbb{R}^{d'}$ and all $x_1, x_2 \in \mathbb{R}^d$,

$$|(\rho B)(t, x_1, y) - (\rho B)(t, x_2, y)| \leq L|x_1 - x_2|.$$

(iii) The functions $f_0(t, x, y, \mathfrak{z}) := \xi(t, x, y, \mathfrak{z})$ and $f_1(t, x, y, \mathfrak{z}) := \eta(t, x, y, \mathfrak{z})$ are continuously differentiable in $x \in \mathbb{R}^d$ for each $(t, y, \mathfrak{z}) \in \mathbb{R}_+ \times \mathbb{R}^{d'} \times \mathfrak{Z}_i$, for $i = 0$ and $i = 1$, respectively, such that

$$\lim_{\varepsilon \downarrow 0} \sup_{t \in [0, T]} \sup_{\mathfrak{z} \in \mathfrak{Z}_i} \sup_{|y| \leq R} \sup_{|x| \leq R, |\bar{x}| \leq R, |x - x'| \leq \varepsilon} |D_x f_i(t, x, y, \mathfrak{z}) - D_x f_i(t, \bar{x}, y, \mathfrak{z})| = 0$$

for every $R > 0$.

Assumption 2.5 Let $m \geq 0$ be an integer.

- (i) The partial derivatives in $x \in \mathbb{R}^d$ of the coefficients $b, B, \sigma, \rho, (\rho B), \eta$ and ξ up to order $m + 1$ are functions such that

$$\sum_{k=1}^{m+1} |D_x^k(b, B, \sigma, \rho, (\rho B))| \leq L \quad \text{for all } t \in [0, T], x \in \mathbb{R}^d, y \in \mathbb{R}^{d'}.$$

- (ii) Moreover,

$$\sum_{k=1}^{m+1} |D_x^k \eta| \leq L \bar{\eta}, \quad \sum_{k=1}^{m+1} |D_x^k \xi| \leq L \bar{\xi},$$

for all $t \in [0, T], x \in \mathbb{R}^d, y \in \mathbb{R}^{d'}$ and $\mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$.

A few remarks are in order to elaborate on the need for the assumptions made above.

Remark 2.1 The function B is required to be bounded by Assumption 2.1. Though this excludes linear drifts in the observation (in particular it excludes the important special case of Kalman-Bucy filtering), it is a necessary assumption in the case of our general model in order to obtain crucial L_p estimates needed later in this paper (see for instance Lemma 4.4). To include the case of B satisfying a linear growth condition one should assume specific conditions on the filtering model, which allow one to transform the Zakai equation into an SPDE without a zero order term. (See [12] on including the case of Kalman-Bucy filtering in nonlinear models.)

Remark 2.2 It is a natural condition in Assumption 2.2 that $\bar{\eta} \in L_2(\mathfrak{Z}_0, \mathcal{Z}_0, \nu_0)$ and $\bar{\xi} \in L_2(\mathfrak{Z}_1, \mathcal{Z}_1, \nu_1)$. This, together with the measurability and growth conditions on ξ and η ensures, in particular, that for any adapted cadlag process $Z = (X, Y)$ the stochastic differentials with respect to the martingale measures in equation (1.1) are well-defined. Note that since the measures ν_0 and ν_1 may not be finite, the boundedness of $\bar{\eta}$ and $\bar{\xi}$ does not imply that $\bar{\eta} \in L_2(\mathfrak{Z}_0, \mathcal{Z}_0, \nu_0)$ and $\bar{\xi} \in L_2(\mathfrak{Z}_1, \mathcal{Z}_1, \nu_1)$. The boundedness condition on $\bar{\xi}$ and $\bar{\eta}$ in Assumption 2.2 serves a different purpose. This, together with Assumption 2.4(i) and the Lipschitz condition on η and ξ in Assumption 2.1(i) implies that for a constant $c = c(\lambda, K_\xi, K_\eta)$ we have for all $\theta \in [0, 1], y \in \mathbb{R}^{d'}, t \in [0, T]$ and $\mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$,

$$c^{-1} |x - \bar{x}| \leq |x - \bar{x} + \theta(f_i(t, x, y, \mathfrak{z}_i) - f(t, \bar{x}, y, \mathfrak{z}_i))| \leq c |x - \bar{x}| \quad \text{for } x, \bar{x} \in \mathbb{R}^d,$$

with $f_0(t, x, y, \mathfrak{z}_0) := \eta(t, x, y, \mathfrak{z}_0)$ and $f_1(t, x, y, \mathfrak{z}_1) := \xi(t, x, y, \mathfrak{z}_1)$. This, together with Assumption 2.4(iii) in particular implies that for all $\theta \in [0, 1], y \in \mathbb{R}^{d'}, t \in [0, T]$ and $\mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$ the mappings

$$\tau^\eta(x) = x + \theta \eta(t, x, y, \mathfrak{z}_0) \quad \text{and} \quad \tau^\xi(x) = x + \theta \xi(t, x, y, \mathfrak{z}_1)$$

are C^1 -diffeomorphisms. This property plays a crucial role in the existence of the conditional density and in obtaining L_p -estimates for it.

Remark 2.3 Note that if τ is a continuously differentiable L -biLipschitz function on \mathbb{R}^d , i.e., for a constant $L \geq 1$ we have

$$L^{-1}|x - y| \leq |\tau(x) - \tau(y)| \leq L|x - y| \quad \text{for all } x, y \in \mathbb{R}^d,$$

then

$$L^{-d} \leq |\det(D\tau(x))| \leq L^d \quad \text{for } x \in \mathbb{R}^d.$$

Proof For $d = 1$ it is obvious, and for $d > 1$ it can be easily shown by using the singular value decomposition for the matrices $D\tau$, $D\tau^{-1}$, or by applying Hadamard’s inequality to their determinants. \square

Recall that \mathcal{F}_t^Y denotes the completion of the σ -algebra generated by $(Y_s)_{s \leq t}$.

Theorem 2.1 *Let Assumptions 2.1, 2.2, 2.4 and 2.5 hold. If $K_1 \neq 0$ in Assumption 2.1, then let additionally Assumption 2.3 hold. Assume the conditional density $\pi_0 = P(X_0 \in dx | \mathcal{F}_0^Y) / dx$ exists almost surely and for some $p \geq 2$ and integer $m \geq 0$ we have $\mathbb{E}|\pi_0|_{W_p^m}^p < \infty$. Then almost surely $P(X_t \in dx | \mathcal{F}_t^Y) / dx$ exists and belongs to W_p^m for every $t \in [0, T]$.*

Moreover, there is a W_p^m -valued weakly cadlag process $\pi = (\pi_t)_{t \in [0, T]}$ such that for each $t \in [0, T]$ almost surely $\pi_t = P(X_t \in dx | \mathcal{F}_t^Y) / dx$. If $K_1 = 0$ and $m \geq 1$, then π is strongly cadlag as W_p^s -valued process for $s \in [0, m)$.

If $a := \sigma\sigma^* / 2$ is uniformly elliptic, i.e., for a constant $c > 0$ we have

$$a^{ij}(t, z)\bar{x}^i\bar{x}^j \geq c|\bar{x}|^2 \quad \text{for all } (t, z) \in [0, T] \times \mathbb{R}^{d+d'} \text{ and } \bar{x} = (\bar{x}^1, \dots, \bar{x}^d) \in \mathbb{R}^d, \tag{2.2}$$

then the following remark states that we can recover in the special case of $p = 2$, in the presence of jumps, a well-known result obtained in [14] for partially observed diffusion processes. Based on calculations in the present paper, we will prove it together with its generalisation for $p \in [2, \infty)$ in a forthcoming article.

Remark 2.4 Let (2.2) and Assumptions 2.1, 2.2 and 2.4 (i)&(ii) hold with $K_1 = 0$. Assume that for an integer $m \geq 0$ the conditional density $\pi_0 = P(X_0 \in dx) / dx$ exists almost surely such that $\mathbb{E}|\pi_0|_{W_2^m}^2 < \infty$, and the derivatives in $x \in \mathbb{R}^d$ of the functions σ, ρ, ξ and η up to order $m + 1$, and of the functions b and B up to order m are functions, in magnitude bounded by a constant. Then for each $t \in [0, T]$ the conditional density $P(X_t \in dx | \mathcal{F}_t^Y) / dx$ exists and belongs to W_2^m almost surely, and there is a W_2^m -valued cadlag process $\pi = (\pi_t)_{t \in [0, T]}$ such that $\pi_t = P(X_t \in dx | \mathcal{F}_t^Y) / dx$ (a.s.) for each $t \in [0, T]$. Moreover, $\pi_t \in W_2^{m+1}$ for $P \otimes dt$ -a.e. $(\omega, t) \in \Omega \times [0, T]$ such that

$$\int_0^T |\pi_t|_{W_2^{m+1}}^2 dt < \infty \quad \text{almost surely.}$$

3 Preliminaries

Recall the notions and notations concerning measure-valued processes, given in the final part of the Introduction, and note that if $(\nu_t)_{t \geq 0}$ is an \mathbb{M} -valued weakly cadlag process then there is a set $\Omega' \subset \Omega$ of full probability and there is uniquely defined (up to indistinguishability) \mathbb{M} -valued processes $(\nu_{t-})_{t \geq 0}$ such that for every $\omega \in \Omega'$

$$\nu_{t-}(\varphi) = \lim_{s \uparrow t} \nu_s(\varphi) \quad \text{for all } \varphi \in C_b(\mathbb{R}^d) \text{ and } t > 0,$$

and for each $\omega \in \Omega'$ we have $\nu_{t-} = \nu_t$, for all but at most countably many $t \in (0, \infty)$. The following result was proven in [7]. In order to formulate it, we define

$$\gamma_t = \exp\left(-\int_0^t B(s, X_s, Y_s) dV_s - \frac{1}{2} \int_0^t |B(s, X_s, Y_s)|^2 ds\right), \quad t \in [0, T],$$

and note that since B is bounded in magnitude, we have $\mathbb{E}\gamma_T = 1$ and it is an \mathcal{F}_t -martingale under P . Thus, the measure Q , defined by $dQ := \gamma_T dP$ is a probability measure, equivalent to P .

Theorem 3.1 *Let Assumption 2.1 hold. If $K_1 \neq 0$, then assume also $\mathbb{E}|X_0|^2 < \infty$. Then there exist measure-valued \mathcal{F}_t^Y -adapted weakly cadlag processes $(P_t)_{t \in [0, T]}$ and $(\mu_t)_{t \in [0, T]}$ such that*

$$P_t(\varphi) = \mu_t(\varphi) / \mu_t(\mathbf{1}), \quad \text{for } \omega \in \Omega, \quad t \in [0, T],$$

$$P_t(\varphi) = \mathbb{E}(\varphi(X_t) | \mathcal{F}_t^Y), \quad \mu_t(\varphi) = \mathbb{E}_Q(\gamma_t^{-1} \varphi(X_t) | \mathcal{F}_t^Y) \quad (\text{a.s.}) \text{ for each } t \in [0, T].$$

We refer to μ_t (resp. P_t) as the unnormalised (resp. normalised) conditional distribution of X_t given \mathcal{F}_t^Y , $t \in [0, T]$.

We introduce the random differential operators

$$\tilde{\mathcal{L}}_t = a_t^{ij}(x) D_{ij} + b_t^i(x) D_i + \beta_t^k \mathcal{M}_t^k, \quad \mathcal{M}_t^k = \rho_t^{ik}(x) D_i + B_t^k(x), \quad k = 1, 2, \dots, d', \tag{3.1}$$

where $\beta_t := B_t(X_t)$ and

$$a_t^{ij}(x) := \frac{1}{2} \sum_{k=1}^{d_1} (\sigma_t^{ik} \sigma_t^{jk})(x) + \frac{1}{2} \sum_{l=1}^{d'} (\rho_t^{il} \rho_t^{jl})(x), \quad \sigma_t^{ik}(x) := \sigma^{ik}(t, x, Y_t),$$

$$\rho_t^{il}(x) := \rho^{il}(t, x, Y_t),$$

$$b_t^i(x) := b^i(t, x, Y_t), \quad B_t^k(x) := B^k(t, x, Y_t)$$

for $\omega \in \Omega$, $t \geq 0$, $x = (x^1, \dots, x^d) \in \mathbb{R}^d$, and $D_i = \partial/\partial x^i$, $D_{ij} = \partial^2/(\partial x^i \partial x^j)$ for $i, j = 1, 2, \dots, d$. Moreover for every $t \geq 0$ and $\mathfrak{z} \in \mathfrak{Z}_1$ we introduce the random operators I_t^ξ and J_t^ξ defined by

$$T_t^\xi \varphi(x, \mathfrak{z}) = \varphi(x + \xi_t(x, \mathfrak{z}), \mathfrak{z})$$

$$I_t^\xi \varphi(x, \mathfrak{z}) = T_t^\xi \varphi(x, \mathfrak{z}) - \varphi(x, \mathfrak{z}), \quad J_t^\xi \phi(x, \mathfrak{z}) = I_t^\xi \phi(x, \mathfrak{z}) - \sum_{i=1}^d \xi_t^i(x, \mathfrak{z}) D_i \phi(x, \mathfrak{z})$$

for functions $\varphi = \varphi(x, \mathfrak{z})$ and $\phi = \phi(x, \mathfrak{z})$ of $x \in \mathbb{R}^d$ and $\mathfrak{z} \in \mathfrak{Z}_1$, and furthermore the random operators I_t^η and J_t^η , defined as I_t^ξ and J_t^ξ , respectively, with $\eta_t(x, \mathfrak{z})$ in place of $\xi_t(x, \mathfrak{z})$, where

$$\xi_t(x, \mathfrak{z}_1) := \xi(t, x, Y_{t-}, \mathfrak{z}_1), \quad \eta_t(x, \mathfrak{z}_0) := \eta(t, x, Y_{t-}, \mathfrak{z}_0)$$

for $\omega \in \Omega, t \geq 0, x \in \mathbb{R}^d$ and $\mathfrak{z}_i \in \mathfrak{Z}_i$ for $i = 0, 1$.

From [5] we know that if the unnormalised conditional distribution μ_t has a density such that $u_t = d\mu_t/dx$ (a.s.) for each $t \in [0, T]$ for an L_p -valued weakly cadlag process $(u_t)_{t \in [0, T]}$ for some $p \geq 2$, then it satisfies for each $\varphi \in C_0^\infty$ almost surely

$$\begin{aligned} (u_t, \varphi) &= (\psi, \varphi) + \int_0^t (u_s, \tilde{\mathcal{L}}_s \varphi) ds + \int_0^t (u_s, \mathcal{M}_s^k \varphi) dV_s^k + \int_0^t \int_{\mathfrak{Z}_0} (u_s, J_s^\eta \varphi) \nu_0(d\mathfrak{z}) ds \\ &+ \int_0^t \int_{\mathfrak{Z}_1} (u_s, J_s^\xi \varphi) \nu_1(d\mathfrak{z}) ds + \int_0^t \int_{\mathfrak{Z}_1} (u_{s-}, I_s^\xi \varphi) \tilde{N}_1(d\mathfrak{z}, ds). \end{aligned} \tag{3.2}$$

for all $t \in [0, T]$. Formally we may write (3.2) as the Cauchy problem

$$\begin{aligned} du_t &= \tilde{\mathcal{L}}_t^* u_t dt + \mathcal{M}_t^{*k} u_t dV_t^k + \int_{\mathfrak{Z}_0} J_t^{\eta*} u_t \nu_0(d\mathfrak{z}) dt \\ &+ \int_{\mathfrak{Z}_1} J_t^{\xi*} u_t \nu_1(d\mathfrak{z}) dt + \int_{\mathfrak{Z}_1} I_t^{\xi*} u_{t-} \tilde{N}_1(d\mathfrak{z}, dt), \\ u_0 &= \psi. \end{aligned} \tag{3.3}$$

for a given ψ .

Definition 3.1 Let integers $m \geq 0$ and $p \geq 2$. Let ψ be an W_p^m -valued \mathcal{F}_0 -measurable random variable. Then we say that a W_p^m -valued \mathcal{F}_t -adapted weakly cadlag process $(u_t)_{t \in [0, T]}$ is a W_p^m -solution of (3.3) with initial condition ψ , if for each $\varphi \in C_0^\infty$ almost surely (3.2) holds for every $t \in [0, T]$.

If $m = 0$, then we call u an L_p -solution instead of a W_p^0 -solution.

As in [5] we are interested in solutions which satisfy

$$\text{ess sup}_{t \in [0, T]} |u_t|_{L_1} < \infty \quad \text{and} \quad \sup_{t \in [0, T]} \int_{\mathbb{R}^d} |y|^2 |u_t(y)| dy < \infty \quad (\text{a.s.}) \tag{3.4}$$

To formulate the following results from [5], Lemma 5.7 and Theorem 2.1 therein, we recall that there exists a cadlag \mathcal{F}_t^Y -adapted process $({}^o\gamma_t)_{t \in [0, T]}$, called the optional projection of $(\gamma_t)_{t \in [0, T]}$ under P with respect to $(\mathcal{F}_t^Y)_{t \in [0, T]}$, such that for every \mathcal{F}_t^Y -stopping time $\tau \leq T$ we have

$$\mathbb{E}(\gamma_\tau | \mathcal{F}_\tau^Y) = {}^o\gamma_\tau \quad \text{almost surely.} \tag{3.5}$$

Since for each t , by known properties of conditional expectations, almost surely

$$\mu_t(\mathbf{1}) = \mathbb{E}_Q(\gamma_t^{-1} | \mathcal{F}_t^Y) = 1/\mathbb{E}(\gamma_t | \mathcal{F}_t^Y) = 1/\gamma_t$$

and P, μ are weakly cadlag in the sense described above, we also have that almost surely $P_t(\varphi) = \mu_t(\varphi)\gamma_t$ for each $t \in [0, T]$ and $\varphi \in C_b$.

Theorem 3.2 *Let Assumptions 2.1, 2.2 and 2.4 hold. If $K_1 \neq 0$, then let additionally Assumption 2.3 hold for some $r > 2$. Assume the conditional density $\pi_0 = P(X_0 \in dx | Y_0)/dx$ exists almost surely and $\mathbb{E}|\pi_0|_{L_p}^p < \infty$ for some $p \geq 2$. Then*

(i) *For each $t \in [0, T]$ the unnormalised conditional density $d\mu_t/dx$ exists almost surely and there is an L_p -valued weakly cadlag process $(u_t)_{t \in [0, T]}$ such that $u_t = d\mu_t/dx$ (a.s.) for each $t \in [0, T]$ and*

$$\mathbb{E} \sup_{t \in [0, T]} |u_t|_{L_p}^p \leq N \mathbb{E}|\pi_0|_{L_p}^p$$

for a constant $N = N(d, d', p, K, K_\xi, K_\eta, L, T, \lambda, \|\bar{\xi}\|_{L_2}, \|\bar{\eta}\|_{L_2})$. Moreover, u is the unique L_2 -solution to (3.3) satisfying the conditions in (3.4).

(ii) *Almost surely the conditional density $P(X_t \in dx | \mathcal{F}_t^Y)/dx$ exists and belongs to L_p for all $t \in [0, T]$. Moreover, there is an L_p -valued weakly cadlag process $(\pi_t)_{t \in [0, T]}$, such that for each $t \in [0, T]$ almost surely $\pi_t = P(X_t \in dx | \mathcal{F}_t^Y)/dx$, as well as almost surely $\pi_t = u_t \gamma_t$ for all $t \in [0, T]$.*

Proof See Lemma 5.5 and Theorem 2.1 in [5]. □

Lemma 3.3 *Let $1 < p < \infty$ and let $(v_t)_{t \in [0, T]}$ be a weakly cadlag L_p -valued process. Assume moreover that for an $m \geq 0$ almost surely $\text{ess sup}_{t \in [0, T]} |v_t|_{W_p^m} < \infty$ and $v_T \in W_p^m$. Then v is (almost surely) a weakly cadlag W_p^m -valued process.*

Proof Let Ω' be the set of those $\omega \in \Omega$ such that $(v_t(\omega))_{t \in [0, T]}$ is weakly cadlag as an L_p -valued function, $v_T(\omega) \in W_p^m$ and $\text{ess sup}_{t \in [0, T]} |v_t(\omega)|_{W_p^m} < \infty$. Then $P(\Omega') = 1$, and for each $\omega \in \Omega'$ there exists a dense subset \mathbb{T}_ω in $[0, T]$ such that $\sup_{t \in \mathbb{T}_\omega} |v_t(\omega)|_{W_p^m} < \infty$. If $\omega \in \Omega'$ and $t \notin \mathbb{T}_\omega, t \neq T$, then there exists a sequence $(t_n)_{n=1}^\infty \subset \mathbb{T}_\omega$ such that $t_n \downarrow t$. Since $\sup_{t \in \mathbb{T}_\omega} |v_t(\omega)|_{W_p^m} < \infty$ there exists a subsequence, also denoted by $(t_n)_{n=1}^\infty$, such that $v_{t_n}(\omega)$ converges weakly in W_p^m to some element $\tilde{v} \in W_p^m$. However, as v is weakly cadlag as an L_p -valued process, we know that $v_{t_n} \rightarrow v_t$ weakly in L_p as $n \rightarrow \infty$ and hence $\tilde{v} = v_t \in W_p^m$. Thus clearly also $\sup_{t \in [0, T]} |v_t(\omega)|_{W_p^m} < \infty$ if $\omega \in \Omega'$. To see that v is weakly cadlag as a W_p^m -valued process, note first that since W_p^m is a reflexive space, which is embedded continuously and densely into L_p , we have that the dual $(L_p)^* = L_q, q = p/(p-1)$, is embedded continuously and densely into $(W_p^m)^*$. Therefore, for each $\varepsilon > 0$ and $\phi \in (W_p^m)^*$ there is an $\phi_\varepsilon \in L_q$ such that $|\phi - \phi_\varepsilon|_{(W_p^m)^*} < \varepsilon$. Fix a $t \in [0, T]$ and a sequence $t_n \downarrow t$. Then

$$\begin{aligned} |(v_{t_n}, \phi) - (v_t, \phi)| &\leq |(v_{t_n}, \phi - \phi_\varepsilon)| + |(v_{t_n}, \phi_\varepsilon) - (v_t, \phi_\varepsilon)| + |(v_t, \phi_\varepsilon - \phi)| \\ &\leq 2\varepsilon \sup_{t \in [0, T]} |v_t|_{W_p^m} + |(v_{t_n}, \phi_\varepsilon) - (v_t, \phi_\varepsilon)|. \end{aligned}$$

Recalling that v is weakly cadlag as an L_p -valued process finishes the proof. \square

The corollary of the following lemma will play an essential role in the proof of the statement on the strong cadlagness of L_p -solutions to the filtering equations, see Proposition 6.4.

Lemma 3.4 *Let ζ be an \mathbb{R}^d -valued function on \mathbb{R}^d such that for an integer $m \geq 1$ it is continuously differentiable up to order m , and*

$$\inf_{\theta \in [0,1]} \inf_{x \in \mathbb{R}^d} |\det(\mathbb{I} + \theta D\zeta(x))| =: \lambda > 0, \quad \max_{0 \leq k \leq m} \sup_{x \in \mathbb{R}^d} |D^k \zeta(x)| =: M_m < \infty. \tag{3.6}$$

Then the following statements hold.

(i) *The function $\tau = x + \theta\zeta(x)$, $x \in \mathbb{R}^d$, is a C^m -diffeomorphism for each θ , such that for all $x \in \mathbb{R}^d$, $\theta \in [0, 1]$,*

$$\lambda' \leq |\det D\tau^{-1}(x)| \leq \lambda'', \quad \text{and} \quad \max_{1 \leq k \leq m} \sup_{x \in \mathbb{R}^d} |D^k \tau^{-1}| \leq M'_m < \infty, \tag{3.7}$$

with constants $\lambda' = \lambda'(d, M_1) > 0$, $\lambda'' = \lambda''(d, M_1)$ and $M'_m = M'_m(d, \lambda, M_m)$.

(ii) *The function $\zeta^*(x) = -x + \tau^{-1}(x)$, $x \in \mathbb{R}^d$, is continuously differentiable up to order m , such that*

$$\begin{aligned} \sup_{\mathbb{R}^d} |\zeta^*| &= \sup_{\mathbb{R}^d} |\zeta|, \tag{3.8} \\ \sup_{\mathbb{R}^d} |D^k \zeta^*| &\leq M_m^* \max_{1 \leq j \leq k} \sup_{\mathbb{R}^d} |D^j \zeta|, \quad \text{for } k = 1, 2, \dots, m, \tag{3.9} \end{aligned}$$

$$\inf_{\theta \in [0,1]} \inf_{\mathbb{R}^d} |\det(\mathbb{I} + \theta D\zeta^*)| \geq \lambda' \inf_{\theta \in [0,1]} \inf_{\mathbb{R}^d} |\det(\mathbb{I} + \theta D\zeta)|, \tag{3.10}$$

with a constant $M_m^* = M_m^*(d, \lambda, M_m)$ and with λ' from (3.7).

(iii) *For the function $\mathbf{c} = \det(\mathbb{I} + D\zeta^*) - 1$ we have*

$$\sup_{x \in \mathbb{R}^d} |D^k \mathbf{c}(x)| \leq N \max_{1 \leq j \leq k+1} \sup_{\mathbb{R}^d} |D^j \zeta|, \tag{3.11}$$

for $0 \leq k \leq m - 1$ with a constant $N = N(d, \lambda, m, M_m)$.

Proof Claims (i) and (ii) are Lemma 6.1 in [5]. To prove (iii) notice that for the function $F(A) = \det A$, considered as the function of the entries A^{ij} of $d \times d$ real matrices A , by Taylor’s formula

$$\mathbf{c} = \det(\mathbb{I} + D\zeta^*) - \det \mathbb{I} = \int_0^1 \frac{\partial}{\partial A^{ij}} F(\mathbb{I} + \theta D\zeta^*) d\theta D_i \zeta^{*j}.$$

\square

Corollary 3.5 For \mathbb{R}^d -valued functions ζ on \mathbb{R}^d we define the operators T^ζ , I^ζ and J^ζ by

$$T^\zeta \varphi(x) = \varphi(x + \zeta(x)) \text{ for } x \in \mathbb{R}^d, \text{ and } I^\zeta \varphi = T^\zeta \varphi - \varphi, \quad J^\zeta \varphi = I^\zeta - \zeta^i D_i \varphi$$

for differentiable functions φ on \mathbb{R}^d . Assume that ζ satisfies the conditions of Lemma 3.4 with $m = 2$. Then $\tau_{\theta\zeta}(x) = x + \theta\zeta(x)$, $x \in \mathbb{R}^d$ is a C^2 -diffeomorphism for each $\theta \in [0, 1]$, and for every $v, \varphi \in C_0^\infty$ we have

$$(v, I^\zeta \varphi) = (I^{\zeta*} v, \varphi), \quad (v, J^\zeta \varphi) = (K_i^\zeta v, D_i \varphi), \tag{3.12}$$

with

$$I^{\zeta*} v(x) = - \int_0^1 D_i \left(v(\tau_{\theta\zeta}^{-1}(x)) \zeta^i(\tau_{\theta\zeta}^{-1}(x)) |\det D\tau_{\theta\zeta}^{-1}(x)| \right) d\theta,$$

$$K_i^\zeta v(x) = \int_0^1 (\theta - 1) D_j \left(v(\tau_{\theta\zeta}^{-1}(x)) \zeta^i(\tau_{\theta\zeta}^{-1}(x)) \zeta^j(\tau_{\theta\zeta}^{-1}(x)) |\det D\tau_{\theta\zeta}^{-1}(x)| \right) d\theta, \quad x \in \mathbb{R}^d,$$

for $i = 1, 2, \dots, d$. Moreover, for every $x \in \mathbb{R}^d$ we have

$$|I^{\zeta*} v(x)| \leq N \sup_{x \in \mathbb{R}^d} (|\zeta(x)| + |D\zeta(x)|) \int_0^1 |v(\tau_{\theta\zeta}^{-1}(x))| + |(Dv)(\tau_{\theta\zeta}^{-1}(x))| d\theta \tag{3.13}$$

$$|K_i^\zeta v(x)| \leq N \sup_{x \in \mathbb{R}^d} (|\zeta(x)|^2 + |D\zeta(x)|^2) \int_0^1 |v(\tau_{\theta\zeta}^{-1}(x))| + |(Dv)(\tau_{\theta\zeta}^{-1}(x))| d\theta \tag{3.14}$$

with a constant $N = N(d, \lambda, M_1, M_2)$.

Proof By Lemma 3.4 $\tau_{\theta\zeta} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a C^2 -diffeomorphism for every $\theta \in [0, 1]$. Using Taylor’s formula we have

$$(v, I^\zeta \varphi) = \int_{\mathbb{R}^d} \int_0^1 v(x) (D_i \varphi)(\tau_{\theta\zeta}(x)) \zeta^i(x) d\theta dx,$$

$$(v, J^\zeta \varphi) = \int_{\mathbb{R}^d} \int_0^1 (1 - \theta) v(x) (D_{ij} \varphi)(\tau_{\theta\zeta}(x)) \zeta^i(x) \zeta^j(x) d\theta dx,$$

and by a change of variables in the calculation of the integrals over \mathbb{R}^d and then integrating by parts we get the equations in (3.12). Hence we get the estimates (3.13) and (3.14) by applying standard derivative rules and using the estimates in (3.7). \square

4 Sobolev estimates

In this section we present some estimates which are needed in the subsequent sections. In the following lemmas note that by lower indices i we mean the derivative with respect to x^i , i.e. $u_i = \frac{\partial}{\partial x^i} u(x)$. For $\varepsilon > 0$ we use the notation k_ε for the Gaussian density function on \mathbb{R}^d with mean 0 and variance ε . For linear functionals Φ , acting on a real vector space V containing $\mathcal{S} = \mathcal{S}(\mathbb{R}^d)$, the rapidly decreasing functions on \mathbb{R}^d , the mollification $\Phi^{(\varepsilon)}$ is defined by

$$\Phi^{(\varepsilon)}(x) = \Phi(k_\varepsilon(x - \cdot)), \quad x \in \mathbb{R}^d.$$

In particular, when $\Phi = \mu$ is a (signed) measure from \mathcal{S}^* , the dual of \mathcal{S} , or $\Phi = f$ is a function from \mathcal{S}^* , then

$$\mu^{(\varepsilon)}(x) = \int_{\mathbb{R}^d} k_\varepsilon(x - y) \mu(dy), \quad f^{(\varepsilon)}(x) = \int_{\mathbb{R}^d} k_\varepsilon(x - y) f(y) dy, \quad x \in \mathbb{R}^d,$$

and, using the formal adjoint L^* , we write

$$(L^* \mu)^{(\varepsilon)}(x) := \int_{\mathbb{R}^d} L_y k_\varepsilon(x - y) \mu(dy), \quad x \in \mathbb{R}^d$$

when L is a linear operator on V such that the integral is well-defined for every $x \in \mathbb{R}^d$. Here the subscript y in L_y indicates that the operator L acts in the y -variable of the function $\bar{k}_\varepsilon(x, y) := k_\varepsilon(x - y)$. For example, if L is a differential operator of the form $a^{ij} D_{ij} + b^i D_i + c$, where a^{ij} , b^i and c are functions defined on \mathbb{R}^d , then

$$(L^* \mu)^{(\varepsilon)}(x) = \int_{\mathbb{R}^d} (a^{ij}(y) \frac{\partial^2}{\partial y^i \partial y^j} + b^i(y) \frac{\partial}{\partial y^i} + c(y)) k_\varepsilon(x - y) \mu(dy).$$

We will often use the following well-known properties of mollifications with k_ε :

- (i) $|\varphi^{(\varepsilon)}|_{L_p} \leq |\varphi|_{L_p}$ for $\varphi \in L_p(\mathbb{R}^d)$, $p \in [1, \infty)$;
- (ii) $\mu^{(\varepsilon)}(\varphi) := \int_{\mathbb{R}^d} \mu^{(\varepsilon)}(x) \varphi(x) dx = \int_{\mathbb{R}^d} \varphi^{(\varepsilon)}(x) \mu(dx) =: \mu(\varphi^{(\varepsilon)})$ for finite (signed) Borel measures μ on $\mathcal{B}(\mathbb{R}^d)$ and $\varphi \in L_p(\mathbb{R}^d)$, $p \geq 1$;
- (iii) $|\mu^{(\delta)}|_{L_p} \leq |\mu^{(\varepsilon)}|_{L_p}$ for $0 \leq \varepsilon \leq \delta$, finite Borel measures μ on \mathbb{R}^d and $p \geq 1$. This property follows immediately from (i) and the ‘‘semigroup property’’ of the Gaussian kernel,

$$k_{r+s}(y - z) = \int_{\mathbb{R}^d} k_r(y - x) k_s(x - z) dx, \quad y, z \in \mathbb{R}^d \text{ and } r, s \in (0, \infty). \tag{4.1}$$

The following generalization of (iii) is also useful: for integers $p \geq 2$ we have

$$\rho_\varepsilon(y) := \int_{\mathbb{R}^d} \prod_{r=1}^p k_\varepsilon(x - y_r) dx = c_{p,\varepsilon} e^{-\sum_{1 \leq r < s \leq p} |y_r - y_s|^2 / (2\varepsilon p)},$$

$$y = (y_1, \dots, y_p) \in \mathbb{R}^{pd}, \tag{4.2}$$

for $\varepsilon > 0$, with a constant $c_{p,\varepsilon} = c_{p,\varepsilon}(d) = p^{-d/2}(2\pi\varepsilon)^{(1-p)d/2}$. This calculation can be found in [5, Sec. 4]. Clearly, for every $r = 1, 2, \dots, p$ and $i = 1, 2, \dots, d$

$$\partial_{y_r^i} \rho_\varepsilon(y) = \frac{1}{\varepsilon p} \sum_{s=1}^p (y_s^i - y_r^i) \rho_\varepsilon(y), \quad y = (y_1, \dots, y_p) \in \mathbb{R}^d, \quad y_r = (y_r^1, \dots, y_r^d) \in \mathbb{R}^d. \tag{4.3}$$

It is easy to see that

$$\sum_{r=1}^p \partial_{y_r^j} \rho_\varepsilon(y) = 0 \quad \text{for } y \in \mathbb{R}^{pd}, j = 1, 2, \dots, d,$$

which we will often use in the form

$$\partial_{y_r^j} \rho_\varepsilon(y) = - \sum_{s \neq r}^p \partial_{y_s^j} \rho_\varepsilon(y) \quad \text{for } r = 1, \dots, p \text{ and } j = 1, 2, \dots, d. \tag{4.4}$$

Moreover, we will use that for $q = 1, 2$, with a constant $N = N(d, p, q)$,

$$\varepsilon^{-q} \sum_{s \neq r} |y_s - y_r|^{2q} \rho_\varepsilon(y) \leq N \rho_{2\varepsilon}(y), \quad y \in \mathbb{R}^{pd}. \tag{4.5}$$

The case of $\alpha = 0$ in the following lemmas in this section is proven in [5] and hence this case will be omitted in the proofs.

The following estimates for $\mu \in \mathfrak{M}$ with density $u = d\mu/dx \in W_p^m$, for an $m \geq 0$ and an even integer $p \geq 2$, will be useful in later sections. In order for the left-hand side of these estimates to be well-defined, we require that

$$K_1 \int_{\mathbb{R}^d} |x|^2 |u(x)| dx < \infty, \tag{4.6}$$

where we use the formal convention that $0 \cdot \infty = 0$, i.e. if $K_1 = 0$, then the second moment of $|\mu(dx)| = |u(x)|dx$ is not required to be finite.

Lemma 4.1 *Let $\sigma = (\sigma^{ik})$ be a Borel function on \mathbb{R}^d with values in $\mathbb{R}^{d \times k}$, such that for some nonnegative constants K_0, K_1 and L*

$$|\sigma(x)| \leq K_0 + K_1|x|, \quad \sum_{k=1}^{m+1} |D^k \sigma(x)| \leq L, \tag{4.7}$$

for all $x \in \mathbb{R}^d$. Set $a^{ij} = \sigma^{ik} \sigma^{jk} / 2$ for $i, j = 1, 2, \dots, d$. Let $\mu \in \mathfrak{M}$ such that it admits a density $u = d\mu/dx \in W_p^m$ for an integer $m \geq 0$ and an even integer $p \geq 2$

and (4.6) holds. Then for $\varepsilon > 0$ we have

$$A^\alpha := p((D^\alpha \mu^{(\varepsilon)})^{p-1}, D^\alpha((a^{ij} D_{ij})^* \mu)^{(\varepsilon)}) + \frac{p(p-1)}{2}((D^\alpha \mu^{(\varepsilon)})^{p-2} D^\alpha((\sigma^{ik} D_i)^* \mu)^{(\varepsilon)}, D^\alpha((\sigma^{jk} D_j)^* \mu)^{(\varepsilon)}) \leq NL^2 |u|_{W_p^m}^p \tag{4.8}$$

for multi-indices $\alpha = (\alpha_1, \dots, \alpha_d)$ such that $0 \leq |\alpha| \leq m$, where N is a constant depending only on d, m and p .

Proof Note first that using

$$\sup_{x \in \mathbb{R}^d} \sum_{k=0}^{m+2} |D^k k_\varepsilon(x)| < \infty, \quad \sup_{x \in \mathbb{R}^d} \sum_{k=0}^{m+2} |D^k \rho_\varepsilon(x)| < \infty, \quad \text{for all } \varepsilon > 0 \tag{4.9}$$

and

$$\int_{\mathbb{R}^d} (1 + K_1|x|^2) |u(x)| dx < \infty, \tag{4.10}$$

as well as the conditions on σ , it is easy to verify that the left-hand side of (4.8) is well-defined. Changing the order of taking derivatives and integrals, then writing integer powers of integrals as iterated integrals and using

$$D_x^\alpha k_\varepsilon(x - y) = (-1)^{|\alpha|} D_y^\alpha k_\varepsilon(x - y),$$

we have

$$\begin{aligned} ((D^\alpha \mu^{(\varepsilon)}(x))^{p-1} &= \int_{\mathbb{R}^{(p-1)d}} \prod_{r=1}^{p-1} D_x^\alpha k_\varepsilon(x - y_r) \mu(dy_1) \dots \mu(dy_{p-1}) \\ &= \int_{\mathbb{R}^{(p-1)d}} (-1)^{(p-1)|\alpha|} D_{y_1}^\alpha \dots D_{y_{p-1}}^\alpha \prod_{r=1}^{p-1} k_\varepsilon(x - y_r) \mu(dy_1) \dots \mu(dy_{p-1}), \\ D^\alpha((a^{ij} D_{ij})^* \mu)^{(\varepsilon)}(x) &= \int_{\mathbb{R}^d} a^{ij}(y_p) \partial_{y_p^i} \partial_{y_p^j} D_x^\alpha k_\varepsilon(x - y_p) \mu(dy_p) \\ &= \int_{\mathbb{R}^d} (-1)^{|\alpha|} a^{ij}(y_p) \partial_{y_p^i} \partial_{y_p^j} D_{y_p}^\alpha k_\varepsilon(x - y_p) \mu(dy_p), \end{aligned}$$

and hence for their product we get

$$(D^\alpha \mu^{(\varepsilon)})^{p-1} D^\alpha((a^{ij} D_{ij})^* \mu)^{(\varepsilon)}(x) = \int_{\mathbb{R}^{pd}} a^{ij}(y_p) \partial_{y_p^i} \partial_{y_p^j} D_y^{p\alpha} \prod_{r=1}^p k_\varepsilon(x - y_r) \mu_p(dy), \tag{4.11}$$

where $D_y^{p\alpha} := D_{y_1}^\alpha \dots D_{y_p}^\alpha$ and $\mu(dy) := \mu(dy_1) \dots \mu(dy_p)$. Similarly,

$$\begin{aligned} (D^\alpha \mu^{(\varepsilon)})^{p-2} D^\alpha((\sigma^{ik} D_i)^* \mu)^{(\varepsilon)} D^\alpha((\sigma^{jk} D_j)^* \mu)^{(\varepsilon)}(x) \\ = \int_{\mathbb{R}^{pd}} \sigma^{ik}(y_{p-1}) \sigma^{jk}(y_p) \partial_{y_{p-1}^i} \partial_{y_p^j} D_y^{p\alpha} \prod_{r=1}^p k_\varepsilon(x - y_r) \mu_p(dy). \end{aligned}$$

Adding this to (4.11), then integrating against dx over \mathbb{R}^d and using (4.2) we obtain

$$A = \int_{\mathbb{R}^{pd}} \left(pa^{ij}(y_p)\partial_{y_p^i}\partial_{y_p^j} + \frac{p(p-1)}{2}\sigma^{ik}(y_{p-1})\sigma^{jk}(y_p)\partial_{y_{p-1}^i}\partial_{y_p^j} \right) D_y^{p\alpha} \rho_\varepsilon(y) \mu_p(dy).$$

Using here the symmetry of $D_y^{p\alpha} \rho_\varepsilon(y)$ and $\mu_p(dy)$ in $y \in \mathbb{R}^{dp}$ and then interchanging differential operators we get

$$A = \int_{\mathbb{R}^{pd}} \left(\sum_{r=1}^p a^{ij}(y_r)D_y^{p\alpha} \partial_{y_r^i}\partial_{y_r^j} + \sum_{1 \leq r < s \leq p} \sigma^{ik}(y_r)\sigma^{jk}(y_s)D_y^{p\alpha} \partial_{y_r^i}\partial_{y_s^j} \right) \rho_\varepsilon(y) \mu_p(dy)$$

Using

$$\partial_{y_r^j} \rho_\varepsilon(y) = - \sum_{s \neq r} \partial_{y_s^j} \rho_\varepsilon(y),$$

see (4.4), we have

$$\sum_{r=1}^p a^{ij}(y_r)D_y^{p\alpha} \partial_{y_r^i}\partial_{y_r^j} \rho_\varepsilon(y) = - \sum_{1 \leq r < s \leq p} (a^{ij}(y_r) + a^{ij}(y_s))D_y^{p\alpha} \partial_{y_r^i}\partial_{y_s^j} \rho_\varepsilon(y),$$

and due to $a^{ij} = \sigma^{ik}\sigma^{jk}/2$ we have

$$\begin{aligned} -2a^{ij}(y_r, y_s) &:= -2(a^{ij}(y_r) + a^{ij}(y_s)) + \sigma^{ik}(y_r)\sigma^{jk}(y_s) + \sigma^{ik}(y_s)\sigma^{jk}(y_r) \\ &= -(\sigma^{ik}(y_r) - \sigma^{ik}(y_s))(\sigma^{jk}(y_r) - \sigma^{jk}(y_s)). \end{aligned}$$

Hence

$$A = -\frac{1}{2} \sum_{r \neq s} \int_{\mathbb{R}^{pd}} a^{ij}(y_r, y_s)D_y^{p\alpha} \partial_{y_r^i}\partial_{y_s^j} \rho_\varepsilon(y) \mu_p(dy), \tag{4.12}$$

that by integration by parts gives

$$= -\frac{1}{2} \sum_{\beta \leq \alpha} \sum_{\gamma \leq \alpha} c_{\beta\gamma}^{\alpha, \alpha} \int_{\mathbb{R}^{pd}} \sum_{r \neq s} a_{\beta\gamma}^{ij}(y_r, y_s) \partial_{y_r^i}\partial_{y_s^j} \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) u_{\bar{\gamma}}(y_s) \Pi_{q \neq r, q \neq s} u_\alpha(y_q) dy, \tag{4.13}$$

where $a_{\beta\gamma}^{ij}(x, r) := \partial_x^\beta \partial_r^\gamma a^{ij}(x, r)$ and $u_\delta(x) := \partial^\delta u(x)$ for $x, r \in \mathbb{R}^d$, for multi-indices β, γ and $\delta, \bar{\delta} := \alpha - \delta$ for multi-indices $\delta \leq \alpha$ (i.e. $\delta_i \leq \alpha_i$ for $i = 1, 2, \dots, d$), $c_\delta^\alpha = \prod_{i=1}^d c_{\delta_i}^{\alpha_i}$ with binomial coefficients c_k^n for integers $0 \leq k \leq n$,

$$u(y) := u(y_1) \dots u(y_p) \text{ for } y = (y_1, \dots, y_p) \in \mathbb{R}^{dp},$$

and $dy = dy_1 \dots dy_p$ is the Lebesgue measure on \mathbb{R}^{pd} . For each $\beta \leq \alpha$ and $\gamma \leq \alpha$ we are going to estimate the integrand

$$f^{\beta\gamma}(y) := \sum_{r \neq s} a_{\beta\gamma}^{ij}(y_r, y_s) \partial_{y_r^i} \partial_{y_s^j} \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) u_{\bar{\gamma}}(y_s) \Pi_{q \neq r, q \neq s} u_\alpha(y_q), \quad y \in \mathbb{R}^{pd},$$

$$\beta \leq \alpha, \quad \gamma \leq \alpha$$

in the integral in (4.13). Because of the symmetry in β and γ , we need only consider the following cases: (i) $|\beta| \geq 1$ and $|\gamma| \geq 1$, (ii) $|\beta| \geq 1$ and $\gamma = 0$ and (iii) $\beta = \gamma = 0$. To proceed with the calculations in each of these cases, for functions $h = h(y)$ and $g = g(y)$ of $y \in \mathbb{R}^{pd}$ we will use the notations $h \sim g$ if the integral of $g - h$ against dy over \mathbb{R}^{pd} is zero. In case (i) by integration by parts we have

$$f^{\beta\gamma} \sim \sum_{j=1}^4 f_j^{\beta\gamma}$$

with

$$\begin{aligned} f_1^{\beta\gamma} &:= \sum_{r \neq s} \partial_{y_r^i} \partial_{y_s^j} a_{\beta\gamma}^{ij}(y_r, y_s) \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) u_{\bar{\gamma}}(y_s) \Pi_{q \neq r, q \neq s} u_\alpha(y_q), \\ f_2^{\beta\gamma} &:= \sum_{r \neq s} \partial_{y_s^j} a_{\beta\gamma}^{ij}(y_r, y_s) \rho_\varepsilon(y) \partial_{y_r^i} u_{\bar{\beta}}(y_r) u_{\bar{\gamma}}(y_s) \Pi_{q \neq r, q \neq s} u_\alpha(y_q), \\ f_3^{\beta\gamma} &:= \sum_{r \neq s} \partial_{y_r^i} a_{\beta\gamma}^{ij}(y_r, y_s) \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) \partial_{y_s^j} u_{\bar{\gamma}}(y_s) \Pi_{q \neq r, q \neq s} u_\alpha(y_q), \\ f_4^{\beta\gamma} &:= \sum_{r \neq s} a_{\beta\gamma}^{ij}(y_r, y_s) \rho_\varepsilon(y) \partial_{y_r^i} u_{\bar{\beta}}(y_r) \partial_{y_s^j} u_{\bar{\gamma}}(y_s) \Pi_{q \neq r, q \neq s} u_\alpha(y_q). \end{aligned} \tag{4.14}$$

It is easy to see that for $j = 1, 2, 3, 4$

$$|f_j^{\beta\gamma}(y)| \leq NL^2 \rho_\varepsilon(y) \sum_{|\delta| \leq m} |u_\delta(y_1)| \dots \sum_{|\delta| \leq m} |u_\delta(y_p)|, \quad (y_1, y_2, \dots, y_p) \in \mathbb{R}^{pd}$$

with a constant $N = N(d, m, p)$. Hence in the case (i) we get

$$\begin{aligned} \int_{\mathbb{R}^{pd}} f^{\beta\gamma}(y) dy &\leq NL^2 \int_{\mathbb{R}^{pd}} \sum_{|\delta| \leq m} |u_\delta(y_1)| \dots \sum_{|\delta| \leq m} |u_\delta(y_p)| \rho_\varepsilon(y) dy \\ &= NL^2 \int_{\mathbb{R}^{pd}} \int_{\mathbb{R}^d} \sum_{|\delta| \leq m} |u_\delta(y_1)| \dots \sum_{|\delta| \leq m} |u_\delta(y_p)| \Pi_{r=1}^p k_\varepsilon(x - y_r) dx dy \\ &\leq N' L^2 \sum_{|\delta| \leq m} \|D^\delta u\|_{L^p}^p. \end{aligned}$$

with constants N and N' depending only on d, m and p . Integrating by parts in the case (ii) we have

$$f^{\beta 0} \sim -f_1^{\beta 0} - f_2^{\beta 0}$$

with

$$f_1^{\beta 0} = \sum_{r \neq s} \partial_{y_r^i} a_{\beta 0}^{ij}(y_r, y_s) \partial_{y_s^j} \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) \Pi_{q \neq r} u_\alpha(y_q)$$

$$f_2^{\beta 0} = \sum_{r \neq s} a_{\beta 0}^{ij}(y_r, y_s) \partial_{y_s^j} \rho_\varepsilon(y) \partial_{y_r^i} u_{\bar{\beta}}(y_r) \Pi_{q \neq r} u_\alpha(y_q).$$

Clearly, for $r \neq s$ we have

$$\partial_{y_r^i} a_{\beta 0}^{ij}(y_r, y_s) = g^{\beta, j}(y_r, y_s) + h^{\beta, j}(y_r),$$

with

$$g^{\beta, j}(y_r, y_s) = \partial_{y_r^i} \partial_{y_r^i} \partial_{y_r^i} \sigma^{ik}(y_r) (\sigma^{jk}(y_r) - \sigma^{jk}(y_s)) + \partial_{y_r^i} \partial_{y_r^i} \sigma^{jk}(y_r) (\sigma^{ik}(y_r) - \sigma^{ik}(y_s)),$$

$$h^{\beta, j}(y_r) = \sum_{1 \leq |\delta|, \delta < \beta(i)} c_\delta^{\beta(i)} \partial_{y_r^\delta} \sigma^{ik}(y_r) \partial_{y_r^i}^{\beta(i) - \delta} \sigma^{jk}(y_r),$$

where the multi-index $\beta(i)$ is defined by $\partial^{\beta(i)} = \partial_{y_r^i}^{\beta(i)}$. Thus

$$f_1^{\beta 0} = f_{11}^{\beta 0} + f_{12}^{\beta 0}$$

with

$$f_{11}^{\beta 0} = \sum_{r=1}^p \sum_{s \neq r} g^{\beta, j}(y_r, y_s) \partial_{y_s^j} \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) \Pi_{q \neq r} u_\alpha(y_q),$$

$$f_{12}^{\beta 0} = \sum_{r=1}^p \sum_{s \neq r} h^{\beta, j}(y_r) \partial_{y_s^j} \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) \Pi_{q \neq r} u_\alpha(y_q). \tag{4.15}$$

Since

$$|g^{\beta, j}(y_r, y_s)| \leq NL^2 |y_r - y_s| \quad j = 1, 2, \dots, p,$$

for some $N = N(d, m, p)$, taking into account (4.3) we have

$$|g^{\beta, j}(y_r, y_s) \partial_{y_s^j} \rho_\varepsilon(y)| \leq \frac{N}{p\varepsilon} \sum_{1 \leq k < l \leq p} |y_k - y_l|^2 \rho_\varepsilon(y) \leq N' \rho_{2\varepsilon}(y)$$

and hence

$$|f_{11}^{\beta 0}| \leq N' \rho_{2\varepsilon}(y) \sum_{|\delta| \leq m} |u_\delta(y_1)| \dots \sum_{|\delta| \leq m} |u_\delta(y_p)|$$

with a constant $N' = N'(d, m, p)$. Remembering (4.4) by integration by parts we obtain

$$f_{12}^{\beta 0} = - \sum_{r=1}^p h^{\beta, j}(y_r) \partial_{y_r^j} \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) \Pi_{q \neq r} u_\alpha(y_q) \sim f_{121}^{\beta 0} + f_{122}^{\beta 0}$$

with

$$f_{121}^{\beta 0} = \sum_{r=1}^p h^{\beta, j}(y_r) \rho_\varepsilon(y) \partial_{y_r^j} u_{\bar{\beta}}(y_r) \Pi_{q \neq r} u_\alpha(y_q),$$

$$f_{122}^{\beta 0} = \sum_{r=1}^p \partial_{y_r^j} h^{\beta, j}(y_r) \rho_\varepsilon(y) u_{\bar{\beta}}(y_r) \Pi_{q \neq r} u_\alpha(y_q).$$

Hence noting that

$$|h^{\beta, j}(y_r)| + |\partial_{y_r^j} h^{\beta, j}(y_r)| \leq NL^2$$

with a constant $N = N(d, m, p)$, we get

$$|f_{121}^{\beta 0} + f_{122}^{\beta 0}| \leq NL^2 \rho_\varepsilon(y) \sum_{|\delta| \leq m} |u_\delta(y_1)| \dots \sum_{|\delta| \leq m} |u_\delta(y_p)|$$

Consequently, for a constant $N' = N'(d, m, p)$,

$$\begin{aligned} \int_{\mathbb{R}^{pd}} f_1^{\beta 0}(y) dy &\leq NL^2 \int_{\mathbb{R}^{pd}} \sum_{|\delta| \leq m} |u_\delta(y_1)| \dots \sum_{|\delta| \leq m} |u_\delta(y_p)| \rho_{2\varepsilon}(y) dy \\ &\leq N' L^2 \sum_{|\delta| \leq m} \| |D^\delta u|^{(2\varepsilon)} \|_{L^p}^p \leq N' L^2 \sum_{|\delta| \leq m} \| |D^\delta u|^{(\varepsilon)} \|_{L^p}^p. \end{aligned} \tag{4.16}$$

Now we are going to estimate the integral of $f_2^{\beta 0}$. If $|\beta| = 1$, then

$$|a_{\beta 0}^{ij}(y_r, y_s)| \leq NL^2 |y_r - y_s|,$$

and taking into account (4.3), we get

$$|f_2^{\beta 0}| \leq NL^2 \rho_{2\varepsilon}(y) \sum_{|\delta| \leq m} |u_\delta(y_1)| \dots \sum_{|\delta| \leq m} |u_\delta(y_p)|$$

with $N = N(d, p, m)$ in the same way as $|f_{11}|$ is estimated. Hence, as above,

$$\int_{\mathbb{R}^{pd}} f_2^{\beta 0}(y) dy \leq NL^2 \sum_{|\delta| \leq m} \|D^\delta u\|_{L^p}^{(2\varepsilon)p} \leq NL^2 \sum_{|\delta| \leq m} \|D^\delta u\|_{L^p}^{(\varepsilon)p} \tag{4.17}$$

for $|\beta| = 1$. If $|\beta| \geq 2$, then

$$a_{\beta 0}^{ij}(y_r, y_s) = g^{\beta,ij}(y_r, y_s) + h^{\beta,ij}(y_r)$$

with

$$g^{\beta,ij}(y_r, y_s) = \partial_{y_r}^\beta \sigma^{ik}(y_r)(\sigma^{jk}(y_r) - \sigma^{jk}(y_s)) + \partial_{y_r}^\beta \sigma^{jk}(y_r)(\sigma^{ik}(y_r) - \sigma^{ik}(y_s))$$

$$h^{\beta,ij}(y_r) = \sum_{1 \leq |\delta|, \delta < \beta} c_\delta^\beta \partial_{y_r}^\delta \sigma^{ik}(y_r) \partial_{y_r}^{\beta-\delta} \sigma^{jk}(y_r).$$

Noticing that for a constant $N = N(d, m, p)$,

$$\sum_{i,j} |g^{\beta,ij}(y_r, y_s)| \leq NL^2 |y_r - y_s|$$

and

$$\sum_{ij} |h^{\beta,ij}(y_r)| + \sum_{ij} |\partial_{y_r^j} h^{\beta,ij}(y_r)| \leq NL^2,$$

we obtain (4.17) for $|\beta| \geq 2$ in the same way as the integral of $f_1^{\beta 0}$ is estimated. It remains to consider the case (iii), i.e., to estimate the integral of f^{00} . Since

$$|a_{00}^{ij}(y_r, y_s)| \leq NL^2 |y_r - y_s|^2$$

with a constant $N = N(d, m, p)$ and

$$\partial_{y_r^i} \partial_{y_s^j} \rho_\varepsilon(y) = \frac{1}{p^2 \varepsilon^2} \sum_{k=1}^p \sum_{l=1}^p (y_k^i - y_r^i)(y_l^j - y_s^j) + \frac{1}{p\varepsilon} \delta_{ij},$$

we have for a constant $N' = N'(d, m, p)$,

$$|a_{00}^{ij}(y_r, y_s) \partial_{y_r^i} \partial_{y_s^j} \rho_\varepsilon(y)| \leq \frac{N}{\varepsilon^2} L^2 \sum_{1 \leq k < l \leq p} |y_k - y_l|^4 \rho_\varepsilon(y)$$

$$+ \frac{N}{\varepsilon} L^2 \sum_{1 \leq k < l \leq p} |y_k - y_l|^2 \rho_\varepsilon(y)$$

$$\leq N' L^2 \rho_{2\varepsilon}(y) \quad \text{for } y = (y_1, \dots, y_p) \in \mathbb{R}^{pd}.$$

Hence

$$|f^{00}(y)| \leq NL^2 \rho_{2\varepsilon}(y) \prod_{r=1}^p |u_\alpha(y_r)|,$$

that gives

$$\int_{\mathbb{R}^{pd}} f^{00}(y) dy \leq NL^2 \|D^\alpha u\|_{L_p}^{(2\varepsilon)p} \leq NL^2 \|D^\alpha u\|_{L_p}^{(\varepsilon)p}$$

with a constant $N = N(d, m, p)$, and we finish the proof of (4.8) by using $|v^{(\varepsilon)}|_{L_p} \leq |v|_{L_p}$ for $v \in L_p(\mathbb{R}^d)$. □

Corollary 4.2 *Let the conditions of Lemma 4.1 hold for integers $m \geq 0$ and $p \geq 2$ even. Then for $\varepsilon > 0$ we have*

$$((D^\alpha \mu^{(\varepsilon)})^{p-1}, D^\alpha ((a^{ij} D_{ij})^* \mu)^{(\varepsilon)}) \leq NL^2 |u|_{W_p^m}^p \tag{4.18}$$

for multi-indices $\alpha = (\alpha_1, \dots, \alpha_d)$ such that $|\alpha| \leq m$, where N is a constant depending only on d, m and p .

Proof It suffices to note that

$$\begin{aligned} & ((D^\alpha \mu^{(\varepsilon)})^{p-2} D^\alpha ((\sigma^{ik} D_i)^* \mu)^{(\varepsilon)}, D^\alpha ((\sigma^{jk} D_j)^* \mu)^{(\varepsilon)}) \\ &= \int_{\mathbb{R}^d} (D^\alpha \mu^{(\varepsilon)})^{p-2}(x) \sum_{k=1}^d |D^\alpha ((\sigma^{ik} D_i)^* \mu)^{(\varepsilon)}(x)|^2 dx \geq 0. \end{aligned} \tag{4.19}$$

Thus we may add (4.19) to the left-hand side of (4.18) and apply Lemma 4.1 to obtain the inequality (4.18). □

Lemma 4.3 *Let $p \geq 2$ and $m \geq 0$ be integers, and let $\sigma = (\sigma^i)$ and b be Borel functions on \mathbb{R}^d with values in \mathbb{R}^d and \mathbb{R} respectively. Assume the partial derivatives of σ and $b\sigma$ up to order $m + 1$ are functions such that there exist constants $K \geq L \geq 1$ such that*

$$\begin{aligned} \sum_{k=0}^{m+1} |D^k b(x)| &\leq K, \quad |\sigma(x)| \leq K_0 + K_1|x|, \\ \sum_{k=1}^{m+1} |D^k \sigma(x)| + \sum_{k=1}^{m+1} |D^k (b\sigma)(x)| &\leq L \end{aligned}$$

for all $x \in \mathbb{R}^d$. Then for finite signed Borel measures μ on \mathbb{R}^d with density $u := d\mu/dx \in W_p^m$, satisfying (4.6), we have

$$R^\alpha := ((D^\alpha \mu^{(\varepsilon)})^{p-2}, D^\alpha ((\sigma^i D_i)^* \mu)^{(\varepsilon)} D^\alpha (b\mu)^{(\varepsilon)}) \leq NKL |u|_{W_p^m}^p \tag{4.20}$$

for $\varepsilon > 0$ and multi-indices α such that $|\alpha| \leq m$, where N is a constant depending only on d, p, m .

Proof First note that by (4.9) and (4.10), as well as the conditions on σ and b , the left-hand sides of (4.20) is well-defined. Interchanging the order of integration and the differential operator D^α , rewriting the product of integrals as multiple integral, using Fubini's theorem and the identity

$$D_x^\alpha k_\varepsilon(x - z) = (-1)^{|\alpha|} D_z^\alpha k_\varepsilon(x - z), \quad x, y \in \mathbb{R}^d,$$

as well as (4.2), for the left-hand side of the inequality (4.20) we have

$$R^\alpha = \int_{\mathbb{R}^{dp}} f_{krr}(y) dy, \tag{4.21}$$

for any $r, k \in \{1, 2, \dots, p\}$ such that $r \neq k$, where

$$f_{krs}(y) := (-1)^{p|\alpha|} b(y_k) \sigma^i(y_r) \partial_{y_s^i} D_y^{p\alpha} \rho_\varepsilon(y) \prod_{j=1}^p u(y_j), \quad y = (y_1, \dots, y_p) \in \mathbb{R}^{pd}$$

for $k, r, s \in \{1, 2, \dots, p\}$, and recall that $dy = dy_1 \dots dy_p$ and $D_y^{p\alpha} = \prod_{j=1}^p D_{y_j}^\alpha$. As in the proof of Lemma 4.1, for real functions f and g we write $f \sim g$ if they have the same (finite) Lebesgue integral against $dy = dy_1 \dots dy_p$ over \mathbb{R}^{pd} . We write $f \leq g$ if the integrals of f and g against dy over \mathbb{R}^d are finite, and the integral of $f - g$ can be estimated by $NKL|u|_{W_p^m}$ for all $u \in W_p^m$ with a constant $N = N(d, m, p)$, independent of u . By integration by parts we have

$$f_{krr} \sim \sum_{\gamma \leq \alpha} \sum_{\beta \leq \alpha} f_{krr}^{\gamma\beta}$$

with

$$f_{krs}^{\gamma\beta}(y) := c_\gamma^\alpha c_\beta^\alpha b_\gamma(y_k) \sigma_\beta^i(y_r) \partial_{y_s^i} \rho_\varepsilon(y) u_{\bar{\gamma}}(y_k) u_{\bar{\beta}}(y_r) \prod_{j \neq k,r} u_\alpha(y_j)$$

If $\beta \neq 0$ then by integration by parts (dropping $\partial_{y_s^i}$ from ρ_ε to the other terms), and using the boundedness of b , its derivatives up to order $m + 1$, and the boundedness of the derivatives of σ up to order $m + 1$, we see that $f_{krr}^{\gamma\beta} \leq 0$ for any $k = 1, 2, \dots, p, r \neq k$ and $\gamma \leq \alpha$. If $\beta = 0$ and $\gamma = 0$, then $f_{krr}^{\gamma\beta}$ can be estimated by an exact repetition of the proof of Lemma 4.2 in [5], by replacing μ therein with $u_\alpha dy$, to yield $f_{krr}^{\gamma\beta} \leq 0$. Consequently,

$$f_{krr} \leq \sum_{0 \neq \gamma \leq \alpha} f_{krr}^{\gamma 0} \quad \text{for every } k = 1, \dots, p \text{ and } r \in \{1, 2, \dots, p\} \setminus \{k\}.$$

Writing $f_{krr}^{\gamma 0}(y) = g_{krr}^\gamma(y) h_k^{\bar{\gamma}}(y)$, with

$$g_{krs}^\gamma(y) := c_\gamma^\alpha b_\gamma(y_k) \sigma^i(y_r) \partial_{y_s^i} \rho_\varepsilon(y), \quad h_k^{\bar{\gamma}}(y) := u_{\bar{\gamma}}(y_k) \prod_{j \neq k} u_\alpha(y_j),$$

we get

$$p(p-1)(p-2)R^\alpha \leq \sum_{0 \neq \gamma \leq \alpha} \sum_{s=1}^p \sum_{r \neq s} \sum_{k \neq s,r} \int_{\mathbb{R}^{pd}} g_{kss}^\gamma(y) h_k^{\bar{\gamma}}(y) dy + NKL|u|_{W_p^m}^p, \tag{4.22}$$

and by (4.4),

$$\begin{aligned} p(p-1)R^\alpha &\leq - \sum_{0 \neq \gamma \leq \alpha} \sum_{k=1}^p \sum_{r \neq k} \sum_{s \neq r} \int_{\mathbb{R}^{pd}} g_{krs}^\gamma(y) h_k^{\bar{\gamma}}(y) dy + NKL|u|_{W_p^m}^p \\ &= - \sum_{0 \neq \gamma \leq \alpha} \sum_{s=1}^p \sum_{r \neq s} \sum_{k \neq s,r} \int_{\mathbb{R}^{pd}} g_{krs}^\gamma(y) h_k^{\bar{\gamma}}(y) dy \\ &\quad - \sum_{0 \neq \gamma \leq \alpha} \sum_{s=1}^p \sum_{r \neq s} \int_{\mathbb{R}^{pd}} g_{srs}^\gamma(y) h_s^{\bar{\gamma}}(y) dy + NKL|u|_{W_p^m}^p \end{aligned} \tag{4.23}$$

with a constant $N = N(d, m, p)$. Summing up (4.22) and (4.23) we obtain

$$\begin{aligned} c_p R^\alpha &\leq \sum_{0 \neq \gamma \leq \alpha} \sum_{s=1}^p \sum_{r \neq s} \sum_{k \neq r,s} \int_{\mathbb{R}^{pd}} (g_{kss}^\gamma(y) - g_{krs}^\gamma(y)) h_k^{\bar{\gamma}}(y) dy + NKL|u|_{W_p^m}^p \\ &\quad - \sum_{0 \neq \gamma \leq \alpha} \sum_{s=1}^p \sum_{r \neq s} \int_{\mathbb{R}^{pd}} g_{srs}^\gamma(y) h_s^{\bar{\gamma}}(y) dy + NKL|u|_{W_p^m}^p \end{aligned} \tag{4.24}$$

where $c_p = p(p-1)^2$, and

$$(g_{kss}^\gamma(y) - g_{krs}^\gamma(y)) h_k^{\bar{\gamma}}(y) = c_\gamma^\alpha b_\gamma(y_k) (\sigma^i(y_s) - \sigma^i(y_r)) \partial_{y_s^i} \rho_\varepsilon(y) u_{\bar{\gamma}}(y_k) \prod_{j \neq k} u_\alpha(y).$$

By the boundedness of $|b_\gamma|$ and the Lipschitz condition on σ , using (4.5) we get

$$(g_{kss}^\gamma - g_{krs}^\gamma) h_k^{\bar{\gamma}} \leq 0, \quad \text{for all } 0 \neq \gamma \leq \alpha, s = 1, 2, \dots, p \text{ and } r \neq s, k \neq r, s.$$

By integration by parts we have for the last term in (4.24),

$$g_{srs}^\gamma h_s^{\bar{\gamma}} \leq 0, \quad \text{for } 0 \neq \gamma \leq \alpha \text{ and } s = 1, \dots, p, r \neq s,$$

which finishes the proof of (4.20). □

Lemma 4.4 *Let b be a real-valued Borel function on \mathbb{R}^d such that for an integer $m \geq 0$ the derivatives of b up to order m are functions, which, together with b , in magnitude are bounded by a constant K . Then for any $p \geq 2$ for functions $u \in W_p^m$ we have*

$$(|D^\alpha u^{(\varepsilon)}|^{p-2} D^\alpha (bu)^{(\varepsilon)}, D^\alpha (bu)^{(\varepsilon)}) \leq NK^2 \sum_{|\beta| \leq m} \|D^\beta u\|_{L^p}^p \tag{4.25}$$

with a constant $N = N(d, m, p)$.

Proof Using the notation $v_\beta := D^\beta v$ for multi-indices β and functions v on \mathbb{R}^d , we have

$$|D^\alpha (bu)^{(\varepsilon)}| = \left| \left(\sum_{\beta+\gamma=\alpha} (b_\beta u_\gamma) \right)^{(\varepsilon)} \right| \leq K \sum_{\gamma \leq \alpha} |u_\gamma|^{(\varepsilon)}$$

Thus for the left-hand side A of the inequality (4.25) by Young’s inequality we get

$$\begin{aligned} A &= \int_{\mathbb{R}^d} |(D^\alpha u)^{(\varepsilon)}|^{p-2} |(D^\alpha (bu))^{(\varepsilon)}|^2 dx \leq K^2 \int_{\mathbb{R}^d} |(D^\alpha u)^{(\varepsilon)}|^{p-2} \sum_{\gamma \leq \alpha} |u_\gamma|^{(\varepsilon)}|^2 dx \\ &\leq \frac{p-2}{p} K^2 |(u_\alpha)^{(\varepsilon)}|_{L^p}^p + \frac{2}{p} K^2 \sum_{\gamma \leq \alpha} |u_\gamma|^{(\varepsilon)}|_{L^p}^p, \end{aligned}$$

which immediately implies (4.25). □

Recall that for vectors $\xi = \xi(x) \in \mathbb{R}^d$, depending on $x \in \mathbb{R}^d$ the linear operators T^ξ , I^ξ and J^ξ are defined by

$$\begin{aligned} T^\xi \varphi(x) &= \varphi(x + \xi(x)) \\ I^\xi \varphi(x) &:= T^\xi \varphi(x) - \varphi(x), \quad J^\xi \psi(x) := I^\xi \psi(x) - \xi^i(x) D_i \psi(x), \end{aligned} \tag{4.26}$$

$x \in \mathbb{R}^d$, acting on functions φ and differentiable functions ψ on \mathbb{R}^d .

Lemma 4.5 *Let $\xi = \xi(x, \mathfrak{z})$ be an \mathbb{R}^d -valued function of $x \in \mathbb{R}^d$ for every $\mathfrak{z} \in \mathfrak{Z}$ for a set \mathfrak{Z} . Assume that for an integer $m \geq 1$ the partial derivatives of ξ in $x \in \mathbb{R}^d$ up to order m are functions on \mathbb{R}^d for each $\mathfrak{z} \in \mathfrak{Z}$, such that for a constant $\lambda > 0$, a function $\bar{\xi}$ on \mathfrak{Z} and constants $K_0, K_1, K_\xi \geq 0$ we have*

$$\begin{aligned} |\xi(x, \mathfrak{z})| &\leq \bar{\xi}(\mathfrak{z})(K_0 + K_1|x|), \quad \bar{\xi}(\mathfrak{z}) \leq K_\xi, \\ \sum_{k=1}^{m+1} |D_x^k \xi(x, \mathfrak{z})| &\leq \bar{\xi}(\mathfrak{z}), \quad |\det(\mathbb{I} + \theta D_x \xi(x, \mathfrak{z}))| \geq \lambda^{-1} \end{aligned} \tag{4.27}$$

for all $x \in \mathbb{R}^d$, $\mathfrak{z} \in \mathfrak{Z}$ and $\theta \in [0, 1]$. Let $p \geq 2$ be an even integer. Then for every finite signed Borel measure μ with density $u = d\mu/dx \in W_p^m$, satisfying (4.6), we have

$$\begin{aligned} C &:= \int_{\mathbb{R}^d} p(D_x^\alpha \mu^{(\varepsilon)})^{p-1} D_x^\alpha (J^{\xi*} \mu)^{(\varepsilon)} dx \\ &\quad + \int_{\mathbb{R}^d} (D_x^\alpha \mu^{(\varepsilon)} + D_x^\alpha (I^{\xi*} \mu)^{(\varepsilon)})^p - (D_x^\alpha \mu^{(\varepsilon)})^p - p(D_x^\alpha \mu^{(\varepsilon)})^{p-1} D_x^\alpha (I^{\xi*} \mu)^{(\varepsilon)} dx \\ &\leq N \bar{\xi}^2(\mathfrak{z}) |u|_{W_p^m}^p \text{ for } \mathfrak{z} \in \mathfrak{Z}, \quad \varepsilon > 0, \end{aligned} \tag{4.28}$$

for multi-indices α , $0 \leq |\alpha| \leq m$ with a constant $N = N(d, p, m, \lambda, K_\xi)$.

Proof Again we note that by (4.9) and (4.10), together with the conditions on ξ , it is easy to verify that C is well-defined. Notice that

$$D_x^\alpha \mu^{(\varepsilon)} + D_x^\alpha (I^{\xi*} \mu)^{(\varepsilon)} = D_x^\alpha (T^{\xi*} \mu)^{(\varepsilon)}$$

and

$$\begin{aligned} & p(D_x^\alpha \mu^{(\varepsilon)})^{p-1} D_x^\alpha (J^{\xi*} \mu)^{(\varepsilon)} - p(D_x^\alpha \mu^{(\varepsilon)})^{p-1} D_x^\alpha (I^{\xi*} \mu)^{(\varepsilon)} \\ &= -p(D_x^\alpha \mu^{(\varepsilon)})^{p-1} D_x^\alpha ((\xi^i D_i)^* \mu)^{(\varepsilon)}, \end{aligned}$$

Hence

$$C = \int_{\mathbb{R}^d} (D_x^\alpha (T^{\xi*} \mu)^{(\varepsilon)})^p - (D_x^\alpha \mu^{(\varepsilon)})^p - p(D_x^\alpha \mu^{(\varepsilon)})^{p-1} D_x^\alpha ((\xi^i D_i)^* \mu)^{(\varepsilon)} dx. \tag{4.29}$$

First we change the order of D_x^α and the integrals and operators T_y^ξ and I_y^ξ acting in the variable $y \in \mathbb{R}^d$, then we use

$$D_x^\alpha k_\varepsilon(x - y) = (-1)^{|\alpha|} D_y^\alpha k_\varepsilon(x - y)$$

to get

$$\begin{aligned} D_x^\alpha (T^{\xi*} \mu)^{(\varepsilon)} &= (-1)^{|\alpha|} \int_{\mathbb{R}^d} T_y^\xi D_y^\alpha k_\varepsilon(x - y) \mu(dy), \\ D_x^\alpha \mu^{(\varepsilon)} &= (-1)^{|\alpha|} \int_{\mathbb{R}^d} D_y^\alpha k_\varepsilon(x - y) \mu(dy), \\ D_x^\alpha ((\xi^i D_i)^* \mu)^{(\varepsilon)} &= (-1)^{|\alpha|} \int_{\mathbb{R}^d} \xi^i(y) \partial_{y_i} D_y^\alpha k_\varepsilon(x - y) \mu(dy). \end{aligned}$$

Thus rewriting the product of integrals as multiple integrals, and using the product measure $\mu_p(dy) := \mu(dy_1) \dots \mu(dy_p)$ on \mathbb{R}^{dp} by Fubini's theorem we get

$$\begin{aligned} (D_x^\alpha (T^{\xi*} \mu)^{(\varepsilon)})^p(x) &= \int_{\mathbb{R}^{pd}} \prod_{r=1}^p (T_{y_r}^\xi D_{y_r}^\alpha k_\varepsilon(x - y_r)) \mu_p(dy) \\ &= \int_{\mathbb{R}^{pd}} \prod_{r=1}^p T_{y_r}^\xi D_{y_r}^{p\alpha} \prod_{r=1}^p k_\varepsilon(x - y_r) \mu_p(dy), \\ (D_x^\alpha \mu^{(\varepsilon)})^p &= \int_{\mathbb{R}^{pd}} \prod_{r=1}^p (D_{y_r}^\alpha k_\varepsilon(x - y_r)) \mu_p(dy) \\ &= \int_{\mathbb{R}^{pd}} D_y^{p\alpha} \prod_{r=1}^p k_\varepsilon(x - y_r) \mu_p(dy) \end{aligned} \tag{4.30}$$

and

$$p(D_x^\alpha \mu^{(\varepsilon)})^{p-1} D_x^\alpha ((\xi^i D_i)^* \mu)^{(\varepsilon)}$$

$$\begin{aligned}
 &= p \int_{\mathbb{R}^{pd}} \prod_{r=1}^{p-1} (D_{y_r}^\alpha k_\varepsilon(x - y_r)) \xi^i(y_p) \partial_{y_p^i} D_{y_p}^\alpha k_\varepsilon(x - y_p) \mu_p(dy) \\
 &= p \int_{\mathbb{R}^{pd}} \xi^i(y_p) \partial_{y_p^i} D_y^{p\alpha} \prod_{r=1}^p k_\varepsilon(x - y_r) \mu_p(dy) \\
 &= \int_{\mathbb{R}^{pd}} \sum_{r=1}^p \xi^i(y_r) \partial_{y_r^i} D_y^{p\alpha} \prod_{r=1}^p k_\varepsilon(x - y_r) \mu_p(dy), \tag{4.31}
 \end{aligned}$$

where again

$$D_y^{p\alpha} := \prod_{r=1}^p D_{y_r}^\alpha \quad \text{for } y = (y_1, \dots, y_p) \in \mathbb{R}^{dp},$$

and the last equation is due to the symmetry of the function $\prod_{r=1}^p D_{y_r}^\alpha k_\varepsilon(x - y_r)$ and the measure $\mu_p(dy)$ in $y = (y_1, \dots, y_p) \in \mathbb{R}^{pd}$. Thus from (4.29) we get

$$C = \int_{\mathbb{R}^d} \int_{\mathbb{R}^{pd}} L_y^\xi D_y^{p\alpha} \prod_{r=1}^p k_\varepsilon(x - y_r) \mu_p(dy) dx$$

with the operator

$$L_y^\xi = \prod_{r=1}^p T_{y_r}^\xi - \mathbb{I} - \sum_{r=1}^p \xi^i(y_r) \partial_{y_r^i},$$

defined by

$$\begin{aligned}
 L_y^\xi \varphi(y) &= \varphi(y_1 + \xi(y_1), \dots, y_p + \xi(y_p)) - \varphi(y) - \sum_{r=1}^p \xi^i(y_r) \partial_{y_r^i} \varphi(y), \\
 y &= (y_1, \dots, y_p) \in \mathbb{R}^{pd}
 \end{aligned}$$

for differentiable functions φ of $y = (y_1, \dots, y_p) \in \mathbb{R}^{pd}$. Using here Fubini’s theorem then changing the order of the operator $L_y^\xi D_y^{p\alpha}$ and the integration against dx , by virtue of (4.2) we have

$$C = \int_{\mathbb{R}^{pd}} L_y^\xi D_y^{p\alpha} \int_{\mathbb{R}^d} \prod_{r=1}^p k_\varepsilon(x - y_r) dx \mu_p(dy) = \int_{\mathbb{R}^{pd}} L_y^\xi D_y^{p\alpha} \rho_\varepsilon(y) \mu_p(dy). \tag{4.32}$$

By Taylor’s formula

$$L_y^\xi D_y^{p\alpha} \rho_\varepsilon(y) = \int_0^1 (1 - \vartheta) \xi^i(y_k) \xi^j(y_l) (\partial_{y_k^i} \partial_{y_l^j} D^{p\alpha} \rho_\varepsilon)(y + \vartheta \bar{\xi}(y)) d\vartheta,$$

where $y = (y_1, \dots, y_p) \in \mathbb{R}^{pd}$, $y_k \in \mathbb{R}^d$ for $k = 1, 2, \dots, p$, and $\bar{\xi}(y) := (\xi(y_1), \dots, \xi(y_p))$ for $y = (y_1, \dots, y_p) \in \mathbb{R}^{dp}$. Thus by changing the order of integrals

and then changing the variables y_k with $y_k + \vartheta \xi(y_k)$ for $k = 1, 2, \dots, p$, from (4.32) we obtain

$$C = \int_0^1 (1 - \vartheta) C(\vartheta) d\vartheta \tag{4.33}$$

with

$$C(\vartheta) = \int_{\mathbb{R}^{pd}} \sum_{k=1}^p \sum_{l=1}^p \hat{\xi}^i(y_k) \hat{\xi}^j(y_l) \partial_{y_k^i} \partial_{y_l^j} D_y^{p\alpha} \rho_\varepsilon(y) \Pi_{r=1}^p \hat{u}(y_r) dy,$$

where, with $\tau_\vartheta(x) := x + \vartheta \xi(x)$,

$$\hat{\xi}^i(x) := \xi^i(\tau_\vartheta^{-1}(x)), \quad \hat{u}(x) = u(\tau_\vartheta^{-1}(x)) |\det D\tau_\vartheta^{-1}(x)|, \quad x \in \mathbb{R}^d, \quad i = 1, 2, \dots, d, \tag{4.34}$$

and $dy := dy_1 dy_2 \dots dy_p$ denotes the Lebesgue measure on \mathbb{R}^{pd} . Clearly,

$$C(\vartheta) = C_1(\vartheta) + C_2(\vartheta)$$

with

$$C_1(\vartheta) = \int_{\mathbb{R}^{pd}} \sum_{k=1}^p \hat{\xi}^i(y_k) \hat{\xi}^j(y_k) \partial_{y_k^i} \partial_{y_k^j} D_y^{p\alpha} \rho_\varepsilon(y) \Pi_{r=1}^p \hat{u}(y_r) dy,$$

$$C_2(\vartheta) = \int_{\mathbb{R}^{pd}} \sum_{k=1}^p \sum_{l \neq k}^p \hat{\xi}^i(y_k) \hat{\xi}^j(y_l) \partial_{y_k^i} \partial_{y_l^j} D_y^{p\alpha} \rho_\varepsilon(y) \Pi_{r=1}^p \hat{u}(y_r) dy.$$

Using (4.4) and the symmetry in y_k and y_l , we have

$$C_1(\vartheta) = -\frac{1}{2} \int_{\mathbb{R}^{pd}} \sum_{k=1}^p \sum_{l \neq k}^p (\hat{\xi}^i(y_k) \hat{\xi}^j(y_k) + \hat{\xi}^i(y_l) \hat{\xi}^j(y_l)) \partial_{y_k^i} \partial_{y_l^j} D_y^{p\alpha} \rho_\varepsilon(y) \Pi_{r=1}^p \hat{u}(y_r) dy,$$

$$C_2(\vartheta) = \frac{1}{2} \int_{\mathbb{R}^{pd}} \sum_{k=1}^p \sum_{l \neq k}^p (\hat{\xi}^i(y_k) \hat{\xi}^j(y_l) + \hat{\xi}^i(y_l) \hat{\xi}^j(y_k)) \partial_{y_k^i} \partial_{y_l^j} D_y^{p\alpha} \rho_\varepsilon(y) \Pi_{r=1}^p \hat{u}(y_r) dy. \tag{4.35}$$

Hence

$$C(\vartheta) = \int_{\mathbb{R}^{pd}} \sum_{r=1}^p \sum_{s \neq r}^p \hat{a}^{ij}(y_r, y_s) \partial_{y_r^i} \partial_{y_s^j} D_y^{p\alpha} \rho_\varepsilon(y) \Pi_{r=1}^p \hat{u}(y_r) dy \tag{4.36}$$

with

$$\hat{a}^{ij}(y_r, y_s) = -\frac{1}{2} (\hat{\xi}^i(y_r) - \hat{\xi}^i(y_s)) (\hat{\xi}^j(y_r) - \hat{\xi}^j(y_s))$$

Notice that the right-hand side of equation (4.36) is the same as the right-hand side of (4.12) with $\hat{\xi}^i$ in place of σ^i for each $i = 1, 2, \dots, d$ and with \hat{u} in place of u . It is easy to verify, see Lemma 3.3 in [6], that for a constant $N = N(d, \lambda, m, K_\xi)$ we have

$$\sum_{k=1}^{m+1} |D_x^k(\tau_\vartheta^{-1}(x))| \leq N, \quad \text{for each } \vartheta \in [0, 1], \mathfrak{z} \in \mathfrak{Z}, x \in \mathbb{R}^d.$$

Thus also for each $\vartheta \in [0, 1]$,

$$\sum_{k=1}^{m+1} |D^k \hat{\xi}(x, \mathfrak{z})| \leq N \bar{\xi}(\mathfrak{z}) \quad \text{for } x \in \mathbb{R}^d, \mathfrak{z} \in \mathfrak{Z}, \tag{4.37}$$

with a constant $N = N(d, m, \lambda, K_\xi)$, i.e., for each $\vartheta \in [0, 1]$ and $\mathfrak{z} \in \mathfrak{Z}$ the function $\hat{\xi}$ of $x \in \mathbb{R}^d$ satisfies the condition (4.7) on σ in Lemma 4.1, with $N \bar{\xi}(\mathfrak{z})$ in place of L . Consequently, copying the calculations which lead from equation (4.12) to the estimate (4.8) in the proof of Lemma 4.1, we obtain

$$C(\vartheta) \leq N \bar{\xi}^2(\mathfrak{z}) |\hat{u}|_{W_p^m}^p \quad \text{for each } \vartheta \in [0, 1], \mathfrak{z} \in \mathfrak{Z}$$

with a constant $N = N(d, m, p, \lambda, K_\xi)$. Note that due to the condition (4.27) there is a constant $N = N(d, p, m, \lambda, K_\xi)$ such that

$$|\hat{u}|_{W_p^m} \leq N |u|_{W_p^m} \quad \text{for all } \vartheta \in [0, 1]. \tag{4.38}$$

Hence by virtue of (4.33) the estimate (4.28) follows. □

Corollary 4.6 *Let the conditions of Lemma 4.5 hold. Then for every finite signed Borel measure μ with density $u = d\mu/dx \in W_p^m$, satisfying (4.6), we have*

$$\int_{\mathbb{R}^d} (D_x^\alpha \mu^{(\varepsilon)})^{p-1} D_x^\alpha (J^{\xi*} \mu)^{(\varepsilon)} dx \leq N \bar{\xi}^2(\mathfrak{z}) |u|_{W_p^m}^p \quad \text{for } \mathfrak{z} \in \mathfrak{Z}, \varepsilon > 0 \tag{4.39}$$

for multi-indices $\alpha, 0 \leq |\alpha| \leq m$ with a constant $N = N(d, p, m, \lambda, K_\xi)$.

Proof By the convexity of the function $f(a) = a^p, a \in \mathbb{R}$, for even integers $p \geq 2$, we know that $(a + b)^p - a^p - pa^{p-1}b \geq 0$ for all $a, b \in \mathbb{R}$. Applying this with $a = D^\alpha u^{(\varepsilon)}$ and $b = D^\alpha (I^{\xi*} u)^{(\varepsilon)}$ shows that (4.28) implies (4.39). □

Lemma 4.7 *Let the conditions of Lemma 4.5 hold. Then for every finite signed Borel measure μ with density $u = d\mu/dx \in W_p^m$, satisfying (4.6), we have*

$$\left| \int_{\mathbb{R}^d} (D^\alpha u^{(\varepsilon)} + D^\alpha (I^{\xi*} \mu)^{(\varepsilon)})^p - (D^\alpha u^{(\varepsilon)})^p dx \right| \leq N \bar{\xi}(\mathfrak{z}) |u|_{W_p^m}^p, \tag{4.40}$$

for a constant $N = N(d, p, m, \lambda, K_\xi)$ for $\mathfrak{z} \in \mathfrak{Z}$, where the argument $x \in \mathbb{R}^d$ is suppressed in the integrand.

Proof Define

$$\begin{aligned}
 F &:= \int_{\mathbb{R}^d} (D^\alpha u^{(\varepsilon)} + D^\alpha (I^{\xi*} u)^{(\varepsilon)})^p - (D^\alpha u^{(\varepsilon)})^p dx \\
 &= \int_{\mathbb{R}^d} (D^\alpha (T^{\xi*} u)^{(\varepsilon)})^p - (D^\alpha u^{(\varepsilon)})^p dx,
 \end{aligned}$$

where we use the operator T defined in (4.26). As in the proof of Lemma 4.5 in [5] we define the operator

$$M_y^\xi = \Pi_{i=1}^p T_{y_i}^\xi - \mathbb{I}$$

where \mathbb{I} is the identity operator. Observe that using Fubini’s theorem and the notation $D_y^{p\alpha} = \Pi_{r=1}^p D_{y_r}^\alpha$, $dy = dy_1 \cdots dy_p$, $y = (y_1, \dots, y_p) \in \mathbb{R}^{pd}$,

$$\begin{aligned}
 F &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^{dp}} \left(D_x^{p\alpha} \Pi_{i=1}^p T_{y_i}^\xi \Pi_{j=1}^p k_\varepsilon(x - y_j) \Pi_{k=1}^p u(y_k) \right. \\
 &\quad \left. - D_x^{p\alpha} \Pi_{j=1}^p k_\varepsilon(x - y_j) \Pi_{k=1}^p u(y_k) \right) dy dx \\
 &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^{dp}} \left(M_y^\xi D_y^{p\alpha} \Pi_{j=1}^p k_\varepsilon(x - y_j) \right) \Pi_{k=1}^p u(y_k) dy dx \\
 &= \int_{\mathbb{R}^{dp}} \left(M_y^\xi D_y^{p\alpha} \rho_\varepsilon(y) \right) \Pi_{k=1}^p u(y_k) dy.
 \end{aligned}$$

Next, note that by Taylor’s formula with $\bar{\xi}(y) = (\xi(y_1), \dots, \xi(y_p)) \in \mathbb{R}^{dp}$,

$$M_y^\xi D_y^{p\alpha} \rho_\varepsilon(y) = \sum_{k=1}^p \int_0^1 (\partial_{y_k} D_y^{p\alpha} \rho_\varepsilon)(y + \vartheta \bar{\xi}(y)) d\vartheta \xi^i(y_k).$$

Thus, by a change of variables, Fubini’s theorem and the functions defined in (4.34),

$$F = \sum_{k=1}^p \int_0^1 \int_{\mathbb{R}^{dp}} \partial_{y_k} D_y^{p\alpha} \rho_\varepsilon(y) \hat{\xi}^i(y_k) \Pi_{j=1}^p \hat{u}(y_j) dy d\vartheta,$$

which by integration by parts gives, with multi-indices $\beta \leq \alpha$, $\bar{\beta} := \alpha - \beta$ and constants c_β^α ,

$$\begin{aligned}
 F &= \sum_{\beta \leq \alpha} c_\beta^\alpha \sum_{k=1}^p \int_0^1 \int_{\mathbb{R}^{dp}} \partial_{y_k} \rho_\varepsilon(y) \hat{\xi}_\beta^i(y_k) \hat{u}_{\bar{\beta}}(y_k) \Pi_{j \neq k}^p \hat{u}_\alpha(y_j) dy d\vartheta \\
 &= \sum_{\beta \leq \alpha} c_\beta^\alpha \sum_{k=1}^p \int_0^1 f_k^\beta(\vartheta) d\vartheta,
 \end{aligned}$$

where for $k = 1, \dots, p, \vartheta \in [0, 1]$ and $\beta \leq \alpha$,

$$f_k^\beta(\vartheta) := \int_{\mathbb{R}^{dp}} \partial_{y_k^i} \rho_\varepsilon(y) \hat{\xi}_\beta^i(y_k) \hat{u}_{\bar{\beta}}(y_k) \Pi_{j \neq k}^p \hat{u}_\alpha(y_j) dy$$

and where $\hat{u}_\gamma(y_k) = D_{y_k}^\gamma \hat{u}(y_k)$ for $\gamma = \alpha, \bar{\beta}$. We consider two cases. In the first case, let $\bar{\beta} < \alpha$ and hence $|\bar{\beta}| \geq 1$. Then by integration by parts, for all $k = 1, \dots, p$ and a constant $N = N(d, p, m, \lambda)$,

$$\begin{aligned} f_k^\beta(\vartheta) &= - \int_{\mathbb{R}^{dp}} \rho_\varepsilon(y) ((\partial_{y_k^i} \hat{\xi}_\beta^i(y_k)) \hat{u}_{\bar{\beta}}(y_k) + \hat{\xi}_\beta^i(y_k) (\partial_{y_k^i} \hat{u}_{\bar{\beta}}(y_k))) \Pi_{j \neq k}^p \hat{u}_\alpha(y_j) dy d\vartheta \\ &\leq N \bar{\xi}(\mathfrak{z}) |u|_{W_p^m}^p, \end{aligned}$$

where we used (4.37) and (4.38). In the second case $\bar{\beta} = \alpha$ so that $\beta = 0$ and we have

$$\sum_{k=1}^p f_k^0 = \sum_{k=1}^p \int_{\mathbb{R}^{pd}} \partial_{y_k^i} \rho_\varepsilon(y) \hat{\xi}^i(y_k) \Pi_{j=1}^p \hat{u}_\alpha(y_j) dy,$$

as well as by using (4.4) and the symmetry in s and k ,

$$\sum_{k=1}^p f_k^0 = - \sum_{k=1}^p \sum_{s \neq k} \int_{\mathbb{R}^{pd}} \partial_{y_k^i} \rho_\varepsilon(y) \hat{\xi}^i(y_s) \Pi_{j=1}^p \hat{u}_\alpha(y_j) dy.$$

Therefore also, with a constant $N = N(d, p, m, \lambda, K_\xi)$,

$$\begin{aligned} & \left| (p-1) \sum_{k=1}^p f_k^0 + p \sum_{k=1}^p f_k^0 \right| \\ &= \left| \sum_{k=1}^p \sum_{s \neq k} \int_{\mathbb{R}^{pd}} \partial_{y_k^i} \rho_\varepsilon(y) (\hat{\xi}^i(y_k) - \hat{\xi}^i(y_s)) \Pi_{j=1}^p \hat{u}_\alpha(y_j) dy \right| \leq N \bar{\eta}(\mathfrak{z}) |u|_{W_p^m}^p, \end{aligned}$$

where we used (4.37) together with (4.5), as well as (4.38). This proves the lemma. \square

5 Solvability of the filtering equations in Sobolev spaces

The following two lemmas are contained in Lemma 5.2 in [5], in the special case of multi-index $\alpha = 0$, i.e., when D^α is the identity operator. Their proof goes in the same way, as that of Lemma 5.2 in [5], if we consider the kernel $D^\alpha k_\varepsilon$ in place of k_ε in the proof of Lemma 5.2. Hence we only provide an outline and refer the reader to the preceding article [5] for full details.

Lemma 5.1 *Let the Assumption 2.1 hold. Let u be an L_p -solution of (3.3), $p \geq 2$, such that (3.2) holds for each $\varphi \in C_0^\infty(\mathbb{R}^d)$ almost surely for all $t \in [0, T]$ and*

assume moreover that $\text{ess sup}_{t \in [0, T]} |u_t|_{L^1} < \infty$. If $K_1 \neq 0$ in Assumption 2.1 (ii), then assume additionally

$$\text{ess sup}_{t \in [0, T]} \int_{\mathbb{R}^d} |y|^2 |u_t(y)| dy < \infty, \quad \text{almost surely.} \tag{5.1}$$

Then for each $\varepsilon > 0$ and integer $m \geq 0$, for any multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$, $|\alpha| \leq m$, for all $x \in \mathbb{R}^d$ almost surely

$$\begin{aligned} D^\alpha u_t^{(\varepsilon)}(x) &= D^\alpha u_0^{(\varepsilon)}(x) + \int_0^t D^\alpha (\tilde{\mathcal{L}}_s^* u_s)^{(\varepsilon)}(x) ds + \int_0^t D^\alpha (\mathcal{M}_s^{*k} u_s)^{(\varepsilon)}(x) dV_s^k \\ &+ \int_0^t \int_{\mathfrak{Z}_0} D^\alpha (J_s^{\eta*} u_s)^{(\varepsilon)}(x) \nu_0(d\mathfrak{z}) ds + \int_0^t \int_{\mathfrak{Z}_1} D^\alpha (J_s^{\xi*} u_s)^{(\varepsilon)}(x) \nu_1(d\mathfrak{z}) ds \\ &+ \int_0^t \int_{\mathfrak{Z}_1} D^\alpha (I_s^{\xi*} u_{s-})^{(\varepsilon)}(x) \tilde{N}_1(ds, d\mathfrak{z}), \end{aligned} \tag{5.2}$$

for all $t \in [0, T]$.

Proof The case of $\alpha = 0$ is Lemma 5.4 in [5]. The case of $\alpha \neq 0$ such that $0 < |\alpha| \leq m$ works exactly in the same way. We first define for a $\psi \in C_0^\infty(\mathbb{R})$ such that $\psi(0) = 1$, $\psi(r) = 0$ for $|r| \geq 2$, for $n \geq 1$, $\psi_n(x) := \psi(|x|/n) \in C_0^\infty(\mathbb{R}^d)$. Setting $\varphi_x(y) := k_{\varepsilon, \alpha}(x - y)\psi_n(y) \in C_0^\infty$ in (3.3), where $k_{\varepsilon, \alpha}(x - y) = D_x^\alpha k_\varepsilon(x - y)$, yields that for each $x \in \mathbb{R}^d$ almost surely

$$\begin{aligned} (u_t, k_{\varepsilon, \alpha}(x - \cdot)\psi_n) &= (u_0, k_{\varepsilon, \alpha}(x - \cdot)\psi_n) + \int_0^t (u_s, \tilde{\mathcal{L}}_s(k_{\varepsilon, \alpha}(x - \cdot)\psi_n)) ds \\ &+ \int_0^t (u_s, \mathcal{M}_s^k(k_{\varepsilon, \alpha}(x - \cdot)\psi_n)) dV_s^k \\ &+ \int_0^t \int_{\mathfrak{Z}_0} (u_s, J_s^\eta(k_{\varepsilon, \alpha}(x - \cdot)\psi_n)) \nu_0(d\mathfrak{z}) ds \\ &+ \int_0^t \int_{\mathfrak{Z}_1} (u_s, J_s^\xi(k_{\varepsilon, \alpha}(x - \cdot)\psi_n)) \nu_1(d\mathfrak{z}) ds \\ &+ \int_0^t \int_{\mathfrak{Z}_1} (u_{s-}, I_s^\xi(k_{\varepsilon, \alpha}(x - \cdot)\psi_n)) \tilde{N}_1(d\mathfrak{z}, ds) \end{aligned} \tag{5.3}$$

for all $t \in [0, T]$. Then we notice that

$$|k_{\varepsilon, \alpha}(x - y)| \leq \sum_{|\gamma| \leq m+2} |D^\gamma k_\varepsilon(x - y)| \leq Nk_{2\varepsilon}(x - y), \tag{5.4}$$

as well as that by Assumption for all $x, y \in \mathbb{R}^d$, $s \in [0, T]$, $\mathfrak{z}_i \in \mathfrak{Z}_i$, $i = 0, 1$ and $n \geq 0$ we have

$$\sup_{x \in \mathbb{R}^d} |D^k \psi_n| = n^{-k} \sup_{\mathbb{R}^d} |D^k \psi| < \infty, \quad \text{for } k \in \mathbb{N}, .$$

$$\begin{aligned}
 |\tilde{\mathcal{L}}_s(k_{\varepsilon,\alpha}(x-y)\psi_n(y))| + \sum_k |\mathcal{M}_s^k(k_{\varepsilon,\alpha}(x-y)\psi_n(y))|^2 &\leq N(K_0^2 + K_1^2|y|^2 + K_1^2|Y_s|^2), \\
 |J_s^\eta(k_{\varepsilon,\alpha}(x-y)\psi_n(y))| &\leq \sup_{v \in \mathbb{R}^d} |D_v^2(k_{\varepsilon,\alpha}(x-v)\psi_n(v))| |\eta_s(y, \mathfrak{z}_0)|^2 \leq N|\eta_s(y, \mathfrak{z}_0)|^2 \\
 &\leq N\bar{\eta}^2(\mathfrak{z}_0)(K_0^2 + K_1^2|y|^2 + K_1^2|Y_s|^2),
 \end{aligned}$$

and

$$\begin{aligned}
 |J_s^\xi(k_{\varepsilon,\alpha}(x-y)\psi_n(y))| + |I_s^\xi(k_{\varepsilon,\alpha}(x-y)\psi_n(y))|^2 \\
 \leq \sup_{v \in \mathbb{R}^d} |D_v^2(k_{\varepsilon,\alpha}(x-v)\psi_n(v))| |\xi_s(y, \mathfrak{z}_1)|^2 + \sup_{v \in \mathbb{R}^d} |D_v(k_{\varepsilon,\alpha}(x-v)\psi_n(v))|^2 |\xi_s(y, \mathfrak{z}_1)|^2 \\
 \leq N|\xi_s(y, \mathfrak{z}_1)|^2 \leq N\bar{\xi}^2(\mathfrak{z}_1)(K_0^2 + K_1^2|y|^2 + K_1^2|Y_s|^2),
 \end{aligned}$$

for a constant $N = N(\varepsilon, m, d, K_0, K_1, K, K_\xi, K_\eta)$. Using

$$\text{ess sup}_{t \in [0, T]} \int_{\mathbb{R}^d} (1 + |y|^2 + |Y_t|^2) |u_t(y)| dy < \infty, \quad (\text{a.s.})$$

together with the estimates above, we can apply Lebesgue’s theorem on Dominated Convergence to get that for all $x \in \mathbb{R}^d$,

$$\begin{aligned}
 (u_t, k_{\varepsilon,\alpha}(x - \cdot)\psi_n) \rightarrow (u_t, k_{\varepsilon,\alpha}(x - \cdot)), \quad (u_0, k_{\varepsilon,\alpha}(x - \cdot)\psi_n) \rightarrow (u_0, k_{\varepsilon,\alpha}(x - \cdot)) \quad \text{and} \\
 \int_0^t (u_s, \mathcal{A}_s(k_{\varepsilon,\alpha}(x - \cdot)\psi_n)) ds \rightarrow \int_0^t (u_s, \mathcal{A}_s k_{\varepsilon,\alpha}(x - \cdot)) ds
 \end{aligned}$$

as $n \rightarrow \infty$, almost surely uniformly in time, as well as that

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \int_0^t (u_s, \mathcal{M}_s^k(k_{\varepsilon,\alpha}(x - \cdot)\psi_n(\cdot))) dV_s^k &= \int_0^t (u_s, \mathcal{M}_s^k k_{\varepsilon,\alpha}(x - \cdot)) dV_s^k, \\
 \lim_{n \rightarrow \infty} \int_0^t \int_{\mathfrak{Z}_1} (u_{s-}, I_s^\xi(k_{\varepsilon,\alpha}(x - \cdot)\psi_n)) \tilde{N}_1(d\mathfrak{z}, ds) &= \int_0^t \int_{\mathfrak{Z}_1} (u_{s-}, I_s^\xi k_{\varepsilon,\alpha}(x - \cdot)) \tilde{N}_1(d\mathfrak{z}, ds)
 \end{aligned}$$

in probability, uniformly in time. Thus, letting $n \rightarrow \infty$ in (5.3) it remains to note that since \mathcal{A} acts in the y variable,

$$\begin{aligned}
 (u_s, \mathcal{A}_s k_{\varepsilon,\alpha}(x - \cdot)) &= \int_{\mathbb{R}^d} u_s(y) \mathcal{A}_s D_x^\alpha k_\varepsilon(x - y) dy \\
 &= D_x^\alpha \int_{\mathbb{R}^d} u_s(y) \mathcal{A}_s k_\varepsilon(x - y) dy = D^\alpha (\mathcal{A}_s^* u_s)^{(\varepsilon)}(x)
 \end{aligned}$$

for all $(\omega, s, x) \in \Omega \times [0, T] \times \mathbb{R}^d$ if $\mathcal{A} = \tilde{\mathcal{L}}, \mathcal{M}^k$ or the identity, as well as for all $(\omega, s, x, \mathfrak{z}_i) \in \Omega \times [0, T] \times \mathbb{R}^d \times \mathfrak{Z}_i$ if $\mathcal{A} = J^\eta$ or $\mathcal{A} = I^\xi, J^\xi$ with $i = 0, 1$ respectively. □

Lemma 5.2 *Let Assumptions 2.1 and 2.2 hold. Let u be an L_p -solution of (3.3), $p \geq 2$ and assume moreover that $\text{ess sup}_{t \in [0, T]} |u_t|_{L_1} < \infty$. If $K_1 \neq 0$ in Assumption 2.1 (ii), then assume additionally (5.1). Then for each $\varepsilon > 0$ and integer $m \geq 0$, for any multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$, $|\alpha| \leq m$, almost surely*

$$\begin{aligned}
 |D^\alpha u_t^{(\varepsilon)}|_{L_p}^p &= |D^\alpha u_0^{(\varepsilon)}|_{L_p}^p + p \int_0^t (|D^\alpha u_s^{(\varepsilon)}|^{p-2} D^\alpha u_s^{(\varepsilon)}, D^\alpha (\tilde{\mathcal{L}}_s^* u_s)^{(\varepsilon)}) ds \\
 &\quad + p \int_0^t (|D^\alpha u_s^{(\varepsilon)}|^{p-2} D^\alpha u_s^{(\varepsilon)}, D^\alpha (\mathcal{M}_s^{k*} u_s)^{(\varepsilon)}) dV_s^k \\
 &\quad + \frac{p(p-1)}{2} \sum_k \int_0^t (|D^\alpha u_s^{(\varepsilon)}|^{p-2}, |D^\alpha (\mathcal{M}_s^{k*} u_s)^{(\varepsilon)}|^2) ds \\
 &\quad + p \int_0^t \int_{\mathfrak{Z}_0} (|D^\alpha u_s^{(\varepsilon)}|^{p-2} D^\alpha u_s^{(\varepsilon)}, D^\alpha (J_s^{\eta*} \mu_s)^{(\varepsilon)}) \nu_0(d\mathfrak{z}) ds \\
 &\quad + p \int_0^t \int_{\mathfrak{Z}_1} (|D^\alpha u_s^{(\varepsilon)}|^{p-2} D^\alpha u_s^{(\varepsilon)}, D^\alpha (J_s^{\xi*} \mu_s)^{(\varepsilon)}) \nu_1(d\mathfrak{z}) ds \\
 &\quad + p \int_0^t \int_{\mathfrak{Z}_1} (|D^\alpha u_{s-}^{(\varepsilon)}|^{p-2} D^\alpha u_{s-}^{(\varepsilon)}, D^\alpha (I_s^{\xi*} u_{s-})^{(\varepsilon)}) \tilde{N}_1(d\mathfrak{z}, ds) \\
 &\quad + \int_0^t \int_{\mathfrak{Z}_1} \int_{\mathbb{R}^d} \left\{ |D^\alpha u_{s-}^{(\varepsilon)} + D^\alpha (I_s^{\xi*} \mu_{s-})^{(\varepsilon)}|^p - |D^\alpha u_{s-}^{(\varepsilon)}|^p \right. \\
 &\quad \left. - p |D^\alpha u_{s-}^{(\varepsilon)}|^{p-2} D^\alpha u_{s-}^{(\varepsilon)} D^\alpha (I_s^{\xi*} \mu_{s-})^{(\varepsilon)} \right\} dx N_1(d\mathfrak{z}, ds) \tag{5.5}
 \end{aligned}$$

holds for all $t \in [0, T]$.

Proof We apply the Itô formula, Theorem 5.1 in [5], to $|D^\alpha u_t^{(\varepsilon)}|_{L_p}^p$. In order to do that, we need to verify that almost surely for each $x \in \mathbb{R}^d$ and α , such that $0 \leq |\alpha| \leq m$,

$$\begin{aligned}
 \int_0^T |D^\alpha (\tilde{\mathcal{L}}_s^* u_s)^{(\varepsilon)}(x)| ds &< \infty, \quad \int_0^T \sum_k |D^\alpha (\mathcal{M}_s^{k*} u_s)^{(\varepsilon)}(x)|^2 ds < \infty, \\
 \int_0^T \int_{\mathfrak{Z}_0} |D^\alpha (J_s^{\eta*} u_s)^{(\varepsilon)}(x)| \nu_0(d\mathfrak{z}) ds &< \infty, \\
 \int_0^T \int_{\mathfrak{Z}_1} |D^\alpha (J_s^{\xi*} u_s)^{(\varepsilon)}(x)| \nu_1(d\mathfrak{z}) ds &< \infty, \\
 \int_0^T \int_{\mathfrak{Z}_1} |D^\alpha (I_s^{\xi*} u_s)^{(\varepsilon)}(x)|^2 \nu_1(d\mathfrak{z}) ds &< \infty,
 \end{aligned}$$

that for every finite set $\Gamma \in \mathcal{B}(\mathbb{R}^d)$, almost surely

$$\begin{aligned}
 \int_\Gamma \int_0^T |D^\alpha (\tilde{\mathcal{L}}_s^* u_s)^{(\varepsilon)}(x)| dx ds &< \infty, \quad \int_\Gamma \left(\int_0^T \sum_k |D^\alpha (\mathcal{M}_s^{k*} u_s)^{(\varepsilon)}(x)|^2 ds \right)^{1/2} dx < \infty, \\
 \int_\Gamma \int_0^T \int_{\mathfrak{Z}_0} |D^\alpha (J_s^{\eta*} u_s)^{(\varepsilon)}(x)| \nu_0(d\mathfrak{z}) dx ds &< \infty,
 \end{aligned}$$

$$\int_{\Gamma} \int_0^T \int_{\mathfrak{Z}_1} |D^\alpha (J_s^{\xi^*} u_s)^{(\varepsilon)}(x)| v_1(d\mathfrak{z}) dx ds < \infty,$$

$$\int_{\Gamma} \left(\int_0^T \int_{\mathfrak{Z}_1} |D^\alpha (I_s^{\xi^*} u_s)^{(\varepsilon)}(x)|^2 v_1(d\mathfrak{z}) ds \right)^{1/2} dx < \infty,$$

as well as that almost surely

$$A := \int_0^T \int_{\mathbb{R}^d} |D^\alpha (\tilde{L}_s^* u_s)^{(\varepsilon)}(x)|^p dx ds < \infty,$$

$$A_\eta := \int_0^T \int_{\mathbb{R}^d} \left| \int_{\mathfrak{Z}_0} D^\alpha (J_s^{\eta^*} u_s)^{(\varepsilon)}(x) v_0(d\mathfrak{z}) \right|^p dx ds < \infty,$$

$$A_\xi := \int_0^T \int_{\mathbb{R}^d} \left| \int_{\mathfrak{Z}_1} D^\alpha (J_s^{\xi^*} u_s)^{(\varepsilon)}(x) v_1(d\mathfrak{z}) \right|^p dx ds < \infty,$$

$$B := \int_0^T \int_{\mathbb{R}^d} \left(\sum_k |D^\alpha (\mathcal{M}_s^{k^*} u_s)^{(\varepsilon)}(x)|^2 \right)^{p/2} dx ds < \infty,$$

$$G := \int_0^T \int_{\mathbb{R}^d} \int_{\mathfrak{Z}_1} |D^\alpha (I_s^{\xi^*} u_s)^{(\varepsilon)}(x, \mathfrak{z})|^p v_1(d\mathfrak{z}) dx ds < \infty,$$

$$H := \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathfrak{Z}_1} |D^\alpha (I_s^{\xi^*} u_s)^{(\varepsilon)}(x, \mathfrak{z})|^2 v_1(d\mathfrak{z}) \right)^{p/2} dx ds < \infty.$$

For $\alpha = 0$ the claim is Lemma 5.4 in [5] and the estimates can be found in the proof of the preceding Lemma 5.2 therein. To prove the case where $0 < |\alpha| \leq m$, we note that for $\mathcal{A} = \tilde{\mathcal{L}}, \mathcal{M}, I^\xi, J^\xi, J^\eta$ we have

$$D^\alpha (\mathcal{A}^* u)^{(\varepsilon)} = \int_{\mathbb{R}^d} D_x^\alpha (\mathcal{A}_y k_\varepsilon(x - y)) u(y) dy = \int_{\mathbb{R}^d} (\mathcal{A}_y k_{\varepsilon, \alpha}(x - y)) u(y) dy,$$

for $k_{\varepsilon, \alpha}(x) = D^\alpha k_\varepsilon(x)$. Hence, a word for word repetition of the proof of Lemma 5.2 & 5.4 in [5], where we replace k_ε by $k_{\varepsilon, \alpha}$ and recall (5.4), yields the desired result. \square

Lemma 5.3 *Let Assumptions 2.1, 2.2, 2.4 and 2.5 hold with an integer $m \geq 0$. Let $(u_t)_{t \in [0, T]}$ be a W_p^m -solution to (3.3), for an even integer $p \geq 2$, such that $\mathbb{E}|u_0|_{W_p^m}^p < \infty$ and almost surely $\text{ess sup}_{t \in [0, T]} |u_t|_{L_1} < \infty$. If $K_1 \neq 0$ in Assumption 2.1, then assume additionally*

$$\text{ess sup}_{t \in [0, T]} \int_{\mathbb{R}^d} |y|^2 |u_t(y)| dy.$$

Then

$$\mathbb{E} \sup_{t \in [0, T]} |u_t|_{W_p^m}^p \leq N \mathbb{E}|u_0|_{W_p^m}^p \tag{5.6}$$

for a constant $N = N(m, d, p, K, K_\eta, K_\xi, L, T, \lambda, |\bar{\xi}|_{L_2(\mathfrak{Z}_1)}, |\bar{\eta}|_{L_2(\mathfrak{Z}_1)})$.

Proof For $m = 0$ the claim is Lemma 5.4 in [5]. We proceed similarly here. For the present case, fix a multi-index α such that $0 \neq |\alpha| \leq m$, and define

$$\begin{aligned} Q_p(\alpha, b, \sigma, \rho, \beta, u, k_\varepsilon) &= p((D^\alpha u^{(\varepsilon)})^{p-1}, D^\alpha(\tilde{\mathcal{L}}^* u)^{(\varepsilon)}) \\ &\quad + \frac{p(p-1)}{2} \sum_k ((D^\alpha u^{(\varepsilon)})^{p-2}, (D^\alpha(\mathcal{M}^{k*} u)^{(\varepsilon)})^2), \\ Q_p^{(0)}(\alpha, \eta(\mathfrak{z}_0), u, k_\varepsilon) &= p((D^\alpha u^{(\varepsilon)})^{p-1}, D^\alpha(J^{\eta(\mathfrak{z}_0)*} u)^{(\varepsilon)}), \\ Q_p^{(1)}(\alpha, \xi(\mathfrak{z}_1), u, k_\varepsilon) &= p((D^\alpha u^{(\varepsilon)})^{p-1}, D^\alpha(J^{\xi(\mathfrak{z}_1)*} u)^{(\varepsilon)}), \\ \mathcal{R}_p(\alpha, \xi(\mathfrak{z}_1), u, k_\varepsilon) &= |D^\alpha u^{(\varepsilon)} + D^\alpha(I^{\xi(\mathfrak{z}_1)*} u)^{(\varepsilon)}|_{L_p}^p - |D^\alpha u^{(\varepsilon)}|_{L_p}^p \\ &\quad - p((D^\alpha u^{(\varepsilon)})^{p-1}, D^\alpha(I^{\xi(\mathfrak{z}_1)*} u)^{(\varepsilon)}), \end{aligned} \tag{5.7}$$

for $u \in W_p^m, \beta \in \mathbb{R}^{d'}$, functions b, σ and ρ on \mathbb{R}^d , with values in $\mathbb{R}^d, \mathbb{R}^{d \times d_1}$ and $\mathbb{R}^{d \times d'}$, respectively, and \mathbb{R}^d -valued functions $\eta(\mathfrak{z}_0)$ and $\xi(\mathfrak{z}_1)$ for each $\mathfrak{z}_i, i = 0, 1$, where $\beta_t = B_t(X_t)$,

$$\tilde{\mathcal{L}} = \frac{1}{2}(\sigma^{il} \sigma^{jl} + \rho^{ik} \rho^{jk}) D_{ij} + \beta^l \rho^{il} D_i + \beta^l B^l, \quad \mathcal{M}^k = \rho^{ik} D_i + B^k, \quad k = 1, 2, \dots, d'. \tag{5.8}$$

By Lemma 5.2 almost surely

$$\begin{aligned} d|D^\alpha u_t^{(\varepsilon)}|_{L_p}^p &= Q_p(\alpha, b_t, \sigma_t, \rho_t, \beta_t, u_t, k_\varepsilon) dt + \int_{\mathfrak{z}_0} Q_p^{(0)}(\alpha, \eta_t(\mathfrak{z}), u_t, k_\varepsilon) \nu_0(d\mathfrak{z}) dt \\ &\quad + \int_{\mathfrak{z}_1} Q_p^{(1)}(\alpha, \xi_t(\mathfrak{z}), u_t, k_\varepsilon) \nu_1(d\mathfrak{z}) dt + \int_{\mathfrak{z}_1} \mathcal{R}_p(\alpha, \xi_t(\mathfrak{z}), u_{t-}, k_\varepsilon) N_1(d\mathfrak{z}, dt) \\ &\quad + d\zeta_1(\alpha, t) + d\zeta_2(\alpha, t), \end{aligned} \tag{5.9}$$

for all $t \in [0, T]$, and

$$\begin{aligned} \zeta_1(\alpha, t) &= p \int_0^t ((D^\alpha u_s^{(\varepsilon)})^{p-1}, D^\alpha(\mathcal{M}_s^{k*} u_s)^{(\varepsilon)}) dV_s^k, \\ \zeta_2(\alpha, t) &= p \int_0^t \int_{\mathfrak{z}_1} ((D^\alpha u_s^{(\varepsilon)})^{p-1}, D^\alpha(I_s^{\xi*} u_s)^{(\varepsilon)}) \tilde{N}_1(d\mathfrak{z}, ds) \quad t \in [0, T] \end{aligned} \tag{5.10}$$

are local martingales under P . We write

$$\int_{\mathfrak{z}_1} \mathcal{R}_p(\alpha, \xi_t(\mathfrak{z}_1), u_{t-}, k_\varepsilon) N_1(d\mathfrak{z}, dt) = \int_{\mathfrak{z}_1} \mathcal{R}_p(\alpha, \xi_t(\mathfrak{z}_1), u_{t-}, k_\varepsilon) \nu_1(d\mathfrak{z}) dt + d\zeta_3(\alpha, t) \tag{5.11}$$

with

$$\begin{aligned} \zeta_3(\alpha, t) &= \int_0^t \int_{\mathfrak{z}_1} \mathcal{R}_p(\alpha, \xi_s(\mathfrak{z}), u_{s-}, k_\varepsilon) N_1(d\mathfrak{z}, ds) \\ &\quad - \int_0^t \int_{\mathfrak{z}_1} \mathcal{R}_p(\alpha, \xi_s(\mathfrak{z}), u_{s-}, k_\varepsilon) \nu_1(d\mathfrak{z}) ds, \end{aligned}$$

which we can justify if we show

$$A := \int_0^T \int_{\mathfrak{Z}_1} |\mathcal{R}_p(\alpha, \xi_s(\mathfrak{z}), u_s, k_\varepsilon)| \nu_1(d\mathfrak{z}) ds < \infty \text{ (a.s.)} \tag{5.12}$$

To this end observe that by Taylor’s formula

$$0 \leq \mathcal{R}_p(\alpha, \xi_t(\mathfrak{z}), u_t, k_\varepsilon) \leq N \int_{\mathbb{R}^d} (D^\alpha u_t^{(\varepsilon)})^{p-2} (D^\alpha (I^{\xi(\mathfrak{z})^*} u_t)^{(\varepsilon)})^2 + (D^\alpha (I^{\xi(\mathfrak{z})^*} u_t)^{(\varepsilon)})^p dx \tag{5.13}$$

with a constant $N = N(d, p)$. Hence

$$\begin{aligned} & \int_{\mathfrak{Z}_1} \mathcal{R}_p(\alpha, \xi_t(\mathfrak{z}), u_t, k_\varepsilon) \nu_1(d\mathfrak{z}) \\ & \leq N \int_{\mathbb{R}^d} (D^\alpha u_t^{(\varepsilon)})^{p-2} |D^\alpha (I^{\xi(\mathfrak{z})^*} u_t)^{(\varepsilon)}|_{L^2(\mathfrak{Z}_1)}^2 + |D^\alpha (I^{\xi(\mathfrak{z})^*} u_t)^{(\varepsilon)}|_{L^p(\mathfrak{Z}_1)}^p dx \\ & \leq N' (|D^\alpha u_t^{(\varepsilon)}|_{L^p}^p + A_1(t) + A_2(t)) \end{aligned}$$

with

$$A_1(t) = \int_{\mathbb{R}^d} |D^\alpha (I^{\xi(\mathfrak{z})^*} u_t)^{(\varepsilon)}|_{L^2(\mathfrak{Z}_1)}^p dx, \quad A_2(t) = \int_{\mathbb{R}^d} |D^\alpha (I^{\xi(\mathfrak{z})^*} u_t)^{(\varepsilon)}|_{L^p(\mathfrak{Z}_1)}^p dx \tag{5.14}$$

and constants N and N' depending only on d and p . By Minkowski’s inequality, using again that $D_x^\alpha I^\xi k_\varepsilon(x - y) = I^\xi D_x^\alpha k_\varepsilon(x - y)$, and using the notation $k_{\varepsilon, \alpha} = D^\alpha k_\varepsilon$ we have

$$\begin{aligned} |D^\alpha u_t^{(\varepsilon)}|_{L^p}^p &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} k_{\varepsilon, \alpha}(x - y) u_t(y) dy \right|^p dx, \\ &\leq \int_{\mathbb{R}^d} |k_{\varepsilon, \alpha}|_{L^p} |u_t(y)| dy|^p \leq |k_{\varepsilon, \alpha}|_{L^p}^p |u_t|_{L^1}^p, \end{aligned} \tag{5.15}$$

$$\begin{aligned} A_1(t) &= \int_{\mathbb{R}^d} \left| \int_{\mathfrak{Z}_1} \left| \int_{\mathbb{R}^d} (k_{\varepsilon, \alpha}(\cdot - y - \xi_t(y, \mathfrak{z})) - k_{\varepsilon, \alpha}(\cdot - y)) u_t(y) dy \right|^2 \nu_1(d\mathfrak{z}) \right|^{p/2} dx \\ &\leq \int_{\mathfrak{Z}_1} \left| \int_{\mathbb{R}^d} |k_{\varepsilon, \alpha}(\cdot - y - \xi_t(y, \mathfrak{z})) - k_{\varepsilon, \alpha}(\cdot - y)| |u_t(y)| dy \right|_{L^p}^2 \nu_1(d\mathfrak{z}) \Big|^{p/2} \\ &\leq \int_{\mathfrak{Z}_1} \left| \int_{\mathbb{R}^d} |Dk_{\varepsilon, \alpha}|_{L^p} \bar{\xi}(\mathfrak{z}_1) (K_0 + K_1|y| + K_1|Y_t|) |u_t(y)| dy \right|_{L^p}^2 \nu_1(d\mathfrak{z}) \Big|^{p/2} \\ &\leq |Dk_{\varepsilon, \alpha}|_{L^p}^p |\bar{\xi}|_{L^2(\mathfrak{Z}_1)}^p \left(\int_{\mathbb{R}^d} (K_0 + K_1|y| + K_1|Y_t|) |u_t(y)| dy \right)^p, \end{aligned} \tag{5.16}$$

and similarly, using Assumption 2.2,

$$A_2(t) = \int_{\mathbb{R}^d} \int_{\mathfrak{Z}_1} \left| \int_{\mathbb{R}^d} (k_{\varepsilon, \alpha}(x - y - \xi(t, y, \mathfrak{z})) - k_{\varepsilon, \alpha}(x - y)) u_t(y) dy \right|^p \nu_1(d\mathfrak{z}) dx$$

$$\begin{aligned} &\leq \int_{\mathfrak{Z}_1} \left| \int_{\mathbb{R}^d} |k_{\varepsilon, \alpha}(\cdot - y - \xi(t, y, \mathfrak{z})) - k_{\varepsilon, \alpha}(\cdot - y)|_{L_p} |u_t(y)| dy \right|^p \nu_1(d\mathfrak{z}) \\ &\leq K_\xi^{p-2} |Dk_{\varepsilon, \alpha}|_{L_p}^p |\bar{\xi}|_{L_2(\mathfrak{Z}_1)}^2 \left(\int_{\mathbb{R}^d} (K_0 + K_1|y| + K_1|Y_t|)|u_t(y)| dy \right)^p. \end{aligned} \tag{5.17}$$

By (5.13)–(5.17) we have a constant $N = N(p, d, \varepsilon, |\bar{\xi}|_{L_2(\mathfrak{Z}_1)}, K_\xi)$ such that

$$A \leq N \int_0^T |D^\alpha u_t^{(\varepsilon)}|_{L_p}^p + N \int_0^T \left(\int_{\mathbb{R}^d} (K_0 + K_1|y| + K_1|Y_t|)|u_t(y)| dy \right)^p dt < \infty \text{ (a.s.)}$$

Next we claim that, with the operator T^ξ defined in (4.26), we have

$$\zeta_2(\alpha, t) + \zeta_3(\alpha, t) = \int_0^t \int_{\mathfrak{Z}_1} (|D^\alpha (T_s^{\xi*} u_s)^{(\varepsilon)}|_{L_p}^p - |D^\alpha u_s^{(\varepsilon)}|_{L_p}^p) \tilde{N}_1(d\mathfrak{z}, ds) =: \zeta(\alpha, t) \tag{5.18}$$

for $t \in [0, T]$. For that purpose, note first that $D^\alpha u^{(\varepsilon)} + D^\alpha (I^{\xi(\mathfrak{z}_1)*} u)^{(\varepsilon)} = D^\alpha (T^{\xi*} u_s)^{(\varepsilon)}$. To see that the stochastic integral $\zeta(\alpha, t)$ is well-defined as an Itô integral note that by Lemma 4.7,

$$\begin{aligned} &\int_0^T \int_{\mathfrak{Z}_1} ||D^\alpha (T_s^{\xi*} u_s)^{(\varepsilon)}|_{L_p}^p - |D^\alpha u_s^{(\varepsilon)}|_{L_p}^p|^2 \nu_1(d\mathfrak{z}) ds \\ &\leq N |\bar{\xi}|_{L_2(\mathfrak{Z}_1)}^2 \int_0^T |u_s|_{W_p^m}^{2p} ds < \infty \text{ (a.s.)} \end{aligned} \tag{5.19}$$

with a constant $N = N(d, p, m, \lambda, K_\xi)$. Since ν_1 is σ -finite, there is an increasing sequence $(\mathfrak{Z}_{1n})_{n=1}^\infty, \mathfrak{Z}_{1n} \in \mathcal{Z}_1$, such that $\nu_1(\mathfrak{Z}_{1n}) < \infty$ for every n and $\cup_{n=1}^\infty \mathfrak{Z}_{1n} = \mathfrak{Z}_1$. Then it is easy to see that

$$\begin{aligned} \bar{\zeta}_{2n}(\alpha, t) &= p \int_0^t \int_{\mathfrak{Z}_{1n}} \mathbf{1}_{\mathfrak{Z}_{1n}}(\mathfrak{z}) ((D^\alpha u_s^{(\varepsilon)})^{p-1}, D^\alpha (I_s^{\xi*} u_s)^{(\varepsilon)}) N(d\mathfrak{z}, ds), \\ \hat{\zeta}_{2n}(\alpha, t) &= p \int_0^t \int_{\mathfrak{Z}_{1n}} \mathbf{1}_{\mathfrak{Z}_{1n}}(\mathfrak{z}) ((D^\alpha u_s^{(\varepsilon)})^{p-1}, D^\alpha (I_s^{\xi*} u_s)^{(\varepsilon)}) \nu_1(d\mathfrak{z}) ds, \\ \bar{\zeta}_{3n}(\alpha, t) &= \int_0^t \int_{\mathfrak{Z}_{1n}} \mathbf{1}_{\mathfrak{Z}_{1n}}(\mathfrak{z}) \mathcal{R}_p(\alpha, \xi_s(\mathfrak{z}), u_{s-}, k_\varepsilon) N_1(d\mathfrak{z}, ds), \\ \hat{\zeta}_{3n}(\alpha, t) &= \int_0^t \int_{\mathfrak{Z}_{1n}} \mathbf{1}_{\mathfrak{Z}_{1n}}(\mathfrak{z}) \mathcal{R}_p(\alpha, \xi_s(\mathfrak{z}), u_{s-}, k_\varepsilon) \nu_1(d\mathfrak{z}) ds \end{aligned}$$

are well-defined, and

$$\zeta_2(\alpha, t) = \lim_{n \rightarrow \infty} (\bar{\zeta}_{2n}(\alpha, t) - \hat{\zeta}_{2n}(\alpha, t)), \quad \zeta_3(\alpha, t) = \lim_{n \rightarrow \infty} \bar{\zeta}_{3n}(\alpha, t) - \lim_{n \rightarrow \infty} \hat{\zeta}_{3n}(\alpha, t),$$

where the limits are understood in probability. Hence

$$\zeta_2(\alpha, t) + \zeta_3(\alpha, t) = \lim_{n \rightarrow \infty} \left(\bar{\zeta}_{2n}(\alpha, t) + \bar{\zeta}_{3n}(\alpha, t) - (\hat{\zeta}_{2n}(\alpha, t) + \hat{\zeta}_{3n}(\alpha, t)) \right)$$

$$\begin{aligned}
 &= \lim_{n \rightarrow \infty} \left(\int_0^t \int_{\mathfrak{Z}_1} \mathbf{1}_{\mathfrak{Z}_{1n}}(\mathfrak{z}) (|D^\alpha (T^{\xi*} u_s)^{(\varepsilon)}|_{L_p}^p - |D^\alpha u_s^{(\varepsilon)}|_{L_p}^p) \tilde{N}_1(d\mathfrak{z}, ds) \right) \\
 &= \zeta(\alpha, t),
 \end{aligned}$$

which completes the proof of (5.18). Consequently, from (5.9)-(5.11) we have

$$\begin{aligned}
 d|D^\alpha u_t^{(\varepsilon)}|_{L_p}^p &= \mathcal{Q}_p(\alpha, b_t, \sigma_t, \rho_t, \beta_t, u_t, k_\varepsilon) dt + \int_{\mathfrak{Z}_0} \mathcal{Q}_p^{(0)}(\alpha, \eta_t(\mathfrak{z}_0), u_t, k_\varepsilon) \nu_0(d\mathfrak{z}) dt \\
 &+ \int_{\mathfrak{Z}_1} \mathcal{Q}_p^{(1)}(\alpha, \xi_t(\mathfrak{z}_1), u_t, k_\varepsilon) + \mathcal{R}_p(\alpha, \xi_t(\mathfrak{z}_1), u_t, k_\varepsilon) \nu_1(d\mathfrak{z}) dt + d\zeta_1(\alpha, t) + d\zeta(\alpha, t).
 \end{aligned} \tag{5.20}$$

By Lemma 4.1, Corollary 4.2, Lemma 4.3 and Lemma 4.4 we have

$$\mathcal{Q}_p(\alpha, b_s, \sigma_s, \rho_s, \beta_s, u_s, k_\varepsilon) \leq N(L^2 + K^2) |u_s|_{W_p^m}^p \tag{5.21}$$

with a constant $N = N(d, p, m)$, and by Lemma 4.5 and Corollary 4.6, using that $\bar{\xi} \leq K_\xi$ and $\bar{\eta} \leq K_\eta$,

$$\mathcal{Q}_p^{(0)}(\alpha, \eta_s(\mathfrak{z}), u_s, k_\varepsilon) \leq N\bar{\eta}^2(\mathfrak{z}) |u_s|_{W_p^m}^p, \quad (\mathcal{Q}_p^{(1)} + \mathcal{R}_p)(\alpha, \xi_s(\mathfrak{z}), u_s, k_\varepsilon) \leq N\bar{\xi}^2(\mathfrak{z}) |u_s|_{W_p^m}^p \tag{5.22}$$

with a constant $N = N(K_\xi, K_\eta, d, p, \lambda, m)$. Thus from (5.20) we obtain that for all α with $|\alpha| \leq m$ almost surely

$$|D^\alpha u_t^{(\varepsilon)}|_{L_p}^p \leq |u_0^{(\varepsilon)}|_{W_p^m}^p + N \int_0^t |u_s|_{W_p^m}^p ds + m_t^\varepsilon \quad \text{for all } t \in [0, T] \tag{5.23}$$

with a constant $N = N(m, p, d, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2(\mathfrak{Z}_1)}, |\bar{\eta}|_{L_2(\mathfrak{Z}_0)})$ and the local martingale $m^\varepsilon(\alpha, t) = \zeta_1(\alpha, t) + \zeta(\alpha, t)$. Summing over all $|\alpha| \leq m$ gives

$$|u_t^{(\varepsilon)}|_{W_p^m}^p \leq |u_0^{(\varepsilon)}|_{W_p^m}^p + N \int_0^t |u_s|_{W_p^m}^p ds + m_t^\varepsilon \quad \text{for all } t \in [0, T] \tag{5.24}$$

with (another) constant $N = N(m, p, d, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2(\mathfrak{Z}_1)}, |\bar{\eta}|_{L_2(\mathfrak{Z}_0)})$ and a local martingale, denoted again by m^ε . For integers $n \geq 1$ set $\tau_{n,k}^\varepsilon = \bar{\tau}_n \wedge \bar{\tau}_k^\varepsilon$, where $(\bar{\tau}_n^\varepsilon)_{n=1}^\infty$ is a localising sequence of stopping times for $m^{(\varepsilon)}$ and

$$\bar{\tau}_n = \inf \left\{ t \in [0, T] : \int_0^t |u_s|_{W_p^m}^p ds \geq n \right\}.$$

Then from (5.24), using also $|D^\alpha u^{(\varepsilon)}|_{L_p} = |(D^\alpha u)^{(\varepsilon)}|_{L_p} \leq |D^\alpha u|_{L_p}$ for multi-indices $\alpha \leq m$ and $\varepsilon > 0$, we get

$$\mathbb{E} |u_{t \wedge \tau_{n,k}^\varepsilon}^{(\varepsilon)}|_{W_p^m}^p \leq \mathbb{E} |u_0|_{W_p^m}^p + N \int_0^t \mathbb{E} |u_{s \wedge \tau_{n,k}^\varepsilon}^{(\varepsilon)}|_{W_p^m}^p ds < \infty \quad \text{for } t \in [0, T] \text{ and integers } n \geq 1.$$

Taking here first $k \rightarrow \infty$ and then $\varepsilon \rightarrow 0$, by Fatou’s lemma followed by Gronwall’s lemma we get

$$\mathbb{E}|u_{t \wedge \tau_n}|_{W_p^m}^p \leq N\mathbb{E}|u_0|_{W_p^m}^p \quad \text{for } t \in [0, T] \text{ and integers } n \geq 1$$

with a constant $N = N(m, p, d, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$. Letting here $n \rightarrow \infty$, by Fatou’s lemma we obtain

$$\sup_{t \in [0, T]} \mathbb{E}|u_t|_{W_p^m}^p \leq N\mathbb{E}|u_0|_{W_p^m}^p. \tag{5.25}$$

To prove (5.6) we define the stopping times

$$\rho_{n,k}^\varepsilon = \bar{\rho}_n \wedge \tilde{\tau}_k^\varepsilon \text{ for integers } k, n \geq 1 \text{ and } \varepsilon > 0, \text{ where}$$

$$\bar{\rho}_n = \inf \left\{ t \in [0, T] : \int_0^t |u_s|_{W_p^m}^{2p} ds \geq n \right\},$$

and $\tilde{\tau}_k^\varepsilon$ defined above. Using the Davis inequality and Lemma 4.3 by standard calculations for every $n \geq 1$ and $|\alpha| \leq m$ we get

$$\begin{aligned} \mathbb{E} \sup_{t \leq T} |\zeta_1(\alpha, t \wedge \rho_{n,k}^\varepsilon)| &\leq 3\mathbb{E} \left(\sum_k \int_0^{T \wedge \rho_{n,k}^\varepsilon} ((D^\alpha u_s^{(\varepsilon)})^{p-1}, D^\alpha (\mathcal{M}_s^{k*} u_s)^{(\varepsilon)})^2 ds \right)^{1/2} \\ &\leq N\mathbb{E} \left(\int_0^{T \wedge \rho_{n,k}^\varepsilon} |u_s|_{W_p^m}^{2p} ds \right)^{1/2} < \infty, \end{aligned} \tag{5.26}$$

and similarly, by using the Davis inequality and Lemma 4.7 we have

$$\mathbb{E} \sup_{s \leq T} |\zeta(\alpha, s \wedge \rho_{n,k}^\varepsilon)| \leq 3\mathbb{E}(\zeta(\alpha, \cdot))^{1/2}(T \wedge \rho_{n,k}^\varepsilon) \leq N\mathbb{E} \left(\int_0^{T \wedge \rho_{n,k}^\varepsilon} |u_s|_{W_p^m}^{2p} ds \right)^{1/2} < \infty, \tag{5.27}$$

with a constant $N = N(m, d, p, K, K_\xi, L, \lambda, |\bar{\xi}|_{L_2(\mathfrak{I}_1)})$, where

$$\langle \zeta(\alpha, \cdot) \rangle(t) = \int_0^t \int_{\mathfrak{I}_1} \|D^\alpha (T^{\xi*} u_s)^{(\varepsilon)}\|_{L_p}^p - |D^\alpha u_s^{(\varepsilon)}|_{L_p}^p \, |^2 \nu_1(\mathfrak{I}) ds, \quad t \in [0, T].$$

Thus, due to (5.25) together with (5.26) and (5.27), from (5.24) we get

$$\begin{aligned} \mathbb{E} \sup_{t \in [0, T]} |u_{t \wedge \rho_{n,k}^\varepsilon}^{(\varepsilon)}|_{W_p^m}^p &\leq N\mathbb{E}|u_0|_{W_p^m}^p + \sum_{|\alpha| \leq m} \mathbb{E} \sup_{t \leq T} |\zeta_1(\alpha, t \wedge \rho_{n,k}^\varepsilon)| \\ &\quad + \sum_{|\alpha| \leq m} \mathbb{E} \sup_{t \leq T} |\zeta(\alpha, t \wedge \rho_{n,k}^\varepsilon)| \\ &\leq N\mathbb{E}|u_0|_{W_p^m}^p + N\mathbb{E} \left(\int_0^T |u_{s \wedge \rho_n}|_{W_p^m}^{2p} ds \right)^{1/2} \end{aligned}$$

with a constant $N = N(m, p, d, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$. Letting here $k \rightarrow \infty$ and then $\varepsilon \rightarrow 0$, we by Fatou’s lemma obtain

$$\mathbb{E} \sup_{t \in [0, T]} |u_{t \wedge \rho_n}|_{W_p^m}^p \leq N \mathbb{E}|u_0|_{W_p^m}^p + N \mathbb{E} \left(\int_0^T |u_{s \wedge \rho_n}|_{W_p^m}^{2p} ds \right)^{1/2} < \infty.$$

Hence standard estimates yield

$$\begin{aligned} \mathbb{E} \sup_{t \in [0, T]} |u_{t \wedge \rho_n}|_{W_p^m}^p &\leq N \mathbb{E}|u_0|_{W_p^m}^p + N \mathbb{E} \left(\sup_{t \in [0, T]} |u_{t \wedge \rho_n}|_{W_p^m}^p \int_0^T |u_{s \wedge \rho_n}|_{W_p^m}^p ds \right)^{1/2} \\ &\leq N \mathbb{E}|u_0|_{W_p^m}^p + \frac{1}{2} \mathbb{E} \sup_{t \in [0, T]} |u_{t \wedge \rho_n}|_{W_p^m}^p + N' \mathbb{E} \int_0^T |u_{s \wedge \rho_n}|_{W_p^m}^p ds \\ &\leq N'' \mathbb{E}|u_0|_{W_p^m}^p + \frac{1}{2} \mathbb{E} \sup_{t \in [0, T]} |u_{t \wedge \rho_n}|_{W_p^m}^p < \infty \end{aligned}$$

with constants N' and N'' only depending on $m, p, d, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2(3_1)}$ and $|\bar{\eta}|_{L_2(3_0)}$. Thus also, we get for all n ,

$$\mathbb{E} \sup_{t \in [0, T]} |u_{t \wedge \rho_n}|_{W_p^m}^p \leq 2N'' \mathbb{E}|u_0|_{W_p^m}^p,$$

and letting here $n \rightarrow \infty$ by Fatou’s lemma we get the desired result. □

The following Proposition 5.4 is Lemma 6.4 in [5]. To formulate it let \mathbb{B}_0^m denote the space of those functions $\psi \in \bigcap_{p \geq 1} \mathbb{W}_p^m$ such that

$$\sum_{k=0}^m \sup_{\omega \in \Omega} \sup_{x \in \mathbb{R}^d} |D^k \psi(x)| < \infty \quad \text{and almost surely } \psi(x) = 0 \text{ for } |x| \geq R,$$

for some constant R depending on ψ . If $m = 0$ then we just write \mathbb{B}_0 in place of \mathbb{B}_0^0 . It is easy to see that \mathbb{B}_0^m is a dense subspace of \mathbb{W}_p^m for every $p \in [1, \infty)$. For $\varepsilon > 0$ let in the following proposition $v^{(\varepsilon)}$ denote the convolution

$$v^{(\varepsilon)}(x) = \int_{\mathbb{R}^d} \chi_\varepsilon(x - y)v(y) dy$$

of a Borel function v on \mathbb{R}^d , where χ is a smooth, symmetric function of unit integral on \mathbb{R}^d , such that $\chi(x) = 0$ for $|x| \geq 1$. Set $\chi_\varepsilon(\cdot) := \varepsilon^{-d} \chi(\cdot/\varepsilon)$ for $\varepsilon > 0$ and let

$$\begin{aligned} \mathcal{M}_t^{\varepsilon k} &= \rho_t^{(\varepsilon)ik} D_i + B_t^{(\varepsilon)k}, \quad k = 1, \dots, d', \\ \tilde{\mathcal{L}}_t^\varepsilon &= a_t^{\varepsilon, ij} D_{ij} + b_t^{(\varepsilon)i} D_i + \beta_t^k \mathcal{M}_t^{\varepsilon k}, \quad \beta_t = B(t, X_t, Y_t), \\ a_t^{\varepsilon, ij} &:= \frac{1}{2} \sum_k (\sigma_t^{(\varepsilon)ik} \sigma_t^{(\varepsilon)jk} + \rho_t^{(\varepsilon)ik} \rho_t^{(\varepsilon)jk}), \quad i, j = 1, 2, \dots, d, \end{aligned}$$

and let I^{ξ^ε} , J^{ξ^ε} and J^{η^ε} be defined as I^ξ , J^ξ and J^η , only with $\xi^{(\varepsilon)}$ and $\eta^{(\varepsilon)}$ instead of ξ and η , respectively.

Consider for $\varepsilon \in (0, 1)$ the equation

$$\begin{aligned}
 du_t^\varepsilon &= \tilde{\mathcal{L}}_t^{\varepsilon*} u_t^\varepsilon dt + \mathcal{M}_t^{\varepsilon k*} u_t^\varepsilon dV_t^k + \int_{\mathfrak{Z}_0} J_t^{\eta^\varepsilon*} u_t^\varepsilon \nu_0(d\mathfrak{z}) dt \\
 &+ \int_{\mathfrak{Z}_1} J_t^{\xi^\varepsilon*} u_t^\varepsilon \nu_1(d\mathfrak{z}) dt + \int_{\mathfrak{Z}_1} I_t^{\xi^\varepsilon*} u_t^\varepsilon \tilde{N}_1(d\mathfrak{z}, dt). \tag{5.28}
 \end{aligned}$$

Proposition 5.4 *Let Assumptions 2.1, 2.2 and 2.4 hold with $K_1 = 0$ and let $p \geq 2$ be even. Assume that the following “support condition” holds: There is some $R > 0$ such that*

$$(b_t(x), B_t(x), \sigma_t(x), \rho_t(x), \eta_t(x, \mathfrak{z}_0), \xi_t(x, \mathfrak{z}_1)) = 0 \tag{5.29}$$

for $\omega \in \Omega$, $t \geq 0$, $\mathfrak{z}_0 \in \mathfrak{Z}_0$, $\mathfrak{z}_1 \in \mathfrak{Z}_1$ and $x \in \mathbb{R}^d$ such that $|x| \geq R$. Let $\psi \in \mathbb{B}_0$ such that almost surely $\psi(x) = 0$ for $|x| \geq R$. Then there exist $\varepsilon_0 > 0$ and a $\bar{R} = \bar{R}(R, K, K_0, K_\xi, K_\eta)$ such that the following statements hold.

(i) For each $\varepsilon \in (0, \varepsilon_0)$, for every integer $r \geq 1$, there exists a W_p^r -solution u^ε to (5.28) with initial condition $u_0^\varepsilon = \psi^{(\varepsilon)}$ and such that

$$\mathbb{E} \sup_{t \in [0, T]} |u_t^\varepsilon|_{W_p^r}^p < \infty \text{ and } u_t^\varepsilon(x) = 0 \text{ almost surely for } |x| \geq \bar{R} \text{ and } t \in [0, T].$$

(ii) There exists a unique L_p -solution u to (3.3) (with non-smoothed coefficients) such that almost surely $u_t(x) = 0$ for dx -almost every $x \in \{x \in \mathbb{R}^d : |x| \geq \bar{R}\}$ for every $t \in [0, T]$ and

$$\mathbb{E} \sup_{t \in [0, T]} |u_t|_{L_p}^p \leq N \mathbb{E} |\psi|_{L_p}^p \tag{5.30}$$

with a constant $N = N(d, p, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$.

(iii) There exists a sequence $(\varepsilon_n)_{n=1}^\infty$, $\varepsilon_n \rightarrow 0$, such that for $n \rightarrow \infty$

$$u_T^{\varepsilon_n} \rightarrow u_T \text{ weakly in } \mathbb{L}_p \text{ and } u^{\varepsilon_n} \rightarrow u \text{ weakly in } \mathbb{L}_{p,q} \text{ for every integer } q \geq 2.$$

Proof See Lemma 6.4 in [5]. □

Lemma 5.5 *Let Assumptions 2.1, 2.2, 2.4 and 2.5 hold with $K_1 = 0$, and assume that the support condition (5.29) of Proposition 5.4 holds for some $R > 0$. Then for an integer $m \geq 0$ and even integer $p \geq 2$ there exists a unique W_p^m -solution $(u_t)_{t \in [0, T]}$ to equation (3.3) with initial condition $u_0 = \psi$, for any $\psi \in B_0^m$ such that $\psi = 0$ for $|x| \geq R$. Moreover, almost surely $u_t(x) = 0$ for dx -almost every $x \in \{x \in \mathbb{R}^d : |x| \geq \bar{R}\}$ for every $t \in [0, T]$ for a constant $\bar{R} = \bar{R}(R, K, K_0, K_\xi, K_\eta)$, and*

$$\mathbb{E} \sup_{t \in [0, T]} |u_t|_{W_p^m}^p \leq N \mathbb{E} |\psi|_{W_p^m}^p \tag{5.31}$$

with a constant $N = N(m, d, p, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$.

Proof By Proposition 5.4 (i) for $\varepsilon > 0$ sufficiently small there exists a W_p^m -valued weakly cadlag \mathcal{F}_t -adapted process $(u_t^\varepsilon)_{t \in [0, T]}$, such that for each $\varphi \in C_0^\infty$ almost surely

$$\begin{aligned} (u_t^\varepsilon, \varphi) &= (\psi^{(\varepsilon)}, \varphi) + \int_0^t (u_s^\varepsilon, \tilde{\mathcal{L}}_s^\varepsilon \varphi) ds \\ &\quad + \int_0^t (u_s^\varepsilon, \mathcal{M}_s^{ek} \varphi) dV_s^k + \int_0^t \int_{\mathfrak{Z}_0} (u_s^\varepsilon, J_s^{\eta^\varepsilon} \varphi) \nu_0(d\mathfrak{z}) ds \\ &\quad + \int_0^t \int_{\mathfrak{Z}_1} (u_s^\varepsilon, J_s^{\xi^\varepsilon} \varphi) \nu_1(d\mathfrak{z}) ds + \int_0^t \int_{\mathfrak{Z}_1} (u_s^\varepsilon, I_s^{\xi^\varepsilon} \varphi) \tilde{N}_1(d\mathfrak{z}, ds) \end{aligned} \tag{5.32}$$

holds for all $t \in [0, T]$. By Proposition 5.4 (ii), since almost surely $u_t^\varepsilon = 0$ for $|x| \geq \bar{R}$ for all $t \in [0, T]$ for a constant $\bar{R} = \bar{R}(R, K, K_0, K_\xi, K_\eta)$, we also have

$$\mathbb{E} \sup_{t \in [0, T]} |u_t^\varepsilon|_{L^1} \leq \bar{R}^{d/q} \mathbb{E} \sup_{t \in [0, T]} |u_t^\varepsilon|_{L^p}^p$$

for $q = p/(p - 1)$. Next, note that the smoothed coefficients $b^{(\varepsilon)}, B^{(\varepsilon)}, \sigma^{(\varepsilon)}, \rho^{(\varepsilon)}, \xi^{(\varepsilon)}$ and $\eta^{(\varepsilon)}$ satisfy Assumptions 2.1, 2.2, 2.5 and Assumption 2.4 (ii) & (iii) with the same constants K_0, L, K_ξ and K_η , independent of ε . By Remark 2.3 and Lemma 6.2 in [5] we have a constant $\varepsilon_0 > 0$ such that for $\varepsilon \in (0, \varepsilon_0)$, for the mappings

$$\tau_{t, \theta, \mathfrak{z}_0}^{\eta^{(\varepsilon)}}(x) = x + \theta \eta_t^{(\varepsilon)}(x, \mathfrak{z}_0), \quad \text{and} \quad \tau_{t, \theta, \mathfrak{z}_1}^{\xi^{(\varepsilon)}}(x) = x + \theta \xi_t^{(\varepsilon)}(x, \mathfrak{z}_1), \quad x \in \mathbb{R}^d,$$

we have that

$$|\det D_x \tau_{t, \theta, \mathfrak{z}_0}^{\eta^{(\varepsilon)}}(x)| \geq \lambda/2 \quad \text{and} \quad |\det D_x \tau_{t, \theta, \mathfrak{z}_1}^{\xi^{(\varepsilon)}}(x)| \geq \lambda/2$$

for all $x \in \mathbb{R}^d, t \in [0, T], \theta \in [0, 1]$ and $\mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$. Hence, for $\varepsilon \in (0, \varepsilon_0)$ the functions $\tau_{t, \theta, \mathfrak{z}_0}^{\eta^{(\varepsilon)}}$ and $\tau_{t, \theta, \mathfrak{z}_1}^{\xi^{(\varepsilon)}}$ are C^1 -diffeomorphisms on \mathbb{R}^d for each $t \in [0, T], \theta \in [0, 1], \mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$, and the functions $f_0(t, x, y, \mathfrak{z}_0) := \eta^{(\varepsilon)}(t, x, y, \mathfrak{z}_0), f_1(t, x, y, \mathfrak{z}_1) := \xi^{(\varepsilon)}(t, x, y, \mathfrak{z}_1), (x, y) \in \mathbb{R}^{d+d'}, t \in [0, T], \mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$ satisfy Assumption 2.4(i) with a constant $\bar{\lambda} = \bar{\lambda}(\lambda, d, K_\xi, K_\eta, K_0, L) > 0$ in place of λ . Hence by Lemma 5.3 for each $\varepsilon \in (0, \varepsilon_0)$ we have

$$\mathbb{E} |u_T^\varepsilon|_{W_p^m}^p + \mathbb{E} \left(\int_0^T |u_t^\varepsilon|_{W_p^m}^p dt \right)^{p/r} \leq \mathbb{E} |u_T^\varepsilon|_{W_p^m}^p + T^{p/r} \mathbb{E} \sup_{t \in [0, T]} |u_t^\varepsilon|_{W_p^m}^p \leq N \mathbb{E} |\psi|_{W_p^m}^p \tag{5.33}$$

for every integer $r \geq 1$, with a constant $N = N(m, d, p, K, K_\eta, K_\xi, L, T, \lambda, |\bar{\xi}|_{L_2(\mathfrak{Z}_1)}, |\bar{\eta}|_{L_2(\mathfrak{Z}_1)})$, independent of ε and r . Letting $(\varepsilon_n)_{n=1}^\infty$ be the sequence from Proposition 5.4 (iii), we know that for $n \rightarrow \infty$

$$u_T^{\varepsilon_n} \rightarrow u_T \quad \text{weakly in } \mathbb{L}_p(\mathcal{F}_T) \quad \text{and} \quad u^{\varepsilon_n} \rightarrow u \quad \text{weakly in } \mathbb{L}_{p,r} \text{ for integers } r > 1,$$

where u is the unique L_p -solution to (3.3) and, if necessary by passing to a subsequence,

$$u_T^{\varepsilon_n} \rightarrow u_T \text{ weakly in } \mathbb{W}_p^m(\mathcal{F}_T) \text{ and } u^{\varepsilon_n} \rightarrow u \text{ weakly in } \mathbb{W}_{p,r}^m \text{ for integers } r > 1.$$

Letting $r \rightarrow \infty$ in (5.33) yields

$$\mathbb{E}|u_T|_{\mathbb{W}_p^m}^p + \mathbb{E} \operatorname{ess\,sup}_{t \in [0, T]} |u_t|_{\mathbb{W}_p^m}^p \leq N \mathbb{E}|\psi|_{\mathbb{W}_p^m}^p.$$

By Lemma 3.3 u is weakly cadlag as W_p^m -valued process. Thus we can replace the essential supremum above by the supremum to obtain (5.31). By Proposition 5.4 (ii) we also have that almost surely $u_t(x) = 0$ for dx -almost every $x \in \{x \in \mathbb{R}^d : |x| \geq \bar{R}\}$ for every $t \in [0, T]$ for a constant $\bar{R} = \bar{R}(R, K, K_0, K_\xi, K_\eta)$. This finishes the proof. \square

Corollary 5.6 *Let Assumptions 2.1, 2.2, 2.4 and 2.5 hold with an integer $m \geq 0$ and $K_1 = 0$. Assume, moreover that the support condition (5.29) holds for some $R > 0$. Then there is a linear operator \mathbb{S} defined on \mathbb{W}_p^m for every $p \geq 2$ such that $\mathbb{S}\psi$ admits a $P \otimes dt$ -modification $u = (u_t)_{t \in [0, T]}$ which is a W_p^m -solution to equation (3.3) with initial condition $u_0 = \psi$ for every $\psi \in \mathbb{W}_p^m$, and*

$$\mathbb{E} \sup_{t \in [0, T]} |u_t|_{\mathbb{W}_p^m}^p \leq N \mathbb{E}|\psi|_{\mathbb{W}_p^m}^p \tag{5.34}$$

with a constant $N = N(m, d, p, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$. Moreover, if $\psi \in \mathbb{W}_p^m$ such that almost surely $\psi(x) = 0$ for $|x| \geq R$, then almost surely $u_t(x) = 0$ for $|x| \geq \bar{R}$ for $t \in [0, T]$ for a constant $\bar{R} = \bar{R}(R, K, K_0, K_\xi, K_\eta)$.

Proof By Corollary 6.5 in [5] we know that there exist linear operators \mathbb{S} and \mathbb{S}_T defined on \mathbb{L}_p for every $p \geq 2$, such that if $\psi \in \mathbb{L}_p$ for a $p \geq 2$ then $\mathbb{S}\psi \in \mathbb{L}_{p,r}$ for every integer $r \geq 2$ and $\mathbb{S}\psi$ admits a $P \otimes dt$ -modification $u = (u_t)_{t \in [0, T]}$ that is an L_p -solution to (3.3), and $\mathbb{S}_T\psi = u_T$. By an abuse of notation we refer to this stochastic modification u whenever we write $\mathbb{S}\psi$ in the following. It remains to show that if $\psi \in \mathbb{W}_p^m$, then u is in particular a W_p^m -solution to (3.3), i.e. it is almost surely a weakly cadlag W_p^m -valued process.

If p is an even integer, then this follows from Lemma 5.5. Assume p is not an even integer. Then let p_0 be the greatest even integer such that $p_0 \leq p$ and let p_1 be the smallest even integer such that $p \leq p_1$. By Lemma 5.5, in particular (5.31), we get that

$$|\mathbb{S}_T\psi|_{\mathbb{W}_{p_i}^m} + |\mathbb{S}\psi|_{\mathbb{W}_{p_i,r}^m} \leq N_i |\psi|_{\mathbb{W}_{p_i}^m} \text{ for } i = 0, 1 \tag{5.35}$$

for every $r \in [1, \infty)$ and constants $N_i = N_i(m, d, p_i, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$, $i = 0, 1$, independent of r . Hence, by a well-known generalization of the Riesz-Thorin interpolation theorem we also get for all $r \geq 1$,

$$|\mathbb{S}_T\psi|_{\mathbb{W}_p^m} + |\mathbb{S}\psi|_{\mathbb{W}_{p,r}^m} \leq N |\psi|_{\mathbb{W}_p^m} \text{ for } i = 0, 1, \tag{5.36}$$

for (another) constant $N = N(m, d, p, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$. Consider a sequence $(\psi^n)_{n=1}^\infty \subset \mathbb{B}_0^m$ such that $\psi^n \rightarrow \psi$ in \mathbb{W}_p^m . For each n , $u^n = \mathbb{S}\psi^n$ is the unique $W_{p_i}^m$ -solution to (3.3), $i = 0, 1$, with initial condition ψ^n . By virtue of (5.36), using that $|\psi^n - \psi|_{\mathbb{W}_p^m} \rightarrow 0$, as $n \rightarrow \infty$ we know that also

$$\begin{aligned} u^n &\rightarrow u \quad \text{weakly in } \mathbb{W}_{p,r}^m \text{ for every integer } r > 1 \text{ and} \\ u_T^n &\rightarrow u_T \quad \text{weakly in } \mathbb{W}_p^m(\mathcal{F}_T), \end{aligned}$$

where $u = \mathbb{S}\psi$ is the unique L_p -solution introduced in the beginning of the proof, satisfying (5.36). To see that u is weakly cadlag as W_p^m -valued process, note that by letting $r \rightarrow \infty$ in (5.36) for $\mathbb{S}\psi = u$ and $\mathbb{S}_T\psi = u_T$ we get

$$\mathbb{E}|u_T|_{W_p^m}^p + \mathbb{E} \operatorname{ess\,sup}_{t \in [0, T]} |u_t|_{W_p^m}^p \leq N \mathbb{E}|\psi|_{W_p^m}^p,$$

with (another) constant $N = N(m, d, p, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$. By Lemma 3.3 we then know that u is (almost surely) a W_p^m -valued weakly cadlag process. Thus we can replace the essential supremum above with the supremum, to obtain (5.34). To prove the claim about the support of u , note that if $\psi(x) = 0$ for $|x| \geq R$, for a constant R , and $\psi^n \rightarrow \psi$ in \mathbb{W}_p^m , then for sufficiently large n we have $\psi^n(x) = 0$ for $|x| \geq 2R$. By Proposition 5.4 (ii) thus also $u_t^n(x) = 0$ for dx -almost every $x \in \{x \in \mathbb{R}^d : |x| \geq \bar{R}\}$ for every $t \in [0, T]$ and n sufficiently large, for a constant $\bar{R} = \bar{R}(R, K, K_0, K_\xi, K_\eta)$. This is clearly preserved in the limit as $n \rightarrow \infty$. This finishes the proof. \square

6 Proof of Theorem 2.1

Let χ be a smooth function on \mathbb{R} such that $\chi(r) = 1$ for $r \in [-1, 1]$, $\chi(r) = 0$ for $|r| \geq 2$, $\chi(r) \in [0, 1]$ and $\sum_{k=1}^{m+2} |d^k/(dr^k)\chi(r)| \leq C$ for all $r \in \mathbb{R}$ and a constant $C \geq 1$. For integers $n \geq 1$ we define the function χ_n by $\chi_n(x) = \chi(|x|/n)$, $x \in \mathbb{R}^d$.

Lemma 6.1 (i) Let $b = (b^i)$ be an \mathbb{R}^d -valued function on \mathbb{R}^m such that for a constant L

$$|b(v) - b(z)| \leq L|v - z| \quad \text{for all } v, z \in \mathbb{R}^m. \tag{6.1}$$

Then for $b_n(z) = \chi(|z|/n)b(z)$, $z \in \mathbb{R}^m$, for integers $n \geq 1$ we have

$$|b_n(z)| \leq 2nL + |b(0)|, \quad |b_n(v) - b_n(z)| \leq C(3L + |b(0)|/n)|v - z| \quad \text{for all } v, z \in \mathbb{R}^m. \tag{6.2}$$

(ii) Let additionally to (i) the function b satisfy

$$\sum_{k=1}^m |D^k b| \leq M, \tag{6.3}$$

for a constant $M > 0$. Then b_n satisfies (6.3) in place of b with $M'(n) = M'(M, C, m, |b(0)|/n)$ in place of M (and where $M'(n)$ is decreasing as function of n).

Proof The proof is left as an easy exercise for the reader. □

To preserve the diffeomorphic property of the mappings

$$\tau_{t,\mathfrak{z}_0,\theta}^\eta(x) = x + \theta\eta_t(x, \mathfrak{z}_0) \quad \text{and} \quad \tau_{t,\mathfrak{z}_1,\theta}^\xi(x) = x + \theta\xi_t(x, \mathfrak{z}_1) \tag{6.4}$$

(for all $\omega \in \Omega, t \in [0, T], \theta \in [0, 1]$ and $\mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$) as a function of $x \in \mathbb{R}^d$, when the functions ξ and η are truncated, we introduce, for each fixed $R > 0$ and $\epsilon > 0$, the function κ_ϵ^R defined on \mathbb{R}^d by

$$\kappa_\epsilon^R(x) = \int_{\mathbb{R}^d} \phi_\epsilon^R(x-y)k(y)dy, \quad \phi_\epsilon^R(x) \begin{cases} 1, & |x| \leq R+1, \\ 1 + \epsilon \log\left(\frac{R+1}{|x|}\right), & R+1 < |x| < (R+1)e^{1/\epsilon}, \\ 0, & |x| \geq (R+1)e^{1/\epsilon}, \end{cases} \tag{6.5}$$

where k is a nonnegative C^∞ mapping on \mathbb{R}^d with support in $\{x \in \mathbb{R}^d : |x| \leq 1\}$.

Lemma 6.2 *Let $\xi : \mathbb{R}^d \mapsto \mathbb{R}^d$ be such that for a constant $L \geq 1$ and for every $\theta \in [0, 1]$ the function $\tau_\theta(x) = x + \theta\xi(x)$ is L -biLipschitz, i.e.*

$$L^{-1}|x - y| \leq |\tau_\theta(x) - \tau_\theta(y)| \leq L|x - y| \tag{6.6}$$

for all $x, y \in \mathbb{R}^d$. Then for any $M > L$ and any $R > 0$ there is an $\epsilon = \epsilon(L, M, R, |\xi(0)|) > 0$ such that with $\kappa^R := \kappa_\epsilon^R$ the function $\xi^R := \kappa^R\xi$ vanishes for $|x| \geq \bar{R}$ for a constant $\bar{R} = \bar{R}(L, M, R, |\xi(0)|) > R, |\xi^R|$ is bounded by a constant $N = N(L, M, R, |\xi(0)|)$, and for every $\theta \in [0, 1]$ the mapping

$$\tau_\theta^R(x) = x + \theta\xi^R(x), \quad x \in \mathbb{R}^d$$

is M -biLipschitz.

Proof This is Lemma 7.3 in [5]. □

We summarize the results of Lemmas 7.1, 7.2 and Remark 7.1 in [5] in the following lemma. For that purpose, define the functions $b^n = (b^{ni}(t, z)), B^n = (B^{nj}(t, z)), \sigma^n = (\sigma^{nij}(t, z)), \eta^n = (\eta^{ni}(t, z, \mathfrak{z}_0))$ and $\xi^n = (\xi^{ni}(t, z, \mathfrak{z}_1))$ by

$$(b^n, B^n, \sigma^n, \rho^n) = (b, B, \sigma, \rho)\chi_n, \quad (\eta^n, \xi^n) = (\eta, \xi)\bar{\chi}_n \tag{6.7}$$

for every integer $n \geq 1$, where χ_n and $\bar{\chi}_n$ are functions on $\mathbb{R}^{d+d'}$ defined by $\chi_n(z) = \chi(|z|/n)$ and $\bar{\chi}_n(x, y) = \kappa^n(x)\chi(|y|/n)$ for $z = (x, y) \in \mathbb{R}^{d+d'}$, with χ used in Lemma 6.1 and with $\kappa^n = \kappa_\epsilon^n$ from Lemma 6.2, such that, by the L -biLipschitzness of the mappings in (6.4), the mappings

$$\tau_{t,\mathfrak{z}_0,\theta}^{\eta^n}(x) = x + \theta\eta_t^n(x, \mathfrak{z}_0) \quad \text{and} \quad \tau_{t,\mathfrak{z}_1,\theta}^{\xi^n}(x) = x + \theta\xi_t^n(x, \mathfrak{z}_1),$$

on \mathbb{R}^d are biLipschitz (for each $\omega \in \Omega, t \in [0, T], \theta \in [0, 1]$ and $\mathfrak{z}_i \in \mathfrak{Z}_i, i = 0, 1$).

Lemma 6.3 *Let Assumptions 2.1, 2.2 and 2.5 hold. If $K_1 \neq 0$ in Assumption 2.1 (ii), then let additionally Assumption 2.3 for some $r > 2$ hold. Assume the initial conditional density $\pi_0 = P(X_0 \in dx | \mathcal{F}_0^Y) / dx$ exists (a.s.) and satisfies $\mathbb{E}|\pi_0|_{\mathbb{W}_p^m}^p < \infty$ for some $p \geq 2$ and integer $m \geq 0$. Then there exist sequences*

$$(X_0^n)_{n=1}^\infty, ((X_t^n, Y_t^n)_{t \in [0, T]})_{n=1}^\infty, \text{ as well as } (\pi_0^n)_{n=1}^\infty \text{ and } ((\pi_t^n)_{t \in [0, T]})_{n=1}^\infty$$

such that the following are satisfied:

- (i) For each $n \geq 1$ the coefficients $b^n, B^n, \sigma^n, \rho^n, \xi^n$ and η^n , defined in (6.7), satisfy Assumptions 2.1 and 2.2 with $K_1 = 0$ and constants $K'_0 = K'_0(n, K, K_0, K_1, K_\xi, K_\eta)$ and $L' = L'(K, K_0, K_1, L, K_\xi, K_\eta)$ in place of K_0 and L , Assumption 2.5 with a constant $K' = K'(K_0, K_1)$ in place of L , as well as Assumption 2.4 with $\lambda' = \lambda'(K_0, K_1, K_\xi, K_\eta, \lambda) > 0$ in place of λ . Moreover, for each $n \geq 1$ they satisfy the support condition (5.29) of Lemma 5.5 for some $R = R(n)$.
- (ii) For each $n \geq 1$ the random variable X_0^n is \mathcal{F}_0 -measurable and satisfies

$$\lim_{n \rightarrow \infty} X_0^n = X_0, \omega \in \Omega, \text{ and } \mathbb{E}|X_0^n|^r \leq N(1 + \mathbb{E}|X_0|^r)$$

for $r \geq 1$ with a constant N independent of n .

- (iii) $Z_t^n = (X_t^n, Y_t^n)$ is the solution to (1.1) with the coefficients $b^n, B^n, \sigma^n, \rho^n, \xi^n$ and η^n in place of b, B, σ, ρ, ξ and η , respectively, and with initial condition $Z_0^n = (X_0^n, Y_0)$.
- (iv) For each $n \geq 1$ we have $\pi_0^n = P(X_0^n \in dx | \mathcal{F}_0^{Y^n}) / dx, \pi_0^n(x) = 0$ for $|x| \geq n + 1$ and

$$\lim_{n \rightarrow \infty} |\pi_0^n - \pi_0|_{\mathbb{W}_p^m} = 0,$$

where $\pi_0 = P(X_0 \in dx | \mathcal{F}_0^Y) / dx$.

- (v) For each $n \geq 1$ there exists an L_r -solution u^n to (3.3), $r = 2, p$, such that u^n is the unnormalised conditional density of X^n given Y^n , almost surely

$$u_t^n(x) = 0 \text{ for } dx\text{-a.e. } x \in \{x \in \mathbb{R}^d : |x| \geq \bar{R}\} \text{ for all } t \in [0, T]$$

with a constant $\bar{R} = \bar{R}(n, K, K_0, K_\xi, K_\eta)$ and

$$\mathbb{E} \sup_{t \in [0, T]} |u_t^n|_{L^p}^p \leq N \mathbb{E} |\pi_0^n|_{L^p}^p \text{ for } n \geq K_0 \tag{6.8}$$

with a constant $N = N(d, d', K, L, K_\xi, K_\eta, T, p, \lambda, \|\xi\|_{L_2}, \|\eta\|_{L_2})$. Moreover,

$$u^n \rightarrow u \text{ weakly in } \mathbb{L}_{r, q} \text{ for } r = p, 2 \text{ and all integers } q > 1,$$

where u is the unnormalised conditional density of X given Y , satisfying (6.8) with the same constant N and (u, π_0) in place of (u^n, π_0^n) .

(vi) Consequently, for each $n \geq 1$ and $t \in [0, T]$ we have

$$\pi_t^n(x) = P(X_t^n \in dx | \mathcal{F}_t^{Y^n})/dx = u_t^n(x) \circ \gamma_t^n, \quad \text{almost surely,}$$

as well as

$$\pi_t(x) = P(X_t \in dx | \mathcal{F}_t^Y)/dx = u_t(x) \circ \gamma_t, \quad \text{almost surely,}$$

where $\circ \gamma_t^n$ and $\circ \gamma_t$ are cadlag positive normalising processes, adapted to $\mathcal{F}_t^{Y^n}$ and \mathcal{F}_t^Y , respectively.

Proof This is Corollary 7.4 in [5]. Note that the condition $n \geq K_0$ in (6.8) comes from Proposition 6.7, as then $|b(0)|/n \leq K_0/n \leq 1$. In other words, the constant N in that equation is the same for all $n \geq K_0$. □

Now we are in the position to prove our main result.

Proof of Theorem 2.1 Step I. Assume first that the support condition (5.29) holds with some $R > 0$ and that the initial conditional density π_0 is such that $\pi_0(x) = 0$ for $|x| \geq R$. By Corollary 5.6 we know that there exists a W_p^m -solution $(u_t)_{t \in [0, T]}$ to (3.3) with initial condition π_0 , satisfying

$$\mathbb{E} \sup_{t \in [0, T]} |u_t|_{W_p^m}^p \leq N \mathbb{E} |\pi_0|_{W_p^m}^p \tag{6.9}$$

with a constant $N = N(m, d, p, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$. Moreover, we have $u_t = 0$ for $|x| \geq \bar{R}$, for a constant $\bar{R} = \bar{R}(R, K, K_0, K_1, K_\xi, K_\eta)$, and hence clearly

$$\sup_{t \in [0, T]} |u_t|_{L_1} \leq \bar{R}^{d/q} \sup_{t \in [0, T]} |u_t|_{L_p} \quad \text{and} \quad \sup_{t \in [0, T]} \int_{\mathbb{R}^d} |y|^2 |u_t(y)| dy < \infty \text{ (a.s.),}$$

with $q = p/(p - 1)$. Since also $\pi_0 = P(X_0 \in dx | \mathcal{F}_0^Y)/dx \in \mathbb{L}_1$, then in particular $\pi_0 \in \mathbb{L}_2$ and hence

$$\mathbb{E} \sup_{t \in [0, T]} |u_t|_{L_2}^2 \leq N \mathbb{E} |\pi_0|_{L_2}^2, \tag{6.10}$$

with a constant $N = N(d, p, T, K, K_\xi, K_\eta, L, \lambda, |\bar{\xi}|_{L_2}, |\bar{\eta}|_{L_2})$. By Lemma 5.5 u is the unique L_2 -solution and therefore by Theorem 3.2, u is in particular the unnormalised conditional density, i.e., $u_t = d\mu_t/dx$ for all $t \in [0, T]$, almost surely, with μ the unnormalised conditional distribution from Theorem 3.1. Thus also for each $t \in [0, T]$,

$$\pi_t = P(X_t \in dx | \mathcal{F}_t^Y)/dx = u_t \circ \gamma_t, \quad \text{almost surely,}$$

where ${}^o\gamma_t$ is the \mathcal{F}_t^Y -optional projection of the normalizing process γ under P introduced in (3.5).

Step II. Finally, we dispense with the assumption that the coefficients and the initial condition are compactly supported. Define the functions $b_n, B_n, \sigma_n, \rho_n, \xi_n$ and η_n as in (6.7). Note that by Lemma 6.3 the truncated coefficients satisfy Assumptions 2.1 and 2.2 with $K_1 = 0$ and constants $K'_0 = K'_0(n, K, K_0, K_1, K_\xi, K_\eta)$ and $L' = L'(K, K_0, K_1, L, K_\xi, K_\eta)$ in place of K_0 and L , the coefficients $b_n, B_n, \sigma_n, \rho_n$ satisfy Assumption 2.5(i) with a constant $K' = K'(m, K_0, K_1)$ in place of L , and moreover that the coefficients η_n and ξ_n satisfy Assumption 2.5(ii) with $K'\bar{\eta}$ and $K'\bar{\xi}$ instead of $\bar{\eta}$ and $\bar{\xi}$ respectively. Furthermore, by Lemma 6.2, for each $n \geq 1$ the coefficients η_n and ξ_n satisfy Assumption 2.4 with a constant $\lambda' = \lambda'(\lambda, K_0, K_1, K_\eta, K_\xi) > 0$ in place of λ . Note that K', L' and λ' do not depend on n . Moreover, for each $n \geq 1$ they satisfy the support condition (5.29) of Lemma 5.5 for some $R = R(n) > 0$. By assumption, $\pi_0 = P(X_0 \in dx | \mathcal{F}_0^Y) / dx$ exists almost surely and $\mathbb{E}|\pi_0|_{W_p^m}^p < \infty$. Then let $(X_0^n)_{n=1}^\infty$ and $(\pi_0^n)_{n=1}^\infty \subset \mathbb{W}_p^m$ be the sequences from Lemma 6.3 such that

$$\lim_{n \rightarrow \infty} |\pi_0^n - \pi_0|_{\mathbb{W}_p^m} = 0, \tag{6.11}$$

$\pi_0^n(x) = 0$ for $|x| \geq R(n)$ and $\pi_0^n = P(X_0^n \in dx | \mathcal{F}_0^Y) / dx$ (a.s.), where (X_0^n, Y_0) is the initial condition to the system (1.1), and $(R(n))_{n=1}^\infty$ is the sequence of positive numbers from the support condition for the coefficients (σ^n, \dots, ξ^n) . By Step I we know that there exists a W_p^m -solution $(u_t^n)_{t \in [0, T]}$ to (3.3) with initial condition π_0^n , which is the unnormalized conditional density of $X^n = (X_t^n)_{t \in [0, T]}$ given $Y^n = (Y_t^n)_{t \in [0, T]}$, where $Z^n = (X^n, Y^n)$ is the solution to (1.1) with initial condition (X_0^n, Y_0) . By Lemma 6.3 (v) we know moreover that

$$u^n \rightarrow u \text{ weakly in } \mathbb{L}_{r, q} \text{ for } r = p, 2 \text{ and all integers } q > 1,$$

where u is the unnormalised conditional density of X given Y from Theorem 3.2, satisfying

$$\mathbb{E} \sup_{t \in [0, T]} |u_t|_{L_2}^2 \leq N \mathbb{E}|\pi_0|_{L_2}^2,$$

with a constant $N = N(d, p, T, K, K_\xi, K_\eta, L, \lambda, \|\bar{\xi}\|_{L_2}, \|\bar{\eta}\|_{L_2})$. Moreover, u is an L_p -solution to (3.3) and by Theorem 3.2 (ii), it is the unique L_2 -solution to (3.3). It remains to show that u is also a W_p^m -solution to (3.3), as well as that it is strongly cadlag as W_p^s -valued process, for $s \in [0, m)$. To prove the former, by (6.9) together with (6.11) we get

$$\begin{aligned} &\mathbb{E}|u_T^n|_{W_p^m}^p + \mathbb{E} \left(\int_0^T |u_t^n|_{W_p^m}^r dt \right)^{p/r} \\ &\leq \mathbb{E}|u_T^n|_{W_p^m}^p + T^{p/r} \mathbb{E} \sup_{t \in [0, T]} |u_t^n|_{W_p^m}^p \end{aligned}$$

$$\leq N\mathbb{E}|\pi_0^n|_{W_p^m}^p \leq N \sup_{n \geq 1} \mathbb{E}|\pi_0^n|_{W_p^m}^p < \infty \tag{6.12}$$

for every $n \geq 1$ and $r > 1$ with a constant N independent of r and n . Hence for $n \rightarrow \infty$ we have

$$u_T^n \rightarrow u_T \text{ weakly in } \mathbb{W}_p^m \text{ and } u^n \rightarrow u \text{ weakly in } \mathbb{W}_{p,r}^m \text{ for any } r > 1,$$

and u satisfies

$$\mathbb{E}|u_T|_{W_p^m}^p + \mathbb{E} \left(\int_0^T |u_t|_{W_p^m}^r dt \right)^{p/r} \leq N\mathbb{E}|\pi_0|_{W_p^m}^p \text{ for all } r \geq 1.$$

Letting $r \rightarrow \infty$ above yields

$$\mathbb{E}|u_T|_{W_p^m}^p + \mathbb{E} \operatorname{ess\,sup}_{t \in [0, T]} |u_t|_{W_p^m}^p \leq N\mathbb{E}|\pi_0|_{W_p^m}^p.$$

By Lemma 3.3 we then know that u is weakly cadlag as an W_p^m -valued process, i.e. it is a W_p^m -solution to (3.3). Clearly, by Lemma 6.3, also for each $t \in [0, T]$

$$\pi_t(x) = P(X_t \in dx | \mathcal{F}_t^Y) / dx = u_t(x) \circ \gamma_t, \text{ almost surely,} \tag{6.13}$$

with $\circ \gamma$ from Theorem 3.2. We now show that if $m \geq 1$ and $K_1 = 0$, then u is strongly cadlag as W_p^s -valued process for $s \in [0, m)$. To this and first we state u is strongly cadlag as an L_p -valued process.

Proposition 6.4 *Let Assumptions 2.1 through 2.5 hold with $K_1 = 0$ and with $m = 1$. Let $p \geq 2$ and let $u = (u_t)_{t \in [0, T]}$ be a W_p^1 -solution to (3.3). Then u is strongly cadlag as an L_p -valued process.*

Proof We apply Theorem 2.2 in [8]. In order to do so, we rewrite equation (3.2) into the form used therein. Clearly, for $v \in W_p^1$ and $\varphi \in C_0^\infty$ we have

$$(v, \mathcal{M}^k \varphi) = (\mathcal{M}^{k*} v, \varphi) \text{ with } \mathcal{M}^{k*} v = -D_i(\rho_s^{ik} v) + B_s^k v, \quad k = 1, 2, \dots, d_1$$

and

$$(v, \tilde{\mathcal{L}}_s \varphi) = -(D_j(a_s^{ij} v), D_i \varphi) - (D_i(b_s^i v), \varphi) + \beta_s^k (\mathcal{M}_s^{*k} v, \varphi).$$

Using Corollary 3.5 with $\eta_t(\cdot, \mathfrak{z})$ and $\xi_t(\cdot, \mathfrak{z})$ in place of ζ we can see that

$$(v, J_t^{\eta} \varphi) = (K_i^{\eta_t} v, D_i \varphi), \quad (v, J_t^{\xi} \varphi) = (K_i^{\xi_t} v, D_i \varphi) \text{ and } (v, I_t^{\xi} \varphi) = (I_t^{\xi*} v, \varphi),$$

for $v \in W_p^1$ and $\varphi \in C_0^\infty$, where $K_i^{\eta_t} v$ and $K_i^{\xi_t} v$ are defined as $K_i^\zeta v$ in Corollary 3.5 with ζ replaced with $\eta_t(\mathfrak{z}_0)$ and $\xi_t(\mathfrak{z}_1)$, respectively, (for $t \in [0, T]$, $\omega \in \Omega$, $\mathfrak{z}_0 \in \mathfrak{Z}_0$,

$\mathfrak{z}_1 \in \mathfrak{Z}_1, i = 1, \dots, d$, and $I_t^{\xi^*}$ is defined as I^{ζ^*} , with ζ replaced by $\xi_t(\mathfrak{z}_1)$. Thus for every $\varphi \in C_0^\infty$ almost surely

$$\begin{aligned} (u_t, \varphi) &= (\psi, \varphi) - \int_0^t (D_j(a_s^{ij} u_s), D_i \varphi) ds - \int_0^t (D_i(b_s^i u_s) + \beta_s^k \mathcal{M}_s^{*k} u_s, \varphi) ds \\ &\quad + \int_0^t (\mathcal{M}_s^{*k} u_s, \varphi) dV_s^k + \int_0^t \int_{\mathfrak{Z}_0} (K_i^{\eta_s} u_s, D_i \varphi) \nu_0(d\mathfrak{z}) ds \\ &\quad + \int_0^t \int_{\mathfrak{Z}_1} (K_i^{\xi_s} u_s, D_i \varphi) \nu_1(d\mathfrak{z}) ds + \int_0^t \int_{\mathfrak{Z}_1} (I_s^{\xi^*} u_s, \varphi) \tilde{N}_1(d\mathfrak{z}, ds) \end{aligned} \tag{6.14}$$

for all $t \in [0, T]$. It is easy to see that almost surely

$$\begin{aligned} \int_0^T |D_i(a_s^{ij} u_s)|_{L^p}^p ds < \infty, \quad \int_0^T |D_i(b_s^i u_s) + \beta_s^k \mathcal{M}_s^{*k} u_s|_{L^p}^p ds < \infty, \\ \int_0^T \int_{\mathbb{R}^d} \left(\sum_k |(\mathcal{M}_s^{k*} u_s)(x)|^2 \right)^{p/2} dx ds < \infty. \end{aligned} \tag{6.15}$$

By estimates (3.13) and (3.14), for all $x \in \mathbb{R}^d$ we have

$$\begin{aligned} |I^{\xi^*} u(x)| &\leq N \bar{\xi} \int_0^1 |u(\tau_{\theta \bar{\xi}}^{-1}(x))| + |(Du)(\tau_{\theta \bar{\xi}}^{-1}(x))| d\theta, \\ |K_i^{\xi} u(x)| &\leq N \bar{\xi}^2 \int_0^1 |u(\tau_{\theta \bar{\xi}}^{-1}(x))| + |(Du)(\tau_{\theta \bar{\xi}}^{-1}(x))| d\theta, \\ |K_i^\eta u(x)| &\leq N \bar{\eta}^2 \int_0^1 |u(\tau_{\theta \bar{\eta}}^{-1}(x))| + |(Du)(\tau_{\theta \bar{\eta}}^{-1}(x))| d\theta \end{aligned}$$

for every $\omega \in \Omega, s \in [0, T], \mathfrak{z}_i \in \mathfrak{Z}_i (i=0,1)$, suppressed in these estimates, with a constant $N = N(d, \lambda, L, K_\eta, K_\xi)$ and with the C^2 -diffeomorphisms

$$\tau_{\theta \eta}(x) = x + \theta \eta(x) \quad \text{and} \quad \tau_{\theta \xi}(x) = x + \theta \xi(x).$$

Hence by Jensen’s inequality, Fubini’s theorem and Minkovski’s inequality we get

$$\int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathfrak{Z}_1} |I_s^{\xi^*} u_s(x)|^2 \nu_1(d\mathfrak{z}) \right)^{p/2} dx ds \leq N |\bar{\xi}|_{L^2(\mathfrak{Z}_1)}^p \int_0^T |u_s|_{W_p^1}^p ds < \infty \text{ (a.s.)} \tag{6.16}$$

with a constant $N = N(p, d, \lambda, L, K_\eta, K_\xi)$. By Jensen’s inequality and Fubini’s theorem we obtain

$$\int_0^T \int_{\mathbb{R}^d} \int_{\mathfrak{Z}_1} |I_s^{\xi^*} u_s(x)|^p \nu_1(d\mathfrak{z}) dx ds \leq N |\bar{\xi}|_{L^2(\mathfrak{Z}_1)}^2 \int_0^T |u_s|_{W_p^1}^p ds < \infty \text{ (a.s.)}, \tag{6.17}$$

and for every $i = 1, 2, \dots, d$

$$\int_0^T \int_{\mathbb{R}^d} \int_{\mathfrak{Z}_1} |K_i^{\xi_s} u_s(x)|^p \nu_1(d\mathfrak{z}) dx ds \leq N |\bar{\xi}|_{L^2(\mathfrak{Z}_1)}^2 \int_0^T |u_s|_{W_p^1}^p ds < \infty \text{ (a.s.)}, \tag{6.18}$$

$$\int_0^T \int_{\mathbb{R}^d} \int_{\mathfrak{Z}_1} |K_i^{\eta_s} u_s(x)|^p \nu_1(d\mathfrak{z}) dx ds \leq N |\bar{\xi}|_{L^2(\mathfrak{Z}_1)}^2 \int_0^T |u_s|_{W_p^1}^p ds < \infty \text{ (a.s.)} \tag{6.19}$$

with a constant $N = N(p, d, \lambda, L, K_\eta, K_\xi)$. Hence, by virtue of Theorem 2.2 in [8] we get from equation (6.14), taking into account (6.15) through (6.19), that $(u_t)_{t \in [0, T]}$ is strongly cadlag as an L_p -valued process. \square

By the above proposition u is a strongly cadlag L_p -valued process, as well as weakly cadlag as an W_p^m -valued process. By interpolation we then have a constant $N = N(d, m, s, p)$ such that for $\theta = s/m$

$$\begin{aligned} |u_t - u_{t_n}|_{W_p^s} &\leq N |u_t - u_{t_n}|_{W_p^m}^\theta |u_t - u_{t_n}|_{L_p}^{1-\theta} \leq 2N \zeta^\theta |u_t - u_{t_n}|_{L_p}^{1-\theta}, \\ |u_{r_n} - u_{r-}|_{W_p^s} &\leq N |u_{r_n} - u_{r-}|_{W_p^m}^\theta |u_{r_n} - u_{r-}|_{L_p}^{1-\theta} \leq 2N \zeta^\theta |u_{r_n} - u_{r-}|_{L_p}^{1-\theta} \end{aligned}$$

for any $t \in [0, T], r \in (0, T]$, any strictly decreasing sequences $t_n \rightarrow t$ and strictly increasing sequences $r_n \rightarrow r$ with $r_n, t_n \in (0, T)$, where u_{r-} denotes the weak limit in W_p^m of u at r from the left, and $\zeta := \sup_{t \in [0, T]} |u_t|_{W_p^m} < \infty$ (a.s.). Letting here $n \rightarrow \infty$ we finish the proof. \square

Acknowledgements The authors are grateful to Nicolai Krylov, whose comments and suggestions greatly improved the present article.

Author Contributions Both authors worked on the manuscript.

Data Availability No datasets were generated or analysed during the current study.

Conflict of interest The authors declare no Conflict of interest.

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

References

1. Bain, A., Crisan, D.: *Fundamentals of Stochastic Filtering*. Springer, New York (2009)
2. Blackwood, S.: *Lévy processes and filtering theory*, Dissertation, University of Sheffield, (2014)

3. Calvia, A., Ferrari, G.: Nonlinear filtering of partially observed systems arising in singular stochastic optimal control. *Appl. Math. Optim.* **85**(2), 1–43 (2022)
4. Dareiotis, K.A., Kumar, C., Sabanis, S.: On tamed Euler approximations of SDEs driven by Lévy noise with applications to delay equations, *SIAM J. Numer. Anal.* (2016)
5. Davie, A., Germ, F., Gyöngy, I.: On partially observed jump diffusions II. The filtering density, *Stochastics and Partial Differential Equations: Analysis and Computations* (2023)
6. De-Léon Contreras, M., Gyöngy, I., Wu, S.: On solvability of integro-differential equations. *Potential Anal.* **55**(3), 443–475 (2021)
7. Germ, F., Gyöngy, I.: On Partially Observed Jump Diffusions I. *The Filtering Equations*, (2022). [arXiv:2205.08286](https://arxiv.org/abs/2205.08286)
8. Gyöngy, I., Wu, S.: Itô's formula for jump processes in L^p -spaces, *Stoch. Process. Appl.*, (2021)
9. Krylov, N.V.: An analytic approach to SPDEs, *stochastic partial differential equations: six perspectives. Math. Surv. Monogr.* **64**, 185–242 (1999)
10. Krylov, N.V.: On divergence form SPDEs with VMO coefficients. *SIAM J. Math. Anal.* **40**(6), 2262–2285 (2009)
11. Krylov, N.V.: On divergence form SPDEs with growing coefficients in W_2^1 spaces without weights. *SIAM J. Math. Anal.* **42**, 609–633 (2010)
12. Krylov, N.V.: Kalman-Bucy filter and SPDEs with growing lower-order coefficients in W_p^1 spaces without weights. III. *J. Math.* **54**(3), 1069–1114 (2010)
13. Krylov, N.V.: Filtering equations for partially observable diffusion processes with Lipschitz continuous coefficients. *The Oxford handbook of nonlinear filtering*, 169–194, Oxford Univ. Press, Oxford, (2011)
14. Krylov, N.V., Rozovskii, B.L.: On conditional distributions of diffusion processes. *Math. USSR Izv.* **12**, 336–356 (1978)
15. Kurtz, T.G., Xiong, J.: Particle representations for a class of nonlinear SPDEs, *Stoch. Processes Appl.* **83** (1999)
16. Kurtz, T.G., Ocone, D.L.: Unique characterization of conditional distributions in nonlinear filtering, *Annals Probab* (1988)
17. Maroulas, V., Pan, X., Xiong, J.: Large deviations for the optimal filter of nonlinear dynamical systems driven by Lévy noise. *Stoch. Process. Appl.* **130**, 203–231 (2020)
18. Qiao, H., Duan, J.: Nonlinear filtering of stochastic dynamical systems with Lévy noises, *Adv. Appl. Probab.* 47-3 (2015)
19. Qiao, H.: Nonlinear filtering of stochastic differential equations driven by correlated Lévy noises, *Stochastics* (2021)
20. Rozovskii, B.L.: On conditional distributions of degenerate diffusion processes. *Theory Probab. Appl.* **25**(1), 147–151 (1980)

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