FILTERING EQUATIONS FOR PARTIALLY OBSERVABLE DIFFUSION PROCESSES WITH LIPSCHITZ CONTINUOUS COEFFICIENTS

N.V. KRYLOV

ABSTRACT. We present several results on smoothness in $L_p$ sense of filtering densities under the Lipschitz continuity assumption on the coefficients of a partially observable diffusion processes. We obtain them by rewriting in divergence form filtering equation which are usually considered in terms of formally adjoint to operators in nondivergence form.

1. INTRODUCTION

For the author, one of the main motivations for developing the theory of stochastic partial differential equations (SPDEs) is its relation to the filtering problem for partially observable diffusion processes.

This problem’s setting is as follows.

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with an increasing filtration $\{\mathcal{F}_t, t \geq 0\}$ of complete, with respect to $(\mathcal{F}, P)$, $\sigma$-fields $\mathcal{F}_t \subset \mathcal{F}$. Denote by $\mathcal{P}$ the predictable $\sigma$-field in $\Omega \times (0, \infty)$ associated with $\{\mathcal{F}_t\}$. Let $d \geq 1$, $d_1 > d$, and $d_2 \geq d_1$ be integers and $w_t$ be a $d_2$-dimensional Wiener process with respect to $\{\mathcal{F}_t\}$. Let $K, T, \delta > 0$ be fixed finite constants.

Consider a $d_1$-dimensional two component process $z_t = (x_t, y_t)$ with $x_t$ being $d$-dimensional and $y_t$ ($d_1 - d$)-dimensional. We assume that $z_t$ is a diffusion process defined as a solution of the system

$$
\begin{align*}
    dx_t &= b(t, z_t)dt + \theta(t, z_t)dw_t, \\
    dy_t &= B(t, z_t)dt + \Theta(t, y_t)dw_t
\end{align*}
$$

(1.1)

with some initial data.

The coefficients of (1.1) are assumed to be vector- or matrix-valued functions of appropriate dimensions defined on $[0, T] \times \mathbb{R}^{d_1}$. Actually $\Theta(t, y)$ is assumed to be independent of $x$, so that it is a function on $[0, T] \times \mathbb{R}^{d_1-d}$ rather than $[0, T] \times \mathbb{R}^{d_1}$ but as always we may think of $\Theta(t, y)$ as a function of $(t, z)$ as well.

The component $x_t$ is treated as unobservable and $y_t$ as the only observations available. The problem is to find a way to compute the density $\pi_t(x)$ of
the conditional distribution of $x_t$ given $y_s, s \leq t$. Finding an equation satisfied by $\pi_t$ (filtering equation) is considered to be a solution of the (filtering) problem. Filtering equations turn out to be particular cases of SPDEs.

The history of filtering equations for diffusion processes is long and its beginning is controversial. Probably, the first filtering equations were published in [St60]. They turned out to be plain wrong. Then in [Ku64] other equations were proposed, see for instance equation (5) of [Ku64]. However, it is hard to make sense of these equations because most likely some terms appeared from stochastic integrals written in the Stratonovich form and the others appeared from the Itô integrals. Perhaps, the author of [Ku64] realized this too and published an attempt to rescue some results of [Ku64] in [Ku67]. This attempt turned successful for simplified models without the so-called cross terms.

Meanwhile, in [Sh66] the correct filtering equations in full generality, yet assuming some regularity of the filtering density, were presented and then in [LS68] they were rigorously proved. This is the reason we propose to call the filtering equations in the case of partially observable diffusion processes Shiryaev’s equations and their particular case without cross terms Kushner’s equations.

In case $d = 1$ the result of [Sh66] is presented in [LS01] on the basis of the famous Fujisaki-Kallianpur-Kunita theorem (see [FKK]) about the filtering equations in a very general setting (much more general than in [LS68]). Some authors even call the filtering equation for diffusion processes the Fujisaki-Kallianpur-Kunita equation.

By adding to the Fujisaki-Kallianpur-Kunita theorem some simple facts from the theory of SPDEs, the a priori regularity assumption was removed in [KR78] and under the Lipschitz and uniform nondegeneracy assumption the $L_2$-version of Theorem 2.6 was proved. The basic result of [KR78] is that $\pi_t \in H^1_2$. It is also proved that if the coefficients are smoother, $\pi_t(x)$ is smoother too. The nondegeneracy assumption is removed in [R90] on the account of assuming that $\theta \theta^*$ is three times continuously differentiable in $x$. It is again proved that $\pi_t \in H^1_2$ and $\pi_t$ is even smoother if the coefficients are smoother.

In [K99] the results of [KR78] were improved, $\theta \theta^*$ is assumed to be twice continuously differentiable in $x$ and it is shown that $\pi_t \in H^2_p$ with any $p \geq 2$.

The above mentioned results of [KR78], [R90], and [K99] use the filtering theory in combination with the theory of SPDEs, the latter being stimulated by certain needs of filtering theory. It turns out that the theory of SPDEs alone can be used to obtain the above mentioned regularity results about $\pi_t$ without knowing anything from the filtering theory itself. It also can be used to solve other problems from the filtering theory.

The first “direct” (only using the theory of SPDEs) proof of regularity of $\pi_t$ is given in [KZ00] in the case that system (1.1) defines a nondegenerate diffusion process and $\theta \theta^*$ is twice continuously differentiable in $x$. It is proved that $\pi_t \in H^2_p$ with any $p \geq 2$ as in [K99]. Advantages of having
arbitrary $p$ are seen from results like our Theorem 2.7. Of course, on the way of investigating $\pi_t$ in [KZ00] filtering equations are derived “directly” in an absolutely different manner than before (on the basis of an idea from [KR81]).

In this article we relax the smoothness assumption in [KZ00] to the assumption that the coefficients of (1.1) are merely Lipschitz continuous, the assumption which is almost always supposed to hold when one deals with systems like (1.1). We find that $\pi_t \in H^1_p$. Thus, under the weakest smoothness assumptions we obtain the best (in the author’s opinion) regularity result on $\pi_t$. In particular, we prove that if the initial data is sufficiently regular, then the filtering density is almost Lipschitz continuous in $x$ and $1/2$ Hölder continuous in $t$. However, we still assume $z_t$ to be nondegenerate. Our approach is heavily based on analytic results. There is also a probabilistic approach developed in [Kn97] and based on explicit formulas for solutions initiated in [Pa79] and later developed in [KR81] and [Kn82] (also see references therein). This approach cannot give as sharp results as ours in our situation.

It seems to the author that under the same assumptions of Lipschitz continuity, by following an idea from [K79] one can solve another problem from filtering theory, the so-called innovation problem, and obtain the equality

$$\sigma\{y_s, s \leq t\} = \sigma\{\tilde{w}_s, s \leq t\},$$

where $\tilde{w}_t$ is the innovation Wiener process of the problem (its definition is reminded in Section 2). Recall that for degenerate diffusion processes the positive solution of the innovation problem is obtained in [Pu84] again on the basis of the theory of SPDEs under the assumption that the coefficients are more regular.

By the way, in our situation, if the coefficients are more regular, the filtering equation can be rewritten in a nondivergence form and then additional smoothness of the filtering density, existence of which is already established in this article, is obtained on the basis of regularity results from [K99].

The article is organized as follows. In Section 2 we state our main results part of which is proved in the same section. In Sections 3 and 4 we prove Theorems 2.6 and 2.8, respectively. Section 5 contains a collection of results from the theory of SPDEs which we use in the previous sections.

As it is done traditionally in filtering theory we consider finite-dimensional driving Wiener processes. However, our results will be based on the theory of SPDEs, outlined in Section 5 with countably many Wiener processes. We leave to the reader to do some trivial modifications in Section 5 in order to be able to apply its results in such cases.

2. Main results

First we state and discuss our assumptions.
Assumption 2.1. The functions $b$, $\theta$, $B$, and $\Theta$ are Borel measurable and bounded functions of their arguments. Each of them satisfies the Lipschitz condition in $z$ with constant $K \in (0, \infty)$.

Introduce
\[
\tilde{\theta}_t(z) = \begin{pmatrix} \theta(t, z) \\ \Theta(t, y) \end{pmatrix}, \quad \tilde{a}_t(z) = \frac{1}{2} \tilde{\theta}_t \tilde{\theta}_t^*(z), \quad \tilde{b}_t(z) = \begin{pmatrix} b(t, z) \\ B(t, z) \end{pmatrix},
\]
(2.1)
\[
\tilde{L}_t(z) = \tilde{a}_{ij}^t(z) \frac{\partial^2}{\partial z_i \partial z_j} + \tilde{b}_i^t(z) \frac{\partial}{\partial z_i},
\]
(2.2)
where $\tilde{\theta}^*$ is the transpose of $\tilde{\theta}$ and the summation convention is imposed.

Remark 2.1. System of equations (1.1) can be now written as
\[
dz_t = \tilde{b}(t, z_t) dt + \tilde{\theta}(t, z_t) dw_t.
\]
(2.3)

Assumption 2.2. The process $z_t$ is uniformly nondegenerate: for any $\lambda, z \in \mathbb{R}^{d_1}$ and $t \in [0, T]$ we have
\[
\tilde{a}_{ij}^t(z) \lambda^i \lambda^j \geq \frac{\delta}{2} |\lambda|^2.
\]
Traditionally, Assumption 2.2 is split into two following assumptions in which some useful objects are introduced. These assumptions were also used in the past to reduce $\tilde{\theta}$ to the so-called triangular form by replacing $w_t$ with a different Brownian motion.

Assumption 2.3. The symmetric matrix $\Theta \Theta^*$ is invertible and
\[
\Psi := (\Theta \Theta^*)^{-\frac{1}{2}}
\]
is a bounded function of $(t, y)$.

Remark 2.2. Assumption 2.3 follows from Assumption 2.2 and, furthermore, $\Psi \leq \delta^{-1}(\delta^{ij})$.

Assumption 2.4. For any $\xi \in \mathbb{R}^d$, $z = (x, y) \in \mathbb{R}^{d_1}$, and $t > 0$, we have
\[
|Q(t, y) \theta^* (t, z) \xi|^2 \geq \frac{\delta}{2} |\xi|^2,
\]
where $Q$ is the orthogonal projector on $\text{Ker} \Theta$. In other words,
\[
(\theta (I - \Theta^* \Psi^2 \Theta) \theta^* \xi, \xi) \geq \frac{\delta}{2} |\xi|^2.
\]
(2.4)

Remark 2.3. From (2.4) we see that $\theta \theta^*$ is uniformly positive definite with constant of positivity $\delta$. Also, it turns out that (2.4) holds under Assumption 2.2.

Indeed, take a $\zeta = (\xi, \Psi \eta) \in \mathbb{R}^d \times \mathbb{R}^{d_1-d}$ with $\eta = -\Psi \Theta \theta^* \xi$ and observe that
\[
2 \delta |\xi|^2 \leq 2(\tilde{a} \zeta, \zeta) = |\tilde{\theta}^* \zeta|^2 = |\theta \xi|^2 + 2(\tilde{\theta}^* \xi, \Theta^* \Psi \eta) + |\Theta^* \Psi \eta|^2
\]
\[
= |\theta \xi|^2 + 2(\Psi \Theta \tilde{\theta}^* \xi, \eta) + |\eta|^2 = |\theta \xi|^2 - |\Psi \Theta \tilde{\theta}^* \xi|^2,
\]
which is even stronger than (2.4).
Remark 2.4. We have seen that Assumptions 2.3 and 2.4 follow from Assumption 2.2. In turn Assumptions 2.3 and 2.4 imply Assumption 2.2 perhaps with a different constant in the latter.

To show this, we take \( \zeta = (\xi, \eta) \in \mathbb{R}^d \times \mathbb{R}^{d_1-d} \) and observe that
\[
2(\tilde{a}\zeta, \zeta) = (\theta\theta^*\xi, \xi) + 2(\Theta\Theta^*\xi, \eta) + (\Theta\Theta^*\eta, \eta)
\]
where \( \tilde{\eta} = \Psi^{-1}\eta \), and \( \varepsilon \in (0,1) \). By using the inequality \( 2(\mu, \nu) + \varepsilon|\mu|^2 \geq -\varepsilon^{-1}|\nu|^2 \) we see that
\[
2(\Psi\Theta\theta^*\xi, \tilde{\eta}) + \varepsilon(\tilde{\eta}, \tilde{\eta}) \geq -\varepsilon^{-1}|\Psi\Theta\theta^*\xi|^2,
\]
and by taking \( N \) such that \( \Psi \leq N(\delta^{ij}) \), for which \( \Theta\Theta^* \geq N^{-2}(\delta^{ij}) \), we conclude
\[
2(\tilde{a}\zeta, \zeta) \geq |\theta^*\xi|^2 - \varepsilon^{-1}||\Psi\Theta\theta^*\xi||^2 + (1-\varepsilon)N^{-2}|\eta|^2
\]
\[
\geq \delta|\xi|^2 + (1-\varepsilon^{-1})||\Psi\Theta\theta^*\xi||^2 + (1-\varepsilon)N^{-2}|\eta|^2,
\]
where the last inequality follows from (2.4). Finally, \( \Psi\Theta\theta^* \) is a bounded function, so that, for a constant \( N_1 \),
\[
2(\tilde{a}\zeta, \zeta) \geq (\delta + N_1(1-\varepsilon^{-1})){|\xi|^2 + (1-\varepsilon)N^{-2}|\eta|^2}.
\]
For \( \varepsilon \) sufficiently close to 1 the last expression is greater than \( \delta_1|\xi|^2 \) with a constant \( \delta_1 > 0 \), which is equivalent to the uniform ellipticity of \( \tilde{a} \).

Before stating the next assumption we remind the reader that, for \( \gamma \in \mathbb{R} \) and \( u \in C^\infty_0(\mathbb{R}^d) \) one introduces \( (1-\Delta)^{-\gamma/2}u \) by means of the Fourier transform. Then, for \( p \in (1,\infty) \), one defines the spaces of Bessel potential \( H^\gamma_p(\mathbb{R}^d) \) as the set of distributions obtained as the closure of \( C^\infty_0(\mathbb{R}^d) \) with respect to the norm
\[
\|u\|_{H^\gamma_p(\mathbb{R}^d)} := \|(1-\Delta)^{\gamma/2}u\|_{L^p(\mathbb{R}^d)}.
\]
One important and highly nontrivial piece of information is that
\[
H^1_p(\mathbb{R}^d) = W^1_p(\mathbb{R}^d) := \{u \in L^p(\mathbb{R}^d) : \nabla u \in L^p(\mathbb{R}^d)\}
\]
and
\[
\|u\|_{H^1_p(\mathbb{R}^d)} \sim \|u\|_{W^1_p(\mathbb{R}^d)} := \|u\|_{L^p(\mathbb{R}^d)} + \|\nabla u\|_{L^p(\mathbb{R}^d)}.
\]
(2.5)

Assumption 2.5. The random vectors \( x_0 \) and \( y_0 \) are independent of the process \( w_t \). The conditional distribution of \( x_0 \) given \( y_0 \) has a density, which we denote by \( \pi_0(x) = \pi_0(\omega, x) \). We have \( p \geq 2 \) and \( \pi_0 \in L^p(\Omega, H^{1-2/p}_p(\mathbb{R}^d)) \) (actually, we need slightly less, see Remark 3.1).

Next we introduce few more notation. Let
\[
\Psi_t = \Psi(t, y_t), \quad \Theta_t = \Theta(t, y_t), \quad a_t(x) = \frac{1}{2}\theta\theta^*(t, x, y_t), \quad b_t(x) = b(t, x, y_t),
\]
\[
\sigma_t(x) = \theta(t, x, y_t)\Theta^*_t\Psi_t, \quad \beta_t(x) = \Psi_t B(t, x, y_t).
\]
In the remainder of the article we use the notation
\[ D_i = \frac{\partial}{\partial x^i} \]
only for \( i = 1, \ldots, d \) and set
\[
L_i(x) = a^{ij}_i(x)D_jD_i + b_i^j(x)D_i,
\]
(2.6)
\[
L^*_i(x)u_t(x) = D_iD_j(a^{ij}_i(x)u_t(x)) - D_i(b_i^j(x)u_t(x)) = D_j(a^{ij}_i(x))D_1u_t(x) + u_t(x)D_i(a^{ij}_i(x)),
\]
(2.7)
\[
\Lambda^k(x)u_t(x) = \beta^k(x)u_t(x) + \sigma^j_k(x)D_iu_t(x),
\]
(2.8)
\[
\Lambda^{k*}(x)u_t(x) = \beta^k_t(x)u_t(x) - D_i(\sigma^j_k(x)u_t(x)) = -\sigma^j_k(x)D_iu_t(x) + (\beta^k_t(x) - D_i\sigma^j_k(x))u_t(x),
\]
(2.9)
where \( t \in [0, T] \), \( x \in \mathbb{R}^d \), \( k = 1, \ldots, d_1 - d \), and as above we use the summation convention over all “reasonable” values of repeated indices, so that the summation in (2.6), (2.7), (2.8), and (2.9) is done for \( i, j = 1, \ldots, d \) (whereas in (2.2) for \( i, j = 1, \ldots, d_1 \)). Observe that Lipschitz continuous functions have bounded generalized derivatives and by
\[
D_i a^{ij}_i, \quad D_i \sigma^j_k
\]
we mean these derivatives. From Remark 2.3 we have that the operator \( L \) defined by (2.6) is uniformly elliptic with constant of ellipticity \( \delta \).

Finally, by \( F^+_t \) we denote the completion of \( \sigma\{y_s : s \leq t\} \) with respect to \( P, \mathcal{F} \).

Let us consider the following initial value problem
\[
d\pi_t(x) = L_i(x)\pi_t(x) \, dt + \Lambda^{k*}(x)\pi_t(x)\Psi^k_t \, dy_t,
\]
(2.10)
where \( \pi_0(x) = \pi_0(x) \),
\[
\pi_t(\omega, x) = \pi_0(\omega, x).
\]
Equation (2.10) is called the Duncan-Mortensen-Zakai or just the Zakai equation.

We understand this equation and the initial condition in the following sense. We are looking for a function \( \pi = \pi_t(x) = \pi_t(\omega, x), \omega \in \Omega, t \in [0, T], x \in \mathbb{R}^d \), such that

(i) For each \( (\omega, t), \pi_t(\omega, x) \) is a generalized function on \( \mathbb{R}^d \),

(ii) We have \( \pi \in L_P(\Omega \times [0, T], \mathcal{P}, H^1_p(\mathbb{R}^d)) \),

(iii) For each \( \varphi \in C_0^\infty(\mathbb{R}^d) \) with probability one for all \( t \in [0, T] \) it holds that
\[
(\pi_t, \varphi) = (\pi_0, \varphi) - \int_0^t (a^{ij}_i D_i \pi_t - b_i^j \pi_t + \pi_t D_i a^{ij}_i, D_j \varphi) \, dt
\]
\[
- \int_0^t (\sigma^j_k D_i \pi_t + (D_i \sigma^j_k - \beta^k_i \pi_t, \varphi) \Psi^k_t (B^r(t, z_t) \, dt + \Theta^{ks}(t, \eta_t) \, dw_t^s),
\]
(2.11)
where by \( (f, \varphi) \) we mean the action of a generalized function \( f \) on \( \varphi \), in particular, if \( f \) is a locally summable,
\[
(f, \varphi) = \int_{\mathbb{R}^d} f(x) \varphi(x) \, dx.
\]
Observe that all expressions in (2.11) are well defined due to the fact that the coefficients of $\bar{\pi}$ and of $D_i \bar{\pi}$ are bounded and appropriately measurable and $\bar{\pi}, D_i \bar{\pi} \in L_p(\Omega \times [0,T], \mathcal{P}, L_p(\mathbb{R}^d))$ (see (2.5)).

Hence, equation (2.10) has the same form as (5.1) and the existence and uniqueness part of Lemma 2.5 below follow from Theorem 5.1 and Remark 3.1. The second assertion of the lemma follows from Theorem 5.4.

In all what follows in the main part of the article we suppose that Assumptions 2.1, 2.2, and 2.5 are satisfied.

Lemma 2.5. There exists a unique solution $\bar{\pi}$ of (2.10) with initial condition $\pi_0$ in the sense explained above. In addition, $\bar{\pi}_t \geq 0$ for all $t \in [0,T]$ (a.s.).

Here is a basic result of filtering theory for partially observable diffusion processes. Its relation to the previously known ones is discussed above.

Theorem 2.6. Let $\bar{\pi}$ be the function from Lemma 2.5. Then

$$0 < \int_{\mathbb{R}^d} \bar{\pi}_t(x) \, dx = (\bar{\pi}_t, 1) < \infty$$

(2.12)

for all $t \in [0,T]$ (a.s.) and for any $t \in [0,T]$ and real-valued, bounded or nonnegative, (Borel) measurable function $f$ given on $\mathbb{R}^d$

$$E[f(x_t)|\mathcal{F}_t^y] = (\bar{\pi}_t, f) \quad (\bar{\pi}_t, 1) \quad (a.s.).$$

(2.13)

Equation (2.13) shows (by definition) that

$$\pi_t(x) := \frac{\bar{\pi}_t(x)}{(\bar{\pi}_t, 1)}$$

is a conditional density of distribution of $x_t$ given $y_s, s \leq t$. Since, generally, $(\bar{\pi}_t, 1) \neq 1$, one calls $\pi_t$ an unnormalized conditional density of distribution of $x_t$ given $y_s, s \leq t$.

The following is a direct corollary of Theorem 5.5.

Theorem 2.7. Let $\pi_0$ be a nonrandom function and $\pi_0 \in H^{1-2/p}_{p-2/p}(\mathbb{R}^d)$ for all $p \geq 2$, which happens for instance, if $\pi_0$ is a Lipschitz continuous function with compact support. Then for any $\varepsilon \in (0,1/2)$ almost surely $\bar{\pi}_t(x)$ is $1/2 - \varepsilon$ Hölder continuous in $t$ with a constant independent of $x$, $\bar{\pi}_t(x)$ is $1-\varepsilon$ Hölder continuous in $x$ with a constant independent of $t$, and the above mentioned (random) constants have all moments.

In filtering theory usually the following theorem is proved before anything else is done. We do not need it for proving the above results and give the proof just to show that the $L_p$-theory of SPDEs allows one to get all basic results from filtering theory.

Historically, $P_t[\beta]$ was introduced by (2.16) and shown to have (a modification possessing) appropriate measurability properties. Then $\bar{\pi}_t$ used to
be defined as the density of conditional distribution of \( x_t \) given \( F_t^\gamma \) divided by an appropriate modification of

\[
E(\rho_t \mid F_t^\gamma), \tag{2.14}
\]

where

\[
\rho_t = \exp\left( -\int_0^t \tilde{\beta}_s \, d\tilde{w}_s - \frac{1}{2} \int_0^t |\tilde{\beta}_s|^2 \, ds \right), \quad \tilde{\beta}_s = \beta_s(x_s), \quad \tilde{w}_t = \int_0^t \Psi_s \Theta_s \, dw_s.
\]

In this case \((\bar{\pi}_t, 1)^{-1}\) turns out to be this same appropriate modification of (2.14) (cf. our (3.20)).

The most surprising statements in Theorem 2.8 are assertions (iv) and (v). In (iv) the difference of two Wiener processes \( \tilde{w}_t \) and \( \tilde{w}_t \) (that the latter is a Wiener process is checked in the proof of Lemma 3.3) is asserted to be a differentiable nontrivial function.

Assertion (v) shows that (2.14), which is a conditional expectation of a martingale, is again a martingale and, moreover, while evaluating it we can just put conditional expectations of \( \tilde{\beta}_s \) given \( F_y_s \) in place of \( \tilde{\beta}_s \) in the expression of \( \rho_t \) with simultaneous replacement of \( \tilde{w} \) with \( \tilde{w} \).

**Theorem 2.8.** (i) The process \((\bar{\pi}_t, 1)\) is continuous in \( t \) (a.s.) and (a.s.)

\[
(\bar{\pi}_t, 1) = (\pi_0, 1) + \int_0^t (\bar{\pi}_s, \beta^k_s) \Psi^{kr}_s B^r(t, z_s) \, ds + \int_0^t (\bar{\pi}_s, \beta^k_s) \Psi^{kr}_s \Theta^{rn}(t, y_s) \, dw^r_s. \tag{2.15}
\]

(ii) The process \( \bar{\pi}_t \) is a continuous \( L_1 \)-valued process (a.s.).

(iii) Introduce \( P_t[\beta] = (P_t[\beta^1], ..., P_t[\beta^{d_1-d}]) \) by

\[
P_t[\beta] = (\bar{\pi}_t, 1)^{-1} \int_{\mathbb{R}^d} \beta(t, x) \bar{\pi}_t(x) \, dx = (\bar{\pi}_t, 1)^{-1} \Psi(t, y_t) \int_{\mathbb{R}^d} B(t, x, y_t) \bar{\pi}_t(x) \, dx.
\]

Then \( P_t[\beta] \) is a jointly measurable bounded \( F_t^\gamma \)-adapted process on \([0, t]\) (a.s.) and for each \( t \in [0, T] \)

\[
P_t[\beta] = E(\beta(t_x) \mid F_t^\gamma) \quad \text{(a.s.)}. \tag{2.16}
\]

(iv) The process

\[
\tilde{w}_t = \tilde{w}_t + \int_0^t (\beta_s(x_s) - P_s[\beta]) \, ds
\]

is a \((d_1 - d)\)-dimensional Wiener process with respect to \( F_t^\gamma \) (the so-called innovation process), where

\[
\tilde{w}_t = \int_0^t \Psi_s \Theta_s \, dw_s.
\]

(v) We have (a.s.) for all \( t \in [0, T] \)

\[
(\bar{\pi}, 1) = \exp \left( \int_0^t P_s[\beta] \, d\tilde{w}_s + \frac{1}{2} \int_0^t |P_s[\beta]|^2 \, ds \right), \tag{2.17}
\]
so that
\[(\pi, 1)^{-1} = \exp \left( - \int_0^t P_s[\beta] \, d\tilde{w}_s - \frac{1}{2} \int_0^t |P_s[\beta]|^2 \, ds \right) \]
is an exponential martingale, and for each \(m > 0\)
\[E \sup_{t \leq T} (\pi, 1)^m + E \sup_{t \leq T} (\pi, 1)^{-m} < \infty. \tag{2.18} \]

3. Proof of Theorem 2.6

We will use some notion and results from the theory of SPDEs, which are recalled in Section 5. From now on we drop \(\mathbb{R}^d\) in notation like \(H^p_\beta(\mathbb{R}^d)\) and \(L_p(\mathbb{R}^d)\).

Remark 3.1. The assumption that \(\pi_0 \in L_p(\Omega, H^{1-2/p}_p)\) is only needed to guarantee (see the proof of Theorem 5.1 of [K99]) that there exists a \(\psi \in H^{1}_p(T)\) such that \(\psi_0 = \pi_0\),
\[
d\psi_t = \Delta \psi_t \, dt = D_i f^i_t \, dt, \quad (f^i_t = D_i \psi_t),
\]
with \(N\) independent of \(\pi_0\).

As is mentioned before Lemma 2.5 by Theorem 5.1 and Remark 3.1, there exists a unique solution \(\bar{\pi} \in H^{1}_p(T)\) of (2.10) with initial condition \(\pi_0\). By Theorem 5.4 \(\bar{\pi}_t \geq 0\) for all \(t \in [0, T]\) (a.s.). By Theorem 5.5 \(\bar{\pi}_t\) is a continuous \(L_p\)-valued process and
\[E \sup_{t \in [0, T]} \|\bar{\pi}_t\|_{L_p}^p \, dt < \infty. \tag{3.1} \]

Now, we prove three auxiliary results.

Lemma 3.2. Let \(\xi_t, \xi^n_t, n = 1, 2, \ldots, t \in [0, T]\), be \(k\)-dimensional continuous semimartingales such that, for any \(t \in [0, T]\), \(\xi^n_t \to \xi_t\) in probability as \(n \to \infty\). Assume that
\[
\xi^n_t = \xi^n_0 + \int_0^t \alpha^n_s \, ds + m^n_t, \quad \xi_t = \xi_0 + \int_0^t \alpha_s \, ds + m_t,
\]
where \(\alpha_t\) and \(\alpha^n_t\) are predictable processes bounded by the same nonrandom constant and \(m_t\) and \(m^n_t\) are martingales such that
\[
\langle m^{ni}, m^{nj} \rangle_t = \int_0^t \gamma_{s}^{nij} \, ds, \quad \langle m^i, m^j \rangle_t = \int_0^t \gamma_s^{ij} \, ds, \quad i, j = 1, \ldots, k,
\]
where \(\gamma^n_t := (\gamma^{nij}_t)\) and \(\gamma_t := (\gamma^{ij}_t)\) are predictable matrix-valued processes bounded by the same nonrandom constant and such that \((\gamma^n_t)^{-1}\) and \((\gamma_t)^{-1}\) exist and are also bounded by the same nonrandom constant. Assume that on \([0, T] \times \mathbb{R}^l \times \mathbb{R}^k\) we are given functions \(f^n_t(x, y)\) and \(f_t(x, y)\) such that they are uniformly bounded and \(f^n \to f\) in measure as \(n \to \infty\). Then \(f^n_t(x, \xi^n_t) \to f_t(x, \xi_t)\) in measure on \(\Omega \times [0, T] \times \mathbb{R}^l\).
Proof. It suffices to show that any subsequence \( \{n'\} \) of integers has a subsequence \( \{n''\} \) such that \( f^n_{n''}(x, \xi_{n''}) \to f_t(x, \xi_t) \) in measure. Since any subsequence \( \{n'\} \) has a subsequence \( \{n''\} \) such that \( f^{n''} \to f \) almost everywhere, by having in mind renumbering if needed, we may assume that for the original sequence we have \( f^n \to f \) almost everywhere. In that case for almost any \( x \in \mathbb{R}^l \), \( f^n(x, y) \to f_t(x, y) \) and, if we prove that for each such \( x \) we have \( f^n(x, \xi^n_t) \to f_t(x, \xi_t) \) in measure on \( \Omega \times [0, T] \), then

\[
E \int_0^T |f^n_t(x, \xi^n_t) - f_t(x, \xi_t)| \, dt \to 0,
\]

which after being integrated with respect to \( x \) would shows that \( f^n_t(x, \xi^n_t) \to f_t(x, \xi_t) \) in measure on \( \Omega \times [0, T] \times \mathbb{R}^l \).

It follows that we only need to prove that, if on \( [0, T] \times \mathbb{R}^k \) we are given functions \( f^n_t(y) \) and \( f_t(y) \) such that they are uniformly bounded and \( f^n \to f \) \((t, y)\)-almost everywhere as \( n \to \infty \), then

\[
E \int_0^T |f^n_t(\xi^n_t) - f_t(\xi_t)| \, dt \to 0. \tag{3.2}
\]

Furthermore, since the coefficients \( \alpha^n, \alpha, \gamma^n, \) and \( \gamma \) are uniformly bounded

\[
\sup_n \sup_{t \in [0, T]} P(|\xi^n_t| + |\xi_t| \geq R) \leq R^{-2} \sup_n \sup_{t \in [0, T]} E(|\xi^n_t|^2 + |\xi_t|^2) \to 0
\]

as \( R \to \infty \). Therefore, if for any \( R \in (0, \infty) \) we know that \( f^n_t(\xi^n_t) \) is independent of \( R \). Then by applying this result in the general case to \( f^n_t(y) I_{|y| < R} \) and \( f_t(y) I_{|y| < R} \) we would obtain that

\[
\lim_{n \to \infty} E \int_0^T |f^n_t(\xi^n_t) - f_t(\xi_t)| \, dt \leq NR^{-2},
\]

where \( N \) is independent of \( R \). This would imply \( (3.2) \) in the general case. This shows that without restricting generality we may assume that for an \( R \in (0, \infty) \) the functions \( f^n_t(y) \) and \( f_t(y) \) vanish if \( |y| \geq R \).

Now observe that the left-hand side of \((3.2)\) is majorated by \( I_n + J_n \), where

\[
I_n = E \int_0^T |f^n_t(\xi^n_t) - f_t(\xi^n_t)| \, dt, \quad J_n = E \int_0^T |f_t(\xi^n_t) - f_t(\xi_t)| \, dt.
\]

We recall a result of [K77] implying that for any \( g \in L_{k+1}([0, T] \times \mathbb{R}^k) \) we have

\[
E \int_0^T (|g_t(\xi^n_t)| + |g_t(\xi_t)|) \, dt \leq N\|g\|_{L_{k+1}([0, T] \times \mathbb{R}^k)},
\]

where \( N \) is independent of \( n \) and \( g \). We apply this result to \( g = f^n - f \) and observe that these functions are uniformly bounded, vanish for \( |y| \geq R \), and tend to zero in measure. Hence, their \( L_{k+1}([0, T] \times \mathbb{R}^k) \)-norms tend to zero. This implies that \( I_n \to 0 \).
Next, notice that for any function $g$
\[ J_n \leq E \int_0^T |g_t(\xi_t^n) - g_t(\xi_t)| \, dt \]
\[ + E \int_0^T |f_t(\xi_t^n) - g_t(\xi_t^n)| \, dt + E \int_0^T |f_t(\xi_t) - g_t(\xi_t)| \, dt \]
implicating that
\[ \lim_{n \to \infty} J_n \leq \lim_{n \to \infty} E \int_0^T |g_t(\xi_t^n) - g_t(\xi_t)| \, dt + N\|f - g\|_{L^{k+1}([0,T] \times \mathbb{R}^d)}, \quad (3.3) \]
where $N$ is independent of $g$. For any $\varepsilon > 0$ we can find a smooth $g$ such that the second term on the right in (3.3) will be less than $\varepsilon$. In addition, the first term vanishes for smooth $g$ since $\xi_t^n \to \xi_t$ in probability for any $t$. Since $\varepsilon$ is arbitrary, it follows that the left-hand side of (3.3) equals zero.

The lemma is proved.

The following result with its proof is an adaptation of Lemma 5.1 of [KZ00] and its proof.

Lemma 3.3. The function $\tilde{\pi}_t$ is $\mathcal{F}_t^y$-adapted.

Proof. Define
\[ \tilde{\beta}_t = \beta_t(x_t) = \Psi_t B(t, z_t), \quad \tilde{w}_t = \int_0^t \Psi_s \, dy_s, \quad \tilde{\pi}_t = \int_0^t \Psi_s \Theta_s \, dw_s. \]
Since $\Psi_t$ is $\mathcal{F}_t^y$-adapted, the process $\tilde{w}_t$ is $\mathcal{F}_t^y$-adapted too. Furthermore, $\Psi_s \Theta_s \Psi_s^T$ is a unit matrix so that by Lévy’s theorem $\tilde{w}_t$ is a Wiener process. We want to change the probability measure so that $\tilde{w}_t$ would become a Wiener process with respect to this new measure. Define
\[ \rho_t = \exp\left(-\int_0^t \tilde{\beta}_s \, d\tilde{w}_s - \frac{1}{2} \int_0^t |\tilde{\beta}_s|^2 \, ds\right), \quad Q(d\omega) = \rho_T(w) P(d\omega). \quad (3.4) \]
The process $\rho_t$ is an exponential local martingale. Since $\tilde{\beta}$ is bounded, $\rho_t$ is square integrable, so that $Q$ is a probability measure. Since
\[ d\tilde{w}_t = \tilde{\beta}_t \, dt + d\tilde{w}_t \]
and $\tilde{w}_t$ is a Wiener process on $(\Omega, \mathcal{F}, P)$, by Girsanov’s theorem, $\tilde{w}_t$, $t \in [0,T]$, is a Wiener process on $(\Omega, \mathcal{F}, Q)$ with respect to the filtration $\{\mathcal{F}_t\}$. As has been noticed before, it is $\mathcal{F}_t^y$-adapted and, obviously,
\[ \mathcal{F}_t^\| \subset \mathcal{F}_t, \]
so that $(\tilde{w}_t, \mathcal{F}_t^\|)$ is a Wiener process. Now rewrite (2.10) as
\[ d\tilde{\pi}_t(x) = L_t^\| (x) \tilde{\pi}_t(x) \, dt + \Lambda_t^\| (x) \tilde{\pi}_t(x) \, d\tilde{w}_t^\|, \quad (3.5) \]
and consider this equation relative to $(\Omega, \mathcal{F}, \mathcal{F}_t^\|, Q)$.

By Theorem 5.1 and Remark 3.1 equation (3.3) with initial data $\pi_0$ has a unique $\mathcal{F}_t^\|$-adapted solution belonging to $\mathcal{H}_p^1(\mathcal{F}^\|, Q, T) \subset \mathcal{H}_p^1(\mathcal{F}, Q, T)$, where by $\mathcal{H}_p^1(\mathcal{F}, Q, T)$ we mean the space $\mathcal{H}_p^1(T)$ constructed on the basis
of the new probability measure $Q$ and filtration $\mathcal{F}_t$. We denote by $\tilde{\pi}_t$ this solution.

We have already mentioned that $\tilde{\pi} \in \mathcal{H}_p^1(\mathcal{F}, P, T)$. We want to derive that $\tilde{\pi}_t$ is $\mathcal{F}_t^\mu$-adapted from the uniqueness by showing that $\tilde{\pi} = \tilde{\pi}$ because both are $\mathcal{F}_t$-adapted solutions of the same equation. The only obstacle is that the norms in $\mathcal{H}_p^1(\mathcal{F}, Q, T)$ and $\mathcal{H}_p^1(T)$ are different. To overcome this obstacle, we are going to use stopping times.

For integers $n$ define

$$
\tau(n) = T \wedge \inf \{ t \geq 0 : \int_0^t \| \tilde{\pi}_s \|_{H_p^1}^p \, ds \geq n \}.
$$

Obviously, $\tau(n)$ are $\mathcal{F}_t^\mu$-stopping times and $\mathcal{F}_t$-stopping times. Furthermore,

$$
\| \tilde{\pi} \|_{H_p^1}^p(\mathcal{F}, P, \tau(n)) = E \int_0^{\tau(n)} \| \tilde{\pi}_s \|_{H_p^1}^p \, ds \leq n < \infty.
$$

This and the equation (cf. (3.5))

$$
d\tilde{\pi}_t(x) = [L_t^*(x)\tilde{\pi}_t(x) + \beta_t^k \Lambda_t^k(x)\tilde{\pi}_t(x)] \, dt + \Lambda_t^k(x)\tilde{\pi}_t(x) \, d\tilde{w}_t^k
$$

show that, $\tilde{\pi} \in \mathcal{H}_p^1(\mathcal{F}, P, \tau(n))$. By the above mentioned uniqueness, $\tilde{\pi}_t = \tilde{\pi}_t$ on $(0, \tau(n))$ (a.e.). Since both functions are continuous in $t \in [0, T]$ (Theorem 5.5 (i)), we have that $\tilde{\pi}_t I_{0 < t \leq \tau(n)}$ and $\tilde{\pi}_t I_{0 < t \leq \tau(n)}$ are indistinguishable, and since one of them is $\mathcal{F}_t^\mu$-adapted, so is the other. We conclude that $\tilde{\pi}_t I_{0 < t \leq \tau(n)}$ is $\mathcal{F}_t^\mu$-adapted, which after letting $n \to \infty$ yields the result. The lemma is proved.

Assertion of the following lemma is a very particular case of one of the assertions of Theorem 2.8. Before stating the lemma we recall that $\tilde{\pi}_t \geq 0$ for all $t \in [0, T]$ (a.s.), so that $(\tilde{\pi}_t, 1)$ is well defined (and may be infinite).

**Lemma 3.4.** We have

$$
E \sup_{t \in [0, T]} (\tilde{\pi}_t, 1)^{1/2} < \infty. 
$$

**Proof.** For $\varphi \in C_0^\infty(\mathbb{R}^d)$ one can rewrite (2.11) as

$$
(\tilde{\pi}_t, \varphi) = (\pi_0, \varphi) + \int_0^t (\tilde{\pi}_s, L_s \varphi) \, ds
$$

$$
+ \int_0^t (\tilde{\pi}_s, \Lambda_s^k \varphi) \Psi_s^{k\tau} (B^\tau(s, z_s) \, ds + \Theta^{\tau n}(s, y_s) \, dw_s^n).
$$

Using (3.1) and an obvious passage to the limit, it is easy to prove that (3.7) holds not only for $\varphi \in C_0^\infty(\mathbb{R}^d)$, but also for $\varphi \in W^2_q$ with $q = p/(p - 1)$.

On $\mathbb{R}^d$ for $m = 1, 2, \ldots$ introduce the functions

$$
\varphi(x = (1 + |x|^2)^{-d}, \quad \varphi_m(x) = \varphi(x/m).
$$
Observe that for a constant $N$ it holds that
\[
|D_i \phi_m| + m |D_i D_j \phi_m| \leq Nm^{-1} \phi_m
\] (3.8)
on $\mathbb{R}^d$ for all $m$. In particular,
\[
2L_i \phi_m \leq N_0 \phi_m, \quad 2|\Psi^{kr}_i B^r(t, z_t) \Lambda_t^k \phi_m| \leq N_0 \phi_m, \quad (3.9)
\]
where $N_0$ is a constant independent of $m$ and the arguments of the functions involved.

By plugging in (3.7) the function $\phi_m$ in place of $\phi$, we obtain
\[
(\bar{\pi}_t, \phi_m) = (\pi_0, \phi_m) + \int_0^t (\bar{\pi}_s, L_s \phi_m) \, ds
\]
\[
+ \int_0^t (\bar{\pi}_s, \Lambda_t^k \phi_m) \Psi^{kr}_s B^r(s, z_s) \, ds + \Theta^{rn}(s, y_s) \, dw^n_s. \quad (3.10)
\]
By using Itô’s formula for transforming
\[
(\bar{\pi}_t, \phi_m)e^{-N_0 t}, \quad (3.11)
\]
and using (3.9) we see that
\[
d[(\bar{\pi}_t, \phi_m)e^{-N_0 t}] = e^{-N_0 t} (\bar{\pi}_t, \Lambda_t^k \phi_m) \Psi^{kr}_t \Theta^{rn}(t, y_t) \, dw^n_t
\]
\[
+ e^{-N_0 t} [(\bar{\pi}_t, L_t \phi_m) + (\bar{\pi}_s, \Lambda_t^k \phi_m) \Psi^{kr}_s B^r(s, z_s) - N_0(\bar{\pi}_s, \phi_m)] \, dt
\]
\[
\leq e^{-N_0 t} (\bar{\pi}_t, \Lambda_t^k \phi_m) \Psi^{kr}_t \Theta^{rn}(t, y_t) \, dw^n_t. \quad (3.12)
\]
It follows that process (3.11) is a supermartingale. It is continuous and nonnegative. Therefore,
\[
E \sup_{t \in [0,T]} e^{-N_0 t} \left( \int_{\mathbb{R}^d} \phi_m \bar{\pi}_t(x) \, dx \right)^{1/2} \leq 2 \left( E \int_{\mathbb{R}^d} \phi_m \bar{\pi}_0(x) \, dx \right)^{1/2} \leq 2.
\]
Upon letting $m \to \infty$ and using the monotone convergence theorem we come to (3.13) and the lemma is proved.

**Proof of Theorem 2.6** Take a nonnegative $\zeta \in C^\infty_0(\mathbb{R}^d)$, which integrates to one and for $n = 1, 2, \ldots$ set
\[
\zeta_n(z) = n^{d_1} \zeta(nz).
\]
Also introduce mollifications of one of the coefficients of (1.1) by
\[
\theta^{(n)}(t, z) = \zeta_n(z) * \theta(t, z),
\]
where the convolutions is taken with respect to $z$.

The function $\zeta$ can be considered as the density of a random variable. If needed, we extend our initial probability space in such a way that it would allow us to introduce a new random $\mathbb{R}^{d_1}$-valued vector $\xi$ having density $\zeta$ and such that $\xi$ is independent of $z_0$ and the process $w_t$, $t \geq 0$.

After that, for $n = 1, 2, \ldots$, we consider the following modification of (1.1):
\[
dx^{(n)}_t = b(t, z_t^{(n)}) \, dt + \theta^{(n)}(t, z_t^{(n)}) \, dw_t
\]
\[
dy^{(n)}_t = B(t, z_t^{(n)}) \, dt + \Theta(t, y_t^{(n)}) \, dw_t \quad (3.12)
\]
with initial data \( x_0^{(n)} = x_0 + n^{-1} \xi, \ y_0^{(n)} = y_0 \) and \( z_t^{(n)} = (x_t^{(n)}, y_t^{(n)}) \). Observe that the conditional distribution of \( x_0^{(n)} \) given \( y_0 \) has a density equal to

\[
\pi_0^{(n)} = \zeta_n * \pi_0.
\]

Since \( \theta(t, x, y) \) is Lipschitz in \( x \) (even in \((x, y)\)) we have \(|\theta(t, z) - \theta^{(n)}(t, z)| \leq Nn^{-1} \), where \( N \) is independent of \( n, t, z \). This shows that system (3.12) satisfies Assumption 2.2 for all large \( n \). In addition \( \theta^{(n)} \) possesses enough smoothness in order for the results of [KZ00] to be applicable. For all large \( n \), it follows that, for any smooth bounded and nonnegative function \( c_t(y) \) on \([0, T] \times \mathbb{R}^{d_1-d} \) and any \( \varphi \in C_0^\infty(\mathbb{R}^{d}) \),

\[
E \varphi(z_T^{(n)}) \exp\left(- \int_0^T c_s(y_s^{(n)}) \, ds\right)
= E \rho_T^{(n)} \int_{\mathbb{R}^d} \varphi(x, y_T^{(n)}) \tilde{\pi}_T^{(n)}(x) \, dx \exp\left(- \int_0^T c_s(y_s^{(n)}) \, ds\right), \tag{3.13}
\]

where \( \tilde{\pi}_t^{(n)} \) is the solution of equation (2.10) corresponding to system (3.12) with initial condition \( \tilde{\pi}_0^{(n)} = \pi_0^{(n)} \) and \( \rho_t^{(n)} \) is introduced as in (3.4) on the basis of (3.12):

\[
\rho_t^{(n)} = \exp\left(- \int_0^t \tilde{\beta}_s^{(n)} \, ds\right),
\]

\[
\tilde{\pi}_t^{(n)} = \int_0^t \tilde{\Psi}_s^{(n)} \tilde{\Theta}_s^{(n)} \, dw_s,
\]

\[
\tilde{\Theta}_t^{(n)} = \tilde{\Theta}(t, y_t^{(n)}), \quad \tilde{\Psi}_t^{(n)} = \tilde{\Psi}(t, y_t^{(n)}).
\]

Later on we will also use the following notation for other coefficients of equation (2.10) corresponding to system (3.12). Introduce

\[
a_t^{(n)}(x) = \frac{1}{2} \theta^{(n)} \theta^{(n)*}(t, x, y_t^{(n)}), \quad b_t^{(n)}(x) = b(t, x, y_t^{(n)}),
\]

\[
\sigma_t^{(n)}(x) = \theta^{(n)}(t, x, y_t^{(n)}) \Theta_t^{(n)*} \Psi_t^{(n)}.
\]

Since we know that \( \tilde{\pi}_t^{(n)} \geq 0 \), it follows from the validity of (3.13) for all \( \varphi \in C_0^\infty(\mathbb{R}^{d}) \), that it is also valid for all Borel nonnegative or bounded \( \varphi \). In particular, for any \( f \in C_0^\infty(\mathbb{R}^d) \) (independent of \( y \)) we have

\[
E f(x_T^{(n)}) \exp\left(- \int_0^T c_s(y_s^{(n)}) \, ds\right)
= E \rho_T^{(n)} \int_{\mathbb{R}^d} f(x) \tilde{\pi}_T^{(n)}(x) \, dx \exp\left(- \int_0^T c_s(y_s^{(n)}) \, ds\right). \tag{3.14}
\]

Our next step is to pass to the limit in (3.14) as \( n \to \infty \). It is a standard fact that for any \( m > 0 \)

\[
\lim_{n \to \infty} E \sup_{t \leq T} |z_t^{(n)} - z_t|^m = 0, \tag{3.15}
\]
which, in particular, implies that the left-hand sides of (3.14) tend to
\[ Ef(x_T) \exp\left(- \int_0^T c_s(y_s) \, ds \right). \]

Furthermore, the process \( \rho_t^{(n)} \) is the solution of the linear equation
\[ d\rho_t^{(n)} = -\rho_t^{(n)} \gamma_t^{(n)} \, dw_t, \]
with initial condition \( \rho_0^{(n)} = 1 \), where
\[ \gamma_t^{(n)} = \Psi(t, y_t^{(n)}) B(t, z_t^{(n)}) \Psi(t, y_t^{(n)}) \Theta(t, y_t^{(n)}). \]
Also introduce
\[ \gamma_t = \Psi(t, y_t) B(t, z_t) \Psi(t, y_t) \Theta(t, y_t) \]
and observe that the processes \( \gamma_t^{(n)} \) and \( \gamma_t \) are bounded.

Furthermore, it follows from (3.15) that for any \( m > 0 \)
\[ \lim_{n \to \infty} E \sup_{t \leq T} |\gamma_t^{(n)} - \gamma_t|^m = 0, \]
which in turn implies that
\[ \lim_{n \to \infty} E \sup_{t \leq T} |\rho_t^{(n)} - \rho_t|^m = 0, \]
where \( \rho_t \) is the solution of the equation \( d\rho_t = -\rho_t \gamma_t \, dw_t \) with initial condition \( \rho_0 = 1 \) and is given in (3.4).

To investigate the limit of the remaining factor on the right in (3.14) we will use Theorem 5.2. By the well-known properties of convolutions
\[ \|\pi_0^{(n)}\|_{H^{1-2/p}}^p \leq \|\pi_0\|_{H^{1-2/p}}^p, \quad \lim_{n \to \infty} E\|\pi_0^{(n)} - \pi_0\|_{H^{1-2/p}}^p = 0. \]
This and Remark 3.1 show that the assumption of Theorem 5.2 regarding the convergence of the initial data for \( \bar{\pi}_t^{(n)} \) and \( \bar{\pi}_t \) is satisfied. Furthermore, there are no free terms in filtering equations. Therefore, it only remains to check the appropriate convergence of the coefficients. Theorem 5.2 requires the following convergences in measure \( P(d\omega)dt \, dx \) to hold on \( \Omega \times [0, T] \times \mathbb{R}^d \):
\[
\begin{align*}
\alpha_t^{(n)}(x) &\to \alpha_t(x), \quad \beta_t^{(n)}(x) \to \beta_t(x), \quad D_i \alpha_t^{(n)ij}(x) \to D_i \alpha_t^{ij}(x), \\
\sigma_t^{(n)}(x) &\to \sigma_t(x), \quad \beta_t^{(n)}(x) \to \beta_t(x), \quad D_i \sigma_t^{(n)ik}(x) \to D_i \sigma_t^{ik}(x).
\end{align*}
\]
Relation (3.15) and the assumption that the coefficients of system (1.1) are Lipschitz continuous show that, actually, apart from cases involving the derivatives of \( a \) and \( \sigma \) all the remaining convergences hold uniformly in \((t,x)\) almost surely. It is easy to see that in order to take care of the terms with derivatives it suffices to check that
\[ D_i \theta_t^{(n)}(t, x, y_t^{(n)}) \to D_i \theta(t, x, y_t) \quad (3.16) \]
in measure for any \( i = 1, \ldots, d \). Observe that by the well known properties of convolutions
\[ D_i \theta_t^{(n)}(t, x, y) \to D_i \theta(t, x, y) \]
for almost all \((t,x,y)\). Therefore, applying Lemma 3.2 shows that (3.16) holds.

Now by Theorem 5.2 and Hölder’s inequality we conclude

\[
\lim_{n \to \infty} E \left| \int_{\mathbb{R}^d} f(x) \bar{\pi}_T^n(x) \, dx - \int_{\mathbb{R}^d} f(x) \bar{\pi}_T(x) \, dx \right|^p = 0.
\]

This along with the above investigation of other terms in (3.14) yields after letting \(n \to \infty\) that

\[
Ef(x_T) \exp(-\int_0^T c_s(y_s) \, ds) = E\rho_T(\bar{\pi}_T, f) \exp(-\int_0^T c_s(y_s) \, ds).
\]

The arbitrariness of \(c\) leads to

\[
E(f(x_T) \mid \mathcal{F}_T^y) = E(\rho_T(\bar{\pi}_T, f) \mid \mathcal{F}_T^y), \quad \text{(a.s.)}
\]

which combined with the \(\mathcal{F}_T^y\)-measurability of \(\bar{\pi}_T\) (Lemma 3.3) shows that

\[
E(f(x_T) \mid \mathcal{F}_T^y) = (\bar{\pi}_T, f)E(\rho_T \mid \mathcal{F}_T^y) \quad \text{(a.s.)} \tag{3.18}
\]

Observe that on the set of \(\omega\) where

\[
E(\rho_T \mid \mathcal{F}_T^y) = 0 \tag{3.19}
\]

we have (a.s.)

\[
E(f(x_T) \mid \mathcal{F}_T^y) = 0.
\]

The arbitrariness of \(f\) shows that on the said set (a.s.)

\[
1 = E(1 \mid \mathcal{F}_T^y) = 0
\]

and consequently (3.19) can only happen with probability zero.

Furthermore, by Theorem 5.4 we have \(\bar{\pi}_t \geq 0\). A standard measure-theoretic argument then shows that (3.18) holds for all nonnegative Borel \(f\) rather than only for \(f \in C_0^\infty(\mathbb{R}^d)\). By taking \(f \equiv 1\) we see that

\[
1 = (\bar{\pi}_T, 1)E(\rho_T \mid \mathcal{F}_T^y) \quad \text{(a.s.)}
\]

implying that

\[
\infty > (\bar{\pi}_T, 1) > 0, \quad E(\rho_T \mid \mathcal{F}_T^y) = (\bar{\pi}_T, 1)^{-1} \quad \text{(a.s.)} \tag{3.20}
\]

Coming back to (3.18) we conclude

\[
E[f(x_T) \mid \mathcal{F}_T^y] = \frac{(\bar{\pi}_T, f)}{(\bar{\pi}_T, 1)} \quad \text{(a.s.)}
\]

for any nonnegative and any bounded Borel \(f\) as well. Obviously, one can replace here \(T\) with any \(t \in [0,T]\) and to prove Theorem 2.6 it only remains to show that (a.s.) relation (2.12) holds for all \(t \in [0,T]\).

The second inequality in (2.12) holds due to Lemma 3.4. To prove the first one it only remains to observe that by the above for each particular \(t \in [0,T]\) with probability one

\[
\int_{\mathbb{R}^d} \bar{\pi}_t^p(x) \, dx > 0
\]
and by Theorem 5.5 the above integral is continuous in $t$ with probability one. The theorem is proved.

4. Proof of Theorem 2.8

To prove (i) we first show that the right-hand sides of (3.10) converge as $n \to \infty$ uniformly in $t \in [0,T]$ in probability to the right-hand side of (2.15). Owing to (3.8) and (3.6)

$$
\int_0^T |(\bar{\pi}_s, L_s \varphi_m)| \, ds \leq NTm^{-1} \sup_{s \in [0,T]} (\bar{\pi}_s, 1) \to 0 \quad \text{(a.s.)},
$$

where $N$ is the constant from (3.8). Similarly one takes care of the term with $ds$ containing the derivatives of $\varphi_m$ in the second integral on the right in (3.10). Observing that by the dominated convergence theorem and again by (3.6)

$$
\int_0^T |(\bar{\pi}_s, |\beta_k^s| |\varphi_m - 1|)| \, ds \to 0 \quad \text{(a.s.)},
$$

we conclude that the usual integrals on the right-hand sides of (3.10) converge as $n \to \infty$ uniformly in $t \in [0,T]$ to the usual integral the right-hand side of (2.15) almost surely.

To show the convergence of the stochastic integrals in (3.10) to the stochastic integral in (2.15) uniform in probability it suffices (and is necessary) to show that the quadratic variation of the differences converges to zero in probability. The said quadratic variation is obviously less than a constant times

$$
\sum_k \int_0^T (\bar{\pi}_s, \Lambda^k(\varphi_m - 1))^2 \, ds,
$$

which tends to zero (a.s.) by the same reasons as above. Thus, indeed the right-hand sides of (3.10) converge as $n \to \infty$ uniformly in $t \in [0,T]$ in probability to the right-hand side of (2.15). The left-hand sides converge for all $t \in [0,T]$ (a.s.) by the monotone convergence theorem. This proves (i).

Assertion (ii) easily follows from the continuity of $(\bar{\pi}_t, 1)$, the continuity of $\bar{\pi}_t$ as an $L^p$-valued process, and Scheffé’s lemma.

In (iii) that $P_t[\beta]$ is bounded follows from the boundedness of $\beta$. The stated measurability properties of $P_t[\beta]$ are obtained by a standard measure-theoretic argument form the fact that if $f(t, x, y) = \alpha(t) \beta(x) \gamma(y)$, where $\alpha, \beta, \gamma$ are smooth functions with compact support, then

$$
\int_{\mathbb{R}^d} f(t, x, y_t) \bar{\pi}_t(x) \, dx = \alpha(t) \gamma(y_t) \int_{\mathbb{R}^d} \beta(x) \bar{\pi}_t(x) \, dx
$$

possesses the measurability properties in (iii) since the last factor is a continuous (a.s.) $\mathcal{F}_t^y$-adapted process.
To prove (2.16) it suffices to use (2.13) which implies that for each \( t \in [0, T] \) and \( y \in \mathbb{R}^{d-1} \)

\[
E(B(t, x, y) \mid \mathcal{F}_t^y) = (\bar{\pi}_t, 1)^{-1} \int_{\mathbb{R}^d} B(t, x, y) \bar{\pi}_t(x) \, dx \quad \text{(a.s.)}
\]

and then plug in here \( y_t \) in place of \( y \) in the argument of \( B \), which is possible because \( B(t, x, y) \) is Lipschitz in \( y \) (even in \( (x, y) \)). This finishes proving assertion (iii).

In (iv) the fact that \( \tilde{w}_t \) is \( \mathcal{F}_t^y \)-measurable easily follows from an equivalent formula for \( \tilde{w}_t \):

\[
\tilde{w}_t = \int_0^t \Psi(s, y_s) \, dy_s - \int_0^t P_s[\beta] \, ds,
\]

where all terms on the right are \( \mathcal{F}_t^y \)-measurable. Furthermore, \( \tilde{w}_t \) turns out to be an \( \mathcal{F}_t^y \)-martingale on \([0, T]\). To check this, take any \( \mathcal{F}_t^y \)-stopping time \( \tau \leq T \) and notice that \( \tau \) is also an \( \mathcal{F}_t \)-stopping time, so that

\[
E \tilde{w}_\tau = E \int_0^\tau (\beta_t(x_t) - P_t[\beta]) \, dt.
\]

By using (2.10) and the fact that, by definition, \( \{t < \tau\} \in \mathcal{F}_t^y \) we see that the right-hand side equals

\[
E \int_0^T I_{t<\tau}(\beta_t(x_t) - P_t[\beta]) \, dt = \int_0^T EI_{t<\tau} \beta_t(x_t) \, dt - \int_0^T EI_{t<\tau} P_t[\beta] \, dt
\]

\[= \int_0^T EI_{t<\tau} \beta_t(x_t) \, dt - \int_0^T EI_{t<\tau} (E(\beta_t(x_t) \mid \mathcal{F}_t^y)) \, dt = 0.
\]

Thus, \( E \tilde{w}_\tau = 0 \) for any \( \mathcal{F}_t^y \)-stopping time \( \tau \leq T \) which combined with the \( \mathcal{F}_t^y \)-adaptedness of \( \tilde{w}_t \) and its continuity in \( t \) is well known to be equivalent to saying that \( \tilde{w}_t \) is an \( \mathcal{F}_t^y \)-martingale on \([0, T]\). Its quadratic variation can be evaluated as the limit of sums of products of increments and is, obviously, equal to the quadratic variation of \( \tilde{w}_t \), which, as we have seen in the proof of Lemma 3.3, is a Wiener process. Therefore, the quadratic variation of \( \tilde{w}_t \) is that of a Wiener process and by Lévy’s theorem \( \tilde{w}_t \) is itself a Wiener process with respect to \( \mathcal{F}_t^y \). This proves assertion (iv).

In (v) inequality (2.18) follows from (2.17), the fact that \( \beta \) is bounded, and the well-known properties of exponential martingales. To prove (2.17) observe that (2.15) in terms of \( P_t[\beta] \) and \( \tilde{w}_t^k \) is rewritten as

\[
d(\bar{\pi}_t, 1) = (\bar{\pi}_t, \beta_t^k(x_t)) \beta_t^k(x_t) \, dt + (\bar{\pi}_t, \beta_t^k) \, d\tilde{w}_t^k
\]

\[= (\bar{\pi}_t, 1)P_t[\beta^k] \beta_t^k(x_t) \, dt + (\bar{\pi}_t, 1)P_t[\beta] \, d\tilde{w}_t^k
\]

\[= (\bar{\pi}_t, 1)|P_t[\beta]|^2 \, dt + (\bar{\pi}_t, 1)P_t[\beta] \, d\tilde{w}_t^k.
\]

Hence, \( (\bar{\pi}_t, 1) \) satisfies the linear equation

\[
d(\bar{\pi}_t, 1) = (\bar{\pi}_t, 1)|P_t[\beta]|^2 \, dt + (\bar{\pi}_t, 1)P_t[\beta] \, d\tilde{w}_t^k,
\]

the unique solution of which with initial data \((\bar{\pi}_0, 1) = (\pi_0, 1) = 1\) is known to be given by (2.17). The theorem is proved.
5. Appendix

The setting in this section is somewhat different from that of Section 1. Let \((\Omega, \mathcal{F}, P)\) be a complete probability space with an increasing filtration \(\{\mathcal{F}_t, t \geq 0\}\) of complete with respect to \((\mathcal{F}, P)\) \(\sigma\)-fields \(\mathcal{F}_t \subset \mathcal{F}\). Denote \(\mathcal{P}\) the predictable \(\sigma\)-field in \(\Omega \times (0, \infty)\) associated with \(\{\mathcal{F}_t\}\). Let \(u_t^k, k = 1, 2, \ldots, \) be independent one-dimensional Wiener processes with respect to \(\{\mathcal{F}_t\}\).

We take a stopping time \(\tau\) and for \(t \leq \tau\) we are considering the following equation in \(\mathbb{R}^d\)

\[
du_t = (L_t u_t - \lambda u_t + D_t f_t^i + f_t^0) \, dt + (\Lambda_t^k u_t + \gamma_t^k) \, dw^k_t, \tag{5.1}
\]

where \(u_t = u_t(x) = u_t(\omega, x)\) is an unknown function,

\[
L_t \psi(x) = D_j (a_{ij}^j(x) D_i \psi(x) + a_i^j(x) \psi(x)) + b_i^j(x) D_i \psi(x) + c_i(x) \psi(x),
\]

\[
\Lambda_t^k \psi(x) = \sigma_{ik}^j(x) D_i \psi(x) + \nu_i^k(x) \psi(x),
\]

the summation convention with respect to \(i, j = 1, \ldots, d\) and \(k = 1, 2, \ldots\) is enforced and detailed assumptions on the coefficients and the free terms will be given later.

Fix a number

\[ p \geq 2 \]

and denote \(L_p = L_p(\mathbb{R}^d)\). We use the same notation \(L_p\) for vector- and matrix-valued or else \(\ell_2\)-valued functions such as \(g_t = (g_t^k)\) in (5.1). For instance, if \(u(x) = (u^1(x), u^2(x), \ldots)\) is an \(\ell_2\)-valued measurable function on \(\mathbb{R}^d\), then

\[
\|u\|_{L_p}^p = \int_{\mathbb{R}^d} |u(x)|^p \, dx = \int_{\mathbb{R}^d} \left( \sum_{k=1}^{\infty} |u^k(x)|^2 \right)^{p/2} \, dx.
\]

As above

\[
D_i = \frac{\partial}{\partial x^i}, \quad i = 1, \ldots, d, \quad \Delta = D_1^2 + \ldots + D_d^2.
\]

By \(Du\) and \(D^2u\) we mean the gradient and the matrix of second order derivatives with respect to \(x\) of a function \(u\) on \(\mathbb{R}^d\).

As above, for \(\gamma \in \mathbb{R}\) by \(H_\gamma^0 = (1 - \gamma/2) L_p\) we denote the space of Bessel potentials. Observe a slight change of notation. Since we will always be dealing with \(\mathbb{R}^d\) we drop this symbol in the notation like \(H_\gamma^0(\mathbb{R}^d)\). Most often in this appendix we will use \(H_\gamma^0\) for \(\gamma = 0, 1\) and use (2.5).

If \(\tau\) is a stopping time, then

\[
\mathbb{H}_p^\gamma(\tau) := L_p([0, \tau], \mathcal{P}, H_\gamma^0), \quad \mathbb{I}_p(\tau) = \mathbb{H}_p^0(\tau).
\]

We also need the space \(\mathcal{H}_p^1(\tau)\), which is the space of functions \(u_t = u_t(\omega, \cdot)\) on \(\{(\omega, t) : 0 \leq t \leq \tau, t < \infty\}\) with values in the space of generalized functions on \(\mathbb{R}^d\) having the following properties:

(i) For any \(T \in [0, \infty)\), we have \(u \in \mathbb{H}_p^1(\tau \wedge T)\) and \(u_0 \in L_p(\Omega, \mathcal{F}_0, L_p)\).
(ii) There exist \( f^i \in L_p(\tau), \; i = 0, \ldots, d \) and \( g = (g^1, g^2, \ldots) \in \mathbb{L}_p(\tau) \) such that for any \( \varphi \in C_0^\infty \) with probability 1 for all finite \( t \leq \tau \) we have

\[
(u_t, \varphi) = (u_0, \varphi) + \int_0^t (- (f^i_s, D_i \varphi) + (f^0_s, \varphi)) \, ds + \sum_{k=1}^\infty \int_0^t (g^k_s, \varphi) \, dw^k_s. \tag{5.2}
\]

The reader can find in \([K99]\) a discussion of (i) and (ii), in particular, the fact that the series in (5.2) converges uniformly in probability on every finite subinterval of \([0, \tau)\). On the other hand, it is worth saying that the above introduced space \( H^1_p(\tau) \) are not quite the same as in \([K99]\). There are three differences. One is that there is a restriction on \( u_0 \) in \([K99]\). However the most important spaces are \( H^1_p(\tau) \) which are defined as the subsets of \( H^1_p(\tau) \) consisting of functions with \( u_0 = 0 \). All other elements of \( H^1_p(\tau) \) are obtained by adding to an element of \( H^1_p(\tau) \) an appropriate continuation for \( t > 0 \) of the initial data. Another issue is that in \([K99]\) we have \( f^i = 0, \; i = 1, \ldots, d, \) and \( f^0 \in H^{-1}_p(\tau) \). Actually, this difference is fictitious because one knows that any \( f \in H^{-1}_p \)

(a) has the form \( D_i f^i + f^0 \) with \( f^i \in L_p \) and

\[
\|f\|_{H^{-1}_p} \leq N \sum_{j=0}^d \|f^j\|_{L_p},
\]

where \( N \) is independent of \( f, f^j \), and on the other hand,

(b) for any \( f \in H^{-1}_p \) there exist \( f^j \in L_p \) such that \( f = D_i f^i + f^0 \) and

\[
\sum_{j=0}^d \|f^j\|_{L_p} \leq N\|f\|_{H^{-1}_p},
\]

where \( N \) is independent of \( f \).

The third difference is that instead of (i) we require \( D^2 u \in \mathbb{H}_p^{-1}(\tau) \) in \([K99]\). However, as it follows from Theorem 3.7 of \([K99]\) and the boundedness of the operator \( D : L_p \rightarrow H_p^{-1} \), this difference disappears if \( \tau \) is a bounded stopping time.

To summarize, the spaces \( H^1_{p,0}(\tau) \) introduced above and in \([K99]\) coincide if \( \tau \) is bounded and we choose a particular representation of the deterministic part of the stochastic differential just for convenience.

In case that property (ii) holds, we write

\[
d u_t = (D_i f^i_t + f^0_t) \, dt + g^k_t \, dw^k_t \tag{5.3}
\]

for \( t \leq \tau \) and this explains the sense in which equation (5.1) is understood. Of course, we still need to specify appropriate assumptions on the coefficients and the free terms in (5.1). Before we go to these assumptions we remind the reader that according to \([K99]\) and the above discussion, for bounded \( \tau \),
one introduces a norm in $H^1_{p,0}(\tau)$ by

$$
\|u\|_{H^1_{p,0}(\tau)} = E \int_0^\tau \left( \sum_{j=1}^d \|D_j u_t\|_{L_p}^p + \sum_{j=0}^d \|f^j_t\|_{L_p}^p + \|g_t\|_{L_p}^p \right) dt
$$

if $u$ satisfies (5.3). By identifying two elements of $H^1_{p,0}(\tau)$ if their difference has a zero $H^1_{p,0}(\tau)$-norm, one obtains a Banach space (see [K99]).

We will also identify two elements $u', u'' \in H^1_p(\tau)$ if and only if the difference $u' - u''$ is in $H^1_{p,0}(\tau)$ and equals zero.

**Assumption 5.1.** (i) The coefficients $a_t^{ij}, a_t^i, b_t^i, \sigma_t^{ik}, c_t,$ and $\nu_t^k$ are measurable with respect to $\mathcal{P} \times B(\mathbb{R}^d)$, where $B(\mathbb{R}^d)$ is the Borel $\sigma$-field on $\mathbb{R}^d$.

(ii) There is a constant $K$ such that for all values of indices and arguments

$$
|a_t^i| + |b_t^i| + |c_t| + |\nu|_{L_2} \leq K, \quad c_t \leq 0.
$$

(iii) There is a constant $\delta > 0$ such that for all values of the arguments and $\xi \in \mathbb{R}^d$

$$
(a_t^{ij} - a_t^{ij})\xi^i \xi^j \geq \delta |\xi|^2, \quad |a_t^{ij}| \leq \delta^{-1},
$$

(5.4)

where $a_t^{ij} = (1/2)(\sigma^i, \sigma^j)_{L_2}$. Finally, the constant $\lambda \geq 0$.

Assumption 5.1 (i) guarantees that equation (5.1) makes perfect sense for any constant $\lambda$ if $u \in H^1_{p}(\tau)$. By the way, adding the term $-\lambda u_t$ with constant $\lambda \geq 0$ is one more technically convenient step. One can always introduce this term, if originally it is absent, by considering $v_t := u e^{\lambda t}$.

**Assumption 5.2.** There is a continuous function $\kappa(\varepsilon)$ defined for $\varepsilon \geq 0$ such that $\kappa(0) = 0$ and

$$
|\sigma_t^i(x) - \sigma_t^i(y)|_{L_2} + |a_t^{ij}(x) - a_t^{ij}(y)| \leq \kappa(|x - y|)
$$

for all $i, j, t, x, y$.

Here are the main results used in the previous sections concerning (5.1). They are taken from [Ki04] and [K09]. Generalization of these results to the case of VMO coefficients $a_t^{ij}$ can be found in [K09].

**Theorem 5.1.** Let $\lambda \geq 0$, let $\tau$ be a stopping time, let $f^j, g \in L_p(\tau)$, and let $\psi$ be a function such that $\psi \in H^1_p(\tau) \cap H^1_p(\tau)$. Then equation (5.1) on $[0, \tau)$ has a unique solution $u \in H^1_p(\tau)$ such that $u_0 = \psi_0$.

Write

$$
d\psi_t = (D_t \alpha_t^i + \alpha_t^0) dt + \beta_t^k dw_t^k.
$$

Then the above solution $u$ satisfies

$$
\lambda^{1/2} \|u\|_{L_p(\tau)} + \|D u\|_{L_p(\tau)}
$$

$$
\leq N \left( \sum_{i=1}^d \|f^i\|_{L_p(\tau)} + \|g\|_{L_p(\tau)} + \sum_{i=1}^d \|\alpha^i\|_{L_p(\tau)} + \|\beta\|_{L_p(\tau)} + \|\psi\|_{H^1_p(\tau)} \right)
$$
\[ + N \lambda^{-1/2} \left( \| f^0 \|_{L_p(\tau)} + \| \alpha^0 \|_{L_p(\tau)} + \| \psi \|_{\mathbb{H}_1^1(\tau)} \right) + N \lambda^{1/2} \| \psi \|_{L_p(\tau)}, \]  

provided that \( \lambda > \lambda_0 \), where the constants \( N, \lambda_0 \geq 0 \) depend only on \( d, p, K, \delta \), and the function \( \kappa \).

Observe that estimate (5.5) shows a good reason for writing the free term in (5.1) in the form \( D_i f^i + f^0 \), because \( f^i, i = 1, \ldots, d \), and \( f^0 \) enter (5.5) differently.

Here is a result about continuous dependence of solutions on the data.

**Theorem 5.2.** Assume that for each \( n = 1, 2, \ldots \) we are given functions \( a_{iij}^n, a_{ii}^n, b_{ii}^n, c_{ii}^n, \sigma_{iik}^n, \nu_{i}^n, f_{ii}^n, g_{ik}^n \), and \( \psi^n \) having the same meaning and satisfying the same assumptions with the same \( \delta, K, \kappa \) as the original ones. Assume that

\[ (a_{iij}^n, a_{ii}^n, b_{ii}^n, c_{ii}^n) \rightarrow (a_{iij}, a_{ii}, b_{ii}, c_{ii}), \]

\[ |\sigma_{iik}^n - \sigma_{iik}|_{\ell_2} + |\nu_{i}^n - \nu_{i}|_{\ell_2} \rightarrow 0 \]

as \( n \rightarrow \infty \) in measure \( P(d\omega)dt \). Also let

\[ d\psi_t^n = (D_i \alpha_{ii}^n + \alpha_{ii}^n) dt + \beta_{ik}^n d\omega_t^k \]

and assume that for a stopping time \( \tau \)

\[ \sum_{j=0}^d \left( \| f^{n_j} - f^j \|_{L_p(\tau)} + \| \alpha^{n_j} - \alpha^j \|_{L_p(\tau)} \right) \]

\[ + \| g^n - g \|_{L_p(\tau)} + \| \beta^n - \beta \|_{L_p(\tau)} + \| \psi^n - \psi \|_{\mathbb{H}_1^1(\tau)} \rightarrow 0 \]

as \( n \rightarrow \infty \). Take \( \lambda \geq \lambda_0 \), take the function \( u \) from Theorem 5.1 and let \( u^n \) be unique solutions of equations (5.1) constructed from \( a_{iij}^n, a_{ii}^n, b_{ii}^n, c_{ii}^n, \sigma_{iik}^n, \nu_{i}^n, f_{ii}^n, g_{ik}^n \), and \( \psi^n \) and having initial values \( \psi_0^n \).

Then for any finite \( T \geq 0 \) we have

\[ \| u^n - u \|_{\mathbb{H}_1^1(\tau \wedge T)} \rightarrow 0, \quad E \sup_{t \leq \tau \wedge T} \| u^n_t - u_t \|_{L_p} \rightarrow 0 \]

as \( n \rightarrow \infty \).

The following result shows that the solution does not depend on \( p \).

**Theorem 5.3.** Let \( p_1, p_2 \in [2, \infty) \) and let the assumptions of Theorem 5.1 be satisfied with \( p = p_1 \) and \( p = p_2 \). Then the solutions corresponding to \( p = p_1 \) and \( p = p_2 \) coincide, that is there is a unique solution \( u \in \mathcal{H}_p^1(\tau) \cap \mathcal{H}_p^1(\tau) \) of equation (5.1) with initial data \( \psi_0 \).

In many situation the following maximum principle is useful.

**Theorem 5.4.** Under the assumptions of Theorem 5.1 suppose that \( \psi_0 \geq 0, f^i = 0, i = 1, \ldots, d, f^0 \geq 0, g = 0 \). Then for the solution \( u \) almost surely we have \( u_t \geq 0 \) for all finite \( t \leq \tau \).
Finally, we used the following embedding theorem (see Corollary 4.12 and Remark 4.14 of [K01]). For \( p \in (0, 1) \), a Banach space \( X \), and a set \( A \subset \mathbb{R}^d \) by \( C^\kappa(A,X) \) we mean Hölder’s space of continuous \( X \)-valued functions on \( A \) with finite norm \( \| \cdot \|_{C^\kappa(A,X)} \) defined by

\[
\| u \|_{C^\kappa(A,X)} = \sup_{s,t \in A} |t - s|^{-\kappa} |u(t) - u(s)|_X, \quad \| u \|_{C(A,X)} = \sup_{t \in A} |u(t)|_X, 
\]

\[
\| u \|_{C^\kappa(A,X)} = \| u \|_{C^\kappa(A,X)} + \| u \|_{C(A,X)}. 
\]

**Theorem 5.5.** Let \( \tau \leq T \), where the constant \( T \in (0, \infty) \) and let \( u \in \mathcal{H}_p^1(\tau) \) satisfy (5.3) with \( f^i \in L_p(\tau), g \in L_p(\tau), \) and \( u_0 \in L_p(\Omega, \mathcal{F}_0, H_p^{1-2/p}) \), Then:

(i) Almost surely \( u_t \) is a continuous function of \( t \) with values in \( L_p \) for all \( t \in [0, \tau] \).

(ii) (case \( p > 2 \)) Assume that for some numbers \( \alpha \) and \( \beta \) we have

\[ \frac{2}{p} < \alpha < \beta \leq 1. \]

Then, for any \( a > 0 \),

\[ E[u]^p \leq N T^{(\beta - \alpha)/p} a^{\beta - 1} I(a), \tag{5.6} \]

\[ E\| u \|^p_{C^{(\alpha/1 - 1/p)}([0, \tau], H_p^{1-\beta})} \leq N E\| u_0 \|^p_{H_p^{1-\beta}} + N T^{p/2 - 1/2} a^{\beta - 1} I(a), \tag{5.7} \]

where the constants \( N \) are independent of \( a, \tau, T, \) and \( u \) and

\[ I(a) := a\| u \|^p_{H_p^1(\tau)} + a^{-1} \| D_i f^i + f^0 \|^p_{H_p^{1-1}(\tau)} + \| g \|^p_{L_p(\tau)}. \]

In particular, if \( p(1 - \beta) > d \), then

\[ E \sup_x |u(\cdot, x)|^p_{C^{1/2 - 1/p}([0, \tau], H_p^{1-\beta})} \leq N T^{(\beta - \alpha)/p} a^{\beta - 1} I(a), \tag{5.8} \]

\[ E \sup_{t \in [0, T]} \| u(t, \cdot) \|^p_{C^{1-\beta - d/p}([0, \tau], H_p^{1-\beta})} \leq N E\| u(0) \|^p_{H_p^{1-\beta}} + N T^{p/2 - 1/2} a^{\beta - 1} I(a). \tag{5.9} \]

Finally, (5.7) also holds if \( p = 2 \) and \( \beta = 1 \).

It is probably worth saying that (5.8) and (5.9) are not stated in [K01]. These are just obvious consequences of (5.6) and (5.7) and the embedding theorem: \( H_p^\gamma \subset C^{\gamma - d/p} \) if \( \gamma - d/p > 0 \) and \( \gamma - d/p \) is not an integer.

**References**

[FKK] M. Fujisaki, G. Kallianpur, and H. Kunita, *Stochastic differential equations for the non linear filtering problem*, Osaka J. Math., Vol. 9 (1972), 19-40.

[Ki04] Kyeong-Hun Kim, *On \( L_p \)-theory of stochastic partial differential equations of divergence form in \( C^1 \) domains*, Probab. Theory Related Fields, Vol. 130 (2004), No. 4, 473-492.

[K77] N.V. Krylov, “Controlled diffusion processes”, Nauka, Moscow, 1977 in Russian; English translation: Springer, 1980.

[K79] N.V. Krylov, *On the equivalence of \( \sigma \)-algebras in the filtering problem of diffusion processes*, Teor. Verojatnost. i Primenen, Vol. 24 (1979), No. 4, 771-780 in Russian; English translation: Theor. Probability Appl., Vol. 24 (1980), No. 4, 772-781.
N.V. KRYLOV

[K99] N.V. Krylov, *An analytic approach to SPDEs*, pp. 185-242 in Stochastic Partial Differential Equations: Six Perspectives, Mathematical Surveys and Monographs, Vol. 64, AMS, Providence, RI, 1999.

[K01] N.V. Krylov, *Some properties of traces for stochastic and deterministic parabolic weighted Sobolev spaces*, Journal of Functional Analysis, Vol. 183, No. 1 (2001), 1-41.

[K09] N.V. Krylov, *On divergence form SPDEs with VMO coefficients*, SIAM J. Math. Anal. Vol. 40 (2009), No. 6, 2262-2285.

[KR78] N.V. Krylov and B.L. Rozovskii, *On conditional distributions of diffusion processes*, Izvestiya Akademii Nauk SSSR, seriya matematicheskaya, Vol. 42, No. 2 (1978), 356-378 in Russian; English translation in Math. USSR Izvestija, Vol. 12 (1978), No. 2, 336-356.

[KR81] N.V. Krylov and B.L. Rozovskii, *On the first integrals and Liouville equations for diffusion processes*, pp. 117-125 in “Stochastic Differential Systems, Proc. 3rd IFIP-WG 7/1 Working Conf., Visegrád, Hungary, Sept. 15-20, 1980”, Lecture Notes in Contr. Inform. Sci., Vol. 36, 1981.

[KZ00] N.V. Krylov and A. Zatezalo, *A direct approach to deriving filtering equations for diffusion processes*, Applied Mathematics and Optimization, Vol. 42, No. 3 (2000), 315-332.

[Kn82] H. Kunita, *First order stochastic partial differential equations*, Stochastic analysis (Katata/Kyoto, 1982), 249-269, North-Holland Math. Library, Vol. 32, North-Holland, Amsterdam, 1984.

[Kn97] H. Kunita, “Stochastic flows and stochastic differential equations”, Reprint of the 1990 original, Cambridge Studies in Advanced Mathematics, Vol. 24, Cambridge University Press, Cambridge, 1997.

[Ku64] H. J. Kushner, *On the differential equations satisfied by conditional probability densities of Markov processes, with applications*, J. Soc. Indust. Appl. Math. Ser. A Control, Vol. 2 (1964), 106-119.

[Ku67] H. J. Kushner, *Dynamical equations for optimal nonlinear filtering*, J. Differential Equations, Vol. 3 (1967), 179-190.

[Pa79] E. Pardoux, *Stochastic partial differential equations and filtering of diffusion processes*, Stochastics, Vol. 3 (1979), No. 2, 127-167.

[Pu84] O.G. Purtukhia, *The innovation problem for degenerate diffusion processes (growing coefficients)*, Uspekhi Mat. Nauk, Vol. 39 (1984), No. 4 (238), 177-178 in Russian; English translation: Russ. Math. Surv., Vol. 39 (1984), No. 4, 137-138.

[LS68] R.Sh. Liptser and A.N. Shiryaev, *Nonlinear filtration of diffusion Markov processes*, (Russian), Trudy Mat. Inst. Steklov, Vol. 104 (1968), 135-180; English translation: Proceedings of the Steklov Institute of Mathematics, Vol. 104 (1968), 163-218.

[LS01] R.Sh. Liptser and A.N. Shiryaev, “Statistics of random processes”, “Nauka”, Moscow, 1974 in Russian; English translation: Vols. I, II, Springer-Verlag, New York, 1977-1978, Second edition 2001.

[R90] B.L. Rozovskii, “Stochastic evolution systems”, Kluwer, Dordrecht, 1990.

[Sh66] A. N. Shiryaev, *On stochastic equations in the theory of conditional Markov process*, Teor. Verojatnost. i Primenen., Vol. 11 (1966), 200-206 in Russian; English translation: Theor. Probability Appl. 11 (1966), 179-184.

[St60] R. L. Stratonovich, *Conditional Markov processes*, Teor. Verojatnost. i Primenen., Vol. 5 (1960), 172-195 in Russian; English translation; Theor. Probability Appl., Vol. 5 (1960), 156-178.

127 Vincent Hall, University of Minnesota, Minneapolis, MN, 55455

E-mail address: krylov@math.umn.edu