The Annals of Probability
2003, Vol. 31, No. 2, 564–591

ON THE SPLITTING-UP METHOD AND STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS

BY ISTVÁN GYÖNGY[1] AND NICOLAI KRYLOV[2]

University of Edinburgh and University of Minnesota

We consider two stochastic partial differential equations

$$du_\varepsilon(t) = (L_r u_\varepsilon(t) + f_r(t))\, dV^r_{\varepsilon t} + (M_k u_\varepsilon(t) + g_k(t)) \circ dY^k_t, \qquad \varepsilon = 0, 1,$$

driven by the same multidimensional martingale $Y = (Y^k)$ and by different increasing processes $V^r_0, V^r_1, r = 1, 2, \dots, d_1$, where L_r and M^k are second- and first-order partial differential operators and $\circ$ stands for the Stratonovich differential. We estimate the moments of the supremum in t of the Sobolev norms of $u_1(t) - u_0(t)$ in terms of the supremum of the differences $|V^r_{0t} - V^r_{1t}|$. Hence, we obtain moment estimates for the error of a multistage splitting-up method for stochastic PDEs, in particular, for the equation of the unnormalized conditional density in nonlinear filtering.

1. Introduction. Stochastic partial differential equations (SPDEs) appear in many real-world applications. There are several methods of finding solutions numerically: for instance, finite difference method, Galerkin's approximation, finite element method and Wiener chaos decomposition (see, e.g., [4, 5, 8, 13, 17] and the references therein). One of the most promising methods is the splitting-up method introduced in the context of SPDEs in [1] and further developed in [2, 3, 14]. Error estimates are given in [3] and [9] in the case of the filtering equations. The methods of these papers are based on semigroup theory and, as it seems to the authors, are not extendible to the general situation of filtering equations. Here we present an approach to proving the rate of convergence for the splitting-up method, which is based on stochastic calculus and not on semigroup theory. This not only allows us to improve some results of [1–3, 9] in the direction of convergence in sup norm, but also to put forth the splitting-up method for general filtering equations.

Let us loosely describe the splitting-up method and our approach to it. In the situation of [3] the splitting-up method is stated in the following way. Assume that we are given independent one-dimensional Wiener processes w^k_t, $k = 1, \dots, d_0$, first-order operators M_k, $k = 1, \dots, d_0$, and a a second-order elliptic operator L acting on functions defined on $\mathbb{R}^d$. Let the coefficients of L and M_k be independent of time and suppose that we want to solve the equation

$$du(t,x) = Lu(t,x)\, dt + M_k u(t,x) \circ dw^k_t, \qquad x \in \mathbb{R}^d,\ t > 0, \tag{1.1}$$

Received December 2000; revised February 2002.
[1]Supported by an Ordway research grant.
[2]Supported in part by NSF Grant DMS-98-76586.
AMS 2000 subject classifications. 60H15, 65M12, 65M15, 93E11.
Key words and phrases. Stochastic partial differential equations, splitting-up.

on $[0, T]$, with some initial data $u_0 = u_0(x)$, where $\circ$ stands for the Stratonovich differential.

Let $T_n := \{t_i = iT/n : i = 0, 1, 2, \dots, n\}$ be a partition of the interval $[0, T]$ for a fixed integer $n \geq 1$. Set $\delta := T/n$ and define the approximation $u_n(t)$ for $t \in T_n$, by $u_n(0) = u_0$,

$$u_n(t_{i+1}, \cdot) = \mathbf{P}_\delta \mathbf{Q}_{t_i t_{i+1}} u_n(t_i, \cdot) \tag{1.2}$$

recursively, where $\{\mathbf{P}_t : t \geq 0\}$ and $\{\mathbf{Q}_{st} : 0 \leq s \leq t\}$ denote the solution operators corresponding to the equations

$$dv(t, x) = Lv(t, x)\, dt, \qquad v(0, x) = v(x) \tag{1.3}$$

and

$$d\tilde{v}(t, x) = M_k \tilde{v}(t, x) \circ dw_t^k, \qquad \tilde{v}(s, x) = v(x), \tag{1.4}$$

respectively. In this way the approximation of (1.1) in each interval $[t_i, t_{i+1}]$ is split into two steps: solving the degenerate SPDE (1.4) and taking its solution at time t_{i+1} as the initial value at time t_i while solving PDE (1.3) again on $[t_i, t_{i+1}]$. In [3] these steps are called *correction* and *prediction* steps, and it is proved that under appropriate conditions

$$\max_{t \in T_n} E\|u(t) - u_n(t)\|_0^2 \leq N/n^2, \tag{1.5}$$

where $\|\cdot\|_0$ is the usual L_2 norm in $\mathbb{R}^d$.

Instead of going back and forth in time, we propose to stretch out the time scale by using the time scales $A_t(n)$ and $B_t(n)$, defined by

$$A_t(n) := \begin{cases} k\delta, & \text{for } t \in \big[2k\delta, (2k+1)\delta\big), \\ t - (k+1)\delta, & \text{for } t \in \big[(2k+1)\delta, (2k+2)\delta\big), \end{cases}$$
$$B_t(n) := A_{t+\delta}(n),$$

and to consider the equation

$$dv_n(t, x) = Lv_n(t, x)\, dA_t(n) + M_k v_n(t, x) \circ dw^k_{B_t(n)}. \tag{1.6}$$

Obviously, $v_n(2t) = u_n(t)$ and $u(t) = \bar{u}_n(2t)$ for $t \in T_n$, where $\bar{u}_n := u(B_t(n), x)$ satisfies

$$d\bar{u}_n(t, x) = L\bar{u}_n(t, x)\, dB_t(n) + M_k \bar{u}_n(t, x) \circ dw^k_{B_t(n)}. \tag{1.7}$$

Equations (1.6) and (1.7) suggest and make possible using stochastic calculus to estimate $E \sup_{t \leq T} \|v_n(2t) - \bar{u}_n(2t)\|_0^p$, which gives an estimate for $E \max_{t \in T_n} \|u_n(t) - u(t)\|_0^p$. One of our results (Theorem 2.3, stated and proved in Section 2) says that for each $T > 0$ and $p > 0$, there is a constant N such that

$$E \max_{t \in T_n} \|u_n(t) - u(t)\|_0^p \le N/n^p \tag{1.8}$$

for all integers $n \ge 1$. By a straightforward modification of the proof of this estimate, we can see that it also holds for the approximation defined by $u_n(t_{i+1}) := \mathbf{Q}_{t_i t_{i+1}} \mathbf{P}_\delta u_n(t_i)$ in place of (1.2).

We thus improve (1.5) by taking the maximum inside the expectation and allowing any $p > 0$ in place of 2. Moreover, we also get estimate (1.8) in the case of time-dependent random operators L and M_k. We also do not require L to be uniformly elliptic. It can just be degenerate elliptic with smooth coefficients. Our assumptions on the smoothness of the coefficients of L and M_k are the same as in [3] when we prove (1.8). Under higher smoothness assumptions, we prove that in (1.8) one can replace the L_2 norm of $u_n(t) - u(t)$ with the H^m norm. Then, if m is large enough, the Sobolev embedding theorems provide estimates of the sup norm in x of $u_n(t) - u(t)$ and its derivatives. Thus, in particular, we estimate $u_n - u$ uniformly in $t \in T_n$ and $x \in \mathbb{R}^d$.

In the explanation of our approach to the splitting-up method, we used the Stratonovich differential in the equations above. In fact, in our results we consider more general equations than (1.1). In particular, in place of the Stratonovich differential $M_k u_k(t,x) \circ dw_t^k$ in (1.1), which is just a short notation for $\frac{1}{2} M_k M_k u(t,x)\,dt + M_k u_k(t,x)\,dw_t^k$ with the stochastic Itô differential $M_k u_k(t,x)\,dw_t^k$, we consider the more general term $L_0 u(t,x)\,dt + M_k u(t,x)\,dw_t^k$ with a second-order differential operator L_0. Correspondingly, in place of (1.4), we consider $d\tilde{v}(t,x) = L_0 \tilde{v}(t,x)\,dt + M_k \tilde{v}(t,x)\,dw_t^k$, and we assume the stochastic parabolicity (see Assumption 2.5) for this equation, which is satisfied in the special case $L_0 := \frac{1}{2} M_k M_k$ of (1.4). In this connection we note that it is well known that, in general, this equation is not solvable if the stochastic parabolicity is not satisfied (see [11]). In particular, it is not well posed when $L_0 = 0$.

We also establish a multistage splitting-up method, by which we mean the following. Assume that L in (1.1) is the sum of a finite number of elliptic operators, say $L = L_1 + L_2$, where L_1 is a second-order elliptic operator and L_2 is a first-order one. Define now the approximation u_n by

$$u_n(t_{i+1}) = \mathbf{P}_\delta^{(2)} \mathbf{P}_\delta^{(1)} \mathbf{Q}_{t_i t_{i+1}} u_n(t_i),$$

such that $v(t) := \mathbf{P}_t^{(i)} v$ denotes the solution of (1.3) with L_i in place of L. By our theorem estimate (1.8) remains valid.

The paper is organized as follows. In Section 2 we introduce our general setting but state the results only for the case of time-independent data. In this way the reader will not be overwhelmed right away with some quite technical details. In this section we also prove Theorem 2.3 on the basis of Theorem 2.1, which, in turn, is proved in Section 4, after we prepare some auxiliary facts in Section 3.

In Section 5 we generalize Theorem 2.1 for time-dependent and random coefficients, which allows us to establish the splitting-up method for general filtering equations in Section 6.

In conclusion, we introduce some notation used everywhere below. Throughout the paper d, d_0, d_1 are fixed integers, K, T are fixed finite positive constants, p is a fixed number in $(0, \infty)$ and

$$D_i = \partial/\partial x^i, \qquad D_{ij} = \partial^2/\partial x^i\, \partial x^j.$$

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and let $\mathcal{F}_t$, $t \geq 0$, be an increasing filtration of sub-σ-fields of $\mathcal{F}$, such that $\mathcal{F}_0$ is complete with respect to $(\mathcal{F}, P)$. By $\mathcal{P}$ we denote the σ-field of predictable subsets of $\Omega \times (0, \infty)$ generated by $\mathcal{F}_t$. We assume that on Ω we are given a continuous $\mathcal{F}_t$-martingale $Y_t = (Y_t^1, \dots, Y_t^{d_0})$.

We always assume the summation convention over repeated integer-valued indices.

2. The case of time-independent coefficients. For $\varepsilon = 0, 1$ and $r = 0, 1, \dots, d_1$ (notice r can be 0), let $V_{t,\varepsilon}^r$ be continuous increasing processes defined for $t \in [0, T]$. Consider the following equation:

$$\begin{aligned} du(t,x) &= \big(L_r u(t,x) + f_r(t,x)\big)\, dV_{t,\varepsilon}^r \\ &\quad + \big(M_k u(t,x) + g_k(t,x)\big)\, dY_t^k \end{aligned} \tag{2.1}$$

for $t \in (0, T]$, $x \in \mathbb{R}^d$ with initial condition $u(0,x) = u_{0\varepsilon}(x)$, where the operators L_r and M_k are written as

$$L_r = a_r^{ij}(t,x) D_{ij} + a_r^i(t,x) D_i + a_r(t,x), \qquad M_k = b_k^i(t,x) D_i + b_k(t,x).$$

To formulate our assumptions, we fix an integer $m \geq 0$.

ASSUMPTION 2.1 (Smoothness of the coefficients). All the coefficients $a_r^{ij}(t,x)$, $a_r^i(t,x)$, $a_r(t,x)$, $b_k^i(t,x)$, $b_k(t,x)$ are predictable for any $x \in \mathbb{R}^d$, and, for any $(\omega, t) \in \Omega \times (0, \infty)$, their derivatives up to order $m + 3$ exist, are continuous and by magnitude are bounded by K.

ASSUMPTION 2.2. The processes $V_{t,\varepsilon}^r$ are predictable $V_{0,\varepsilon}^r = 0$, $V_{t,\varepsilon}^0 =: V_t^0$ is independent of ε, and there is a predictable increasing process V_t such that

$$V_0 = 0, \qquad V_T \leq K, \qquad \sum_{r,\varepsilon} dV_{t,\varepsilon}^r + d\langle Y \rangle_t \leq dV_t \tag{2.2}$$

in the sense of measures on $[0, T]$.

REMARK 2.1. Actually (2.2) is always satisfied with $V_t = \sum_{r,\varepsilon} V_{t,\varepsilon}^r + \langle Y \rangle_t$, provided that this process is bounded by K on $[0, T]$. Also notice that we single out one of $V_{t,\varepsilon}^r$ with $r = 0$ in order to show later that we do not need Assumption 5.1 to be imposed on all the operators L_r.

Equation (2.1) is supposed to be parabolic in the usual stochastic sense.

ASSUMPTION 2.3. For any $\omega \in \Omega$, $\varepsilon = 0, 1$, $x, \lambda \in \mathbb{R}^d$, we have

$$2a_r^{ij}(t,x)\lambda^i\lambda^j \, dV_{t,\varepsilon}^r - b_k^i(t,x)b_l^j(t,x)\lambda^i\lambda^j \, d\langle Y^k, Y^l\rangle_t \ge 0$$

in the sense of measures on $[0,T]$. (Recall that the summation convention is used over repeated integer-valued indices and that $r = 0, 1, \dots, d_1$.)

We investigate the convergence of not only functions themselves but also of their derivatives in L_2. Therefore, we need the spaces H^n of L_2 functions whose generalized derivatives up to order n are also in L_2. There are several ways to introduce the norm and the inner product in H^n. We choose the following:

$$(u,v)_n := \sum_{|\alpha|\le n} (D^\alpha u, D^\alpha v)_0,$$

where $(\cdot\,,\cdot)_0$ is the inner product in L_2 and $\alpha = (\alpha_1, \dots, \alpha_d)$ are multi-indices,

$$|\alpha| := \alpha_1 + \dots + \alpha_d, \qquad D^\alpha := D_1^{\alpha_1} \cdots D_d^{\alpha_d}.$$

ASSUMPTION 2.4. For each $\omega \in \Omega$, the functions $f_r(t) = f_r(t,\cdot)$ are weakly continuous as H^{m+3}-valued functions, $g_k(t) = g_k(t,\cdot)$ are weakly continuous as H^{m+4}-valued functions, and the initial conditions $u_{0\varepsilon}$ satisfy $u_{0\varepsilon} \in L_2(\Omega, \mathcal{F}_0, H^{m+3})$. Furthermore, f_r and g_k are predictable, and

$$E \sup_{t\in[0,T]} \|f\|_{m+3}^p + E \sup_{t\in[0,T]} \|g\|_{m+4}^p + E\|u_0\|_{m+3}^p \le K,$$

where $\|f\|_{m+3}^2 = \sum_r \|f_r(t)\|_{m+3}^2$ and $\|g\|_{m+4}^2 = \sum_k \|g_k(t)\|_{m+4}^2$.

DEFINITION 2.1. By a solution of (2.1) with initial data u_0, we mean an L_2-valued predictable function $u(t) = u(t,\cdot)$ defined on $\Omega \times [0,T]$ such that

$$P\left(\int_0^T \|u(t)\|_1^2 \, dt < \infty\right) = 1,$$

and for any $\phi \in C_0^\infty$, the equation

$$\begin{aligned}\big(u(t,\cdot),\phi\big)_0 &= \big(u(0,\cdot),\phi\big)_0 \\ &\quad + \int_0^t \big[-\big(a_r^{ij} D_i u(s), D_j\phi\big)_0 \\ &\qquad + \big((a_r^i - a_{rx^j}^{ij}) D_i u(s) + a_r u(s) + f_r(s), \phi\big)_0\big] \, dV_{s,\varepsilon}^r \\ &\quad + \int_0^t \big(b_k^i D_i u(s) + b_k u(s) + g_k(s), \phi\big)_0 \, dY_s^k\end{aligned}$$

holds for all $t \in [0,T]$ at once with probability 1.

We know from (Itô's formula) [6] that for any solution u there exists a solution $\bar{u}$ such that $\bar{u}(t,\cdot)$ is a continuous L_2-valued function for each ω and for any $\phi \in C_0^\infty$, the equation $(u(t,\cdot),\phi)_0 = (\bar{u}(t,\cdot),\phi)$ holds for all $t \in [0,T]$ at once with probability 1. This is the reason why henceforth we only consider L_2-continuous versions of solutions.

THEOREM 2.1. *Under Assumptions* 2.1–2.4, *for* $\varepsilon = 0, 1$, (2.1) *with initial condition* $u_{0\varepsilon}$ *has a unique solution* $u_\varepsilon(t)$. *Furthermore*, $u_\varepsilon(t)$ *is weakly continuous in* H^{m+3} *for each* ω *and*

$$E \sup_{t\in[0,T]} \|u_\varepsilon(t)\|_{m+3}^p \le N,$$

where N *depends only on* d, d_0, d_1, K, p, m *and* T.

This theorem is a particular case of Theorem 3.1. The following is the basic tool of proving our estimate of convergence for the splitting-up method.

THEOREM 2.2. *Let* a_r^{ij}, a_r^i, a_r, b_k^i, b_k, b, f_r, *and* g_k *be independent of* t. *Then under Assumptions* 2.1–2.4, *there is a constant* N *depending only on* d, d_0, d_1, K, p, m *and* T, *such that*

$$E \sup_{t\in[0,T]} \|u_1(t) - u_0(t)\|_m^p \le N(E\|u_{01} - u_{00}\|_m^p + A^p), \tag{2.3}$$

where

$$A = \sup_{\omega\in\Omega} \max_{t\in[0,T]} \max_r |V_{t,1}^r - V_{t,0}^r|.$$

Theorem 2.2 is proved in Section 4. Now we give its application to the splitting-up method along the lines discussed in the Introduction. In $(0,T]\times\mathbb{R}^d$ we consider the following equation:

$$\begin{aligned} du(t,x) = \sum_{r=1}^{d_1} \big(L_r u(t,x) + f_r(t,x)\big)\,dt + \big(L_0 u(t,x) + f_0(t,x)\big)\,dV_t^0 \\ + \big(M_k u(t,x) + g_k(t,x)\big)\,dY_t^k, \end{aligned} \tag{2.4}$$

with the same operators L_r and M_k as above and initial condition $u(0,x) = u_0(x)$.

ASSUMPTION 2.5. Assumptions 2.1 and 2.4 are satisfied. The process V_t^0 is predictable continuous increasing and starting at 0. We have $V_T^0 + \langle Y\rangle_T \le K$. The matrices (a_r^{ij}) are nonnegative, and, for any $\omega\in\Omega$, $x, \lambda \in \mathbb{R}^d$, we have

$$2a_0^{ij}(t,x)\lambda^i\lambda^j\,dV_t^0 - b_k^i(t,x)b_r^j(t,x)\lambda^i\lambda^j\,d\langle Y^k, Y^r\rangle_t \ge 0$$

in the sense of measures on $[0,T]$.

By $u(t)$ we denote the unique solution of (2.4) with initial condition $u(0,x) = u_0(x)$, which exists owing to Theorem 2.1.

Next set $T_n := \{t_i := iT/n : i = 0, 1, 2, \dots, n\}$, $\delta := T/n$ for an integer $n \geq 1$, and define the approximation $u^{(n)}$, by $u^{(n)}(0) := u_0$,

$$u^{(n)}(t_{i+1}) := \mathbf{P}_\delta^{(d_1)} \cdots \mathbf{P}_\delta^{(2)} \mathbf{P}_\delta^{(1)} \mathbf{Q}_{t_i t_{i+1}} u^{(n)}(t_i), \qquad i = 0, 1, 2, \dots, n-1, \tag{2.5}$$

where $\mathbf{P}_t^{(\gamma)} \psi := v(t)$, $\gamma = 1, 2, \dots, d_1$, and $\mathbf{Q}_{st} \psi := \tilde{v}(t)$ denote the solutions of the equations

$$dv(t,x) = \big(L_\gamma v(t,x) + f_\gamma(x)\big)\,dt, \qquad t \geq 0,$$

$$d\tilde{v}(t,x) = \big(L_0 \tilde{v}(t,x) + f_0(x)\big)\,dV_t^0 + \big(M_k \tilde{v}(t,x) + g_k(x)\big)\,dY_t^k, \qquad t \geq s,$$

respectively, with initial conditions $v(0,x) = \psi(x)$ and $\tilde{v}(s,x) = \psi(x)$, respectively.

THEOREM 2.3. *Let a_r^{ij}, a_r^i, a_r, b_k^i, b_k, b, f_r and g_k be independent of t. Then under Assumption* 2.5, *there is a constant N depending only on d, d_0, d_1, K, p, m and T, such that*

$$E \max_{t \in T_n} \|u^{(n)}(t) - u(t)\|_m^p \leq N n^{-p}$$

for all $n \geq 1$.

PROOF. Set $d' := d_1 + 1$, fix an integer $n \geq 1$ and let $\delta := T/n$. According to our idea, we change time by using the following function:

$$\kappa(t) := \begin{cases} t - k d_1 \delta, & \text{for } t \in [kd'\delta, (kd'+1)\delta],\ k = 0, 1, \dots, \\ (k+1)\delta, & \text{for } t \in [(kd'+1)\delta, (k+1)d'\delta],\ k = 0, 1, \dots, \\ 0, & \text{for } t \leq 0. \end{cases}$$

Define

$$\bar{Y}^k(t) := Y_{\kappa(t)}^k, \qquad \bar{\mathcal{F}}_t = \mathcal{F}_{\kappa(t)}, \qquad \bar{V}_{t,0}^0 = \bar{V}_{t,1}^0 := V_{\kappa(t)}^0,$$

$$\bar{V}_{t,0}^r := \kappa(t), \qquad \bar{V}_{t,1}^r := \kappa(t - r\delta) \qquad \text{for } r = 1, 2, \dots, d_1.$$

Consider the equations

$$du_\varepsilon(t) = \big(L_r u_\varepsilon(t) + f_r\big)\,d\bar{V}_{t,\varepsilon}^r + \big(M_k u_\varepsilon(t) + g_k\big)\,d\bar{Y}_t^k, \qquad \varepsilon = 0, 1, \tag{2.6}$$

with $u_0(0,x) = u_1(0,x) = u_0(x)$. It is easy to see that Assumptions 2.2 and 2.3 also hold with $\bar{Y}^k$ and $\bar{V}_\varepsilon^r$ ($\varepsilon = 0, 1$) in place of Y^k and V_ε^r, respectively. Thus, by Theorem 2.1 the solutions u_0 and u_1 exist, and by virtue of Theorem 2.2, there is a constant N depending only on d, d_0, d_1, p, m, K and T, such that

$$E \sup_{t \in [0, Td']} \|u_1(t) - u_0(t)\|_m^p \leq N \sup_{t \in [0, Td']} \sup_{r \leq d_1} |\kappa(t + r\delta) - \kappa(t)|^p = N T^p n^{-p},$$

which implies the theorem, since clearly $u_0(d't) = u(t)$ and $u_1(d't) = u^{(n)}(t)$ for $t \in T_n$. $\square$

REMARK 2.2. We can define the approximation $u^{(n)}$ by splitting up in any order; that is, we can define $u^{(n)}$ by

$$u^{(n)}(t_{i+1}) := \mathbf{P}_\delta^{(d_1)} \cdots \mathbf{P}_\delta^{(l+1)} \mathbf{Q}_{t_i t_{i+1}} \mathbf{P}_\delta^{(l)} \cdots \mathbf{P}_\delta^{(2)} \mathbf{P}_\delta^{(1)} u^{(n)}(t_i)$$

in place of (2.5). Then one can easily see from its proof that Theorem 2.3 remains valid.

3. Auxiliary results. First, we consider the equation

$$du(t,x) = \big(Lu(t,x) + f(t,x)\big)\, dV_t + \big(M_k u(t,x) + g_k(t,x)\big)\, dY_t^k \tag{3.1}$$

for $t \in (0,T]$, $x \in \mathbb{R}^d$ with initial condition $u(0,x) = u_0(x)$, where $T \in (0,\infty)$ is a fixed number and the operators L and M_k are written as

$$L = a^{ij}(t,x) D_{ij} + a^i(t,x) D_i + a(t,x), \qquad M_k = b_k^i(t,x) D_i + b_k(t,x).$$

For convenience, we enumerate some further assumptions regarding (3.1). Fix an integer $m = 1, 2, \dots$ and remember that by K we denote a fixed positive constant.

ASSUMPTION 3.1 (Smoothness of the coefficients). All the coefficients $a^{ij}(t,x)$, $a^i(t,x)$, $a(t,x)$, $b_k^i(t,x)$, $b_k(t,x)$ are predictable for any $x \in \mathbb{R}^d$, and, for any $(\omega,t) \in \Omega \times (0,\infty)$, their derivatives up to order m and for a^{ij} up to order $2 \vee m$ exist, are continuous and by magnitude are bounded by K.

ASSUMPTION 3.2 [Stochastic parabolicity of (3.1)]. The process V_t is increasing, continuous, predictable, $V_0 = 0$, and $V_T \le K$. We have $d\langle Y\rangle_t \le dV_t$ and for any $x, \lambda \in \mathbb{R}^d$, in the sense of measures on $[0,T]$,

$$2a^{ij}\lambda^i\lambda^j\, dV_t - b_k^i b_r^j \lambda^i \lambda^j\, d\langle Y^k, Y^r\rangle_t \ge 0.$$

ASSUMPTION 3.3. In (3.1) the function f is predictable H^m valued, g_k are predictable H^{m+1} valued, u_0 is H^m valued and $\mathcal{F}_0$ measurable. Furthermore, for $l \le m$ and

$$K_l(t) := \int_0^t \{\|f(s)\|_l^2 + \|g(s)\|_{l+1}^2\}\, dV_s,$$

where $f(s) = f(s,\cdot)$, $g(s) = g(s,\cdot)$ and $\|g(s)\|_{l+1}^2 := \sum_k \|g^k(s)\|_{l+1}^2$, we have

$$E\|u_0\|_m^p + EK_m^{p/2}(T) < \infty.$$

Solutions of (3.1) are always understood according to Definition 2.1.

THEOREM 3.1. *Under Assumptions* 3.1–3.3 *there exists a unique solution of* (3.1) *with initial condition* u_0. *In addition*, $u(t)$ *is weakly continuous in* H^m *for each* ω *and, for any integer* $l \in [0,m]$,

$$E \sup_{t\in[0,T]} \|u(t)\|_l^p \le NE\|u_0\|_l^p + NEK_l^{p/2}(T), \tag{3.2}$$

where N *depends only on* d, d_0, K, m, p *and* T.

PROOF. If $p=2$, the theorem is quite similar to Theorem 3.1 of [12] and can be proved by the same method. The only difference is that $V_t = t$ and Y_t is a d_1-dimensional Wiener process in [12]. Actually one can also obtain our Theorem 3.1 for $p=2$ quite formally from Theorem 3.1 of [12]. Indeed, replacing V_t with $V_t + t$ [and multiplying the corresponding coefficients by $dV_t/(dV_t + dt)$] allows us to assume that V_t is strictly increasing. After that a time change reduces the whole situation to the one with $V_t = t$. To deal with Y_t, one uses the fact that any continuous martingale can be written as a stochastic integral against a Wiener process.

For $p \neq 2$, we reproduce part of the proof of Theorem 3.1 of [12]. It is worth noting that in [12] $L_p(\mathbb{R}^d)$ norms of solutions are estimated. Although we could do the same in our situation, we do not know how to apply these estimates to derive error estimates for $L_p(\mathbb{R}^d)$ norms for the splitting-up method. Nevertheless, we know how to derive error estimates for expectations of the pth powers of $L_2(\mathbb{R}^d)$ norms. This is why we only state and prove those estimates in our theorem.

As in the proof of Theorem 3.1 of [12], by adding into the equation $\varepsilon \Delta u\, dV_t$ if necessary, we may assume that $\|u(t)\|_{m+1}^2$ is integrable over $\Omega \times [0,T]$ against $dP \times dV_t$. Then, by using Itô's formula and integrating by parts, we get that, if $u(t)$ is our solution, then

$$
\begin{aligned}
d\sum_{|\alpha|\le l} \|D^\alpha u(t)\|_0^2 &\le N\big(\|u(t)\|_l^2 + \|f(t)\|_l^2 + \|g\|_{l+1}^2\big)\, dV_t \\
&\quad + 2\sum_{|\alpha|\le l} \big(D^\alpha u(t), D^\alpha[M_k u(t) + g_k(t)]\big)_0\, dY_t^k.
\end{aligned}
\tag{3.3}
$$

Here, due to Assumption 3.1 and the Leibnitz formula,

$$
\big(D^\alpha u, D^\alpha(b_k^i D_i u)\big)_0 = \tfrac{1}{2}\int_{\mathbb{R}^d} b_k^i D_i |D^\alpha u|^2\, dx + \sum_{|\beta|+|\gamma|=|\alpha|} (D^\alpha u, c_\alpha^{\beta\gamma} D^\gamma u)_0,
$$

where $c_\alpha^{\beta\gamma}$ are bounded functions. Integrating by parts, we see that

$$
\big|\big(D^\alpha u(t), D^\alpha[M_k u(t) + g_k(t)]\big)_0\big| \le N\|u(t)\|_l^2 + N\|u(t)\|_l \|g(t)\|_l.
$$

Now we write (3.3) in the integral form, raise both parts to the $p/2$th power and use the Burkholder–Davis–Gundy inequality. We also use that, if $p \ge 2$, then, by Hölder's inequality,

$$
\left(\int_0^\tau \|u\|_l^2\, dV_t\right)^{p/2} \le \delta^q \sup_{t\le\tau} \|u\|_l^p + \delta^{-2/p} N \int_0^\tau \|u\|_l^p\, dV_t
\tag{3.4}
$$

for any $\delta \in (0,1)$, $q \in \mathbb{R}$ and stopping time $\tau \le T$, where the first term on the right-hand side can even be dropped. Finally, we notice that (3.4) holds for $p \in (0,2)$

as well with $q = 2/(2-p)$ since, by Young's inequality for any $\delta > 0$,

$$\left(\int_0^\tau \|u\|_l^2\,dV_t\right)^{p/2} \le \sup_{t\le\tau}\|u\|^{(2-p)p/2}\left(\int_0^\tau \|u\|_l^p\,dV_t\right)^{p/2}$$
$$\le \delta^{2/(2-p)}\sup_{t\le\tau}\|u\|_l^p + \delta^{-2/p}N\int_0^\tau \|u\|_l^p\,dV_t.$$

Then we obtain that, for any stopping time $\tau \le T$,

$$E\sup_{t\le\tau}\|u(t)\|_l^p$$
$$\le 2E\|u_0\|_l^p + \tfrac{1}{4}E\sup_{t\le\tau}\|u(t)\|_l^p + NE\int_0^\tau \|u(t)\|_l^p\,dV_t$$
$$+ NEK_l^{p/2}(\tau) + NE\left(\int_0^\tau \left(\|u(t)\|_l^4 + \|u(t)\|_l^2\|g(t)\|_{l+1}^2\right)dV_t\right)^{p/4}.$$

The last term is less than

$$NE\sup_{t\le\tau}\|u(t)\|_l^{p/2}\left(\int_0^\tau \left(\|u(t)\|_l^2 + \|g(t)\|_{l+1}^2\right)dV_t\right)^{p/4}$$
$$\le \tfrac{1}{4}E\sup_{t\le\tau}\|u(t)\|_l^p + NE\int_0^\tau \|u(t)\|_l^p\,dV_t + NEK_l^{p/2}(\tau).$$

Thus,

$$E\sup_{t\le\tau}\|u(t)\|_l^p \le 4E\|u_0\|_l^p + NE\int_0^\tau \|u(t)\|_l^p\,dV_t + NEK_l^{p/2}(\tau),$$

and (3.2) follows by the stochastic version of Gronwall's inequality. The theorem is proved. $\square$

We are going to use Theorem 3.1 for estimating the difference of solutions of two equations of type (3.1). Namely, let

$$(a_\varepsilon^{ij}, a_\varepsilon^i, a_\varepsilon, f_\varepsilon, b_{k\varepsilon}^i, b_{k\varepsilon}, g_{k\varepsilon}, u_{0\varepsilon}), \tag{3.5}$$

where $\varepsilon = 0, 1$, be two sets of data satisfying Assumptions 3.1–3.3 for $\varepsilon = 0, 1$. Continue these data linearly with respect to ε on [0, 1] so that we can now use the same notation (3.5) for any $\varepsilon \in [0, 1]$. Let L_ε and $M_{k\varepsilon}$ be the operators L and M_k constructed on the basis of $a_\varepsilon^{ij}, a_\varepsilon^i, a_\varepsilon$ and $b_{k\varepsilon}^i, b_{k\varepsilon}$. We will be interested in the difference $u_0 - u_1$, where u_ε is defined as the unique solution of

$$\begin{aligned} du_\varepsilon(t,x) &= \left(L_\varepsilon u_\varepsilon(t,x) + f_\varepsilon(t,x)\right)dV_t \\ &\quad + \left(M_{k\varepsilon}u_\varepsilon(t,x) + g_{k\varepsilon}(t,x)\right)dY_t^k, \end{aligned} \tag{3.6}$$

with initial data $u_{0\varepsilon}$. Notice that Assumption 3.2 is satisfied for L_ε and $M_{k\varepsilon}$ with any $\varepsilon \in [0,1]$. This follows from the fact that

$$b^i_{k\varepsilon} b^j_{r\varepsilon} \lambda^i \lambda^j \, d\langle Y^k, Y^r\rangle_t$$

is a nonnegative quadratic, hence convex function of ε. Therefore, Theorem 3.1 implies the following:

LEMMA 3.2. *The function u_ε exists, is unique and*

$$\sup_{\varepsilon\in[0,1]} E \sup_{t\in[0,T]} \|u_\varepsilon(t)\|_m^p \le \sup_{\varepsilon\in[0,1]} NE\big(\|u_{0\varepsilon}\|_m^p + K_m^{p/2}(T)\big), \tag{3.7}$$

where N depends only on d, d_0, K, m, p and T.

Now comes an estimate of $u_1 - u_0$.

THEOREM 3.3. *Let $m \ge 3$ and $p \ge 1$. Then, for any integer $l \ge 0$,*

$$E \sup_{t\in[0,T]} \|u_1(t) - u_0(t)\|_l^p \le \sup_{\varepsilon\in[0,1]} E \sup_{t\in[0,T]} \|v_\varepsilon(t)\|_l^p, \tag{3.8}$$

where v_ε is the unique solution of the following equation obtained by formal differentiation of (3.6):

$$\begin{aligned} dv_\varepsilon(t,x) &= \big(L_\varepsilon v_\varepsilon(t,x) + L' u_\varepsilon(t,x) + f'(t,x)\big)\, dV_t \\ &\quad + \big(M_{k\varepsilon} v_\varepsilon(t,x) + M'_k u_\varepsilon(t,x) + g'_k(t,x)\big)\, dY^k_t, \end{aligned} \tag{3.9}$$

with initial condition u'_0, where the primed functions are introduced according to $w' = w_1 - w_0$. Furthermore,

$$\sup_{\varepsilon\in[0,1]} E \sup_{t\in[0,T]} \|v_\varepsilon(t)\|_{m-2}^p < \infty. \tag{3.10}$$

PROOF. Owing to (3.7), the functions $\tilde f = L' u_\varepsilon + f'$ and $\tilde g_k = M'_k u_\varepsilon + g'_k$ satisfy Assumption 3.3 with $m-2 \ge 1$ in place of m. Hence, the existence and uniqueness of v_ε and estimate (3.10) follow from Theorem 3.1.

While proving (3.8) for a fixed l, we may and will assume that the right-hand side is finite. Furthermore, notice that to prove (3.8) it suffices to show that v_ε is the derivative of u_ε in an appropriate space. To make this precise, for a function w_ε and h such that $\varepsilon, \varepsilon + h \in [0,1]$ define $\delta_h w_\varepsilon = (w_{\varepsilon+h} - w_\varepsilon)/h$. It turns out that it suffices to show that, for any $\varepsilon \in [0,1]$,

$$E \sup_{t\in[0,T]} \|\delta_h u_\varepsilon(t) - v_\varepsilon(t)\|_0^p \to 0, \tag{3.11}$$

whenever $h \to 0$ in such a way that $\varepsilon + h \in [0,1]$.

Indeed, assume that (3.11) holds and let $R_n := n^l(n - \Delta)^{-l}$, $n > 0$. Notice that $\|R_n h\|_l \le N\|h\|_0$ for $h \in L_2$, where N is independent of h. Therefore, (3.11) implies that, for any $n > 0$,

$$E \sup_{t\in[0,T]} \|\delta_h R_n u_\varepsilon(t) - R_n v_\varepsilon(t)\|_l^p \to 0.$$

Since $p \ge 1$, it easily follows that

$$E \sup_{t\in[0,T]} \|R_n u_0(t) - R_n u_1(t)\|_l^p \le \sup_{\varepsilon\in[0,1]} E \sup_{t\in[0,T]} \|R_n v_\varepsilon(t)\|_l^p.$$

By using the Fourier transform, one proves $\|R_n h\|_l \le \|h\|_l$ for $h \in H^l$, and also that if $h \in L_2$ and

$$N_0 := \liminf_{n\to\infty} \|R_n h\|_l < \infty,$$

then $h \in H^l$ and $\|h\|_l \le N_0$. After these observations to get (3.8), it only remains to use Fatou's lemma.

Now we prove (3.11). Simple manipulations show that the function

$$r_{\varepsilon h}(t) := \delta_h u_\varepsilon(t) - v_\varepsilon(t)$$

satisfies

$$\begin{aligned} dr_{\varepsilon h}(t) &= \big[L_\varepsilon r_{\varepsilon h}(t) + L'\big(u_{\varepsilon+h}(t) - u_\varepsilon(t)\big)\big]\,dV_t \\ &\quad + \big[M_k r_{\varepsilon h}(t) + M_k'\big(u_{\varepsilon+h}(t) - u_\varepsilon(t)\big)\big]\,dY_t^k. \end{aligned}$$

Hence, by Theorem 3.1, for a constant N independent of ε and h,

$$E \sup_{t\in[0,T]} \|\delta_h u_\varepsilon(t) - v_\varepsilon(t)\|_0^p \le NE\bigg(\int_0^T \|u_{\varepsilon+h}(t) - u_\varepsilon(t)\|_2^2\,dV_t\bigg)^{p/2},$$

which by the interpolation inequality $\|h\|_2 \le N\|h\|_0^{1/3}\|h\|_3^{2/3}$, Hölder's inequality and (3.7) is less than a constant times

$$\bigg(E \sup_{t\in[0,T]} \|u_{\varepsilon+h}(t) - u_\varepsilon(t)\|_0^p\bigg)^{1/3}.$$

Finally, observe that $q_{\varepsilon h}(t) := u_{\varepsilon+h}(t) - u_\varepsilon(t)$ satisfies

$$\begin{aligned} dq_{\varepsilon h}(t) &= [L_\varepsilon q_{\varepsilon h}(t) + hL'u_{\varepsilon+h}(t) + hf'(t)]\,dV_t \\ &\quad + [M_k q_{\varepsilon h}(t) + hM_k' u_{\varepsilon+h}(t) + hg_k'(t)]\,dY_t^k \end{aligned}$$

and, by Theorem 3.1 and (3.7),

$$E \sup_{t\in[0,T]} \|u_{\varepsilon+h}(t) - u_\varepsilon(t)\|_0^p \le Nh^p \to 0$$

as $h \to 0$. This proves (3.11) and finishes the proof of the theorem. □

4. Proof of Theorem 2.2. Remember that V_t is introduced in Assumption 2.2 and let

$$V^r_{t,\varepsilon} = \varepsilon V^r_{t,1} + (1-\varepsilon)V^r_{t,0}, \qquad \rho^r_{t\varepsilon} = dV^r_{t,\varepsilon}/dV_t \ \ (\leq 1),$$

$$L_\varepsilon = \rho^r_{t\varepsilon} L_r, \qquad f_\varepsilon = \rho^r_{t\varepsilon} f_r, \qquad M_{k\varepsilon} = M_k, \qquad g_{k\varepsilon} = g_k.$$

Then (2.1) becomes (3.6). Next define

$$a^{ij}_\varepsilon = \rho^r_{t\varepsilon} a^{ij}_r, \qquad a^i_\varepsilon = \rho^r_{t\varepsilon} a^i_r, \qquad a_\varepsilon = \rho^r_{t\varepsilon} a_r.$$

Notice that

$$a^{ij}_\varepsilon \lambda^i \lambda^j \, dV_t = a^{ij}_r \lambda^i \lambda^j \, dV^r_{t,\varepsilon}.$$

It follows that the assumptions of our equation (3.6), stated before Theorem 3.3, are satisfied with $m+3$ in place of m. This theorem implies that in order to prove Theorem 2.2 it suffices to show that, for any $\varepsilon \in [0,1]$,

$$E \sup_{t\in[0,T]} \|v_\varepsilon(t)\|^p_m \leq N(E\|u_{01} - u_{00}\|^p_m + A^p),$$

where $v_\varepsilon(t)$ satisfies $v_\varepsilon(0) = u_{01} - u_{00}$ and is the unique solution of (3.9). The latter in our case becomes

$$dv_\varepsilon(t) = L_r v_\varepsilon(t)\, dV^r_{t,\varepsilon} + \big(L_r u_\varepsilon(t) + f_r(t)\big)\, dA^r_t + M_k v_\varepsilon(t)\, dY^k_t, \tag{4.1}$$

where $A^r_t = V^r_{t,1} - V^r_{t,0}$ and, of course, $u_\varepsilon(t)$ is the unique solution of (2.1) with the above-defined $V^r_{t,\varepsilon}$ and initial data $u_{0\varepsilon} = \varepsilon u_{01} + (1-\varepsilon)u_{00}$.

Next we need two lemmas. Remember that H^{-1} is the space of distributions which is dual to H^1 and there is a natural way to extend $(v,u)_0$ by continuity from $v, u \in L_2$ to $v \in H^{-1}$, $u \in H$. This extension of the inner product in L_2 is denoted by $\langle v, u\rangle$ or $\langle u, v\rangle$. Similarly, for any positive integer m the inner product $(\cdot,\cdot)_m$ in H^m can be extended by continuity to a duality $\langle\cdot,\cdot\rangle_m$ between H^{m-1} and H^{m+1}. Set

$$q^{kl}_t := d\langle Y^k, Y^l\rangle_t / dV_t, \qquad \tilde{a}^{ij}_\varepsilon := a^{ij}_\varepsilon - \tfrac{1}{2} b^i_k b^j_l q^{kl}_t.$$

Define the quadratic forms

$$[v]^2_m = [v]^2_m(t) = (\tilde{a}^{ij}_\varepsilon D_i v, D_j v)_m + C_m \|v\|^2_m, \qquad v \in H^{m+1}, \tag{4.2}$$

where $C_0 = 0$ and, if $m \geq 1$, C_m is a constant to be specified later in such a way that the right-hand side of (4.2) is nonnegative, so that notation (4.2) makes sense. We polarize $[v]^2_m$ to define the corresponding bilinear forms

$$4[v,w]_m = [v+w]^2_m - [v-w]^2_m, \qquad v, w \in H^{m+1}.$$

To simplify the notation, write

$$v_\alpha = D^\alpha v, \qquad v_{\alpha i} = D^\alpha D_i v, \qquad v_{\alpha ij} = D^\alpha D_{ij} v.$$

Then

$$(u, v)_m = \sum_{|\alpha| \le m} (u_\alpha, v_\alpha)_0.$$

Quite often we deal with finite sums $\sum_{\alpha\beta} a^{\alpha\beta} v_\alpha v_\beta$ with uniformly bounded coefficients $a^{\alpha\beta}$. Let $\mathcal{H}$ denote the set of such forms. For $\xi, \eta \in \mathcal{H}$ we write $\xi \sim \eta$ if there is a form

$$\zeta = \sum_{|\alpha| \le m} v_\alpha P^\alpha v, \qquad P^\alpha v = \sum_{|\beta| \le m} a^{\alpha\beta} v_\beta, \tag{4.3}$$

such that the integrals (over $\mathbb{R}^d$) of $\xi - \eta$ and ζ are the same and $|a^{\alpha\beta}|$ can be estimated in terms of d, d_0, d_1, m and K. Forms of type ζ are particularly interesting because their integrals are estimated through a constant under control times $\|v\|_m^2$.

LEMMA 4.1. *There is a constant C_m with $C_0 = 0$ depending only on K, d, d_0, d_1 and m such that the right-hand side of* (4.2) *is nonnegative. Furthermore, for $m \ge 1$, any multi-indices α, β, γ satisfying $\alpha = \beta + \gamma$, $|\beta| \ge 1$ and $|\alpha| \le m$, and any $v \in H^{m+1}$, we have $(\tilde{a}_\varepsilon^{ij} D_i v)_\alpha v_{\alpha j} \sim \tilde{a}_\varepsilon^{ij} v_{\alpha i} v_{\alpha j}$ and*

$$I^{\alpha\beta\gamma} := \tilde{a}^{ij}_{\varepsilon\beta} v_{\gamma i} v_{\alpha j} \sim \tilde{a}^{ij}_{\varepsilon\beta} v_{\gamma i j} v_\alpha \sim 0. \tag{4.4}$$

PROOF. Notice that the assertion of the lemma holds true for $m = 0$ due to Assumption 2.3 saying that $\tilde{a}_\varepsilon$ is a nonnegative matrix (V_t-a.e.). For $m \ge 1$ and $m \ge |\alpha| \ge 1$, use the Leibnitz formula to get

$$(\tilde{a}_\varepsilon^{ij} D_i v)_\alpha v_{\alpha j} = \tilde{a}_\varepsilon^{ij} v_{\alpha i} v_{\alpha j} + \sum_{\beta+\gamma=\alpha, |\beta| \ge 1} c^{\alpha\beta\gamma} I^{\alpha\beta\gamma},$$

where $c^{\alpha\beta\gamma}$ are certain constants. Since the first term on the right-hand side is nonnegative, it only remains to prove (4.4).

Integrating by parts allows us to carry the derivative with respect to x^j from v_α to $\tilde{a}^{ij}_{\varepsilon\beta} v_{\gamma i}$. Observe that $\tilde{a}^{ij}_{\varepsilon\beta j}$ is bounded by a constant, under control, since $|\beta| + 1 \le m + 1$. It follows that $I^{\alpha\beta\gamma} \sim -\tilde{a}^{ij}_{\varepsilon\beta} v_{\gamma i j} v_\alpha$, and it only remains to prove $I^{\alpha\beta\gamma} \sim 0$.

If $|\beta| \ge 2$ in $I^{\alpha\beta\gamma}$, then $v_{\gamma i j}$ is the derivative of v of order at most m. In this case, $I^{\alpha\beta\gamma} \sim 0$ and we may concentrate on $|\beta| = 1$. In that case, due to $\tilde{a}_\varepsilon^{ij} = \tilde{a}_\varepsilon^{ji}$, we have

$$I^{\alpha\beta\gamma} = \tilde{a}^{ij}_{\varepsilon\beta} v_{\gamma i} D^\beta v_{\gamma j} = \tfrac{1}{2} \tilde{a}^{ij}_{\varepsilon\beta} D^\beta (v_{\gamma i} v_{\gamma j}),$$

and integrating by parts shows that $I^{\alpha\beta\gamma} \sim 0$ again. The lemma is proved. □

In particular, we now have $|[v, w]_m| \le [v]_m [w]_m$ (dV_t-a.e.) for all $v, w \in H^{m+1}$, $m \ge 0$.

LEMMA 4.2. *There exists a constant N depending only on d, d_0, d_1, m and K, such that, for any $v \in H^{m+1}$, $u \in H^{m+3}$, $h \in H^{m+2}$, $\varepsilon \in [0,1]$:*

(i) *for any r, k, we have*

$$|(v, L_r h)_m| + |\langle L_r v, h\rangle_m| + |(v, L_r M_k u)_m| + |(M_k v, L_r u)_m| \le N\|v\|_m(\|h\|_{m+2} + \|u\|_{m+3}); \tag{4.5}$$

(ii) *almost everywhere with respect to dV_t,*

$$p(v,v) := 2\langle v, L_r v\rangle_m \rho^r_{t\varepsilon} + (M_k v, M_r v)_m q^{kr}_t + 2[v]^2_m \le N\|v\|^2_m; \tag{4.6}$$

(iii) *for any i almost everywhere with respect to dV_t,*

$$|q_i(v,u)| \le N\|u\|_{m+3}([v]_m + \|v\|_m), \tag{4.7}$$

where

$$q_i(v,u) = (\langle L_r v, L_i u\rangle_m + \langle v, L_i L_r u\rangle_m)\rho^r_{t\varepsilon} + (M_k v, L_i M_r u)_m q^{kr}_t.$$

PROOF. One can easily get estimate (4.5) by Cauchy's inequality combined with integration by parts. The proof of (ii) is very similar to that of Lemma 2.1 in [12]. We may (and will) assume that $v \in H^{m+2}$. Then the left-hand side of inequality (4.6) minus $2[v]^2_m$ is the integral over $\mathbb{R}^d$ of

$$Q := \sum_{|\alpha|\le m} \{2\rho^r_{t\varepsilon} v_\alpha (L_r v)_\alpha + q^{kr}_t (M_k v)_\alpha (M_r v)_\alpha\} =: \sum_{|\alpha|\le m} Q^\alpha.$$

By integrating by parts, we obtain

$$2v a^i_\varepsilon v_i \sim a^i_\varepsilon (v^2)_i \sim -a^i_{\varepsilon i} v^2 \sim 0,$$

and similarly, for $|\alpha| \le m$,

$$v_\alpha (a^i_\varepsilon v_i)_\alpha \sim v_\alpha a^i_\varepsilon v_{\alpha i} \sim 0, \qquad v_\alpha (a_\varepsilon v)_\alpha \sim 0,$$

$$(b^i_k v_i)_\alpha (b_r v)_\alpha \sim 0, \qquad (b_k v)_\alpha (b_r v)_\alpha \sim 0.$$

Hence, upon defining $L^0_r v = a^{ij} D_{ij} v$ and $M^0_k v = b^i_k D_i v$, we get

$$Q^\alpha \sim \{2\rho^r_{t\varepsilon} v_\alpha (L^0_r v)_\alpha + q^{kr}_t (M^0_k v)_\alpha (M^0_r v)_\alpha\}. \tag{4.8}$$

If $m = 0$, then the only possible value for α is 0 and the integral on the right-hand side of (4.8) equals $-2[v]^2_0$, which implies (4.6). Therefore, in the remaining part of the proof we assume that $m \ge 1$.

For $m \ge |\alpha| \ge 1$ define $\Gamma(\alpha)$ as the set of couples of multi-indices (β, γ) such that $|\beta| = 1$ and $\alpha = \beta + \gamma$ and define the constants $c^{\alpha\beta\gamma}$ from the equality

$$D^\alpha(\phi\psi) = \phi D^\alpha \psi + \sum_{\Gamma(\alpha)} c^{\alpha\beta\gamma} (D^\beta \phi) D^\gamma \psi + \cdots,$$

where the missing terms are those that contain the derivatives of ψ of order at most $|\alpha|-2$. Then, for $m \geq |\alpha| \geq 1$, owing to $q_t^{kr}=q_t^{rk}$, we obtain

$$q_t^{kr}(M_k^0 v)_\alpha (M_r^0 v)_\alpha = q_t^{kr}(b_k^i v_i)_\alpha (b_r^j v_j)_\alpha q_t^{kr}$$
$$\sim q_t^{kr} b_k^i v_{\alpha i} b_r^j v_{\alpha j} + 2q_t^{kr} \sum_{\Gamma(\alpha)} c^{\alpha\beta\gamma} b_{k\beta}^i v_{\gamma i} b_r^j v_{\alpha j}.$$

Upon remembering that b_k^i are twice differentiable and $|\beta|+1=2$ and $|\gamma|+1=|\alpha|\leq m$, we get

$$q_t^{kr} b_{k\beta}^i v_{\gamma i} b_r^j v_{\alpha j} \sim -q_t^{kr} b_{k\beta}^i v_{\gamma ij} b_r^j v_\alpha = -\tfrac{1}{2} q_t^{kr} (b_r^j b_k^i)_\beta v_{\gamma ij} v_\alpha.$$

Furthermore,

$$2\rho_{t\varepsilon}^r v_\alpha (L_r^0 v)_\alpha \sim 2 v_\alpha a_\varepsilon^{ij} v_{\alpha ij} + 2\sum_{\Gamma(\alpha)} c^{\alpha\beta\gamma} v_\alpha a_{\varepsilon\beta}^{ij} v_{\gamma ij}.$$

After these computations (4.8) and Lemma 4.1 yield

$$Q_\alpha \sim -2 v_{\alpha i} \tilde{a}_\varepsilon^{ij} v_{\alpha j} + 2\sum_{\Gamma(\alpha)} c^{\alpha\beta\gamma} \tilde{a}_{\varepsilon\beta}^{ij} v_{\gamma ij} v_\alpha \sim -2(\tilde{a}_\varepsilon^{ij} D_i v)_\alpha v_{\alpha j}.$$

Thus,

$$p(v,v) = \int_{\mathbb{R}^d} Q\,dx + 2[v]_m^2 = \sum_{|\alpha|\leq m} (v_\alpha, P^\alpha v)_0, \tag{4.9}$$

where P^α are some operators as in (4.3). This proves (4.6).

To prove (4.7), we polarize (4.9) and get

$$\langle L_r v, w\rangle_m \rho_{t\varepsilon}^r + \langle v, L_r w\rangle_m \rho_{t\varepsilon}^r + (M_k v, M_r w)_m q_t^{kr} + 2[v,w]_m$$
$$= \tfrac{1}{2} \sum_{|\alpha|\leq m} [(v_\alpha, P^\alpha w)_0 + (w_\alpha, P^\alpha v)_0].$$

We plug in $w = L_i u$ to obtain

$$q_i(v,u) + \langle v, (L_r L_i - L_i L_r)u\rangle_m \rho_{t\varepsilon}^r$$
$$+ \big(M_k v, (M_r L_i - L_i M_r)u\big)_m q_t^{kr} + 2[v, L_i u]_m$$
$$= \tfrac{1}{2} \sum_{|\alpha|\leq m} [(D^\alpha v, P^\alpha L_i u)_0 + (D^\alpha L_i u, P^\alpha v)_0].$$

Hence, we obtain (4.7) by Cauchy's inequality and by integration by parts, after noticing that $(L_r L_i - L_i L_r)$ and $(M_r L_i - L_i M_r)$ are third- and second-order operators, respectively. The lemma is proved. □

LEMMA 4.3. *Define*

$$J_t = J_{t\varepsilon} = \int_0^t \big(v_\varepsilon(s), L_i u_\varepsilon(s) + f_i(s)\big)_m \, dA_s^i.$$

Then there exists a constant N depending only on d, d_0, d_1, K, p, m and T such that, for any stopping time $\tau \le T$,

$$\begin{aligned} &E \sup_{t\le\tau} \bigg(J_{t\varepsilon} - \int_0^t [v_\varepsilon]_m^2(s)\, dV_s \bigg)_+^{p/2} \\ &\quad \le \tfrac{1}{8} E \sup_{t\le\tau} \|v_\varepsilon(t)\|_m^p + N\bigg(A^p + E\int_0^\tau \|v_\varepsilon(t)\|_m^p \, dV_t \bigg). \end{aligned} \tag{4.10}$$

PROOF. We want to estimate $J_{t\varepsilon}$ through A without using the variations of A_t^i. Therefore, we integrate by parts with respect to s or alternatively use Itô's formula (see [6]). We also remember that the coefficients of L_r and f_r are independent of t. Then we obtain

$$J_t = \big(v_\varepsilon(t), L_i u_\varepsilon(t) + f_i(t)\big)_m A_t^i - J_{1t} - \cdots - J_{4t}, \tag{4.11}$$

where J_{it} are defined by the following formulas in which we drop the argument s whenever it does not lead to any confusion:

$$\begin{aligned} J_{1t} &= \int_0^t A_s^i \{\langle L_r v_\varepsilon, L_i u_\varepsilon + f_i\rangle_m + \langle v_\varepsilon, L_i(L_r u_\varepsilon + f_r)\rangle_m\}\, dV_{s,\varepsilon}^r, \\ J_{2t} &= \int_0^t A_s^i \big(M_k v_\varepsilon, L_i(M_r u_\varepsilon + g_r)\big)_m \, d\langle Y^k, Y^r\rangle_s, \\ J_{3t} &= \int_0^t A_s^i \big\{(M_k v_\varepsilon, L_i u_\varepsilon + f_i)_m + \big(v_\varepsilon, L_i(M_r u_\varepsilon + g_r)\big)_m\big\}\, dY_s^k, \\ 2J_{4t} &= 2\int_0^t A_s^i (L_j u_\varepsilon + f_j, L_i u_\varepsilon + f_i)_m \, dA_s^j \\ &= \int_0^t (L_j u_\varepsilon + f_j, L_i u_\varepsilon + f_i)_m \, d(A_s^i A_s^j). \end{aligned}$$

By Lemma 4.2 and Young's inequality,

$$\begin{aligned} J_{1t} + J_{2t} &\le NA \int_0^t \{\|u_\varepsilon\|_{m+3}[v_\varepsilon]_m + \|v_\varepsilon\|_m(\|f\|_{m+2} + \|g\|_{m+3} + \|u_\varepsilon\|_{m+3})\}\, dV_s \\ &\le \int_0^t [v_\varepsilon]_m^2 \, dV_s + \int_0^t \|v_\varepsilon\|_m^2 \, dV_s \\ &\quad + NA^2 \int_0^t \{\|u_\varepsilon\|_{m+3}^2 + \|f\|_{m+2}^2 + \|g\|_{m+3}^2\}\, dV_s. \end{aligned}$$

Next notice that, by Lemma 4.2,

$$\begin{aligned} &\big|(M_k v_\varepsilon, L_i u_\varepsilon + f_i)_m + \big(v_\varepsilon, L_i(M_r u_\varepsilon + g_r)\big)_m\big| \\ &\qquad \le N\|v_\varepsilon\|_m(\|u_\varepsilon\|_{m+3} + \|f\|_{m+1} + \|g\|_{m+2}). \end{aligned}$$

Therefore, by the Burkholder–Davis–Gundy inequality,

$$\begin{aligned} E\sup_{t\le\tau}|J_{3t}|^{p/2} &\le NA^{p/2}E\biggl(\int_0^\tau \|v_\varepsilon\|_m^2(\|u_\varepsilon\|_{m+3}^2+\|f\|_{m+1}^2+\|g\|_{m+2}^2)\,dV_t\biggr)^{p/4} \\ &\le NA^{p/2}E\sup_{t\in[0,T]}\bigl(\|u_\varepsilon(t)\|_{m+3}^{p/2}+\|f(t)\|_{m+1}^{p/2}+\|g(t)\|_{m+2}^{p/2}\bigr) \\ &\qquad\times\biggl(\int_0^\tau \|v_\varepsilon\|_m^2\,dV_t\biggr)^{p/4}. \end{aligned}$$

We use Cauchy's inequality, (3.7), the argument about (3.4) and our assumptions and infer that

$$E\sup_{t\le\tau}|J_{3t}|^{p/2}\le NA^p+\tfrac{1}{16}E\sup_{t\le\tau}\|v_\varepsilon(t)\|_m^p+NE\int_0^\tau\|v_\varepsilon\|_m^p\,dV_t.$$

It follows that the left-hand side of (4.10) is less than

$$\tfrac{1}{8}E\sup_{t\le\tau}\|v_\varepsilon(t)\|_m^p+NA^p+NE\int_0^\tau\|v_\varepsilon\|_m^p\,dV_t+NE\sup_{t\le\tau}|J_{4t}|^{p/2},$$

and to prove the lemma it only remains to estimate J_{4t}.

We integrate by parts again and find that

$$2J_{4t}=\bigl(L_ju_\varepsilon(t)+f_j,L_iu_\varepsilon(t)+f_i\bigr)_mA_t^iA_t^j-R_{1t}-R_{2t}-R_{3t}, \tag{4.12}$$

where

$$\begin{aligned} R_{1t}&=2\int_0^t A_s^iA_s^j\langle L_j(L_ru_\varepsilon+f_r),L_iu_\varepsilon+f_i\rangle_m\,dV_{s,\varepsilon}^r, \\ R_{2t}&=\int_0^t A_s^iA_s^j\bigl(L_j(M_ku_\varepsilon+g_k),L_i(M_ru_\varepsilon+g_r)\bigr)_m\,d\langle Y^k,Y^r\rangle_s, \\ R_{3t}&=2\int_0^t A_s^iA_s^j\bigl(L_j(M_ku_\varepsilon+g_k),L_iu_\varepsilon+f_i\bigr)_m\,dY_s^k. \end{aligned}$$

Since $\langle L_j(L_ru_\varepsilon+f_r),L_iu_\varepsilon+f_i\rangle_m$ is readily estimated through $\|u_\varepsilon\|_{m+3}^2+\|f\|_{m+1}^2$, we see that

$$E\sup_{t\le\tau}|R_{1t}+R_{2t}|^{p/2}\le NA^p.$$

Furthermore, the Burkholder–Davis–Gundy inequality obviously implies that the same estimate holds for R_{3t}. Hence, $E\sup_{t\le\tau}|J_{4t}|^{p/2}\le NA^p$. The lemma is proved. □

PROOF OF THEOREM 2.2. Applying the differential operator D^α to both sides of (4.1), using Itô's formula (see [6]) for $\|D^\alpha v_\varepsilon(t)\|_0^2$ and summing over all $|\alpha| \le m$, we get

$$d\|v_\varepsilon(t)\|_m^2 = 2\langle v_\varepsilon(t), L_r v_\varepsilon(t)\rangle_m \, dV_{t,\varepsilon}^r + 2\big(v_\varepsilon(t), L_r u_\varepsilon(t) + f_r(t)\big)_m \, dA_t^r$$
$$+ \big(M_k v_\varepsilon(t), M_r v_\varepsilon(t)\big)_m \, d\langle Y^k, Y^r\rangle_t + 2\big(v_\varepsilon(t), M_k v_\varepsilon(t)\big)_m \, dY_t^k.$$

By using Lemma 4.2(ii), we obtain

$$d\|v_\varepsilon(t)\|_m^2 \le -2[v_\varepsilon]_m^2 \, dV_t + N\|v_\varepsilon\|_m^2 \, dV_t + 2\, dJ_t + 2(v_\varepsilon, M_k v_\varepsilon)_m \, dY_t^k,$$

where J_t is defined in Lemma 4.3. Here, as before, integrating by parts implies that $|(v_\varepsilon, M_k v_\varepsilon)_m| \le N\|v_\varepsilon\|_m^2$. Hence, by Lemma 4.3 and the Burkholder–Davis–Gundy inequality,

$$\begin{aligned} E\sup_{t\le\tau}\|v_\varepsilon(t)\|_m^p &\le NE\|u_{01}-u_{00}\|_m^p + 4E\sup_{t\le\tau}\biggl(J_{t\varepsilon} - \int_0^t [v_\varepsilon]_m^2 \, dV_s\biggr)_+^{p/2} \\ &\quad + NE\biggl(\int_0^\tau \|v_\varepsilon(t)\|_m^4 \, dV_t\biggr)^{p/4} \\ &\le NE\|u_{01}-u_{00}\|_m^2 + \tfrac{1}{2}E\sup_{t\le\tau}\|v_\varepsilon(t)\|_m^p \\ &\quad + NA^p + NE\int_0^\tau \|v_\varepsilon(t)\|_m^p \, dV_t + NE\biggl(\int_0^\tau \|v_\varepsilon(t)\|_m^4 \, dV_t\biggr)^{p/4} \end{aligned}$$

for any stopping time $\tau \le T$. The last term here is estimated through [see (3.4)]

$$\begin{aligned} &NE\sup_{t\le\tau}\|v_\varepsilon(t)\|_m^{p/2}\biggl(\int_0^\tau \|v_\varepsilon(t)\|_m^2 \, dV_t\biggr)^{p/4} \\ &\qquad \le \tfrac{1}{4}E\sup_{t\le\tau}\|v_\varepsilon(t)\|_m^p + NE\int_0^\tau \|v_\varepsilon(t)\|_m^p \, dV_t, \end{aligned}$$

which implies

$$E\sup_{t\le\tau}\|v_\varepsilon(t)\|_m^p \le NE\|u_{01}-u_{00}\|_m^p + NA^p + NE\int_0^\tau \|v_\varepsilon(t)\|_m^p \, dV_t.$$

Now we get

$$E\sup_{t\le\tau}\|v_\varepsilon(t)\|_m^p \le NE\|u_{01}-u_{00}\|_m^p + NA^p$$

by a stochastic version of Gronwall's lemma. If $p \ge 1$, this finishes the proof of (2.3) owing to Theorem 3.3.

To deal with $p \in (0,1)$, we notice that a careful analysis of the above proof of (2.3) shows that

$$\begin{aligned} &E\sup_{t\le\tau}\|u_1(t)-u_0(t)\|_m^2 \\ &\quad \le NE\|u_{01}-u_{00}\|_m^2 \\ &\qquad + NA^2 E\Bigl\{\|u_{01}\|_{m+3}^2 + \|u_{00}\|_{m+3}^2 + \sup_{t\le\tau}\bigl(\|f(t)\|_{m+3} + \|g(t)\|_{m+4}\bigr)^2\Bigr\} \end{aligned}$$

for any stopping time $\tau \le T$, and, furthermore (a.s.),

$$E\left\{\sup_{t\le\tau}\|u_1(t)-u_0(t)\|_m^2\Big|\mathcal{F}_0\right\}$$
$$\le N\|u_{01}-u_{00}\|_m^2+NA^2\big(\|u_{01}\|_{m+3}^2+\|u_{00}\|_{m+3}^2\big)$$
$$+NA^2E\left\{\sup_{t\le\tau}\big(\|f(t)\|_{m+3}+\|g(t)\|_{m+4}\big)^2\Big|\mathcal{F}_0\right\}.$$

A standard transformation of such inequalities (see, for instance, the derivation of Theorem 3.6.8 from Lemma 3.6.3 of [10]) shows that, for any $\delta\in(0,1)$ (a.s.),

$$E\left\{\sup_{t\le\tau}\|u_1(t)-u_0(t)\|_m^{2\delta}\Big|\mathcal{F}_0\right\}$$
$$\le N\|u_{01}-u_{00}\|_m^{2\delta}+NA^{2\delta}\big(\|u_{01}\|_{m+3}^{2\delta}+\|u_{00}\|_{m+3}^{2\delta}\big)$$
$$+NA^{2\delta}E\left\{\sup_{t\le\tau}\big(\|f(t)\|_{m+3}+\|g(t)\|_{m+4}\big)^{2\delta}\Big|\mathcal{F}_0\right\}.$$

Upon taking here $\delta=p/2$ and taking the expectations of both parts of the last inequality, we arrive at (2.3). The theorem is proved. □

5. The case of time-dependent coefficients. Here we consider (2.1), keeping Assumptions 2.1–2.4 and assuming that the following condition also holds, in which

$$h(t,x)=\big(a_\gamma^{ij}(t,x),a_\gamma^i(t,x),a_\gamma(t,x),f_\gamma(t,x):\gamma=1,2,\dots,d_1,\ i,j=1,\dots,d\big).$$

In this section we stipulate that Greek integer-valued indices run through $1,2,\dots,d_1$.

ASSUMPTION 5.1. There exists a continuous $\mathcal{F}_t$-martingale

$$Z_t=(Z_t^1,\dots,Z_t^{d_2}),$$

and for any $x\in\mathbb{R}^d$ there exist bounded predictable functions

$$h_r(t,x)=\big(a_{\gamma r}^{ij}(t,x),\ a_{\gamma r}^i(t,x),\ a_{\gamma r}(t,x),\ f_{\gamma r}(t,x)\big)$$

defined on $\Omega\times(0,T]$ for $r=0,1,\dots,d_2$, such that:

(i) $d\langle Z\rangle_t\le dV_t$,
(ii) $h(t,x)=h(0,x)+\int_0^t h_0(s,x)\,dV_s+\int_0^t h_r(s,x)\,dZ_s^r$,

for all ω and t, where, as usual, the summation in r is carried over all possible values, which in this case are $1,2,\dots,d_2$. Furthermore, h_r are continuously differentiable with respect to x up to order $m+1$ and $|D^\beta h_r|\le K$ for $|\beta|\le m+1$.

THEOREM 5.1. *Under Assumptions* 2.1–2.4 *and* 5.1 *there is a constant* N *depending only on* d, d_0, d_1, d_2, K, p, m *and* T, *such that*

$$E \sup_{t\in[0,T]} \|u_1(t) - u_0(t)\|_m^p \le N(E\|u_{01} - u_{00}\|_m^p + A^p).$$

PROOF. Obviously, we need only show that Lemma 4.3 remains valid. Define

$$L_{\gamma r} = a^{ij}_{\gamma r} D_{ij} + a^i_{\gamma r} D_i + a_{\gamma r}$$

and observe that, since $A^0_t \equiv 0$ and now the coefficients of L_γ and f_γ depend on time, there will be three additional terms $-J_{5t} - J_{6t} - J_{7t}$ on the right-hand side of (4.11) with

$$J_{5t} = \int_0^t A^\gamma_s (v_\varepsilon, L_{\gamma 0} u_\varepsilon + f_{\gamma 0})_m \, dV_s,$$

$$J_{6t} = \int_0^t A^\gamma_s (M_k v_\varepsilon, L_{\gamma r} u_\varepsilon + f_{\gamma r})_m \, d\langle Y^k, Z^r\rangle_s,$$

$$J_{7t} = \int_0^t A^\gamma_s (v_\varepsilon, L_{\gamma r} u_\varepsilon + f_{\gamma r})_m \, dZ^r_s.$$

By following already familiar lines, we conclude that

$$\begin{aligned} E\sup_{t\le\tau} |J_{5t}|^{p/2} &\le N A^{p/2} E \sup_{t\le\tau} \|v_\varepsilon\|_m^{p/2} \sup_{t\le\tau} (\|u_\varepsilon\|_{m+2} + K)^{p/2} \\ &\le \tfrac{1}{64} E \sup_{t\le\tau} \|v_\varepsilon\|_m^p + N A^p. \end{aligned}$$

The same estimate holds for J_{6t} since

$$(M_k v_\varepsilon, L_{\gamma p} u_\varepsilon + f_{\gamma p})_m = (v_\varepsilon, M_k^* L_{\gamma p} u_\varepsilon + M_k^* f_{\gamma p})_m,$$

where M_k^* is the formal adjoint of M_k and we can use that the coefficients of $L_{\gamma p}$ and $f_{\gamma p}$ are $m+1$ times differentiable.

As far as J_{7t} is concerned, it suffices to add that

$$\begin{aligned} &E\biggl(\int_0^\tau |A^\gamma_s (v_\varepsilon, L_{\gamma r} u_\varepsilon + f_{\gamma r})_m|^2 \, dV_s \biggr)^{p/4} \\ &\qquad \le N A^{p/2} E \sup_{t\le\tau} \|v_\varepsilon\|_m^{p/2} \sup_{t\le\tau} (\|u_\varepsilon\|_{m+2} + K)^{p/2}. \end{aligned}$$

The only remaining changes to make in the proof of Lemma 4.3 now are related to the fact that in (4.12) there will be the terms $-R_{4t} - R_{5t} - R_{6t} - R_{7t}$ with

$$R_{4t} = 2\int_0^t A^\gamma_s A^\mu_s (L_{\mu 0} u_\varepsilon + f_{\mu 0}, L_\gamma u_\varepsilon + f_\gamma)_m \, dV_s,$$

$$R_{5t} = \int_0^t A^\gamma_s A^\mu_s (L_{\mu r} u_\varepsilon + f_{\mu r}, L_{\gamma i} u_\varepsilon + f_{\gamma i})_m \, d\langle Z^r, Z^i\rangle_s,$$

$$R_{6t} = 2\int_0^t A_s^\gamma A_s^\mu (L_{\mu r} u_\varepsilon + f_{\mu r}, L_\gamma M_k u_\varepsilon)_m \, d\langle Z^r, Y^k\rangle_s,$$

$$R_{7t} = 2\int_0^t A_s^\gamma A_s^\mu (L_{\mu r} u_\varepsilon + f_{\mu r}, L_\gamma u_\varepsilon + f_\gamma)_m \, dZ_s^r.$$

Almost obviously all these terms can be estimated in the same way as in the proof of Lemma 4.3. By this comment we finish the proof of Theorem 5.1. □

By using the above theorem, we can extend our result on splitting-up approximations, Theorem 2.3, to SPDEs with time-dependent coefficients. Let us consider the solution $u(t)$ of (2.4) in $(0, T] \times \mathbb{R}^d$, with initial condition $u(0, x) = u_0(x)$, and remember that $T_n := \{t_i = iT/n : i = 0, 1, 2, \dots, n\}$.

Since now L_r, f_r, M_k, g_k may depend on t, it is convenient to exhibit their dependence on t following the example $L_r(t)$. For $\gamma = 1, 2, \dots, d_1$ and $s \in [0, T]$, let $\mathbf{P}_t^\gamma(s)\varphi$ denote the solution of the equation

$$dv(t) = \big(L_\gamma(s)v(t) + f_\gamma(s)\big)\, dt, \qquad t \ge 0,\ v(0) = \varphi. \tag{5.1}$$

Notice that the coefficients of L_γ and f_γ are "frozen" at time s. Then $u^{(n)}(t)$ for $t \in T_n$ is defined recursively as follows: $u^{(n)}(0) = u_0$,

$$u^{(n)}(t_{i+1}) := \mathbf{P}_\delta^{d_1}(t_{i+1}) \cdots \mathbf{P}_\delta^{2}(t_{i+1})\mathbf{P}_\delta^{1}(t_{i+1})\mathbf{Q}_{t_i t_{i+1}}\big(u^{(n)}(t_i)\big) \tag{5.2}$$

for $i = 0, 1, 2, \dots, n-1$, where $\delta = T/n$ and $\mathbf{Q}_{st}\varphi$ denotes the solution of the equation

$$d\tilde v(t) = \big(L_0(t)\tilde v(t) + f_0(t)\big)\, dV_t^0 + \big(M_k(t)\tilde v(t) + g_k(t)\big)\, dY_t^k, \qquad t \ge s,\ \tilde v(s) = \varphi.$$

THEOREM 5.2. *Under Assumptions* 2.5 *and* 5.1, *there is a constant N depending only on d, d_0, d_1, d_2, K, p, m and T, such that*

$$E \max_{t\in T_n} \|u^{(n)}(t) - u(t)\|_m^p \le N n^{-p}$$

for all integers $n \ge 1$.

PROOF. The proof is almost exactly the same as that of the corresponding statement, Theorem 2.3, in the time-independent case. We define $d' := d_1 + 1$, $\kappa(t)$, $\bar V_{t,\varepsilon}^r$ and $\bar Y^k(t)$ in the same way. Consider the counterparts of (2.6)

$$du_\varepsilon(t) = \big(L_r(\kappa(t))u_\varepsilon(t) + f_r(\kappa(t))\big)\, d\bar V_{t,\varepsilon}^r + \big(M_k(\kappa(t))u_\varepsilon(t) + g_k(\kappa(t))\big)\, d\bar Y_t^k$$

for $\varepsilon = 0$, 1, with initial data $u_\varepsilon(0) = u_0$.

Then it is almost obvious that the assumptions of Theorem 5.1 are satisfied with the same constant K and with $d'T$ in place of T. We apply this theorem and after that, as in the proof of Theorem 2.3, it only remains to observe that $u_0(d't) = u(t)$ and $u_1(d't) = u^{(n)}(t)$ for $t \in T_n$. The theorem is proved. □

REMARK 5.1. We can define the approximation $u^{(n)}$ by

$$u^{(n)}(t_{i+1}) := \mathbf{P}_\delta^{d_1}(t_{i+1}) \cdots \mathbf{P}_\delta^{l+1}(t_{i+1})\mathbf{Q}_{t_i t_{i+1}}\mathbf{P}_\delta^{l}(t_i) \cdots \mathbf{P}_\delta^{2}(t_i)\mathbf{P}_\delta^{1}(t_i)u^{(n)}(t_i)$$

in place of (5.2), where $1 \le l \le d_1$ is a fixed integer. By obvious modifications of the above proof, one can show that Theorem 5.2 also holds for this approximation.

REMARK 5.2. One can also define a splitting-up approximation for the solution of (2.4) by

$$u^{(n)}(t_{i+1}) := \mathbf{P}^{d_1}_{t_i t_{i+1}} \cdots \mathbf{P}^{2}_{t_i t_{i+1}}\mathbf{P}^{1}_{t_i t_{i+1}}\mathbf{Q}_{t_i t_{i+1}}\big(u^{(n)}(t_i)\big)$$

in place of (5.2), where $v(t) := \mathbf{P}^{\gamma}_{st}\varphi$ denotes the solution of the equation

$$dv(t) = \big(L_\gamma v(t) + f_\gamma(t)\big)\,dt, \qquad t \ge s,\ v(s) = \varphi. \tag{5.3}$$

By a straightforward modification of the proof of Theorem 5.2, one can see that it also remains true for this approximation. We prefer the splitting-up approximation defined by (5.2), because, in practice, it is usually more convenient to solve the time-independent PDE (5.1) than to solve the time-dependent PDE (5.3).

Let $C^l = C^l(\mathbb{R}^d)$ denote the Banach space of functions $f = f(x)$, $x \in \mathbb{R}^d$, having continuous derivatives up to order l, such that $\|f\|_{C^l} := \sup_{x\in\mathbb{R}^d}\sum_{|\beta|\le l}|D^\beta f(x)| < \infty$. We get the following corollary from the previous theorem by Sobolev's theorem on embedding of H^m into C^l.

COROLLARY 5.3. *If Assumptions* 2.5 *and* 5.1 *hold with* $m > l + d/2$ *and nonnegative integer* l, *then, for some* $N = N(d, d_0, d_1, d_2, K, p, m)$,

$$E \max_{t\in T_n}\|u^{(n)}(t) - u(t)\|_X^p \le Nn^{-p}$$

for all $n \ge 1$, *where* $X := C^l$ *and* $\|\ \|_X$ *denotes the norm in* X.

The next corollary can be obtained easily by a standard application of the Borel–Cantelli lemma.

COROLLARY 5.4. *If Assumptions* 2.5 *and* 5.1 *hold with* $p > \kappa$ *for some* $\kappa > 1$, *then there is a random variable* ξ, *such that almost surely*

$$\max_{t\in T_n}\|u^{(n)}(t) - u(t)\|_X \le \xi n^{-1+1/\kappa}$$

for all $n \ge 1$, *where* X *is* H^m *or where* $X := C^l$ *if* $m > l + d/2$.

6. An application to nonlinear filtering. Partially observable stochastic dynamical systems are often modeled by a pair $Z_t := (X_t, Y_t)$ of multidimensional stochastic processes satisfying some stochastic differential equations with given coefficients. Here X_t is a d-dimensional process, called the unobservable component, or signal process, and Y_t is a d_0-dimensional process, called the observation process. In a fairly general situation, the evolution of these processes is governed by the equations

$$\begin{aligned} dX_t &= h(t, X_t, Y_t)\,dt + \sigma(t, X_t, Y_t)\,dw_t + \rho(t, X_t, Y_t)\,dW_t, \qquad X_0 = \xi, \\ dY_t &= H(t, X_t, Y_t)\,dt + dW_t, \qquad Y_0 = \eta, \end{aligned} \tag{6.1}$$

where $h(t, x, y) \in \mathbb{R}^d$, $\sigma(t, x, y) \in \mathbb{R}^{d\times\bar{d}}$, $\rho(t, x, y) \in \mathbb{R}^{d\times d_0}$, $H(t, x, y) \in \mathbb{R}^{d_0}$ and (w_t, W_t) is a $(\bar{d} + d_0)$-dimensional Wiener process, independent of the $\mathcal{F}_0$-measurable random vectors ξ, η. The coefficients h, σ, ρ, H are assumed to be bounded and globally Lipschitz in $(x, y) \in \mathbb{R}^{d+d_0}$, uniformly in $t \in [0, T]$.

The classic problem of nonlinear filtering is to compute at time t the best mean square estimate for $\varphi(X_t)$ from the observations $\{Y_s : 0 \le s \le t\}$ for any given bounded smooth functions φ. In other words, one wants to compute the conditional expectation

$$E\big(\varphi(X_t)|Y_s, 0 \le s \le t\big) = \int \varphi(x)P(t, dx)$$

from the data $P(0, dx)$, h, σ, ρ, H and the observation $\{Y_s, s \le t\}$ for a given function φ, where $P(t, dx)$ denotes the conditional distribution of X_t, given $\{Y_s, s \le t\}$.

From [12] one obtains the following result. To formulate it, set $\alpha_0^{ij} := \frac{1}{2}(\rho\rho^*)^{ij}$, $\alpha_1^{ij} := \frac{1}{2}(\sigma\sigma^*)^{ij}$ and $a^{ij} := \alpha_0^{ij} + \alpha_1^{ij}$ $(i, j = 1, 2, \dots, d)$, where ρ^*, σ^* denote the transpose of the matrices ρ, σ.

THEOREM 6.1. *Let $m \ge 1$ be an integer. Assume that* (i) *a^{ij} have uniformly bounded derivatives in x up to order $m + 2$,* (ii) *h and ρ have uniformly bounded derivatives in x up to order $m + 1$ and H have uniformly bounded derivatives in x up to order m and* (iii) *the conditional distribution of ξ given η has a density p_0 (with respect to Lebesgue measure), which belongs to H^m. Then the conditional density $\pi_t(x) := P(t, dx)/dx$ exists and*

$$\pi_t(x) = p(t, x)/\big(p(t), 1\big)_0,$$

where $p = p(t, x)$ is the unique solution of the equation

$$\begin{aligned} dp(t, x) &= \big\{D_{ij}\big(a^{ij}(t, x, Y_t)p(t, x)\big) + D_i\big(h^i(t, x, Y_t)p(t, x)\big)\big\}\,dt \\ &\quad + \big\{H^k(t, x, Y_t)p(t, x) + D_i\big(\rho^{ik}(t, x, Y_t)p(t, x)\big)\big\}\,dY_t^k, \end{aligned} \tag{6.2}$$

with initial condition p_0. Moreover, $\{p(t) : t \in [0, T]\}$ is a continuous H^{m-1}-valued stochastic process and a weakly continuous H^m-valued stochastic process.

This theorem describes the analytical properties of the conditional density π_t and presents a way of computing the estimate for $\varphi(X_t)$, via (6.2), called the Zakai equation (or the Duncan–Mortensen–Zakai equation) for the unnormalized conditional density p_t.

To implement this result in practice, one has to develop numerical methods to approximate the solution of (6.2) and needs to control the error of the approximations. Therefore, various methods of approximation have intensively been studied in the literature.

Notice that for (6.2) the condition of stochastic parabolicity (Assumption 3.2) requires that the matrix $2a^{ij} - (\rho\rho^*)^{ij} = (\sigma\sigma^*)^{ij}$ be nonnegative definite. Clearly, this is always satisfied. The degenerate case, $\sigma = 0$, is of special interest. In this case, the representation of the solution of (6.2) by the method of characteristics gives a relatively simple formula, which does not involve conditional expectation (see [12]). Using this representation, one can obtain an approximation for the solution of (6.2) with $a^{ij} = \alpha_0^{ij}$, and the error can also be estimated (see [3]). This motivates the idea of splitting up (6.2) into the equations

$$du(t,x) = L_0(t, Y_t)u(t,x)\,dt + M_k(t, Y_t)u(t,x)\,dY_t^k \tag{6.3}$$

and

$$dv(t,x) = L_1(t, Y_t)v(t,x)\,dt, \tag{6.4}$$

where

$$\begin{aligned} L_0(t,y)\phi(x) &:= D_{ij}\big(\alpha_0^{ij}(t,x,y)\phi(x)\big), \\ L_1(t,y)\phi(x) &:= D_{ij}\big(\alpha_1^{ij}(t,x,y)\phi(x)\big) + D_i\big(h^i(t,x,y)\phi(x)\big), \\ M_k(t,Y_t)\phi(x) &:= H^k(t,x,y)\phi(x) + D_i\big(\rho^{ik}(t,x,y)\phi(x)\big). \end{aligned}$$

Let $\mathbf{P}_t(t_i)\varphi$ denote the solution, starting from φ, of (6.4) with coefficients frozen at $t = t_i$, $Y_t = Y_{t_i}$, where $t_i := Ti/n$. Define the approximations $p_n(t_i)$, $\bar{p}_n(t_i)$ for $t_i \in T_n := \{Ti/n : i := 0, 1, 2, \dots, n\}$ by $p_n(0) = \bar{p}_n(0) := p_0$,

$$p_n(t_{i+1}) := \mathbf{P}_\delta(t_{i+1})\mathbf{Q}_{t_i t_{i+1}} p_n(t_i), \qquad \bar{p}_n(t_{i+1}) := \mathbf{Q}_{t_i t_{i+1}}\mathbf{P}_\delta(t_i)\bar{p}_n(t_i)$$

for $i = 0, 1, 2, \dots, n-1$, where $\delta = T/n$ and $\mathbf{Q}_{st}\varphi$ denotes the solution of (6.3) for $t \ge s$, with initial condition $v(s) = \varphi$. To apply Theorem 5.2 to these approximations, we need the following assumptions for a fixed integer $m \ge 0$ and real number $p \ge 0$.

ASSUMPTION 6.1. The coefficients $\alpha_0 = (\alpha_0^{ij})$ and $\alpha_1 = (\alpha_1^{ij})$ have continuous derivatives in x up to order $m+5$, $h = (h^i)$ and $\rho = (\rho^{ik})$ have continuous derivatives in x up to order $m+4$ and $H = (H^{ik})$ has continuous derivatives in x up to order $m+3$. All these derivatives are bounded by the constant K.

ASSUMPTION 6.2. The derivatives in x of α_1 and h up to order $m+2$ and $m+1$, respectively, have continuous first-order derivatives in t and continuous second-order derivatives in y, which are bounded by the constant K.

ASSUMPTION 6.3. Almost surely $p_0 \in H^{m+3}$ and $E\|p_0\|_{m+3}^p \le K$.

THEOREM 6.2. *Under Assumptions* 6.1–6.3, *there exists a constant* N *depending only on* d, d_0, $\bar{d}$, K, p, m *and* T, *such that*

$$E\max_{t\in T_n}\|p_n(t)-p(t)\|_X^p \le Nn^{-p}, \qquad E\max_{t\in T_n}\|\bar{p}_n(t)-p(t)\|_X^p \le Nn^{-p} \tag{6.5}$$

for all integers $n \ge 1$, *where* $\|\cdot\|_X$ *denotes the norm in* $X := H^m$.

PROOF. We rewrite (6.2) in the form of (2.4) as follows:

$$\begin{aligned} dp(t,x) &= L_0(t,Y_t)p(t,x)\,dt + L_1(t,Y_t)p(t,x)\,dt \\ &\quad + M_k(t,Y_t)p(t,x)\,dW_t^k, \end{aligned} \tag{6.6}$$

where

$$L_r(t,Y_t)\phi(x) := a_r^{ij}(t,x)D_{ij}\phi(x) + a_r^i(t,x)D_i\phi(x) + a_r(t,x)\phi(x),$$
$$M_k(t,Y_t)\phi(x) := b_k^i(t,x)D_i\phi(x) + b_k(t,x)\phi(x),$$

with random coefficients

$$a_0^{ij}(t,x) := \alpha_0^{ij}(t,x,Y_t),$$
$$a_0^i(t,x) := 2D_j\alpha_0^{ij}(t,x,Y_t) + H^k\rho^{ik}(t,x,Y_t),$$
$$a_0(t,x) := D_{ij}\alpha_0^{ij}(t,x,Y_t) + H^kD_i\rho^{ik}(t,x,Y_t) + H^kH^k(t,x,Y_t),$$
$$a_1^{ij}(t,x) := \alpha_1^{ij}(t,x,Y_t), \qquad a_1^i(t,x) := 2D_j\alpha_1^{ij}(t,x,Y_t) + h^i(t,x,Y_t),$$
$$a_1(t,x) := D_{ij}\alpha_1^{ij}(t,x,Y_t) + D_ih^i(t,x,Y_t),$$
$$b_k^i(t,x) := \rho^{ik}(t,x,Y_t), \qquad b_k(t,x) := D_i\rho^{ik}(t,x,Y_t) + H^k(t,x,Y_t).$$

Clearly, (6.6) satisfies Assumption 2.5 with $V_t^0 := t$ and $Y_t^k := W_t^k$, and Assumption 5.1 holds by virtue of the well-known Itô–Wentzell formula. Hence, we can finish the proof by applying Theorem 5.2 and Remark 5.1 to (6.6). □

By Sobolev's embedding and by the Borel–Cantelli lemma, we obtain the following corollary.

COROLLARY 6.3. *If Assumptions* 6.1–6.3 *hold with* $m > d/2 + l$, *where* $l \ge 0$ *is an integer*, *then estimates* (6.5) *also hold with* $X := C^l(\mathbb{R}^d)$ *in place of* H^m. *If*

Assumptions 6.1 *and* 6.2 *hold and* $E\|p_0\|_{m+3}^p < \infty$ *for some* $p > \kappa$ *and* $\kappa > 1$, *then there is a finite random variable* ξ, *such that almost surely*

$$\max_{t\in T_n} \|p_n(t) - p(t)\|_X \le \xi n^{-1+1/\kappa}, \qquad \max_{t\in T_n} \|\bar{p}_n(t) - p(t)\|_X \le \xi n^{-1+1/\kappa}$$

for all $n \ge 1$, *with* $X := H^m$, *and if* $m > l + d/2$, *then also with* $X := C^l$.

REMARK 6.1. In [3] a version of Theorem 6.2 is given in the time-homogeneous situation, when the coefficients of (6.1) are independent of Y_t, $p = 2$, $m = 0$, and with max's in (6.5) being outside of expectations. However, the number of derivatives required in [3] is smaller. We believe that the latter is actually due to some kind of confusion, since in [3] the authors use a theorem from [12] stated for the equations in the usual form, and (6.2) is written in conjugate form.

REMARK 6.2. One could easily consider the most general form of the signal-observation equations (6.1). In particular, we can put a uniformly nondegenerate smooth matrix-valued function $G(t, Y_t)$ in front of dW_t. Then, under natural assumptions on the smoothness of G, one can get a result similar to Theorem 6.2. We have chosen not to deal with these generalizations just for simplicity of notation. Finally, we note that by using weighted Sobolev spaces in place of H^m one can extend our results to the case of SPDEs with unbounded coefficients. These kinds of SPDEs are important from the point of view of applications, in particular, in nonlinear filtering (see, e.g., [7, 15, 16] and the references therein). However, for the sake of simplicity of presentation, we did not want to cover the case of unbounded coefficients in this paper.

Acknowledgments. The first author gratefully acknowledges the warm hospitality and the inspiring scientific atmosphere he enjoyed while visiting the School of Mathematics of the University of Minnesota, where this research was done. The authors are thankful to the referee for careful reading and useful suggestions.

REFERENCES

[1] BENSOUSSAN, A., GLOWINSKI, R. and RASCANU, A. (1989). Approximation of Zakai equation by the splitting-up method. *Stochastic Systems and Optimization (Warsaw 1988). Lecture Notes in Control Inform. Sci.* **136** 257–265. Springer, New York.

[2] BENSOUSSAN, A., GLOWINSKI, R. and RASCANU, A. (1992). Approximation of some stochastic differential equations by the splitting up method. *Appl. Math. Optim.* **25** 81–106.

[3] FLORCHINGER, P. and LE GLAND, F. (1991). Time-discretization of the Zakai equation for diffusion processes observed in correlated noise. *Stochastics Stochastics Rep.* **35** 233–256.

[4] GERMANI, A. and PICCIONI, M. (1988). Semi-discretization of stochastic partial differential equations on $\mathbb{R}^d$ by a finite element technique. *Stochastics* **23** 131–148.

[5] GRECKSCH, W. and KLOEDEN, P. E. (1996). Time-discretized Galerkin approximations of parabolic stochastic PDEs. *Bull. Austral. Math. Soc.* **54** 79–85.
[6] GYÖNGY, I. and KRYLOV, N. (1982). On stochastic equations with respect to semimartingales II. Itô formula in Banach spaces. *Stochastics* **6** 153–174.
[7] GYÖNGY, I. and KRYLOV, N. (1992). Stochastic partial differential equations with unbounded coefficients and applications III. *Stochastics Stochastics Rep.* **40** 75–115.
[8] GYÖNGY, I. and NUALART, D. (1997). Implicit scheme for stochastic parabolic partial differential equations driven by space–time white noise. *Potential Anal.* **7** 725–757.
[9] ITO, K. and ROZOVSKII, B. (2000). Approximation of the Kushner equation for nonlinear filtering. *SIAM J. Control Optim.* **38** 893–915.
[10] KRYLOV, N. (1995). *Introduction to the Theory of Diffusion Processes*. Amer. Math. Soc., Providence, RI.
[11] KRYLOV, N. and ROZOVSKII, B. (1981). Stochastic evolution equations. *J. Soviet Math.* **16** 1233–1277.
[12] KRYLOV, N. and ROZOVSKII, B. (1986). On the characteristics of degenerate second order parabolic Itô equations. *J. Soviet Math*. **32** 336–348.
[13] LOTOTSKY, S. V. (1996). Problems in statistics of stochastic differential equations. Ph.D. dissertation, Univ. Southern California.
[14] NAGASE, N. (1995). Remarks on nonlinear stochastic partial differential equations: An application of the splitting-up method. *SIAM J. Control Optim.* **33** 1716–1730.
[15] ROZOVSKII, B. (1987). On the kinematic dynamo problem in random flow. In *Probability Theory and Mathematical Statistics* **2** 509–516. VNU, Utrecht.
[16] ROZOVSKII, B. (1990). *Stochastic Evolution Systems*: *Linear Theory and Applications to Nonlinear Filtering.* Kluwer, Dordrecht.
[17] YOO, H. (1998). An analytic approach to stochastic differential equations and its applications. Ph.D. dissertation, Univ. Minnesota.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF EDINBURGH
KING'S BUILDINGS
EDINBURGH EH9 3JZ
UNITED KINGDOM
E-MAIL: gyongy@maths.ed.ac.uk

DEPARTMENT OF MATHEMATICS
127 VINCENT HALL
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
E-MAIL: krylov@math.umn.edu