SPACE-TIME APPROXIMATION OF STOCHASTIC p-LAPLACE TYPE SYSTEMS

DOMINIC BREIT AND MARTINA HOFMANOVÁ

ABSTRACT. We consider systems of stochastic evolutionary equations of the p-Laplace type. We establish convergence rates for a finite-element based space-time approximation, where the error is measured in a suitable quasi-norm. Under natural regularity assumptions on the solution, our main result provides linear convergence in space and convergence of order α in time for all $\alpha \in (0, \frac{1}{2})$. The key ingredient of our analysis is a random time-grid, which allows us to compensate for the lack of time regularity.

1. INTRODUCTION

We study the space-time discretization of stochastic evolutionary PDEs of the type

$$\begin{cases} \mathrm{d}\mathbf{u} &= \operatorname{div}\mathbf{S}(\nabla\mathbf{u})\,\mathrm{d}t + \varPhi(\mathbf{u})\,\mathrm{d}W \\ \mathbf{u}(0) &= \mathbf{u}_0 \end{cases}, \tag{1.1}$$

in a bounded Lipschitz domain $\mathcal{O} \subset \mathbb{R}^d$ and a finite time interval $[0,T]$. Here $\mathbf{S}$ is a nonlinear operator with p-growth (see (2.1) for a precise definition), for instance the p-Laplacian

$$\mathbf{S}(\nabla\mathbf{u}) = \big(\kappa + |\nabla\mathbf{u}|\big)^{p-2}\nabla\mathbf{u} \tag{1.2}$$

where $p \in (1,\infty)$ and $\kappa \geq 0$. We assume that W is a cylindrical Wiener process in a Hilbert space defined on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ and $\varPhi$ has linear growth (see Section 2 for more details). This system can be understood as a model for a large class of problems important for applications. We particularly mention the flow of non-Newtonian fluids (the literature devoted to the deterministic setting is very extensive, for the corresponding stochastic counterparts we refer the reader to [5, 30, 34]).

The deterministic analogue of (1.1), which reads as

$$\begin{cases} \partial_t\mathbf{u} &= \operatorname{div}\mathbf{S}(\nabla\mathbf{u}) \\ \mathbf{u}(0) &= \mathbf{u}_0 \end{cases}, \tag{1.3}$$

seems to be well understood. Existence of a unique weak solution follows from the classical monotone operator theory. Also the regularity of solutions is well-known, see [12, 17, 25, 31]. For the numerical approximation of (1.3) one approximates the time-derivative by a difference quotient and solves in every (discrete) time-step a stationary problem. The latter one finally has to be approximated by a finite element method. In order to understand the convergence properties of the scheme it proved beneficial to introduce the nonlinear quantity

$$\mathbf{F}(\boldsymbol{\xi}) = (\kappa + |\boldsymbol{\xi}|)^{\frac{p-2}{2}}\boldsymbol{\xi}, \quad \boldsymbol{\xi} \in \mathbb{R}^{d\times D},$$

Date: October 2, 2019.

Key words and phrases. Parabolic stochastic PDEs, Nonlinear Laplace-type systems, Finite Element Methods, space-time discretization.

Financial support by the German Science Foundation DFG via the Collaborative Research Center SFB1283 is gratefully acknowledged.

which linearizes the approximation error in a certain sense. Starting from the paper [1] it has been known that the correct way to express the error is through the quasi-norm

$$\|\mathbf{F}(\nabla\mathbf{u}) - \mathbf{F}(\nabla\mathbf{v})\|^2_{L^2(\mathcal{O})} \sim \int_{\mathcal{O}} (\kappa + |\nabla\mathbf{u}| + |\nabla\mathbf{v} - \nabla\mathbf{u}|)^{p-2}|\nabla\mathbf{u} - \nabla\mathbf{v}|^2\,\mathrm{d}x$$

as a distance of functions $\mathbf{u},\mathbf{v} \in W^{1,p}(\mathcal{O})$. This led to optimal convergence results for finite element based space-time approximations of (1.3) for $p \geq 2$ in [2] and in [13] for all $1 < p < \infty$. More precisely, it was shown that

$$\max_{1\leq m\leq M} \|\mathbf{u}(t_m) - \mathbf{u}_{h,m}\|_2^2 + \tau\sum_{m=1}^M \|\mathbf{F}(\nabla\mathbf{u}(t_m)) - \mathbf{F}(\nabla\mathbf{u}_{h,m})\|_2^2 \leq c\left(h^2 + \tau^2\right)$$

where $\mathbf{u}_{h,m}$ denotes the discrete solution with space discretization parameter h and time-discretization $\tau = \frac{T}{M}$ at time points $t_m = \tau m$, $m = 1,\dots,M$. The constant c depends on the geometry of $\mathcal{O}$, on p, as well some quantities involving $\mathbf{u}$ arising from the following regularity properties:

$$\begin{aligned}&\mathbf{F}(\nabla\mathbf{u}) \in L^2(0,T;W^{1,2}(\mathcal{O}))),\ \mathbf{F}(\nabla\mathbf{u}) \in W^{1,2}(0,T;L^2(\mathcal{O})),\\&\nabla\mathbf{u} \in L^\infty(0,T;L^2(\mathcal{O})),\ \partial_t\mathbf{u} \in L^\infty(0,T;L^2(\mathcal{O})).\end{aligned}$$

Regarding the stochastic problem there is a lot of literature for the linear case. The literature dedicated to the regularity theory for linear SPDEs (e.g. (1.2) with $p = 2$) is quite extensive, we refer to [22], [23] and [24] and the references therein. One can show that

$$\mathbf{u} \in L^2(\Omega; C^\alpha([0,T]; W^{k,2}(\mathcal{O}))) \quad \forall\alpha \in (0, \tfrac{1}{2}) \tag{1.4}$$

provided the initial datum is smooth and Φ satisfies appropriate assumptions (here, the order $k \in \mathbb{N}$ is mainly determined by the smoothness of Φ). There is also a growing literature for the numerical approximation of linear SPDEs, see e.g. [26, 32, 33]. One possible way is an implicit Euler scheme. Given a finite dimensional subspace $V_h \subset W^{1,2}_0(\mathcal{O})$ and an initial datum $\mathbf{u}_{h,0}$ one computes $\mathbf{u}_{h,m}$ such that

$$\begin{aligned}&\int_{\mathcal{O}} \mathbf{u}_{h,m}\cdot\boldsymbol{\varphi}\,\mathrm{d}x + \tau\int_{\mathcal{O}} \nabla\mathbf{u}_{h,m} : \nabla\boldsymbol{\varphi}\,\mathrm{d}x\\&\qquad = \int_{\mathcal{O}} \mathbf{u}_{h,m-1}\cdot\boldsymbol{\varphi}\,\mathrm{d}x + \int_{\mathcal{O}} \Phi(\mathbf{u}_{h,m-1})\,\Delta_m W\cdot\boldsymbol{\varphi}\,\mathrm{d}x, \quad m = 1,\dots,M,\end{aligned}$$

for every $\boldsymbol{\varphi} \in V_h$, where $\Delta_m W = W(t_m) - W(t_{m-1})$. Based on the regularity (1.4) one can show that the approximation is of order one with respect to the space-discretization and of order α with respect to the time-discretization. The correct error estimator is

$$\mathbb{E}\bigg[\max_{1\leq m\leq M}\int_{\mathcal{O}} |\mathbf{u}(t_m) - \mathbf{u}_{h,m}|^2\,\mathrm{d}x + \tau\sum_{m=1}^M\int_{\mathcal{O}} |\nabla\mathbf{u}(t_m) - \nabla\mathbf{u}_{h,m}|^2\,\mathrm{d}x\bigg].$$

For nonlinear stochastic problems like (1.1) with $p \neq 2$ there is a lot of literature regarding the existence of PDE weak solutions. The popular variational approach by Pardoux [27] provides an existence theory for a quite general class of stochastic evolutionary equations. Existence of (unique) PDE strong solutions to generalized p-Laplace stochastic PDEs with $p \geq 2$ has been proved in [18]. Regularity results for (1.1) with $1 < p < \infty$ have been shown in [4]. In particular, the first author proves that the unique weak solution satisfies[1]

$$\nabla\mathbf{u} \in L^2_{w^*}(\Omega; L^\infty(0,T;L^2(\mathcal{O}))), \quad \mathbf{F}(\nabla\mathbf{u}) \in L^2(\Omega; L^2(0,T;W^{1,2}(\mathcal{O}))), \tag{1.5}$$

at least locally in space and hence $\mathbf{u}$ is a strong solution (in the analytical sense, see Definition 2.3). Hölder-regularity in time of $\mathbf{u}$ (with values in some L^p-space) can be shown directly

[1] Here $L^2_{w^*}(\Omega; L^\infty(0,T;L^2(\mathcal{O})))$ is the space of weak*-measurable mappings $h : \Omega \to L^\infty(0,T;L^2(\mathcal{O}))$ such that $\mathbb{E}\,\mathrm{esssup}_{0\leq t\leq T}\|h\|^2_{L^2(\mathcal{O})} < \infty$.

from the equation once having established the existence of second derivatives. In the linear case a boot-strap argument yields the same for space-derivatives and so (1.4) holds. Both arguments do not apply for $\mathbf{F}(\nabla\mathbf{u})$ if $p \neq 2$. In fact, the best one can hope for is

$$\mathbf{F}(\nabla\mathbf{u}) \in L^2(\Omega; W^{\alpha,2}(0,T; L^2(\mathcal{O}))) \quad \forall \alpha \in (0, \tfrac{1}{2}), \tag{1.6}$$

cf. [6]. The time-regularity is obviously much lower than in the deterministic case (due to the roughness of the driving Wiener process) or the linear stochastic case (due to the limited space-regularity). In fact, the mapping $t \mapsto \mathbf{F}(\nabla\mathbf{u}(t))$ is not expected to be continuous in time (with values in $L^2(\mathcal{O})$). Hence, the natural error estimator

$$\mathbb{E}\bigg[\max_{1\leq m\leq M}\int_{\mathcal{O}}|\mathbf{u}(t_m)-\mathbf{u}_{h,m}|^2\,\mathrm{d}x+\tau\sum_{m=1}^{M}\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(t_m))-\mathbf{F}(\nabla\mathbf{u}_{h,m})|^2\,\mathrm{d}x\bigg]$$

is not well-defined. The same problem appears in the deterministic case if the time regularity of $\mathbf{F}(\nabla\mathbf{u})$ is too low as a consequence of irregular data, cf, [7].

In order to overcome this problem we introduce a random time-grid. To be precise, we consider random time-points $\mathfrak{t}_m$ which are distributed uniformly in $[m\tau-\tau/4, m\tau+\tau/4]$ where $\tau = T/M$. These time points are defined on a probability space $(\hat{\Omega}, \hat{\mathfrak{F}}, \hat{\mathbb{P}})$ which is possibly different from $(\Omega, \mathfrak{F}, \mathbb{P})$, the space where the Brownian motion in (1.1) is defined. We obtain a corresponding error estimator by taking the expectation $\hat{\mathbb{E}}$ with respect to $(\hat{\Omega}, \hat{\mathfrak{F}}, \hat{\mathbb{P}})$. Based on the regularity in (1.5) and (1.6) we prove that

$$\begin{aligned}\hat{\mathbb{E}}\otimes\mathbb{E}\bigg[\max_{1\leq m\leq M}\int_{\mathcal{O}}|\mathbf{u}(\mathfrak{t}_m)-\mathbf{u}_{h,m}|^2\,\mathrm{d}x+\sum_{m=1}^{M}\tau_m\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(\mathfrak{t}_m))-\mathbf{F}(\nabla\mathbf{u}_{h,m})|^2\,\mathrm{d}x\bigg]\\ \leq c(h^2+\tfrac{h^2}{\tau}+\tau^{2\alpha}),\end{aligned}$$

where $\tau_m = \mathfrak{t}_m - \mathfrak{t}_{m-1}$, see Theorem 3.1. In other words, we understand our scheme as a random variable on a product space $\hat{\Omega} \times \Omega$, where $\hat{\omega} \in \hat{\Omega}$ accounts for the randomness introduced through our random time-grid whereas $\omega \in \Omega$ is the randomness coming from the Brownian motion in (1.1). As a matter of fact the convergence rates from the linear problem still hold: we have convergence order one in space and α in time.

As space-time discretizations of evolutionary SPDEs with monotone coefficients were already studied in [19, 20], let us explain what are the main differences with the result that we put forward in the present paper. The class of equations considered in [19, 20] is more general and contains (1.1) as a special case, at least if $p > \frac{2d}{d+2}$. Indeed, the approach in [19, 20] is based on the Gelfand triple $V \hookrightarrow H \hookrightarrow V^*$, which in our case of (1.1) corresponds to $H = L^2(\mathcal{O})$ and $V = W_0^{1,p}(\mathcal{O})$ as the main part of the equation, i.e. given by $\operatorname{div}\mathbf{S}(\nabla\mathbf{u})$, takes values in V^*. Clearly, the situation $p < \frac{2d}{d+2}$ cannot be included as the embedding $W_0^{1,p}(\mathcal{O}) \hookrightarrow L^2(\mathcal{O})$ fails, nevertheless, this drawback does not occur in our approach.

In order to establish the convergence rates in [20], the authors take an additional assumption upon space regularity of the solution which consequently implies a better time regularity. To be more precise, they suppose that

$$\sup_{0\leq t\leq T}\mathbb{E}\|\mathbf{u}\|_{\mathcal{H}}^2+\mathbb{E}\int_0^T\|\mathbf{u}(t)\|_{\mathcal{V}}^2\,\mathrm{d}t\leq C \tag{1.7}$$

for some separable Hilbert spaces $\mathcal{V} \hookrightarrow \mathcal{H} \hookrightarrow V$. This can certainly be expected in case of equations or in the case of systems with Uhlenbeck-structure, that is,

$$\mathbf{S}(\boldsymbol{\xi}) = \nu(|\boldsymbol{\xi}|)\boldsymbol{\xi},$$

which are nondegenerate, i.e. $\kappa > 0$. Indeed, solutions to the corresponding deterministic counterpart can be shown to take values C^∞. In the degenerate case, i.e. if $\kappa = 0$, the situation becomes critical as generally only α-Hölder continuity of $\nabla\mathbf{u}$ holds true for some

$\alpha < 1$. Optimal values for α were obtained in the elliptic setting for $d = 2$ (see [21]). For systems without Uhlenbeck-structure there even exists examples with unbounded solutions in the elliptic case (see [11]). So, in this situation (1.7) implies restrictive conditions on p as in fact only the case $\mathcal{H} = W^{2,2}$ and $V = W^{1,p}$ seems realistic.

Our proof of convergence rates is not based on the assumption (1.7), which allows us to overcome the drawback explained above and to deal with general systems with p-growth. The method relies rather on space and time regularity of $\mathbf{F}(\nabla\mathbf{u})$ which is the natural quantity, the regularity of which can be studied via energy methods. Let us finally mention that our convergence rates are the same as in [20].

In the papers [8, 9] space-time discretizations of stochastic Navier–Stokes equations are considered. The paper [8] contains a convergence analysis whereas in [9] convergence rates – similar to our results – were shown for the two-dimensional space-periodic problem. It would be of great interest to combine this with the results from this paper and to study numerical approximations for generalized Newtonian fluids as done in the deterministic case in [3].

2. Mathematical framework

We now give the precise assumptions on the system (1.1).

2.1. **Nonlinear operator S.** We assume that $\mathbf{S} : \mathbb{R}^{d\times D} \to \mathbb{R}^{d\times D}$ is of class $C^0(\mathbb{R}^{d\times D}) \cap C^1(\mathbb{R}^{d\times D} \setminus \{0\})$ and satisfies

$$\lambda(\kappa + |\boldsymbol{\xi}|)^{p-2}|\boldsymbol{\zeta}|^2 \le \mathrm{D}\mathbf{S}(\boldsymbol{\xi})(\boldsymbol{\zeta},\boldsymbol{\zeta}) \le \Lambda(\kappa + |\boldsymbol{\xi}|)^{p-2}|\boldsymbol{\zeta}|^2 \tag{2.1}$$

for all $\boldsymbol{\xi},\boldsymbol{\zeta} \in \mathbb{R}^{d\times D}$ with some positive constants λ, Λ, some $\kappa \ge 0$ and $p \in (1,\infty)$. It is well known from the deterministic setting (and was already discussed in [4] in the stochastic setting) that an important role for this system is played by the function

$$\mathbf{F}(\boldsymbol{\xi}) = (\kappa + |\boldsymbol{\xi}|)^{\frac{p-2}{2}}\boldsymbol{\xi}.$$

The following two properties are essential for our analysis (see [14] and [15] for a proof).

Lemma 2.1. *Let $p \in (1,\infty)$ and let $\mathbf{S}$ satisfy (2.1).*

(a) *For all $\boldsymbol{\xi},\boldsymbol{\eta} \in \mathbb{R}^{d\times D}$ we have*

$$|\mathbf{F}(\boldsymbol{\xi}) - \mathbf{F}(\boldsymbol{\eta})|^2 \sim \big(\mathbf{S}(\boldsymbol{\xi}) - \mathbf{S}(\boldsymbol{\eta})\big) : (\boldsymbol{\xi} - \boldsymbol{\eta}),$$

where the constants hidden in $\sim$ only depend on p.[2]

(b) *For every $\varepsilon > 0$ and all $\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\zeta} \in \mathbb{R}^{d\times D}$ it holds*

$$\big|\big(\mathbf{S}(\boldsymbol{\xi}) - \mathbf{S}(\boldsymbol{\eta})\big) : (\boldsymbol{\xi} - \boldsymbol{\zeta})\big| \le \varepsilon\big(\mathbf{S}(\boldsymbol{\xi}) - \mathbf{S}(\boldsymbol{\eta})\big) : (\boldsymbol{\xi} - \boldsymbol{\eta}) + c(\varepsilon,p)\big(\mathbf{S}(\boldsymbol{\xi}) - \mathbf{S}(\boldsymbol{\zeta})\big) : (\boldsymbol{\xi} - \boldsymbol{\zeta}).$$

2.2. **Stochastic noise.** Let $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\ge 0}, \mathbb{P})$ be a stochastic basis with a complete, right-continuous filtration. The process W is a cylindrical Wiener process, that is, $W(t) = \sum_{k\ge 1}\beta_k(t)e_k$ with $(\beta_k)_{k\ge1}$ being mutually independent real-valued standard Wiener processes relative to $(\mathfrak{F}_t)_{t\ge0}$ and $(e_k)_{k\ge1}$ a complete orthonormal system in a separable Hilbert space $\mathfrak{U}$. To give the precise definition of the diffusion coefficient $\varPhi$, consider $\mathbf{z} \in L^2(G)$ and let $\varPhi(\mathbf{z}) : \mathfrak{U} \to L^2(\mathcal{O})$ be defined by $\varPhi(\mathbf{z})e_k = \mathbf{g}_k(\cdot, \mathbf{z}(\cdot))$. In particular, we suppose that $\mathbf{g}_k \in C^1(\mathcal{O} \times \mathbb{R}^D)$ and the following conditions

$$\sum_{k\ge1} |\mathbf{g}_k(x,\boldsymbol{\xi})|^2 \le c(1 + |\boldsymbol{\xi}|^2), \qquad \sum_{k\ge1} |\nabla_{\boldsymbol{\xi}}\mathbf{g}_k(x,\boldsymbol{\xi})|^2 \le c, \tag{2.2}$$

$$\sum_{k\ge1} |\nabla_x \mathbf{g}_k(x,\boldsymbol{\xi})|^2 \le c(1 + |\boldsymbol{\xi}|^2), \tag{2.3}$$

[2] We write $f \sim g$ provided $f \le cg \le Cf$ for some constants c, $C > 0$.

for all $x \in \mathcal{O}$ and $\boldsymbol{\xi} \in \mathbb{R}^D$. The conditions imposed on $\varPhi$, particularly the first assumption from (2.2), imply that

$$\varPhi : L^2(\mathcal{O}) \to L_2(\mathfrak{U}; L^2(\mathcal{O})),$$

where $L_2(\mathfrak{U}; L^2(\mathcal{O}))$ denotes the collection of Hilbert-Schmidt operators from $\mathfrak{U}$ to $L^2(\mathcal{O})$. Thus, given a progressively measurable process $\mathbf{u} \in L^2(\Omega; L^2(0,T; L^2(\mathcal{O})))$, the stochastic integral

$$t \mapsto \int_0^t \varPhi(\mathbf{u})\,\mathrm{d}W$$

is a well defined process taking values in $L^2(\mathcal{O})$ (see [10] for a detailed construction). Moreover, we can multiply by test functions to obtain

$$\left\langle \int_0^t \varPhi(\mathbf{u})\,\mathrm{d}W, \boldsymbol{\varphi} \right\rangle = \sum_{k\geq 1} \int_0^t \langle \mathbf{g}_k(\mathbf{u}), \boldsymbol{\varphi}\rangle\,\mathrm{d}\beta_k, \qquad \boldsymbol{\varphi} \in L^2(\mathcal{O}).$$

The initial datum may be random in general, i.e. $\mathfrak{F}_0$-measurable, and we assume at least $\mathbf{u}_0 \in L^2(\Omega; L^2(\mathcal{O}))$.

2.3. **The concept of solution and preliminary results.** In this subsection we recall the definition of weak and strong solution as well as the basic existence, uniqueness and regularity results established in [4].

Definition 2.2 (Weak solution)**.** An $(\mathfrak{F}_t)$-progressively measurable function

$$\mathbf{u} \in L^2(\Omega; C([0,T]; L^2(\mathcal{O}))) \cap L^p(\Omega; L^p(0,T; W_0^{1,p}(\mathcal{O})))$$

is called a weak solution to (1.1) if for every $\boldsymbol{\varphi} \in C_c^\infty(\mathcal{O})$ and all $t \in [0,T]$ it holds true $\mathbb{P}$-a.s.

$$\begin{aligned}\int_{\mathcal{O}} \mathbf{u}(t)\cdot\boldsymbol{\varphi}\,\mathrm{d}x + \int_{\mathcal{O}}\int_0^t \mathbf{S}(\nabla\mathbf{u}(\sigma)) : \nabla\boldsymbol{\varphi}\,\mathrm{d}x\,\mathrm{d}\sigma& \\ = \int_{\mathcal{O}} \mathbf{u}_0\cdot\boldsymbol{\varphi}\,\mathrm{d}x + \int_{\mathcal{O}}\int_0^t \varPhi(\mathbf{u})\,\mathrm{d}W\cdot\boldsymbol{\varphi}\,\mathrm{d}x.&\end{aligned}$$

Definition 2.3 (Strong solution)**.** A weak solution to (1.1) is called a strong solution provided

$$\operatorname{div}\mathbf{S}(\nabla\mathbf{u}) \in L^1(\Omega; L^1_{\mathrm{loc}}((0,T)\times\mathcal{O}))$$

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

$$\mathbf{u}(t) = \mathbf{u}_0 + \int_0^t \operatorname{div}\,\mathbf{S}(\nabla\mathbf{u})\,\mathrm{d}\sigma + \int_0^t \varPhi(\mathbf{u})\,\mathrm{d}W$$

$\mathbb{P}$-a.s.

The following result, which is taken from [4, Theorem 4, Theorem 5], is the starting point of our analysis (we remark that the case $\kappa = 0$ is not included in [5] but can be obtained by approximation).

Theorem 2.4. *Assume $\mathbf{u}_0 \in L^2(\Omega; W_0^{1,2}(\mathcal{O}))$ is an $\mathfrak{F}_0$-measurable random variable. Suppose further that (2.1), (2.2) and (2.3) hold. Then there is a unique weak solution $\mathbf{u}$ to (1.1) which is a strong solution and satisfies*

$$\mathbb{E}\bigg[\sup_{t\in(0,T)}\int_{\mathcal{O}'}|\nabla\mathbf{u}(t)|^2\,\mathrm{d}x + \int_0^T\int_{\mathcal{O}'}|\nabla\mathbf{F}(\nabla\mathbf{u})|^2\,\mathrm{d}x\,\mathrm{d}t\bigg] < \infty$$

for all $\mathcal{O}' \Subset \mathcal{O}$.

It is expected that the theorem above holds even globally in space provided the boundary of $\mathcal{O}$ is smooth enough - at least under appropriate compability conditions such as zero boundary values for the noise (see the counterexamples by Krylov [22] on the global regularity for SPDEs in connection with this). A proof however, is still missing.

2.4. **Discretization in space.** Let $\mathcal{O} \subset \mathbb{R}^d$ be a connected, open domain with polyhedral boundary. We assume that $\partial\mathcal{O}$ is Lipschitz continuous. For an open, bounded (non-empty) set $U \subset \mathbb{R}^d$ we denote by h_U the diameter of U, and by ρ_U the supremum of the diameters of inscribed balls. We denote by $\mathscr{T}_h$ be a simplicial subdivision of $\mathcal{O}$ with

$$h = \max_{\mathcal{S}\in\mathscr{T}_h} h_{\mathcal{S}}$$

is non-degenerate:

$$\max_{\mathcal{S}\in\mathscr{T}_h} \frac{h_{\mathcal{S}}}{\rho_{\mathcal{S}}} \leq \gamma_0. \tag{2.4}$$

For $\mathcal{S} \in \mathscr{T}_h$ we define the set of neighbors $N_{\mathcal{S}}$ and the neighborhood S_T by

$$N_{\mathcal{S}} := \{\mathcal{S}' \in \mathscr{T}_h \,:\, \overline{\mathcal{S}'} \cap \overline{\mathcal{S}} \neq \emptyset\}, \quad \mathcal{M}_{\mathcal{S}} := \text{interior} \bigcup_{\mathcal{S}'\in N_{\mathcal{S}}} \overline{\mathcal{S}'}.$$

Note that for all $\mathcal{S}, \mathcal{S}' \in \mathscr{T}_h$: $\mathcal{S}' \subset \overline{\mathcal{M}_{\mathcal{S}}} \Leftrightarrow \mathcal{S} \subset \overline{\mathcal{M}_{\mathcal{S}'}} \Leftrightarrow \overline{\mathcal{S}} \cap \overline{\mathcal{S}'} \neq \emptyset$. Due to our assumption on $\mathcal{O}$ the $\mathcal{M}_{\mathcal{S}}$ are connected, open bounded sets.

It is easy to see that the non-degeneracy (2.4) of $\mathscr{T}_h$ implies the following properties, where the constants are independent of h:

(a) $|\mathcal{M}_{\mathcal{S}}| \sim |\mathcal{S}|$ for all $\mathcal{S} \in \mathscr{T}_h$.
(b) There exists $m_1 \in \mathbb{N}$ such that $\#N_{\mathcal{S}} \leq m_1$ for all $\mathcal{S} \in \mathscr{T}_h$.

For $\mathcal{O} \subset \mathbb{R}^d$ and $\ell \in \mathbb{N}_0$ we denote by $\mathscr{P}_\ell(\mathcal{O})$ the polynomials on $\mathcal{O}$ of degree less than or equal to ℓ. Moreover, we set $\mathscr{P}_{-1}(\mathcal{O}) := \{0\}$. Let us characterize the finite element space V_h as

$$V_h := \{\mathbf{v} \in L^1_{\mathrm{loc}}(\mathcal{O}) \,:\, \mathbf{v}|_{\mathcal{S}} \in (\mathscr{P}_1(\mathcal{S}))^D \ \forall \mathcal{S} \in \mathscr{T}_h\}.$$

We will now state assumption on an interpolation operator between the continuous and discrete function spaces (satisfied e.g. by the Scott-Zhang operator [29]). More precisely, we assume the following.[3]

Assumption 2.5. *Let $\Pi_h : (W^{1,1}(\mathcal{O}))^D \to V_h$ be such that the following holds.*

(a) *There holds uniformly in $\mathcal{S} \in \mathscr{T}_h$ and $\mathbf{v} \in (W^{1,1}(\mathcal{O}))^D$*

$$\fint_{\mathcal{S}} |\Pi_h \mathbf{v}| \,\mathrm{d}x + \fint_{\mathcal{S}} |h_{\mathcal{S}} \nabla \Pi_h \mathbf{v}| \,\mathrm{d}x \leq c\, h_{\mathcal{S}} \fint_{\mathcal{M}_{\mathcal{S}}} |\nabla \mathbf{v}| \,\mathrm{d}x.$$

(b) *For all $\mathbf{v} \in (\mathscr{P}_1(\mathcal{O}))^D$ it holds*

$$\Pi_h \mathbf{v} = \mathbf{v}.$$

An easy consequence of this assumption is the inequality

$$\fint_{\mathcal{S}} \Big| \frac{\mathbf{v} - \Pi_h \mathbf{v}}{h_{\mathcal{S}}} \Big|^2 \mathrm{d}x + \fint_{\mathcal{S}} |\nabla \mathbf{v} - \nabla \Pi_h \mathbf{v}|^2 \,\mathrm{d}x \leq c \fint_{\mathcal{M}_{\mathcal{S}}} |\nabla \mathbf{v}|^2 \,\mathrm{d}x \tag{2.6}$$

for all $v \in (W^{1,2}(\mathcal{M}_{\mathcal{S}}))^D$. The following crucial estimate is shown in [16, Theorem 5.7]

Lemma 2.5. *Let Π_h satisfy Assumption 2.5. Let $\mathbf{v} \in (W^{1,p}(\mathcal{O}))^D$ then for all $T \in \mathscr{T}_h$ holds*

$$\fint_{\mathcal{S}} \big|\mathbf{F}(\nabla \mathbf{v}) - \mathbf{F}(\nabla \Pi_h \mathbf{v})\big|^2 \,\mathrm{d}x \leq c \inf_{\mathbf{Q}\in\mathbb{R}^{D\times d}} \fint_{\mathcal{M}_{\mathcal{S}}} \big|\mathbf{F}(\nabla \mathbf{v}) - \mathbf{F}(\mathbf{Q})\big|^2 \,\mathrm{d}x,$$

with c depending only on p and γ_0. In particular, if $\mathbf{F}(\nabla \mathbf{v}) \in (W^{1,2}(\mathcal{O}))^{D\times d}$ then

$$\fint_{\mathcal{S}} \big|\mathbf{F}(\nabla \mathbf{v}) - \mathbf{F}(\nabla \Pi_h \mathbf{v})\big|^2 \,\mathrm{d}x \leq c\, h_{\mathcal{S}}^2 \fint_{\mathcal{M}_{\mathcal{S}}} \big|\nabla \mathbf{F}(\nabla \mathbf{v})\big|^2 \,\mathrm{d}x. \tag{2.7}$$

[3] We denote by $\fint_A f \,\mathrm{d}x = |A|^{-1} \int_A f \,\mathrm{d}x$ the mean value of a integrable function f over the set A.

3. Finite element based space-time approximation

With the preparations from the previous section at hand, we are able to formulate our algorithm for the space-time approximation of (1.1). We construct a random partition of $[0,T]$ with average mesh size $\tau = T/M$ as follows. Let $\mathfrak{t}_0 = 0$ and let $\mathfrak{t}_m$ for $m = 1, \dots, M$ be independent random variables such that $\mathfrak{t}_m$ is distributed uniformly in $[m\tau - \tau/4, m\tau + \tau/4]$. We assume that the random variables τ_m are defined on a probability space $(\hat{\Omega}, \hat{\mathfrak{F}}, \hat{\mathbb{P}})$ and we denote by $\hat{\mathbb{E}}$ the corresponding expectation. Let $\tau_m = \mathfrak{t}_m - \mathfrak{t}_{m-1}$ and observe that τ_m is a random variable satisfying $\tau/2 \le \tau_m \le 3\tau/2$. Let $\mathbf{u}_{h,0} := \Pi_h \mathbf{u}_0$ and for every $m \in \{1, \dots, M\}$ find $\mathbf{u}_{h,m} \in L^2(\Omega; V_h)$ such that for every $\boldsymbol{\varphi} \in V_h$ it holds true $\mathbb{P}$-a.s.

$$\begin{aligned}\int_{\mathcal{O}} \mathbf{u}_{h,m} \cdot \boldsymbol{\varphi} \, \mathrm{d}x + \tau_m \int_{\mathcal{O}} \mathbf{S}(\nabla \mathbf{u}_{h,m}) : \nabla \boldsymbol{\varphi} \, \mathrm{d}x \\ = \int_{\mathcal{O}} \mathbf{u}_{h,m-1} \cdot \boldsymbol{\varphi} \, \mathrm{d}x + \int_{\mathcal{O}} \Phi(\mathbf{u}_{h,m-1}) \, \Delta_m W \cdot \boldsymbol{\varphi} \, \mathrm{d}x,\end{aligned} \tag{3.1}$$

where $\Delta_m W = W(\mathfrak{t}_m) - W(\mathfrak{t}_{m-1})$.

3.1. **Error analysis.** In this subsection we establish convergence with rates of the above defined algorithm.

Theorem 3.1. *Let $\mathbf{u}$ be the unique weak solution to (1.1) in the sense of Definition 2.2 where $\mathbf{u}_0 \in L^2(\Omega, W^{1,2}_0(\mathcal{O}))$ is $\mathfrak{F}_0$-measurable. Suppose that (2.2) hold. Finally, assume that*

$$\mathbf{F}(\nabla \mathbf{u}) \in L^2(\Omega; L^2(0,T; W^{1,2}(\mathcal{O}))), \nabla \mathbf{u} \in L^2_{w^*}(\Omega; L^\infty(0,T; L^2(\mathcal{O}))), \tag{3.2}$$

$$\mathbf{F}(\nabla \mathbf{u}) \in L^2(\Omega; W^{\alpha,2}(0,T; L^2(\mathcal{O}))), \ \mathbf{u} \in L^2(\Omega; C^\alpha([0,T]; L^2(\mathcal{O}))), \tag{3.3}$$

where $\alpha \in (0, \frac{1}{2})$. Then we have

$$\begin{aligned}\hat{\mathbb{E}} \otimes \mathbb{E}\bigg[\max_{1 \le m \le M} \int_{\mathcal{O}} |\mathbf{u}(\mathfrak{t}_m) - \mathbf{u}_{h,m}|^2 \, \mathrm{d}x + \sum_{m=1}^M \tau_m \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathfrak{t}_m)) - \mathbf{F}(\nabla \mathbf{u}_{h,m})|^2 \, \mathrm{d}x \bigg] \\ \le c \left(h^2 + \tfrac{h^2}{\tau} + \tau^{2\alpha} \right),\end{aligned}$$

where $\mathbf{u}_{h,m}$ is the numerical solution to (1.1) given by (3.1) .

Remark 3.2. For $h \le c\tau^{\alpha + \frac{1}{2}}$ we gain the optimal convergence order of α.

The rest of the paper is devoted to the proof of Theorem 3.1.

Proof of Theorem 3.1. Define the error $\mathbf{e}_m = \mathbf{u}(\mathfrak{t}_m) - \mathbf{u}_{h,m}$. Subtracting (3.1) from the weak formulation of (1.1) we obtain

$$\begin{aligned}\int_{\mathcal{O}} \mathbf{e}_m \cdot \boldsymbol{\varphi} \, \mathrm{d}x + \int_{\mathfrak{t}_{m-1}}^{\mathfrak{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}) - \mathbf{S}(\nabla \mathbf{u}_{h,m}) \Big) : \nabla \boldsymbol{\varphi} \, \mathrm{d}x \, \mathrm{d}\sigma \\ = \int_{\mathcal{O}} \mathbf{e}_{m-1} \cdot \boldsymbol{\varphi} \, \mathrm{d}x + \int_{\mathcal{O}} \bigg(\int_{\mathfrak{t}_{m-1}}^{\mathfrak{t}_m} \Phi(\mathbf{u}) \, \mathrm{d}W - \Phi(\mathbf{u}_{h,m-1}) \, \Delta_m W \bigg) \cdot \boldsymbol{\varphi} \, \mathrm{d}x.\end{aligned}$$

for every $\boldsymbol{\varphi} \in V_h$ or equivalently

$$\begin{aligned}\int_{\mathcal{O}} \mathbf{e}_m \cdot \boldsymbol{\varphi} \, \mathrm{d}x + \tau_m \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(t_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m}) \Big) : \nabla \boldsymbol{\varphi} \, \mathrm{d}x \\ = \int_{\mathfrak{t}_{m-1}}^{\mathfrak{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(t_m)) - \mathbf{S}(\nabla \mathbf{u}(\sigma)) \Big) : \nabla \boldsymbol{\varphi} \, \mathrm{d}x \, \mathrm{d}\sigma + \int_G \mathbf{e}_{m-1} \cdot \boldsymbol{\varphi} \, \mathrm{d}x \\ + \int_{\mathcal{O}} \bigg(\int_{\mathfrak{t}_{m-1}}^{\mathfrak{t}_m} \Phi(\mathbf{u}) \, \mathrm{d}W - \Phi(\mathbf{u}_{h,m-1}) \, \Delta_m W \bigg) \cdot \boldsymbol{\varphi} \, \mathrm{d}x.\end{aligned}$$

Setting $\boldsymbol{\varphi} = \mathbf{w}_m - \mathbf{u}_{h,m}$ (with $\mathbf{w}_m \in V_h$ to be chosen later) we gain

$$\begin{aligned}
&\int_{\mathcal{O}} |\mathbf{e}_m|^2 \,\mathrm{d}x + \tau_m \int_G \Big(\mathbf{S}(\nabla \mathbf{u}(t_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m})\Big) : \nabla \mathbf{e}_m \,\mathrm{d}x \\
&= \int_{\mathcal{O}} \mathbf{e}_m \cdot (\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x + \tau_m \int_G \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m})\Big) : \nabla(\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x \\
&+ \int_{\mathcal{O}} \mathbf{e}_m \cdot (\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x + \tau_m \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m})\Big) : \nabla(\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x \\
&= \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}(\sigma))\Big) : \nabla(\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x \,\mathrm{d}\sigma + \int_{\mathcal{O}} \mathbf{e}_{m-1} \cdot (\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x \\
&+ \int_{\mathcal{O}} \bigg(\int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \varPhi(\mathbf{u}) \,\mathrm{d}W - \varPhi(\mathbf{u}_{h,m-1}) \,\Delta_m W \bigg) \cdot (\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x \\
&+ \int_{\mathcal{O}} \mathbf{e}_m \cdot (\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x + \tau_m \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m})\Big) : \nabla(\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x.
\end{aligned}$$

Such that

$$\begin{aligned}
&\int_{\mathcal{O}} \mathbf{e}_m \cdot (\mathbf{e}_m - \mathbf{e}_{m-1}) \,\mathrm{d}x + \tau_m \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m})\Big) : \nabla \mathbf{e}_m \,\mathrm{d}x \\
&= \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}(\sigma))\Big) : \nabla(\mathbf{w}_m - \mathbf{u}(\mathrm{t}_m)) \,\mathrm{d}x \,\mathrm{d}\sigma + \int_{\mathcal{O}} \mathbf{e}_{m-1} \cdot (\mathbf{w}_m - \mathbf{u}(\mathrm{t}_m)) \,\mathrm{d}x \\
&+ \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}(\sigma))\Big) : \nabla(\mathbf{u}(\mathrm{t}_m) - \mathbf{u}_{h,m}) \,\mathrm{d}x \,\mathrm{d}\sigma \\
&+ \int_{\mathcal{O}} \bigg(\int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \varPhi(\mathbf{u}) \,\mathrm{d}W - \varPhi(\mathbf{u}_{h,m-1}) \,\Delta_m W \bigg) \cdot (\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x \\
&+ \int_{\mathcal{O}} \mathbf{e}_m \cdot (\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x + \tau_m \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m})\Big) : \nabla(\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x.
\end{aligned}$$

Now, we apply the identity $\mathbf{a} \cdot (\mathbf{a} - \mathbf{b}) = \frac{1}{2}\big(|\mathbf{a}|^2 - |\mathbf{b}|^2 + |\mathbf{a} - \mathbf{b}|^2\big)$ (which holds for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, here with $\mathbf{a} = \mathbf{e}_m$ and $\mathbf{b} = \mathbf{e}_{m-1}$) to the first term to gain

$$\begin{aligned}
&\int_{\mathcal{O}} \frac{1}{2}\big(|\mathbf{e}_m|^2 - |\mathbf{e}_{m-1}|^2 + |\mathbf{e}_m - \mathbf{e}_{m-1}|^2\big) \,\mathrm{d}x + \tau_m \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m})\Big) : \nabla \mathbf{e}_m \,\mathrm{d}x \\
&= \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}(\sigma))\Big) : \nabla(\mathbf{w}_m - \mathbf{u}(\mathrm{t}_m)) \,\mathrm{d}x \,\mathrm{d}\sigma \\
&+ \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}(\sigma))\Big) : \nabla(\mathbf{u}(\mathrm{t}_m) - \mathbf{u}_{h,m}) \,\mathrm{d}x \,\mathrm{d}\sigma \\
&+ \int_{\mathcal{O}} \bigg(\int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \varPhi(\mathbf{u}) \,\mathrm{d}W - \varPhi(\mathbf{u}_{h,m-1}) \,\Delta_m W \bigg) \cdot (\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x \\
&+ \int_{\mathcal{O}} (\mathbf{e}_m - \mathbf{e}_{m-1}) \cdot (\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x + \tau_m \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}_{h,m})\Big) : \nabla(\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x \\
&= I_1 + \cdots + I_5.
\end{aligned}$$

Applying Lemma 2.1 we have

$$\begin{aligned}
I_1 \leq \varepsilon &\int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{w}_m)\Big) : \nabla(\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m) \,\mathrm{d}x \,\mathrm{d}\sigma \\
&+ c_\varepsilon \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} \Big(\mathbf{S}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{S}(\nabla \mathbf{u}(\sigma))\Big) : \nabla(\mathbf{u}(\mathrm{t}_m) - \mathbf{u}(\sigma)) \,\mathrm{d}x \,\mathrm{d}\sigma
\end{aligned}$$

$$\leq c\,\varepsilon\,\tau_m \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{w}_m)|^2 \,\mathrm{d}x + c_\varepsilon \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{u}(\sigma))|^2 \,\mathrm{d}x \,\mathrm{d}\sigma$$

and similarly

$$I_2 \leq \varepsilon\tau_m \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{u}_{h,m})|^2 \,\mathrm{d}x + c_\varepsilon \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{u}(\sigma))|^2 \,\mathrm{d}x \,\mathrm{d}\sigma,$$

$$I_5 \leq \varepsilon\tau_m \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{u}_{h,m})|^2 \,\mathrm{d}x + c_\varepsilon\tau_m \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{w}_m))|^2 \,\mathrm{d}x,$$

for every $\varepsilon > 0$. Young's inequality yields

$$I_4 \leq \varepsilon \int_{\mathcal{O}} |\mathbf{e}_m - \mathbf{e}_{m-1}|^2 \,\mathrm{d}x + c_\varepsilon \int_{\mathcal{O}} |\mathbf{w}_m - \mathbf{u}(\mathrm{t}_m)|^2 \,\mathrm{d}x.$$

Plugging all together and choosing ε small enough (to absorb the corresponding terms to the left hand side) we have shown

$$\begin{aligned}
&\int_{\mathcal{O}} |\mathbf{e}_m|^2 \,\mathrm{d}x + c \int_{\mathcal{O}} |\mathbf{e}_m - \mathbf{e}_{m-1}|^2 \,\mathrm{d}x + c\tau_m \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{u}_{h,m})|^2 \,\mathrm{d}x \\
&\quad \leq c \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{u}(\sigma))|^2 \,\mathrm{d}x \,\mathrm{d}\sigma + \int_{\mathcal{O}} |\mathbf{e}_{m-1}|^2 \,\mathrm{d}x \\
&\qquad + c \int_{\mathcal{O}} \left(\int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \varPhi(\mathbf{u}) \,\mathrm{d}W - \varPhi(\mathbf{u}_{h,m-1}) \,\Delta_m W \right) \cdot (\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x \\
&\qquad + c \int_{\mathcal{O}} |\mathbf{u}(\mathrm{t}_m) - \mathbf{w}_m|^2 \,\mathrm{d}x + c\,\tau_m \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{w}_m)|^2 \,\mathrm{d}x
\end{aligned}$$

for every $\mathbf{w}_m \in V_h$. Now we choose $\mathbf{w}_m = \Pi_h \mathbf{u}(\mathrm{t}_m)$ and gain by (2.6), (2.7) and the assumptions on $\mathbf{F}(\nabla \mathbf{u})$

$$\begin{aligned}
&\int_{\mathcal{O}} |\mathbf{e}_m|^2 \,\mathrm{d}x + c \int_{\mathcal{O}} |\mathbf{e}_m - \mathbf{e}_{m-1}|^2 \,\mathrm{d}x + c\,\tau_m \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{u}_m)|^2 \,\mathrm{d}x \\
&\quad \leq c \int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m)) - \mathbf{F}(\nabla \mathbf{u}(\sigma))|^2 \,\mathrm{d}x \,\mathrm{d}\sigma + \int_{\mathcal{O}} |\mathbf{e}_{m-1}|^2 \,\mathrm{d}x \\
&\qquad + c\,h^2 \int_{\mathcal{O}} |\nabla \mathbf{u}(\mathrm{t}_m)|^2 \,\mathrm{d}x + c\,\tau_m h^2 \int_{\mathcal{O}} |\nabla \mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_m))|^2 \,\mathrm{d}x \\
&\qquad + c \int_{\mathcal{O}} \left(\int_{\mathrm{t}_{m-1}}^{\mathrm{t}_m} \varPhi(\mathbf{u}) \,\mathrm{d}W - \varPhi(\mathbf{u}_{h,m-1}) \,\Delta_m W \right) \cdot (\mathbf{w}_m - \mathbf{u}_{h,m}) \,\mathrm{d}x.
\end{aligned}$$

Iterating this inequality yields

$$\begin{aligned}
&\int_{\mathcal{O}} |\mathbf{e}_m|^2 \,\mathrm{d}x + c \sum_{n=1}^{m} \int_{\mathcal{O}} |\mathbf{e}_m - \mathbf{e}_{m-1}|^2 \,\mathrm{d}x + c \sum_{n=1}^{m} \tau_n \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_n)) - \mathbf{F}(\nabla \mathbf{u}_{h,n})|^2 \,\mathrm{d}x \\
&\quad \leq c \sum_{n=1}^{m} \int_{\mathrm{t}_{n-1}}^{\mathrm{t}_n} \int_{\mathcal{O}} |\mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_n)) - \mathbf{F}(\nabla \mathbf{u}(\sigma))|^2 \,\mathrm{d}x \,\mathrm{d}\sigma + \int_{\mathcal{O}} |\mathbf{e}_0|^2 \,\mathrm{d}x \\
&\qquad + c \sum_{n=1}^{m} \int_{\mathcal{O}} \left(\int_{\mathrm{t}_{n-1}}^{\mathrm{t}_n} \varPhi(\mathbf{u}) \,\mathrm{d}W - \varPhi(\mathbf{u}_{h,n-1}) \,\Delta_n W \right) \cdot (\mathbf{w}_n - \mathbf{u}_{h,n}) \,\mathrm{d}x \\
&\qquad + c \frac{h^2}{\tau} \sum_{n=1}^{m} \tau_n \int_{\mathcal{O}} |\nabla \mathbf{u}(\mathrm{t}_n)|^2 \,\mathrm{d}x + c\,h^2 \sum_{n=1}^{m} \tau_n \int_{\mathcal{O}} |\nabla \mathbf{F}(\nabla \mathbf{u}(\mathrm{t}_n))|^2 \,\mathrm{d}x.
\end{aligned}$$

It remains to estimate the stochastic term which we call $\mathscr{M}_m$ and write as

$$\begin{aligned}\mathscr{M}_m &= \sum_{n=1}^{m}\int_{\mathcal{O}}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\big(\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\big)\,\mathrm{d}W\cdot(\mathbf{w}_n-\mathbf{u}_{h,n})\,\mathrm{d}x\\
&= \sum_{n=1}^{m}\int_{\mathcal{O}}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\big(\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\big)\,\mathrm{d}W\cdot(\mathbf{u}(\mathfrak{t}_n)-\mathbf{u}_{h,n})\,\mathrm{d}x\\
&\quad+\sum_{n=1}^{m}\int_{\mathcal{O}}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\big(\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\big)\,\mathrm{d}W\cdot(\mathbf{w}_n-\mathbf{u}(\mathfrak{t}_n))\,\mathrm{d}x\\
&= \sum_{n=1}^{m}\int_{\mathcal{O}}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\big(\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\big)\,\mathrm{d}W\cdot\mathbf{e}_{n-1}\,\mathrm{d}x\\
&\quad+\sum_{n=1}^{m}\int_{\mathcal{O}}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\big(\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\big)\,\mathrm{d}W\cdot(\mathbf{e}_n-\mathbf{e}_{n-1})\,\mathrm{d}x\\
&\quad+\sum_{n=1}^{m}\int_{\mathcal{O}}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\big(\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\big)\,\mathrm{d}W\cdot(\mathbf{w}_n-\mathbf{u}(\mathfrak{t}_n))\,\mathrm{d}x\\
&= \mathscr{M}_{m,1}+\mathscr{M}_{m,2}+\mathscr{M}_{m,3}.\end{aligned}$$

For $\mathscr{M}_{m,1}$ we gain by the Burgholder-Davis-Gundy inequality

$$\begin{aligned}\mathbb{E}\bigg[\max_{1\leq m\leq M}\big|\mathscr{M}_{m,1}\big|\bigg] &\leq c\,\mathbb{E}\bigg[\sum_{n=1}^{M}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\|\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\|^2_{L_2(\mathfrak{U},L^2(\mathcal{O}))}\|\mathbf{e}_{n-1}\|^2_{L^2(\mathcal{O})}\,\mathrm{d}t\bigg]^{\frac12}\\
&\leq c\,\mathbb{E}\bigg[\max_{1\leq n\leq M}\|\mathbf{e}_n\|_{L^2(\mathcal{O})}\bigg(\sum_{n=1}^{M}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\|\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\|^2_{L_2(\mathfrak{U},L^2(\mathcal{O}))}\,\mathrm{d}t\bigg)^{\frac12}\bigg]\\
&\leq \varepsilon\,\mathbb{E}\bigg[\max_{1\leq n\leq M}\|\mathbf{e}_n\|^2_{L^2(\mathcal{O})}\bigg]+c_\varepsilon\,\mathbb{E}\bigg[\sum_{n=1}^{M}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\|\mathbf{u}-\mathbf{u}_{h,n-1}\|^2_{L^2(\mathcal{O})}\,\mathrm{d}t\bigg]\\
&\leq 2\varepsilon\,\mathbb{E}\bigg[\max_{1\leq n\leq M}\|\mathbf{e}_n\|^2_{L^2(\mathcal{O})}\bigg]\\
&\quad+c_\varepsilon\,\mathbb{E}\bigg[\sum_{n=1}^{M}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\|\mathbf{u}-\mathbf{u}(\mathfrak{t}_{n-1})\|^2_{L^2(\mathcal{O})}\,\mathrm{d}t\bigg]+c_\varepsilon\,\mathbb{E}\bigg[\sum_{n=1}^{M}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\|\mathbf{u}(\mathfrak{t}_{n-1})-\mathbf{u}_{h,n-1}\|^2_{L^2(\mathcal{O})}\,\mathrm{d}t\bigg]\end{aligned}$$

Here, we also used (2.2) as well as Young's inequality for $\varepsilon>0$ arbitrary. Applying (2.6) and (3.3) we gain

$$\mathbb{E}\bigg[\max_{1\leq m\leq M}\big|\mathscr{M}_{m,1}\big|\bigg]\leq 2\varepsilon\,\mathbb{E}\bigg[\max_{1\leq n\leq M}\|\mathbf{e}_n\|^2_{L^2(\mathcal{O})}\bigg]+c_\varepsilon\tau+c_\varepsilon\,\mathbb{E}\bigg[\sum_{n=1}^{M}\tau_n\|\mathbf{e}_{n-1}\|^2_{L^2(\mathcal{O})}\bigg].$$

The remaining two terms are estimated using similar arguments as follows

$$\begin{aligned}\mathbb{E}\bigg[\max_{1\leq m\leq M}|\mathscr{M}_{m,2}|\bigg]&\leq\mathbb{E}\bigg[\sum_{n=1}^{M}\bigg(\varepsilon\|\mathbf{e}_n-\mathbf{e}_{n-1}\|^2_{L^2(\mathcal{O})}+c_\varepsilon\bigg\|\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\big(\Phi(\mathbf{u})-\Phi(\mathbf{u}_{h,n-1})\big)\,\mathrm{d}W\bigg\|^2_{L^2(\mathcal{O})}\bigg)\bigg]\\
&\leq\varepsilon\mathbb{E}\bigg[\sum_{n=1}^{M}\|\mathbf{e}_n-\mathbf{e}_{n-1}\|^2_{L^2(\mathcal{O})}\bigg]+c_\varepsilon\,\mathbb{E}\bigg[\sum_{n=1}^{M}\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\|\mathbf{u}-\mathbf{u}_{h,n-1}\|^2_{L^2(\mathcal{O})}\,\mathrm{d}t\bigg]\\
&\leq\varepsilon\mathbb{E}\bigg[\sum_{n=1}^{M}\|\mathbf{e}_n-\mathbf{e}_{n-1}\|^2_{L^2(\mathcal{O})}\bigg]+c_\varepsilon\tau+c_\varepsilon\,\mathbb{E}\bigg[\sum_{n=1}^{M}\tau_n\|\mathbf{e}_{n-1}\|^2_{L^2(\mathcal{O})}\bigg],\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\bigg[\max_{1\le m\le M}|\mathscr{M}_{m,3}|\bigg] &\le c\,\mathbb{E}\bigg[\sum_{n=1}^{M}\bigg(\|\mathbf{w}_n-\mathbf{u}(\mathfrak{t}_n)\|_{L^2(\mathcal{O})}^2+\bigg\|\int_{\mathfrak{t}_{n-1}}^{\mathfrak{t}_n}\big(\Phi(\mathbf{u})-\Phi(\mathbf{u}_{n-1})\big)\,\mathrm{d}W\bigg\|_{L^2(\mathcal{O})}^2\bigg)\bigg]\\
&\le c\,\frac{h^2}{\tau}+c\tau+c\,\mathbb{E}\bigg[\sum_{n=1}^{M}\tau_n\|\mathbf{e}_{n-1}\|_{L^2(\mathcal{O})}^2\bigg].
\end{aligned}
$$

Hence we may apply the discrete Gronwall lemma, choose ε sufficiently small and apply (2.6) to $\mathbf{e}_0$ to deduce

$$
\begin{aligned}
&\mathbb{E}\bigg[\max_{1\le m\le M}\int_{\mathcal{O}}|\mathbf{e}_m|^2\,\mathrm{d}x+\sum_{m=1}^{M}\tau_m\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(\mathfrak{t}_m))-\mathbf{F}(\nabla\mathbf{u}_m)|^2\,\mathrm{d}x\bigg]\\
&\quad\le c\mathbb{E}\bigg[\sum_{m=1}^{M}\int_{\mathfrak{t}_{m-1}}^{\mathfrak{t}_m}\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(\mathfrak{t}_m))-\mathbf{F}(\nabla\mathbf{u}(\sigma))|^2\,\mathrm{d}x\,\mathrm{d}\sigma\bigg]+c\,\frac{h^2}{\tau}+c\tau\\
&\qquad+c\,h^2\,\mathbb{E}\bigg[\sum_{m=1}^{M}\tau_m\int_{\mathcal{O}}|\nabla\mathbf{F}(\nabla\mathbf{u}(\mathfrak{t}_m))|^2\,\mathrm{d}x\bigg].
\end{aligned}
$$

Now we observe that due to the construction of the points $\mathfrak{t}_m$, $m=1,\dots,M-1$, as independent uniformly distributed random variables, the expectation $\hat{\mathbb{E}}$ can be computed explicitly as follows

$$
\begin{aligned}
&\hat{\mathbb{E}}\otimes\mathbb{E}\bigg[\sum_{m=1}^{M}\int_{\mathfrak{t}_{m-1}}^{\mathfrak{t}_m}\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(\mathfrak{t}_m))-\mathbf{F}(\nabla\mathbf{u}(\sigma))|^2\,\mathrm{d}x\,\mathrm{d}\sigma\bigg]\\
&=\hat{\mathbb{E}}\otimes\mathbb{E}\bigg[\int_0^{\mathfrak{t}_1}\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(\mathfrak{t}_1))-\mathbf{F}(\nabla\mathbf{u}(\sigma))|^2\,\mathrm{d}x\,\mathrm{d}\sigma\bigg]\\
&\quad+\sum_{m=2}^{M}\hat{\mathbb{E}}\otimes\mathbb{E}\bigg[\int_{\mathfrak{t}_{m-1}}^{\mathfrak{t}_m}\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(\mathfrak{t}_m))-\mathbf{F}(\nabla\mathbf{u}(\sigma))|^2\,\mathrm{d}x\,\mathrm{d}\sigma\bigg]\\
&=\frac{2}{\tau}\mathbb{E}\int_{\tau-\tau/4}^{\tau+\tau/4}\int_0^{\xi}\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(\xi))-\mathbf{F}(\nabla\mathbf{u}(\sigma))|^2\,\mathrm{d}x\,\mathrm{d}\sigma\,\mathrm{d}\xi\\
&\quad+\frac{4}{\tau^2}\sum_{m=2}^{M}\int_{(m-1)\tau-\tau/4}^{(m-1)\tau+\tau/4}\mathbb{E}\int_{m\tau-\tau/4}^{m\tau+\tau/4}\int_{\zeta}^{\xi}\int_{\mathcal{O}}|\mathbf{F}(\nabla\mathbf{u}(\xi))-\mathbf{F}(\nabla\mathbf{u}(\sigma))|^2\,\mathrm{d}x\,\mathrm{d}\sigma\,\mathrm{d}\xi\,\mathrm{d}\zeta\\
&\le c\tau^{2\alpha}\mathbb{E}\|\mathbf{F}(\nabla\mathbf{u})\|_{W^{\alpha,2}(0,\tau+\tau/4;L^2(\mathcal{O}))}^2\\
&\quad+\frac{c\tau^{2\alpha}}{\tau}\sum_{m=2}^{M}\int_{(m-1)\tau-\tau/4}^{(m-1)\tau+\tau/4}\mathrm{d}\zeta\,\mathbb{E}\|\mathbf{F}(\nabla\mathbf{u})\|_{W^{\alpha,2}((m-1)\tau-\tau/4,m\tau+\tau/4;L^2(\mathcal{O}))}^2\\
&\le c\tau^{2\alpha}\mathbb{E}\|\mathbf{F}(\nabla\mathbf{u})\|_{W^{\alpha,2}(0,T;L^2(\mathcal{O}))}^2.
\end{aligned}
$$

Similarly, for the last term we deduce

$$
\begin{aligned}
&\hat{\mathbb{E}}\otimes\mathbb{E}\bigg[\sum_{m=1}^{M}\tau_m\int_{\mathcal{O}}|\nabla\mathbf{F}(\nabla\mathbf{u}(\mathfrak{t}_m))|^2\,\mathrm{d}x\bigg]\\
&=\frac{2}{\tau}\int_{\tau-\tau/4}^{\tau+\tau/4}\xi\,\mathbb{E}\int_{\mathcal{O}}|\nabla\mathbf{F}(\nabla\mathbf{u}(\xi))|^2\,\mathrm{d}x\,\mathrm{d}\xi\\
&\quad+\frac{4}{\tau^2}\sum_{m=2}^{M}\int_{(m-1)\tau-\tau/4}^{(m-1)\tau+\tau/4}\int_{m\tau-\tau/4}^{m\tau+\tau/4}(\xi-\zeta)\mathbb{E}\int_{\mathcal{O}}|\nabla\mathbf{F}(\nabla\mathbf{u}(\xi))|^2\,\mathrm{d}x\,\mathrm{d}\xi\,\mathrm{d}\zeta\\
&\le c\,\mathbb{E}\|\nabla\mathbf{F}(\nabla\mathbf{u})\|_{L^2(0,\tau+\tau/4;L^2(\mathcal{O}))}^2
\end{aligned}
$$

$$+\frac{c}{\tau}\sum_{m=1}^{M}\int_{(m-1)\tau-\tau/4}^{(m-1)\tau+\tau/4}\mathrm{d}\zeta\,\mathbb{E}\|\nabla\mathbf{F}(\nabla\mathbf{u})\|^2_{L^2((m-1)\tau-\tau/4,m\tau+\tau/4;L^2(\mathcal{O}))}$$
$$\leq c\,\mathbb{E}\|\nabla\mathbf{F}(\nabla\mathbf{u})\|^2_{L^2(0,T;L^2(\mathcal{O}))}.$$

Consequently, in view of (3.2) and (3.3), the proof is complete. □

References

[1] J. W. Barrett, W.B. Liu: Finite element approximation of the p-Laplacian. Math. Comput. 61 (204), 523–537. (1993)

[2] J. W. Barrett, W. B. Liu: Finite element approximation of the parabolic p-Laplacian. SIAM J. Numer. Anal. 31 (2), 413–428. (1994)

[3] L. C. Berselli, L. Diening, M. Růžička: Optimal error estimates for a semi-implicit Euler scheme for incompressible fluids with shear dependent viscosities. SIAM J. Num. Anal. 47, 2177–2202. (2012)

[4] D. Breit: Regularity theory for nonlinear systems of SPDEs. Manus. Math. 146, 329–349 (2015).

[5] D. Breit: Existence theory for stochastic power law fluids. J. Math. Fluid Mech. 17, 295–326. (2015)

[6] D. Breit, M. Hofmanová: On time regularity of stochastic evolution equations with monotone coefficients. C. R. Acad. Sci. Paris Ser. I 354, 33–37. (2016)

[7] D. Breit, P. R. Mensah: Space-time approximation of parabolic systems with variable growth. Preprint at arXiv:1804.06097.

[8] Z. Brzeźniak, E. Carelli, A. Prohl: Finite-element-based discretizations of the incompressible Navier–Stokes equations with multiplicative random forcing. IMA J. Num. Anal. 33, 771–824. (2013)

[9] E. Carelli, J. A. Prohl: Rates of convergence for discretizations of the stochastic incompressible Navier-Stokes equations. SIAM J. Numer. Anal. 50(5), 2467–2496. (2012)

[10] G. Da Prato, J. Zabczyk: Stochastic Equations in Infinite Dimensions, Encyclopedia Math. Appl., vol. 44, Cambridge University Press, Cambridge. (1992)

[11] E. DeGiorgi: Un esempio di estrimali discontinue per un problema variazionale di tipo ellittico. Boll. UMI 4, 135–137. (1968)

[12] E. DiBenedetto, A. Friedmann, Regularity of solutions of nonlinear degenerate parabolic systems, J. Reine Angew. Math. 349, pp. 83–128. (1984)

[13] L. Diening, C. Ebmeyer, M. Růžička: Optimal convergence for the implicit space-time discretization of parabolic systems with p-structure. SIAM J. Numer. Anal. 45(2), 457–472. (2007)

[14] L. Diening, F. Ettwein: Fractional estimates for non-differentiable elliptic systems with general growth. Forum Math. 20 (3), 523–556. (2008)

[15] L. Diening, M. Růžička: Non-Newtonian fluids and function spaces, Nonlinear Analysis, Function Spaces and Applications, Vol. 8, 95-143. (2007)

[16] L. Diening and M. Růžička: Interpolation operators in Orlicz Sobolev spaces. Num. Math. 107 (1), 107–129. (2007)

[17] F. Duzaar, R. Mingione, K. Steffen: Parabolic Systems with Polynomial Growth and Regularity. Memoires of the American Mathematical Society 214, no. 1005. (2011)

[18] B. Gess: Strong Solutions for Stochastic Partial Differential Equations of Gradient Type. J. Funct. Anal. 263 (8), 2355–2383. (2012)

[19] I. Gyöngy, A. Millet: On Discretization Schemes for Stochastic Evolution Equations. Potential Analysis 23, 99–134. (2005)

[20] I. Gyöngy, A. Millet: Rate of Convergence of space-time Discretization for Stochastic Evolution Equations. Potential Analysis 30, 29–64. (2009)

[21] T. Iwaniec, J. J. Manfredi: Regularity of p-harmonic functions on the plane. Rev. Mat. Iberoamericana 5 (1-2), 1–19. (1989)

[22] N. V. Krylov: A $W^{n,2}$-theory of the Dirichlet problem for SPDEs in general smooth domains. Probab. Theory Related Fields 98 (3), 389–421. (1994)

[23] N. V. Krylov, B. L. Rozovskii: On the Cauchy problem for linear stochastic partial differetial equations. Izv. Akad. Nauk. SSSR Ser. Mat. 41 (6), 1329-1347; English transl. Math. USSR Izv. 11. (1977)

[24] N. V. Krylov, B. L. Rozovskii: Stochastic evolution equations. Itogi Nauki i Tekhniki. Ser. Sovrem. Probl. Mat. 14, VINITI, Moscow, 71–146; English transl. J. Sov. Math., 16 (4) (1981), 1233–1277. (1979)

[25] Ladyzhenskaya, O.A., Solonnikov, V.A., Ural'tseva, N.N.: Linear and quasi-linear equations of parabolic type. Translations of Mathematical Monographs. 23. Amer. Math. Soc., Providence, RI. (1967)

[26] G. J. Lord, C. Powell, T. Shardlow: An Introduction to Computational Stochastic PDEs. Cambridge Texts in Applied Mathematics. Cambridge University Press. (2014)

[27] E. Pardoux: Equations aux dérivées Partielles stochastiques non linéaires monotones. Etude de solutions fortes de type Itô, Ph.D. thesis, Université Paris Sud. (1975)

[28] C. Prévôt, M. Röckner: A concise course on stochastic partial differential equations. Lecture Notes in Mathematics, 1905. Springer, Berlin. (2007)

[29] L R. Scott, S. Zhang: Finite element interpolation of nonsmooth functions satisfying boundary conditions. Math. Comp. 54 (190), 483–493. (1990)

[30] Y. Terasawa, N. Yoshida: Stochastic power-law fluids: existence and uniqueness of weak solutions. Ann. Appl. Probab. 21 (5), 1827–1859. (2011) 219–240.

[31] Wiegner, M.: On C^{α}-regularity of the gradient of solutions of degenerate parabolic systems. Ann. Mat. Pura Appl. 145 (IV), 385–405. (1986)

[32] Y. Yan: Semidiscrete Galerkin Approximation for a Linear Stochastic Parabolic Partial Differential Equation Driven by an Additive Noise. Num. Math. 44, 829–847. (2004)

[33] Y. Yan: Galerkin finite element methods for stochastic parabolic partial differential equations, SIAM J. Numer. Anal. 43, 1363–1384. (2005)

[34] N. Yoshida: Stochastic Shear thickening fluids: strong convergence of the Galerkin approximation and the energy inequality. Ann. Appl. Probab. 22 (3), 1215–1242. (2012)

(D. Breit) Department of Mathematics, Heriot-Watt University, Riccarton Edinburgh EH14 4AS, UK

E-mail address: d.breit@hw.ac.uk

(M. Hofmanová) Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany

E-mail address: hofmanova@math.uni-bielefeld.de