Numerische Mathematik

A randomized operator splitting scheme inspired by stochastic optimization methods

Monika Eisenmann[1] · **Tony Stillfjord**[1]

Abstract

In this paper, we combine the operator splitting methodology for abstract evolution equations with that of stochastic methods for large-scale optimization problems. The combination results in a randomized splitting scheme, which in a given time step does not necessarily use all the parts of the split operator. This is in contrast to deterministic splitting schemes which always use every part at least once, and often several times. As a result, the computational cost can be significantly decreased in comparison to such methods. We rigorously define a randomized operator splitting scheme in an abstract setting and provide an error analysis where we prove that the temporal convergence order of the scheme is at least 1/2. We illustrate the theory by numerical experiments on both linear and quasilinear diffusion problems, using a randomized domain decomposition approach. We conclude that choosing the randomization in certain ways may improve the order to 1. This is as accurate as applying e.g. backward (implicit) Euler to the full problem, without splitting.

Mathematics Subject Classification 65C99 · 65M12 · 90C15 · 65M55

1 Introduction

The main objective of this paper is to combine two successful strategies from the literature: the first being operator splitting schemes for evolution equations on general, infinite dimensional frameworks and the second being stochastic optimization methods. Operator splitting schemes are an established tool in the field of numerical analysis of evolution equations and have a wide range of applications. Stochastic

✉ Monika Eisenmann
   monika.eisenmann@math.lth.se

   Tony Stillfjord
   tony.stillfjord@math.lth.se

1   Centre for Mathematical Sciences, Lund University, P.O. Box 118, 22100 Lund, Sweden

optimization methods have proven to be efficient at solving large-scale optimization problems, where it is infeasible to evaluate full gradients. They can drastically decrease the computational cost in e.g. machine learning settings. The link between these two seemingly disparate areas is that an iterative method applied to an optimization problem can also be seen as a time-stepping method applied to a gradient flow connected to the optimization problem. In particular, stochastic optimization methods can then be interpreted as randomized operator splitting schemes for such gradient flows. In this context, we introduce a general randomized splitting method that can be applied directly to evolution equations, and provide a rigorous convergence analysis.

Abstract evolution equations of the type

$$
\begin{cases}
u'(t) + A(t)u(t) = f(t), & t \in (0, T], \\
u(0) = u_0
\end{cases}
$$

are an important building block for modeling processes in physics, biology and social sciences. Standard examples which appear in a variety of applications are fluid flow problems, where we model how a flow evolves on a given domain over time, compare [1, 26] and [37, Section 1.3]. The operator $A(t)$ can denote, for example, a non-linear diffusion operator such as the p-Laplacian or a porous medium operator.

Deterministic operator splitting schemes as discussed in more detail in [16] are a powerful tool for this type of equation. An example is given by a domain decomposition scheme, where we split the domain into sub-domains. Instead of solving one expensive problem on the entire domain, we deal with cheaper problems on the sub-domains. This is particularly useful in modern computer architectures, as the sub-problems may often be solved in parallel.

Moreover, evolution equations are tightly connected to unconstrained optimization problems, because the solution of $\min_u F(u)$ is a stationary point of the gradient flow $u'(t) = -\nabla F(u(t))$. The latter is an evolution equation on an infinite time horizon with $A = -\nabla F$ and $f = 0$. In the large-scale case, such optimization problems benefit from stochastic optimization schemes. The most basic such method, the stochastic gradient descent, was first introduced already in [32], but since then it has been extended and generalized in many directions. See, e.g., the review article [3] and the references therein.

Via the gradient flow interpretation, we can see these optimization methods as time-stepping schemes where a randomly chosen sub-problem is considered in each time step. In essence, it is therefore a *randomized* operator splitting scheme. The difference between the works mentioned above and ours is that we apply these stochastic optimization techniques to solve the evolution equation itself rather than just finding its stationary state.

We consider nonlinear evolution equations in an abstract framework similar to [7, 10, 11] where operators of a monotone type have been studied. Deterministic splitting schemes for such equations has been considered in e.g. [14, 15, 17, 29]. A particular kind of splitting schemes which is most closely related to our work, domain decomposition methods, have been studied in [6, 7, 13, 30, 31]. In this paper, we

extend this framework of deterministic splitting schemes to a setting of randomized methods.

Outside of the context of optimization, other kinds of randomized methods have already proved themselves to be useful for solving evolution equations. Starting in [34, 35] explicit schemes for ordinary differential equations have been randomized. This approach has been further extended in [2, 4, 18, 22, 24]. In [8], it has been extended both to implicit methods and to partial differential equations and in [23] to finite element approximations. While these works considered certain randomizations in their schemes, they are conceptually different from our approach. Their main idea is to approximate any appearing integrals through

$$\int_{t_{n-1}}^{t_n} f(t) \, dt \approx f(\xi_n) \quad \text{and} \quad \int_{t_{n-1}}^{t_n} A(t)v \, dt \approx A(\xi_n)v,$$

where ξ_n is a random variable that takes on values in $[t_{n-1}, t_n]$. This ansatz coincides with a Monte Carlo integration idea. In this paper, we use a different approach where we decompose the operator in a randomized fashion. More precisely, we approximate data

$$f = \frac{1}{s} \sum_{\ell=1}^{s} f_\ell \quad \text{and} \quad A = \frac{1}{s} \sum_{\ell=1}^{s} A_\ell$$

by

$$f_B = \frac{1}{|B|} \sum_{\ell \in B} f_\ell \quad \text{and} \quad A_B = \frac{1}{|B|} \sum_{\ell \in B} A_\ell$$

where the batch $B \subset \{1, \ldots, s\}$ is chosen randomly. The stochastic approximations f_B and A_B of the original data f and A are cheaper to evaluate in applications. This is less related to Monte Carlo integration and more similar to stochastic optimization methods, compare [3, 9]. Similar ideas have been considered in [19, 20, 28], where a random batch method for interacting particle systems has been studied. Moreover, very recently and during the preparation of this work, a similar approach has also been applied to the optimal control of linear time invariant (LTI) dynamical systems in [38]. While the convergence rate provided there is essentially the same as what we establish in our main result Theorem 5.2, our setting is more general and allows for nonlinear operators on infinite dimensional spaces rather than finite dimensional matrices. We also consider the error of the time stepping method that is used to approximate the solution to $u'(t) + A_B(t)u(t) = f_B(t)$, while the error bounds in [38] assume that this evolution equation is solved exactly.

This paper is organized as follows. In Sect. 2, we begin by explaining our abstract framework. This includes both the precise assumptions that we make and the definition of our time-stepping scheme. We give a more concrete application of the abstract framework in Sect. 3. With the setting fixed, we first prove in Sect. 4 that the scheme and its solution are indeed well-defined. We prove the convergence of the scheme in

expectation in Sect. 5. These theoretical convergence results are illustrated by numerical experiments with two-dimensional linear and quasilinear nonlinear and linear diffusion problem in Sect. 6. Finally, we collect some more technical auxiliary results in Appendix A.

2 Setting

In the following, we introduce a theoretical framework for the randomized operator splitting. This setting is similar to the one in [7].

Assumption 2.1 *Let $(H, (\cdot, \cdot)_H, \| \cdot \|_H)$ be a real, separable Hilbert space and let $(V, \| \cdot \|_V)$ be a real, separable, reflexive Banach space, which is continuously and densely embedded into H. Moreover, there exists a semi-norm $|\cdot|_V$ on V and a $C_V \in (0, \infty)$ such that $|\cdot|_V \leq C_V \| \cdot \|_V$.*

Denoting the dual space of V by V^* and identifying the Hilbert space H with its dual space, the spaces from Assumption 2.1 form a Gelfand triple and fulfill, in particular,

$$V \overset{d}{\hookrightarrow} H \cong H^* \overset{d}{\hookrightarrow} V^*.$$

Assumption 2.2 *Let the spaces H and V be given as stated in Assumption 2.1. Furthermore, for $T \in (0, \infty)$ as well as $p \in [2, \infty)$, let $\{A(t)\}_{t \in [0,T]}$ be a family of operators $A(t) \colon V \to V^*$ that satisfy the following conditions:*

 (i) *The mapping $Av \colon [0, T] \to V^*$ given by $t \mapsto A(t)v$ is continuous almost everywhere in $(0, T)$ for all $v \in V$.*
 (ii) *The operator $A(t) \colon V \to V^*$, $t \in [0, T]$, is radially continuous, i.e., the mapping $s \mapsto \langle A(t)(v + sw), w \rangle_{V^* \times V}$ is continuous on $[0, 1]$ for all $v, w \in V$.*
 (iii) *There exists $\kappa_A \in [0, \infty)$ and $\eta_A \in [0, \infty)$, which do not depend on t, such that the operator $A(t) + \kappa_A I \colon V \to V^*$, $t \in [0, T]$, fulfills the monotonicity-type condition*

$$\langle A(t)v - A(t)w, v - w \rangle_{V^* \times V} + \kappa_A \|v - w\|_H^2 \geq \eta_A |v - w|_V^p$$

 for all $v, w \in V$.
 (iv) *The operator $A(t) \colon V \to V^*$, $t \in [0, T]$, is uniformly bounded such that there exists $\beta_A \in [0, \infty)$, which does not depend on t, with*

$$\|A(t)v\|_{V^*} \leq \beta_A \left(1 + \|v\|_V^{p-1} \right)$$

 for all $v \in V$.

Assumption 2.3 *The function f is an element of the Bochner space $L^2(0, T; H)$, and the initial value $u_0 \in H$, where H is the Hilbert space from Assumption 2.1.*

Remark 1 We note that Assumption 2.2 *(iii)* implies that the operator $A(t) + \kappa_A I \colon V \to V^*$, $t \in [0, T]$, fulfills a uniform semi-coercivity condition. That is, there exist constants $\mu_A, \lambda_A \in [0, \infty)$, which do not depend on t, such that

$$\langle A(t)v, v\rangle_{V^* \times V} + \kappa_A \|v\|_H^2 + \lambda_A \geq \mu_A |v|_V^p$$

for all $v \in V$. This follows by taking $w = 0$ in *(iii)*, since then

$$\langle A(t)v, v\rangle_{V^* \times V} + \kappa_A \|v\|_H^2 \geq \langle A(t)0, v\rangle_{V^* \times V} + \eta_A |v|_V^p,$$

and by the Cauchy-Schwarz inequality and the weighted Young's inequality (Lemma A.2),

$$\langle A(t)0, v\rangle_{V^* \times V} \geq -\|A(t)0\|_{V^*} \|v\|_V \geq -\left(\frac{\|A(t)0\|_{V^*}^q}{\varepsilon^{\frac{q}{p}} q} + \varepsilon \|v\|_V^p \right)$$

with $\frac{1}{p} + \frac{1}{q} = 1$ and $\varepsilon > 0$. Since $|v|_V \leq C_V \|v\|_V$, we can absorb the second term and take $\lambda_A = \varepsilon^{-\frac{q}{p}} q^{-1} \|A(t)0\|_{V^*}^q$ and $\mu_A = \eta_A - \varepsilon$ after choosing an ε such that $\mu_A \geq 0$. This also shows that the constants λ_A and μ_A are not unique. We can, e.g., increase the coercivity constant at the cost of a larger constant term λ_A. Both these terms enter into our error bounds, which can thus be tuned slightly.

In the case that $A(t)0 = 0$, the constant term disappears and we have $\mu_A = \eta_A$. If $A(t)0 \neq 0$, one could recover this situation by the transformation $(A, f) \to (\tilde{A}, \tilde{f})$ with $\tilde{A}(t)u = A(t)u - A(t)0$, $\tilde{f}(t) = f(t) - A(t)0$. But in the case that $A(t)0 \in V^* \backslash H$ this can cause issues since we require that $f(t) \in H$. Moreover, it might lead to difficulties in solving the nonlinear equations of the form $(I - h_n \tilde{A}(t_n))u^n = u^{n-1} + h_n \tilde{f}(t_n)$. We therefore do not apply such a transformation in this paper.

Assumptions 2.1–2.3, are requirements on the problem that we want to solve. The following Assumptions 2.4–2.5 are needed to state the approximation scheme for the given problem.

Assumption 2.4 *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a complete probability space and let $\{\xi_n\}_{n \in \mathbb{N}}$ be a family of mutually independent random variables. Further, let the filtration $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ be given by*

$$\mathcal{F}_0 := \sigma\big(\mathcal{N} \in \mathcal{F} : \mathcal{P}(\mathcal{N}) = 0\big)$$
$$\mathcal{F}_n := \sigma\big(\sigma\big(\xi_i : i \in \{1, \dots, n\}\big) \cup \mathcal{F}_0\big), \quad n \in \mathbb{N},$$

where σ denotes the generated σ-algebra.

In the following, we denote the expectation with respect to the probability distribution of ξ for a random variable X in the Bochner space $L^1(\Omega; H)$ by $\mathbb{E}_\xi[X]$. Moreover, we abbreviate the total expectation by

$$\mathbb{E}_n[X] = \mathbb{E}_{\xi_1}[\mathbb{E}_{\xi_2}[\dots \mathbb{E}_{\xi_n}[X]\dots]].$$

We denote the space of Hölder continuous functions on $[0, T]$ with Hölder coefficient $\gamma \in (0, 1)$ and values in H by $C^\gamma([0, T]; H)$. For notational convenience we include the case $\gamma = 1$ and denote the space of Lipschitz continuous functions by $C^1([0, T]; H)$.

Assumption 2.5 *Let Assumptions 2.1–2.4 be fulfilled. Assume that for almost every* $\omega \in \Omega$ *there exists a real Banach space* $V_{\xi(\omega)}$ *such that* $V \overset{d}{\hookrightarrow} V_{\xi(\omega)} \overset{d}{\hookrightarrow} H$, $\bigcap_{\omega \in \Omega} V_{\xi(\omega)} = V$ *and there exists a semi-norm* $|\cdot|_{V_{\xi(\omega)}}$ *on* $V_{\xi(\omega)}$ *and a* $C_{V_{\xi(\omega)}} \in (0, \infty)$ *such that* $|\cdot| \leq C_{V_{\xi(\omega)}} \|\cdot\|_{V_{\xi(\omega)}}$. *Moreover, the mapping from* $\omega \mapsto V_{\xi(\omega)}$ *is measurable in the sense that for every* $v \in H$ *the set* $\{\omega \in \Omega : v \in V_{\xi(\omega)}\}$ *is an element of the complete generated* σ-*algebra*

$$\mathcal{F}_\xi := \sigma\big(\sigma(\xi) \cup \sigma\big(\mathcal{N} \in \mathcal{F} : \mathcal{P}(\mathcal{N}) = 0\big)\big).$$

Further, let the family of operators $\{A_{\xi(\omega)}(t)\}_{\omega \in \Omega, t \in [0, T]}$ *be such that for almost every* $\omega \in \Omega$, $\{A_{\xi(\omega)}(t)\}_{t \in [0, T]}$ *fulfills Assumption 2.2 with the spaces* $V_{\xi(\omega)}$, H *and* $V_{\xi(\omega)}^*$ *and corresponding constants* $\kappa_{\xi(\omega)}$, $\eta_{\xi(\omega)}$, $\beta_{\xi(\omega)}$. *These give rise to the semi-coercivity constants* $\mu_{\xi(\omega)}$ *and* $\lambda_{\xi(\omega)}$ *as in Remark 1. Moreover, the mapping* $A_\xi(t)v : \Omega \to V^*$ *is* $\mathcal{F}_\xi$-*measurable and the equality* $\mathbb{E}_\xi[A_\xi(t)v] = A(t)v$ *is fulfilled in* V^* *for* $v \in V$. *The mappings* $\kappa_\xi, \eta_\xi, \mu_\xi, \beta_\xi, \lambda_\xi : \Omega \to [0, \infty)$ *are measurable and there exist* $\kappa, \lambda \in [0, \infty)$ *which fulfill* $\kappa_\xi \leq \kappa$ *almost surely and* $\mathbb{E}_\xi[\lambda_\xi] \leq \lambda$.

Further, let the family $\{f_{\xi(\omega)}\}_{\omega \in \Omega}$ *be given such that* $f_{\xi(\omega)} \in L^2(0, T; H)$. *Moreover, the mapping* $f_\xi(t) : \Omega \to H$ *is* $\mathcal{F}_\xi$-*measurable and* $\mathbb{E}_\xi[f_\xi(t)] = f(t)$ *is fulfilled in* H *for almost all* $t \in (0, T)$.

Under the setting explained in the above assumptions, we consider the initial value problem

$$\begin{cases} u'(t) + A(t)u(t) = f(t) & \text{in } V^*, \quad t \in (0, T], \\ u(0) = u_0 & \text{in } H. \end{cases} \tag{1}$$

For a non-uniform temporal grid $0 = t_0 < t_1 < \cdots < t_N = T$, a step size $h_n = t_n - t_{n-1}$, $h = \max_{n \in \{1, \ldots, N\}} h_n$, and a family of random variables $\{f^n\}_{n \in \{1, \ldots, N\}}$ such that $f^n : \Omega \to H$ is $\mathcal{F}_{\xi_n}$-measurable, we consider the scheme

$$\begin{cases} U^n - U^{n-1} + h_n A_{\xi_n}(t_n)U^n = h_n f^n & \text{in } V_{\xi_n}^*, \quad n \in \{1, \ldots, N\}, \\ U^0 = u_0 & \text{in } H. \end{cases} \tag{2}$$

Note that $U^n : \Omega \to H$ is a random variable and therefore some statements involving it below only hold almost surely. Whenever there is no risk of misinterpretation, we omit writing almost surely for the sake of brevity.

When proving that the scheme is well-defined and establishing an a priori bound, it is sufficient to assume that $\{f_{\xi_n}\}_{n \in \{1, \ldots, N\}}$ are integrable with respect to the temporal parameter. In that case, we can choose for example

$$f^n = \frac{1}{h_n} \int_{t_{n-1}}^{t_n} f_{\xi_n}(t) \, \mathrm{d}t \quad \text{in } H \text{ almost surely.} \tag{3}$$

When considering our error bounds, we assume more regularity for the functions $\{f_{\xi_n}\}_{n \in \{1, \ldots, N\}}$ and demand continuity with respect to the temporal parameter. In this case, we may also use

$$f^n = f_{\xi_n}(t_n) \quad \text{in } H \text{ almost surely.} \tag{4}$$

We will focus on this second choice for the error bounds in Sect. 5.

3 Application: Domain decomposition

One main application that is allowed by our abstract framework is a domain decomposition scheme for a nonlinear fluid flow problem. Domain decomposition schemes are well-known for deterministic operator splittings. However, to the best of our knowledge, it has not been studied in the context of a randomized operator splitting scheme.

3.1 Deterministic domain decomposition

To exemplify our abstract Eq. (1), we consider a (nonlinear) parabolic differential equation. In the following, let $\mathcal{D} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a bounded domain with a Lipschitz boundary $\partial \mathcal{D}$. For $p \in [2, \infty)$, we consider the parabolic p-Laplacian with homogeneous Dirichlet boundary conditions

$$\begin{cases} \partial_t u(t,x) - \nabla \cdot (\alpha(t) |\nabla u(t,x)|^{p-2} \nabla u(t,x)) = \tilde{f}(t,x), & (t,x) \in (0,T) \times \mathcal{D}, \\ u(t,x) = 0, & (t,x) \in (0,T) \times \partial \mathcal{D}, \\ u(0,x) = u_0(x), & x \in \mathcal{D}, \end{cases} \tag{5}$$

for $\alpha \colon [0,T] \to \mathbb{R}$ and $u_0 \colon \mathcal{D} \to \mathbb{R}$. The notation $\tilde{f}$ is used to differentiate between the function $\tilde{f} \colon (0,T) \times \mathcal{D} \to \mathbb{R}$ and the abstract function f on $(0,T)$ that it gives rise to through $[f(t)](x) = \tilde{f}(t,x)$. We consider a domain decomposition scheme similar to [13] for $p = 2$ and to [6, 7] for $p \in [2, \infty)$. For the sake of completeness, we recapitulate the setting here also with a different boundary condition.

For $s \in \mathbb{N}$, let $\{\mathcal{D}_\ell\}_{\ell=1}^s$ be a family of overlapping subsets of $\mathcal{D}$. Let each subset have a Lipschitz boundary and let the union of them fulfill $\bigcup_{\ell=1}^s \mathcal{D}_\ell = \mathcal{D}$. On the sub-domains $\{\mathcal{D}_\ell\}_{\ell=1}^s$, let the partition of unity $\{\chi_\ell\}_{\ell=1}^s \subset W^{1,\infty}(\mathcal{D})$ be given such that the following criteria are fulfilled

$$\chi_\ell(x) > 0 \text{ for all } x \in \mathcal{D}_\ell, \quad \chi_\ell(x) = 0 \text{ for all } x \in \mathcal{D} \backslash \mathcal{D}_\ell, \quad \sum_{\ell=1}^s \chi_\ell = 1$$

for $\ell \in \{1, \ldots, s\}$. With the help of the functions $\{\chi_\ell\}_{\ell \in \{1,\ldots,s\}}$, it is now possible to introduce suitable functional spaces $\{V_\ell\}_{\ell \in \{1,\ldots,s\}}$. We use the weighted Lebesgue space $L^p(\mathcal{D}_\ell, \chi_\ell)^d$ that consists of all measurable functions $v = (v_1, \ldots, v_d) \colon \mathcal{D}_\ell \to \mathbb{R}^d$ such that

$$\|(v_1, \ldots, v_d)\|_{L^p(\mathcal{D}_\ell, \chi_\ell)^d} = \left(\int_{\mathcal{D}_\ell} \chi_\ell |(v_1, \ldots, v_d)|^p \, \mathrm{d}x \right)^{\frac{1}{p}}$$

is finite. In the following, let the pivot space $(H, (\cdot, \cdot)_H, \|\cdot\|_H)$ be the space $L^2(\mathcal{D})$ of square integrable functions on $\mathcal{D}$ with the usual norm and inner product. The spaces V and V_ℓ, $\ell \in \{1, \ldots, s\}$, are given by

$$V = \mathrm{clos}_{\|\cdot\|_V} \left(C_0^\infty(\mathcal{D}) \right) = W_0^{1,p}(\mathcal{D}) \quad \text{and} \quad V_\ell = \mathrm{clos}_{\|\cdot\|_{V_\ell}} \left(C_0^\infty(\mathcal{D}) \right),$$

with respect to the norms

$$\|\cdot\|_V = \|\cdot\|_H + \|\nabla \cdot\|_{L^p(\mathcal{D})^d} \quad \text{and} \quad \|\cdot\|_{V_\ell} = \|\cdot\|_H + \|\nabla \cdot\|_{L^p(\mathcal{D}_\ell, \chi_\ell)^d} \quad (6)$$

and semi-norms

$$|\cdot|_V = \|\nabla \cdot\|_{L^p(\mathcal{D})^d} \quad \text{and} \quad |\cdot|_{V_\ell} = \|\nabla \cdot\|_{L^p(\mathcal{D}_\ell, \chi_\ell)^d}.$$

Note that a bootstrap argument involving the Sobolev embedding theorem shows that the norm given in (6) is equivalent to the standard norm in the space. We can now introduce the operators $A(t) \colon V \to V^*$, $A_\ell(t) \colon V_\ell \to V_\ell^*$, $\ell \in \{1, \ldots, s\}$, $t \in [0, T]$, given by

$$\langle A(t)u, v \rangle_{V^* \times V} = \int_{\mathcal{D}} \alpha(t) |\nabla u|^{p-2} \nabla u \cdot \nabla v \, \mathrm{d}x, \quad u, v \in V,$$

$$\langle A_\ell(t)u, v \rangle_{V_\ell^* \times V_\ell} = \int_{\mathcal{D}_\ell} \chi_\ell \alpha(t) |\nabla u|^{p-2} \nabla u \cdot \nabla v \, \mathrm{d}x, \quad u, v \in V_\ell.$$

Similarly, we define the right-hand sides $f_\ell \colon [0, T] \to H$, $\ell \in \{1, \ldots, s\}$, where $f_\ell(t) = \chi_\ell f(t)$ in H for almost every $t \in (0, T)$.

Lemma 3.1 *Let the parameters of Eq. (5) be given such that $\alpha \in C([0, T]; \mathbb{R})$, $u_0 \in L^2(\mathcal{D})$ and $\tilde{f} \in L^2((0, T) \times \mathcal{D})$. Then the setting described above fulfills Assumptions 2.1–2.3.*

Let the partition of unity $\{\chi_\ell\}_{\ell=1}^s \subset W^{1,\infty}(\mathcal{D})$ fulfill that for every function χ_ℓ there exists $\varepsilon_0 \in (0, \infty)$ such that $\mathcal{D}_\ell^\varepsilon = \{x \in \mathcal{D}_\ell : \chi_\ell(x) \geq \varepsilon\}$ is a Lipschitz domain for all $\varepsilon \in (0, \varepsilon_0)$. Then V and V_ℓ, $\ell \in \{1, \ldots, s\}$, are reflexive Banach spaces and $V = \bigcap_{\ell=1}^s V_\ell$. Further, the family of operators $\{A_\ell(t)\}_{t \in [0,T]}$, $\ell \in \{1, \ldots, s\}$ fulfills Assumption 2.2 with the spaces V_ℓ, H and V_ℓ^*. Moreover, $\sum_{\ell=1}^s A_\ell(t)v = A(t)v$ is fulfilled in V^* for $v \in V$ for almost every $t \in (0, T)$ and corresponding constants $\kappa_A = \kappa_\ell = \lambda_A = \lambda_\ell = 0$, $\mu_A = \mu_\ell = \eta_A = \eta_\ell = 1$.

Finally, the family $\{f_\ell\}_{\ell \in \{1,\ldots,s\}}$ fulfills $f_\ell \in L^2(0, T; H)$ and $\sum_{\ell=1}^s f_\ell(t) = f(t)$ in H for almost all $t \in (0, T)$.

Proof The space $H = L^2(\mathcal{D})$ is a real, separable Hilbert space, while $V = W_0^{1,p}(\mathcal{D})$ is a real, separable Banach space that is densely embedded into H. Thus, they fulfill Assumption 2.1. Analogously to [6, Lemma 3], the spaces V and V_ℓ, $\ell \in \{1, \ldots, s\}$, are reflexive Banach spaces and since $C_0^\infty(\mathcal{D})$ is dense in H and $C_0^\infty(\mathcal{D}) \subseteq V \subset V_\ell$ it follows that V and V_ℓ are dense in H. It remains to prove that $\bigcap_{\ell=1}^s V_\ell = V$ is fulfilled. First, we notice that $\|w\|_{L^p(\mathcal{D}_\ell, \chi_\ell)^d} \leq \|w\|_{L^p(\mathcal{D})^d}$ for every $w \in L^p(\mathcal{D})^d$. Thus, it follows that $V \subseteq V_\ell$ for every $\ell \in \{1, \ldots, s\}$ and in particular $V \subseteq \bigcap_{\ell=1}^s V_\ell$. The other inclusion $\bigcap_{\ell=1}^s V_\ell \subseteq V$ requires more attention. For $\varepsilon \in (0, \infty)$, we introduce the set $\mathcal{D}_\ell^\varepsilon = \{x \in \mathcal{D} : \chi_\ell(x) \geq \varepsilon\}$. By assumption the sets $\mathcal{D}_\ell^\varepsilon$ have Lipschitz boundary for ε small enough. We consider the spaces of restricted functions

$$C_0^\infty(\mathcal{D})|_{\mathcal{D}_\ell^\varepsilon} = \{u \in C^\infty(\mathcal{D}_\ell^\varepsilon) : u|_{\partial \mathcal{D}_\ell^\varepsilon \cap \partial \mathcal{D}} = 0\} \quad \text{and} \quad V_\ell^\varepsilon = \{u|_{\mathcal{D}_\ell^\varepsilon} : u \in V_\ell\}.$$

If a weight function χ_ℓ fulfills $0 < \varepsilon < \chi_\ell \leq 1 < \infty$ on the whole domain $\mathcal{D}$, it follows that the weighted Lebesgue space $L^p(\mathcal{D}_\ell^\varepsilon, \chi_\ell)^d$ coincides with the space $L^p(\mathcal{D}_\ell^\varepsilon)^d$ (see, e.g., [25, Chapter 3]). Thus, we obtain $V_\ell^\varepsilon = W^{1,p}(\mathcal{D}_\ell^\varepsilon)$. The continuity of the trace operator (see, e.g., [27, Theorem 15.23]), implies that

$$\overline{C_0^\infty(\mathcal{D})|_{\mathcal{D}_\ell^\varepsilon}}^{\|\cdot\|_{V_\ell}} = \{u \in W^{1,p}(\mathcal{D}_\ell^\varepsilon) : u|_{\partial \mathcal{D}_\ell^\varepsilon \cap \partial \mathcal{D}} = 0\}.$$

This shows that $u \in V_\ell$ is zero on $\partial \mathcal{D}_\ell^\varepsilon \cap \partial \mathcal{D}$ for every $\varepsilon \in (0, \infty)$ small enough. As ε can be chosen arbitrarily small, it follows that $u \in V_\ell$ fulfills $v|_{\partial \mathcal{D} \cap \partial \mathcal{D}_\ell} = 0$. In combination with [6, Lemma 1], we obtain that $\bigcap_{\ell=1}^s V_\ell = W_0^{1,p}(\mathcal{D}) = V$.

Similar to the argumentation of [6, Lemma 4], it follows that the families of operators $\{A(t)\}_{t\in[0,T]}$ and $\{A_\ell(t)\}_{t\in[0,T]}$, $\ell \in \{1, \ldots, s\}$, fulfills Assumption 2.2 with respect to the corresponding spaces with $\kappa_A = \kappa_\ell = \lambda_A = \lambda_\ell = 0$, $\mu_A = \mu_\ell = \eta_A = \eta_\ell = 1$.

Assumption 2.3 is fulfilled as $\tilde{f} \in L^2((0, T) \times \mathcal{D})$ means that the abstract function f belongs to $L^2(0, T; L^2(\mathcal{D}))$. Thus, as $\chi_\ell \in W^{1,\infty}(\mathcal{D})$, it follows that $f_\ell = \chi_\ell f \in L^2(0, T; H)$ and $\sum_{\ell=1}^s f_\ell(t) = f(t)$ in H for almost every $t \in (0, T)$. □

3.2 Randomized scheme

For a randomized splitting in combination with a domain decomposition, different approaches can be applied. One possibility is to choose a random support of the weight functions $\{\chi_\ell\}_{\ell \in \{1, \ldots, s\}}$. This could possibly be done efficiently using priority queue techniques similar to those in [36]. In this paper, we instead fix the weight functions, but choose a random part of the operator in every time step. For the operator $A(t) = \sum_{\ell=1}^s A_\ell(t)$ and a right hand side $f(t) = \sum_{\ell=1}^s f_\ell(t)$, we introduce a random variable $\xi : \Omega \to 2^{\{1,\ldots,s\}}$ such that $[A_\xi(t)](\omega) = \sum_{\ell \in \xi(\omega)} A_\ell(t)/\tau_\ell$ and $[f_\xi(t)](\omega) = \sum_{\ell \in \xi(\omega)} f_\ell(t)/\tau_\ell$ with

$$\tau_\ell = \sum_{B \in 2^{\{1,\ldots,s\}} : \ell \in B} \mathcal{P}(\Omega_{\xi=B}) \quad \text{with} \quad \Omega_{\xi=B} = \{\omega \in \Omega : \xi(\omega) = B\}.$$

The value τ_ℓ is the proper scaling factor which ensures that $\mathbb{E}_\xi[A_\xi(t)] = A(t)$ and $\mathbb{E}_\xi[f_\xi(t)] = f(t)$. We tacitly assume that $\tau_\ell > 0$, because otherwise we would be in a situation where at least one $A_\ell(t)$ is never chosen. Such a strategy would obviously not work. We set $V_{\xi(\omega)} = \bigcap_{\ell \in \xi(\omega)} V_\ell$.

Lemma 3.2 *Let $\{\xi_n\}_{n \in \{1,\dots,N\}}$ fulfill Assumption 2.4 such that $\xi_n \colon \Omega \to 2^{\{1,\dots,s\}}$ and $\xi_n^{-1}(B) \in \mathcal{F}_{\xi_n}$ for all $B \subset 2^{\{1,\dots,s\}}$ and $n \in \{1,\dots,N\}$. Under the setting above, Assumption 2.5 is fulfilled.*

Proof In the following proof, we drop the index n to keep the notation simpler. The embedding and norm properties are fulfilled as verified in the previous lemma. It remains to verify the measurability condition. We need to verify that for every $v \in H$, the set $\{\omega \in \Omega : v \in V_{\xi(\omega)}\} \in \mathcal{F}_\xi = \sigma\big(\sigma(\xi) \cup \sigma(\mathcal{N} \in \mathcal{F} : \mathcal{P}(\mathcal{N}) = 0)\big)$. For fixed $v \in H$, we set $B_v = \{\ell \in \{1,\dots,s\} : v \in V_\ell\} \in 2^{\{1,\dots,s\}}$. Then it follows that

$$\{\omega \in \Omega : v \in V_{\xi(\omega)}\} = \{\omega \in \Omega : \xi(\omega) \in 2^{B_v}\} = \xi^{-1}\big(2^{B_v}\big) \in \mathcal{F}_\xi.$$

Moreover, we need to verify that the mapping $\omega \mapsto A_{\xi(\omega)}(t)v$ is measurable for every $v \in H$. This can be seen from the decomposition $A_\xi(t)v = S_{A(t)v} \circ \xi$ where $S_{A(t)v} \colon 2^{\{1,\dots,s\}} \to V^*$ is given through $S_{A(t)v}(B) = \sum_{\ell \in B} A_\ell(t)v$. As $\xi^{-1}(B) \in \mathcal{F}_\xi$ for all $B \subset 2^{\{1,\dots,s\}}$ and $S_{A(t)v}^{-1}(X) \subset 2^{\{1,\dots,s\}}$ for any open set $X \subset V^*$, the mapping $\omega \mapsto A_{\xi(\omega)}(t)v$ is measurable. Analogously, it can be proved that the mapping $\omega \mapsto f_{\xi(\omega)}(t)$ is measurable. In Lemma 3.1, we already verified that an operator $A_{\xi(w)}$ fulfills the conditions from Assumption 2.2. Thus, it only remains to prove the expectation property from Assumption 2.5. This is fulfilled as

$$\mathbb{E}_\xi[A_\xi(t)v] = \sum_{B \in 2^{\{1,\dots,s\}}} \mathcal{P}(\Omega_{\xi=B}) \sum_{\ell \in B} \frac{1}{\tau_\ell} A_\ell(t)v$$

$$= \sum_{\ell=1}^s \frac{1}{\tau_\ell} A_\ell(t)v \sum_{B \in 2^{\{1,\dots,s\}} : \ell \in B} \mathcal{P}(\Omega_{\xi=B}) = \sum_{\ell=1}^s A_\ell(t)v = A(t)v \quad \text{in } V^*$$

holds true for $v \in V$ and for almost every $t \in [0, T]$. The same algebraic manipulation in H instead of V^* shows that $\mathbb{E}_\xi[f_\xi(t)] = f(t)$. $\square$

4 Solution is well-defined

In the coming section, we show that our scheme (2) is well-defined. This includes that first of all the scheme possesses a unique solution. We consider a purely deterministic Eq. (1). However, as the numerical scheme is randomized, the solution U^n of (2) is a mapping of the type $U^n \colon \Omega \to H$. Thus, we also need to make sure that it is a measurable function. These facts are verified in Lemma 4.1. Moreover, we provide an integrability result in the form of an a priori bound in Lemma 4.2.

Lemma 4.1 *Let Assumptions 2.1–2.5 be fulfilled. Further, let the random variables $f^n \colon \Omega \to H$ be given such that they are $\mathcal{F}_{\xi_n}$-measurable for every $n \in \{1,\dots,N\}$.*

Then for $\kappa h_n \leq \kappa h < 1$ there exists a unique $\mathcal{F}_n$-measurable function $U^n \colon \Omega \to H$ such that $U^n(\omega) \in V_{\xi_n(\omega)}$ and (2) is fulfilled for every $n \in \{1, \ldots, N\}$.

Proof For $\omega \in \Omega$, we find that the operator $I + h_n A_{\xi_n(\omega)}(t_n) \colon V_{\xi_n(\omega)} \to V^*_{\xi_n(\omega)}$ is monotone, radially continuous and coercive. Thus, it is surjective, compare [33, Theorem 2.18]. Moreover, for $U_1, U_2 \in V_{\xi_n(\omega)}$ with $\big(I + h_n A_{\xi_n(\omega)}(t_n)\big)U_1 = \big(I + h_n A_{\xi_n(\omega)}(t_n)\big)U_2$, it follows that

$$
\begin{aligned}
0 &= \big\langle \big(I + h_n A_{\xi_n(\omega)}(t_n)\big)U_1 - \big(I + h_n A_{\xi_n(\omega)}(t_n)\big)U_2, U_1 - U_2 \big\rangle_{V^*_{\xi_n(\omega)} \times V_{\xi_n(\omega)}} \\
&\geq \big(1 - h_n \kappa\big)\|U_1 - U_2\|_H^2.
\end{aligned}
$$

Thus, it follows that $\|U_1 - U_2\|_H = 0$ and $I + h_n A_{\xi_n(\omega)}(t_n)$ is injective for $\kappa h_n < 1$ and, in particular, bijective.

It remains to verify that $U^n \colon \Omega \to H$ is well-defined. We define the auxiliary function $g \colon \Omega \times H \to V^*$ such that

$$
(\omega, U) \mapsto \begin{cases} h_n f^n(\omega) + U^{n-1} - \big(I + h_n A_{\xi_n(\omega)}(t_n)\big)U, & U \in V_{\xi_n(\omega)} \\ e, & U \in H \backslash V_{\xi_n(\omega)}, \end{cases}
$$

where $e \in V^*$ with $\|e\|_{V^*} = 1$. In the following, we want to apply Lemma A.3 to the function g to prove that U^n is measurable. Applying [33, Lemma 2.16], it follows that for fixed $\omega \in \Omega$, the function $v \mapsto \langle g(\omega, v), w \rangle_{V^* \times V}$ is continuous for all $v, w \in V_{\xi_n(\omega)}$. It remains to verify that for fixed $v \in H$ and $w \in V$, the function $\omega \mapsto \langle g(\omega, v), w \rangle_{V^* \times V}$ is measurable. Let B be an open set in V^*. It then follows that

$$
\begin{aligned}
\big(g(\cdot, v)\big)^{-1}(B) \\
&= \{\omega \in \Omega : g(\omega, v) \in B\} \\
&= \{\omega \in \Omega : v \in V_{\xi_n(\omega)}, h_n f^n(\omega) + U^{n-1} - \big(I + h_n A_{\xi_n(\omega)}(t_n)\big)v \in B\} \\
&\quad \cup \{\omega \in \Omega : v \in H \backslash V_{\xi_n(\omega)}, e \in B\} \\
&= \big(\{\omega \in \Omega : v \in V_{\xi_n(\omega)}\} \cap \{\omega \in \Omega : h_n f^n(\omega) + U^{n-1} - \big(I + h_n A_{\xi_n(\omega)}(t_n)\big)v \in B\}\big) \\
&\quad \cup \big(\{\omega \in \Omega : v \in H \backslash V_{\xi_n(\omega)}\} \cap \{\omega \in \Omega : e \in B\}\big) \\
&=: (T_1 \cap T_2) \cup T_3.
\end{aligned}
$$

As the function $\omega \mapsto h_n f^n(\omega) + U^{n-1} - \big(I + h_n A_{\xi_n(\omega)}(t_n)\big)v$ is measurable, it follows that $T_2 \subset \Omega$ is measurable. The sets T_1 and T_3 are measurable by assumption. Thus, it follows that $\omega \mapsto g(\omega, v)$ and therefore $\omega \mapsto \langle g(\omega, v), w \rangle_{V^* \times V}$ is measurable.

As argued above for every $\omega \in \Omega$, there exists a unique element $U^n(\omega)$ such that $g(\omega, U^n(\omega)) = 0$. Thus, we can now apply Lemma A.3 to prove that $U^n \colon \Omega \to H$ is $\mathcal{F}_n$-measurable. □

Lemma 4.2 *Let Assumptions 2.1–2.5 be fulfilled. Further, let the random variables $f^n \colon \Omega \to H$ be given such that they are $\mathcal{F}_{\xi_n}$-measurable and $\mathbb{E}_{\xi_n}\big[\|f^n\|_H^2\big] < \infty$ for*

every $n \in \{1, \ldots, N\}$. Then for $2\kappa h_n \leq 2\kappa h < 1$ the solution $\{U^n\}_{n \in \{1, \ldots, N\}}$ of (2) fulfills the a priori bound

$$\mathbb{E}_n\big[\|U^n\|_H^2\big] + \sum_{i=1}^n \mathbb{E}_i\big[\|U^i - U^{i-1}\|_H^2\big] + 2\sum_{i=1}^n h_i \mathbb{E}_i\big[\mu_{\xi_i}|U^i|_{V_{\xi_i}}^2\big]$$

$$\leq C\Big(2\|u_0\|^2 + 4T\lambda + 5CT\sum_{i=1}^N h_i \mathbb{E}_{\xi_i}\big[\|f^i\|_H^2\big]\Big),$$

where $C = \frac{1}{1-2h\kappa}\exp\big(\frac{2\kappa T}{1-2h\kappa}\big)$ for all $n \in \{1, \ldots, N\}$.

The proof of this lemma is very similar to the proof of our main result Theorem 5.1 and therefore omitted. The main necessary modification is to directly test (2) with U^n and use the semi-coercivity from Remark 1.

5 Stability and convergence in expectation

With the previous sections in mind, we can now turn our attention to the main results of this paper. We provide error bounds for the scheme (2) measured in expectation. First, we give a stability result in Theorem 5.1. The aim of this bound is to show how two solutions of the same scheme with respect to different right-hand sides and initial values differ. This stability result can then be used to prove the desired error bounds in Theorem 5.2 by using well-chosen data that agrees with the exact solution at the grid points. Note that in contrast to other works (e.g. [10, 11]), we measure $f(t) - A(t)u(t)$ in the H-norm. This can be interpreted as a stricter regularity assumption. The advantage is that certain error terms disappear in expectation, compare the second bound in Lemma A.4.

Theorem 5.1 *Let Assumptions 2.1–2.5 be fulfilled. Further, let the random variable $f^n \colon \Omega \to H$ be given such that it is $\mathcal{F}_{\xi_n}$-measurable and $\mathbb{E}_{\xi_n}\big[\|f^n\|_H^2\big] < \infty$ for every $n \in \{1, \ldots, N\}$. Let $\{U^n\}_{n \in \{1, \ldots, N\}}$ be the solution of (2) and let $\{V^n\}_{n \in \{1, \ldots, N\}}$ be the solution of*

$$\begin{cases} V^n - V^{n-1} + h_n A_{\xi_n}(t_n)V^n = h_n g^n & \text{in } V_{\xi_n}^*, \quad n \in \{1, \ldots, N\}, \\ V^0 = v_0 & \text{in } H, \end{cases} \tag{7}$$

for $v_0 \in H$ and $g^n \colon \Omega \to H$ such that it is $\mathcal{F}_{\xi_n}$-measurable and $\mathbb{E}_{\xi_n}\big[\|g^n\|_H^2\big] < \infty$ for every $n \in \{1, \ldots, N\}$. Then for $2\kappa h_n \leq 2\kappa h < 1$, it follows that

$$\mathbb{E}_n\big[\|U^n - V^n\|_H^2\big] + \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}_i\big[\|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2\big]$$

$$+ 2\sum_{i=1}^{n} h_i \mathbb{E}_i\big[\eta_{\xi_i}|U^i - V^i|_{V_{\xi_i}}^p\big]$$

$$\leq 2C\|u_0 - v_0\|^2 + 4C\sum_{i=1}^{N} h_i^2 \mathbb{E}_i\big[\|f^i - g^i\|_H^2\big] + 5C^2 T\sum_{i=1}^{N} h_i \big\|\mathbb{E}_{\xi_i}\big[f^i - g^i\big]\big\|_H^2$$

for $C = \frac{1}{1-2h\kappa}\exp\left(\frac{2\kappa T}{1-2\kappa T}\right)$ *and* $n \in \{1, \dots, N\}$.

Proof We start by subtracting (7) from (2) and testing with $U^i - V^i$ to get

$$\big((U^i - V^i) - (U^{i-1} - V^{i-1}), U^i - V^i\big)$$
$$+ h_n\langle A_{\xi_i}(t_i)U^i - A_{\xi_i}(t_i)V^i, U^i - V^i\rangle_{V_{\xi_i}^* \times V_{\xi_i}} = h_n(f^i - g^i, U^i - V^i). \quad (8)$$

For the first term of this equality, we use the identity $(a - b, a) = \frac{1}{2}(\|a\|^2 - \|b\|^2 + \|a - b\|^2)$ for $a, b \in H$ to find that

$$\big((U^i - V^i) - (U^{i-1} - V^{i-1}), U^i - V^i\big)$$
$$= \frac{1}{2}\big(\|U^i - V^i\|_H^2 - \|U^{i-1} - V^{i-1}\|_H^2 + \|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2\big).$$

Due to the monotonicity condition from Assumption 2.2 (iii), we obtain

$$\langle A_{\xi_i}(t_i)U^i - A_{\xi_i}(t_i)V^i, U^i - V^i\rangle_{V_{\xi_i}^* \times V_{\xi_i}} + \kappa_{\xi_i}\|U^i - V^i\|_H^2 \geq \eta_{\xi_i}|U^i - V^i|_{V_{\xi_i}}^p.$$

It remains to find a bound for the right-hand side of (8). Applying Cauchy-Schwarz's inequality and the weighted Young's inequality for products (Lemma A.2 with $\varepsilon = 1$), it follows that

$$h_i\big(f^i - g^i, U^i - V^i\big)$$
$$= h_i\big(f^i - g^i, U^{i-1} - V^{i-1}\big) + h_i\big(f^i - g^i, U^i - V^i - (U^{i-1} - V^{i-1})\big)$$
$$\leq h_i\big(f^i - g^i, U^{i-1} - V^{i-1}\big) + h_i\|f^i - g^i\|_H\|U^i - V^i - (U^{i-1} - V^{i-1})\|_H$$
$$\leq h_i\big(f^i - g^i, U^{i-1} - V^{i-1}\big) + h_i^2\|f^i - g^i\|_H^2 + \frac{1}{4}\|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2.$$

Combining the previous statements, we find

$$0 = (U^i - V^i - (U^{i-1} - V^{i-1}), U^i - V^i)$$
$$+ h_i\langle A_{\xi_i}(t_i)U^i - A_{\xi_i}(t_i)V^i, U^i - V^i\rangle_{V_{\xi_i}^* \times V_{\xi_i}} - h_i\big(f^i - g^i, U^i - V^i\big)$$

$$\geq \frac{1}{2}\left(\|U^i - V^i\|_H^2 - \|U^{i-1} - V^{i-1}\|_H^2 + \|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2\right)$$
$$- h_i \kappa_{\xi_i} \|U^i - V^i\|_H^2 + h_i \eta_{\xi_i} |U^i - V^i|_{V_{\xi_i}}^p$$
$$- h_i\left(f^i - g^i, U^{i-1} - V^{i-1}\right) - h_i^2 \|f^i - g^i\|_H^2 - \frac{1}{4}\|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2.$$

After rearranging the terms and multiplying both sides of the inequality with the factor 2, we obtain the following bound

$$\|U^i - V^i\|_H^2 - \|U^{i-1} - V^{i-1}\|_H^2 + \frac{1}{2}\|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2$$
$$+ 2h_i \eta_{\xi_i} |U^i - V^i|_{V_{\xi_i}}^p$$
$$\leq 2h_i \kappa_{\xi_i} \|U^i - V^i\|_H^2 + 2h_i\left(f^i - g^i, U^{i-1} - V^{i-1}\right) + 2h_i^2 \|f^i - g^i\|_H^2.$$

By first taking the $\mathbb{E}_{\xi_i}$-expectation of this inequality and then applying also the $\mathbb{E}_{i-1}$-expectation, we find that

$$\mathbb{E}_i\left[\|U^i - V^i\|_H^2\right] - \mathbb{E}_{i-1}\left[\|U^{i-1} - V^{i-1}\|_H^2\right] + \frac{1}{2}\mathbb{E}_i\left[\|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2\right]$$
$$+ 2h_i \mathbb{E}_i\left[\eta_{\xi_i} |U^i - V^i|_{V_{\xi_i}}^p\right]$$
$$\leq 2h_i \mathbb{E}_i\left[\kappa_{\xi_i} \|U^i - V^i\|_H^2\right] + 2h_i \mathbb{E}_{i-1}\left[\left(\mathbb{E}_{\xi_i}\left[f^i - g^i\right], U^{i-1} - V^{i-1}\right)\right]$$
$$+ 2h_i^2 \mathbb{E}_{\xi_i}\left[\|f^i - g^i\|_H^2\right].$$

After combining the previous two inequalities and summing up from $i = 1$ to $n \in \{1, \ldots, N\}$, we obtain

$$\mathbb{E}_n\left[\|U^n - V^n\|_H^2\right] + \frac{1}{2}\sum_{i=1}^n \mathbb{E}_i\left[\|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2\right]$$
$$+ 2\sum_{i=1}^n h_i \mathbb{E}_i\left[\eta_{\xi_i} |U^i - V^i|_{V_{\xi_i}}^p\right] \tag{9}$$
$$\leq \|u_0 - v_0\|_H^2 + 2\kappa \sum_{i=1}^n h_i \mathbb{E}_i\left[\|U^i - V^i\|_H^2\right]$$
$$+ 2\sum_{i=1}^n h_i \mathbb{E}_{i-1}\left[\left(\mathbb{E}_{\xi_i}\left[f^i - g^i\right], U^{i-1} - V^{i-1}\right)\right] + 2\sum_{i=1}^N h_i^2 \mathbb{E}_{\xi_i}\left[\|f^i - g^i\|_H^2\right],$$

where we only made the right-hand side bigger by summing to the final value N. In the following, denote $i_{\max} \in \{1, \ldots, N\}$ such that $\max_{i \in \{1,\ldots,N\}} \mathbb{E}_i\left[\|U^i - V^i\|_H^2\right] = \mathbb{E}_{i_{\max}}\left[\|U^{i_{\max}} - V^{i_{\max}}\|_H^2\right]$. By Lemma A.3, it follows that $U^{i-1} - V^{i-1}$ is $\mathcal{F}_{i-1}$-measurable and thus independent of the $\mathcal{F}_{\xi_i}$-measurable random variable $f^i - g^i$.

Therefore, we find that

$$2 \sum_{i=1}^{n} h_i \mathbb{E}_{i-1} \left[\left(\mathbb{E}_{\xi_i} [f^i - g^i], U^{i-1} - V^{i-1} \right) \right]$$

$$\leq 2 \sum_{i=1}^{n} h_i \left\| \mathbb{E}_{\xi_i} [f^i - g^i] \right\|_H \mathbb{E}_{i-1} \left[\| U^{i-1} - V^{i-1} \|_H \right]$$

$$\leq 2 \left(\mathbb{E}_{i_{\max}} \left[\| U^{i_{\max}} - V^{i_{\max}} \|_H^2 \right] \right)^{\frac{1}{2}} \sum_{i=1}^{N} h_i \left\| \mathbb{E}_{\xi_i} [f^i - g^i] \right\|_H.$$

To keep the presentation compact, we abbreviate

$$B_1 = \sum_{i=1}^{N} h_i^2 \mathbb{E}_i \left[\| f^i - g^i \|_H^2 \right] \quad \text{and} \quad B_2 = \sum_{i=1}^{N} h_i \left\| \mathbb{E}_{\xi_i} [f^i - g^i] \right\|_H.$$

Setting

$$x_n = \mathbb{E}_n \left[\| U^n - V^n \|_H^2 \right] + \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}_i \left[\| U^i - V^i - (U^{i-1} - V^{i-1}) \|_H^2 \right]$$

$$+ 2 \sum_{i=1}^{n} h_i \mathbb{E}_i \left[\eta_{\xi_i} | U^i - V^i |_{V_{\xi_i}}^p \right],$$

we have $2\kappa \sum_{i=1}^{n} h_i \mathbb{E}_i \left[\| U^i - V^i \|_H^2 \right] \leq 2\kappa \sum_{i=1}^{n} h_i x_i$. We can now apply Grönwall's inequality (Lemma A.1) to (9). It follows that

$$x_n \leq C \left(\| u_0 - v_0 \|^2 + 2 B_1 + 2 \left(\mathbb{E}_{i_{\max}} \left[\| U^{i_{\max}} - V^{i_{\max}} \|_H^2 \right] \right)^{\frac{1}{2}} B_2 \right), \qquad (10)$$

for $C = \frac{1}{1 - 2 h \kappa} \exp \left(\frac{2 \kappa T}{1 - 2 h \kappa} \right)$. As this inequality holds for every $n \in \{1, \ldots, N\}$, it is also fulfilled for $i_{\max}$. Thus, it follows that

$$\mathbb{E}_{i_{\max}} \left[\| U^{i_{\max}} - V^{i_{\max}} \|_H^2 \right]$$

$$\leq C \left(\| u_0 - v_0 \|^2 + 2 B_1 + 2 \left(\mathbb{E}_{i_{\max}} \left[\| U^{i_{\max}} - V^{i_{\max}} \|_H^2 \right] \right)^{\frac{1}{2}} B_2 \right).$$

We can now use that $x^2 \leq 2ax + b^2$ implies that $x \leq 2a + b$ for $a, b, x \in [0, \infty)$ and find

$$\left(\mathbb{E}_{i_{\max}} \left[\| U^{i_{\max}} - V^{i_{\max}} \|_H^2 \right] \right)^{\frac{1}{2}} \leq C^{\frac{1}{2}} \left(\| u_0 - v_0 \|^2 + 2 B_1 \right)^{\frac{1}{2}} + 2 C B_2.$$

Inserting this bound in (10) and applying Young's inequality (Lemma A.2 for $\varepsilon = 1$), we then obtain

$$\mathbb{E}_n\big[\|U^n - V^n\|_H^2\big] + \frac{1}{2}\sum_{i=1}^n \mathbb{E}_i\big[\|U^i - V^i - (U^{i-1} - V^{i-1})\|_H^2\big]$$

$$+ 2\sum_{i=1}^n h_i \mathbb{E}_i\big[\eta_{\xi_i}|U^i - V^i|_{V_{\xi_i}}^2\big]$$

$$\leq C\Big(\|u_0 - v_0\|^2 + 2B_1 + 2C^{\frac{1}{2}}\big(\|u_0 - v_0\|^2 + 2B_1\big)^{\frac{1}{2}}B_2 + 4CB_2^2\Big)$$

$$\leq C\Big(\|u_0 - v_0\|^2 + 2B_1 + \big(\|u_0 - v_0\|^2 + 2B_1\big) + CB_2^2 + 4CB_2^2\Big)$$

$$= 2C\big(\|u_0 - v_0\|^2 + 2B_1\big) + 5C^2B_2^2.$$

It only remains to insert

$$B_2^2 = \Big(\sum_{i=1}^N h_i \big\|\mathbb{E}_{\xi_i}\big[f^i - g^i\big]\big\|_H\Big)^2 \leq T\sum_{i=1}^N h_i \big\|\mathbb{E}_{\xi_i}\big[f^i - g^i\big]\big\|_H^2,$$

to finish the proof.   □

Theorem 5.2 *Let Assumptions 2.1–2.5 be fulfilled. Further, let $f_{\xi_n} \in C([0, T]; H)$ almost surely and $f^n = f_{\xi_n}(t_n) \in L^2(\Omega; H)$ for all $n \in \{1, \ldots, N\}$. Let $\{U^n\}_{n\in\{1,\ldots,N\}}$ be the solution of (2) and u be the solution of (1) that fulfills $u' \in C^\gamma([0, T]; H)$, $\gamma \in (0, 1]$. Moreover, let $A_{\xi_n}(t_n)u(t_n) \in L^2(\Omega; H)$ be fulfilled.*
   Then for $2\kappa h_n \leq 2\kappa h < 1$ and $e^n = U^n - u(t_n)$, it follows that

$$\mathbb{E}_n\big[\|e^n\|_H^2\big] + \frac{1}{2}\sum_{i=1}^n \mathbb{E}_i\big[\|e^i - e^{i-1}\|_H^2\big] + 2\sum_{i=1}^n h_i \mathbb{E}_i\big[\eta_{\xi_i}|e^i|_{V_{\xi_i}}^p\big]$$

$$\leq 8hC\sum_{i=1}^N h_i \mathbb{E}_{\xi_i}\big[\big\|f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - (f(t_i) - A(t_i)u(t_i))\big\|_H^2\big]$$

$$+ 4h^{1+2\gamma}C|u'|_{C^\gamma([0,T];H)}^2 T + 5h^{2\gamma}C^2|u'|_{C^\gamma([0,T];H)}^2 T^2,$$

where $C = \frac{1}{1-2h\kappa}\exp\big(\frac{2\kappa T}{1-2h\kappa}\big)$ for all $n \in \{1, \ldots, N\}$.

Proof We use $\{V^n\}_{n\in\{1,\ldots,N\}}$ given by

$$\begin{cases} V^n - V^{n-1} + h_n A_{\xi_n}(t_n)V^n = h_n g^n & \text{in } V_{\xi_n}^*, \quad n \in \{1, \ldots, N\}, \\ V^0 = u_0 & \text{in } H, \end{cases}$$

where

$$g^n = \frac{1}{h_n}\big(u(t_n) - u(t_{n-1})\big) + A_{\xi_n}(t_n)u(t_n) \in L^2(\Omega; H).$$

With this particular choice of g^n, we can now show that $V^n = u(t_n)$ for every $n \in \{1, \ldots, N\}$. Given the initial value u_0, the solution V^1 is then given by

$$
\begin{aligned}
V^1 &= u_0 + h_1 g^1 - h_1 A_{\xi_1}(t_1) V^1 \\
&= u_0 + \big(u(t_1) - u(t_0)\big) + h_1 A_{\xi_1}(t_1) u(t_1) - h_1 A_{\xi_1}(t_1) V^1 \\
&= u(t_1) + h_1 A_{\xi_1}(t_1) u(t_1) - h_1 A_{\xi_1}(t_1) V^1.
\end{aligned}
$$

Therefore, it follows that

$$
(I + h_1 A_{\xi_1}(t_1)) V^1 = (I + h_1 A_{\xi_1}(t_1)) u(t_1) \quad \text{in } V_{\xi_1}^*.
$$

Since $I + h_1 A_{\xi_1}(t_1)$ is injective, we find $V^1 = u(t_1)$ in V_{ξ_1}. Recursively, it follows that $V^n = u(t_n)$ in V_{ξ_n} for all other $n \in \{1, \ldots, N\}$. Together with the stability estimate from Theorem 5.1 we find for $e^n = U^n - V^n = U^n - u(t_n)$ that

$$
\mathbb{E}_n\big[\|e^n\|_H^2\big] + \frac{1}{2} \sum_{i=1}^n \mathbb{E}_i\big[\|e^i - e^{i-1}\|_H^2\big] + 2 \sum_{i=1}^n h_i \mathbb{E}_i\big[\eta_{\xi_i} |e^i|_{V_{\xi_i}}^2\big] \le 4 C B_1 + 5 C^2 B_2^2,
$$

where

$$
B_1 = \sum_{i=1}^N h_i^2 \mathbb{E}_i\big[\|f^i - g^i\|_H^2\big], \quad B_2 = \sum_{i=1}^N h_i \big\|\mathbb{E}_{\xi_i}\big[f^i - g^i\big]\big\|_H,
$$

$$
C = \frac{1}{1 - 2h\kappa} \exp\Big(\frac{2\kappa T}{1 - 2\kappa T}\Big).
$$

Applying Lemma A.4 for $u' \in C^\gamma([0, T]; H)$, it follows that

$$
\begin{aligned}
B_1 &\le h \sum_{i=1}^N h_i \mathbb{E}_{\xi_i}\Big[\Big\| f_{\xi_i}(t_i) - A_{\xi_i}(t_i) u(t_i) - \frac{1}{h_i} \int_{t_{i-1}}^{t_i} \big(f(t) - A(t) u(t)\big)\, dt \Big\|_H^2\Big] \\
&\le 2h \sum_{i=1}^N h_i \mathbb{E}_{\xi_i}\big[\| f_{\xi_i}(t_i) - A_{\xi_i}(t_i) u(t_i) - (f(t_i) - A(t_i) u(t_i))\|_H^2\big] \\
&\quad + 2h^{1+2\gamma} |u'|_{C^\gamma([0,T];H)}^2 T
\end{aligned}
$$

and

$$
\begin{aligned}
B_2^2 &\le T \sum_{i=1}^N h_i \Big\|\mathbb{E}_{\xi_i}\Big[f_{\xi_i}(t_i) - A_{\xi_i}(t_i) u(t_i) - \frac{1}{h_i} \int_{t_{i-1}}^{t_i} \big(f(t) - A(t) u(t)\big)\, dt \Big]\Big\|_H^2 \\
&\le h^{2\gamma} |u'|_{C^\gamma([0,T];H)}^2 T^2.
\end{aligned}
$$

Altogether, we obtain

$$
\mathbb{E}_n\big[\|e^n\|_H^2\big] + \frac{1}{2}\sum_{i=1}^n \mathbb{E}_i\big[\|e^i - e^{i-1}\|_H^2\big] + 2\sum_{i=1}^n h_i \mathbb{E}_i\big[\eta_{\xi_i}|e^i|_{V_{\xi_i}}^2\big]
$$

$$
\leq 8hC\sum_{i=1}^N h_i \mathbb{E}_{\xi_i}\Big[\big\|f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - (f(t_i) - A(t_i)u(t_i))\big\|_H^2\Big]
$$

$$
+ 4h^{1+2\gamma}C|u'|_{C^\gamma([0,T];H)}^2 T + 5h^{2\gamma}C^2|u'|_{C^\gamma([0,T];H)}^2 T^2.
$$

$\square$

Remark 2 The main results can all be modified to a slightly different setting, where the right-hand side $f(t)$ takes values in V^* and where the family $\{\xi_n\}_{n\in\mathbb{N}}$ of random variables does not have to be mutually independent. In return, this setting requires slightly stronger assumptions on the operator $A(t)$. First, we assume additionally that there exists a constant $c_V \in (0,\infty)$ such that $\|\cdot\|_V \leq c_V\big(\|\cdot\|_H + |\cdot|_V\big)$ is fulfilled. To generalize the a priori bound from Lemma 4.2 and the stability results from Theorem 5.1, we need to assume that μ_A from Assumption 2.2 (v) and η_A from Assumption 2.2 (iii) are strictly positive, respectively. Moreover, if there exist $\gamma \in (0,1]$ and $C \in [0,\infty)$ such that

$$
\sum_{i=1}^N h_i \mathbb{E}_i\Big[\big\|f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - (f(t_i) - A(t_i)u(t_i))\big\|_{V_{\xi_i}^*}^2\Big] \leq Ch^{2\gamma}
$$

is fulfilled and $u' \in C^\gamma([0,T];H)$, we obtain similar error bounds. We omit the proofs, which are very similar to the ones presented above.

6 Numerical experiments

To illustrate the theoretical convergence results for the randomized scheme in practice, we apply it to Eq. (5) as discussed in Sect. 3. This boundary, initial-value problem fits our setting as already explained there. We also consider what happens when we replace the nonlinear diffusion term with linear diffusion, and a smoother exact solution.

In both cases, we consider the problem on the spatial domain $\mathcal{D} = [-1,1] \times [-1,1]$ which we split into rectangular sub-domains $\mathcal{D}_\ell$, $\ell \in \{1,\ldots,s\}$, with M_x rectangles along the x-axis and M_y rectangles along the y-axis. We choose $\mathcal{D}_\ell$ such that they have an overlap of 0.2 on all internal sides. This means that, for example, with $M_x = M_y = 3$, we have $s = M_x M_y = 9$ sub-domains with, e.g., $\mathcal{D}_1 = [-1,-0.267] \times [-1,-0.267]$, $\mathcal{D}_2 = [-0.467,0.467] \times [-1,-0.267]$ and $\mathcal{D}_5 = [-0.467,0.467] \times [-0.467,0.467]$. Note that they are not uniform in size, because the sub-domains adjacent to the outer edge of $\mathcal{D}$ have no overlap on one or two sides.

We have to choose a strategy for which sub-problems to select in each time step, i.e. specify the probabilities $\mathcal{P}(\Omega_{\xi=B})$ for $B \subset 2^{\{1,\ldots,s\}}$. We consider two strategies. In the first, we simply use $\mathcal{P}(\Omega_{\xi=\{\ell\}}) = 1/s$. Thus every sub-domain is equally probable to

be chosen. As a minor variation, we instead select a set of k sub-domains by drawing with replacement according to the uniform probabilities.

In the second strategy, we make use of a predictor. In addition to the stochastic approximation, we compute a deterministic approximation Z^n using the backward Euler method, but on a coarser spatial mesh. The idea is that while this approximation is less accurate, it should be significantly cheaper to compute and still resemble the true solution. In the nth time step, we compute $\Psi_n = |Z^{n-1}| + |Z^n| + |\tilde{f}(t_n, \cdot)| > 10^{-3}$. This function is either 0 or 1 and indicates where in the domain something is actually happening. For each sub-domain, we then check whether it is "sufficiently active" or not by evaluating $\|\Psi_n \chi_l\| \geq \rho \|\Psi_n\|$ for a parameter $\rho \in (0, 1)$. We select the set of those sub-domains which pass the test with probability $1 - \rho$ and the set of all the other sub-domains with probability ρ.

We note that the errors for the first strategy are noticeably larger than those of the second strategy. In the following, we will use fewer sub-domains for the first strategy for that reason. More precisely, we use $M_x = 3$ and $M_y = 1$ for first strategy and $M_x = 3$ and $M_y = 3$ for the second strategy. Furthermore, we can observe that the second strategy works better with more sub-domains, since it essentially adaptively groups them into only two larger sub-domains; the active set and the inactive set. Increasing the number of sub-domains increases the fidelity such that the choice of whether each sub-domain is active or not becomes easier, albeit at a higher computational cost. If the spatial discretization is using finite elements, the limit case would be when every element is its own subdomain. This is what is considered in [36] for a deterministic scheme, where it is, indeed, observed that the overhead costs can be prohibitive even when using very efficient data structures.

We only report errors here, since this is the focus of the paper. A natural next step would be to investigate also the computation times and the efficiency of the schemes compared to deterministic schemes. Since the randomized methods need to solve equation systems of smaller size, they are expected to outperform the deterministic schemes. However, this depends on many factors, such as the problem size, the number of subdomains, the behaviour of the exact solution and the random strategy used. Further, for such a comparison to be useful, it has to be performed with equally optimized and parallelized code for both the randomized and deterministic cases. Such advanced software engineering is out of the scope of this article. Nevertheless, when applying our non-parallelized and not fully optimized code to the linear diffusion problem using the first strategy, we observed a factor 2 speed-up that was independent of the number of time steps.

6.1 A nonlinear example

In our first experiment, we use the problem parameters $T = 1$, $p = 4$ and $\alpha(t) \equiv 1$. Further, we choose the source term $\tilde{f}$ such that the exact solution is given by $u(t, x, y) = \tilde{u}(x - r\cos(2\pi t), y - r\sin(2\pi t))$ with $r = 1/2$,

$$\tilde{u}(x, y) = \left[0.03 - \frac{10^{3/8}}{4}(x^2 + y^2)^{\frac{4}{3}}\right]_+^{\frac{3}{4}}$$

and $[\cdot]_+ = \max\{\cdot, 0\}$. This describes a localized pulse that starts centered at $(0.5, 0)$ and which then rotates around the origin at the constant distance r. The shape of the pulse is inspired by the closed-form Barenblatt solution to $\partial_t u = \nabla \cdot (|\nabla u(t, x)|^{p-2}\nabla u)$, see e.g. [21]. At $t = 0$, this solution is a Dirac delta, which then expands into a cone-shaped peak for $t > 0$. Our pulse is this solution frozen at the time $t = 0.001$. We note that due to the sharp interface where the pulse meets the x-y-plane and to the sharp peak, u is of low regularity.

We discretize the problem in space using central finite differences, such that the approximation of the p-Laplacian is 2nd-order accurate. We use 100 computational nodes in each spatial dimension, for a total of $10\,000$ degrees of freedom. Thus, the temporal error dominates the spatial error when considering the full error in the following. For the temporal discretization, we use the scheme (2), along with one of the two strategies outlined above. For the first strategy, we try $k = 1$ and $k = 2$. For the second, we evaluate the different parameters $\rho = 0.01, 0.05, 0.1, 0.2$. We compute approximations for the different (constant) time steps $h_n = 2^{-5}, 2^{-6}, \ldots, 2^{-13}$ and estimate their corresponding errors at the final time by running the method with 50 random iterations and averaging. That is, we approximate

$$\mathbb{E}_N\left[\|e^N\|_H^2\right] \approx \frac{1}{50}\sum_{j=1}^{50} \|U_j^N - U_{\text{ref}}\|_H^2,$$

where U_j^N is the numerical approximation on the j-th path and U_{ref} is the exact solution $u(t_N, \cdot, \cdot)$ evaluated at the spatial grid.

Figure 1 shows the resulting relative errors vs. the time steps, with the first strategy in the upper plot and the second strategy in the lower. We observe that both strategies result in errors that decrease as $\mathcal{O}(h^{1/2})$, in line with Theorem 5.2.

6.2 A linear example

As a second experiment, we consider a linear version of the previous problem. We use the same parameters as in the previous section, except that we set $p = 2$ and $\alpha(t) = 0.1$, and that the rotating pulse is now Gaussian rather than a sharp peak. More precisely, the exact solution is given by

$$u(t, x, y) = e^{-100(x - r\cos(2\pi t))^2 - 100(y - r\sin(2\pi t))^2}.$$

The resulting errors are shown in Fig. 2. Again, we note that the first, uniform, strategy converges as $\mathcal{O}(h^{1/2})$, in line with Theorem 5.2. The second strategy with $\rho = 0.01$ performs significantly better and the error behaves like $\mathcal{O}(h)$ in the first part of the plot. This is essentially the same behaviour as if we would apply backward Euler to the full problem, but the method only updates the approximation on the most relevant sub-domains and is therefore cheaper to evaluate. This improved convergence order is
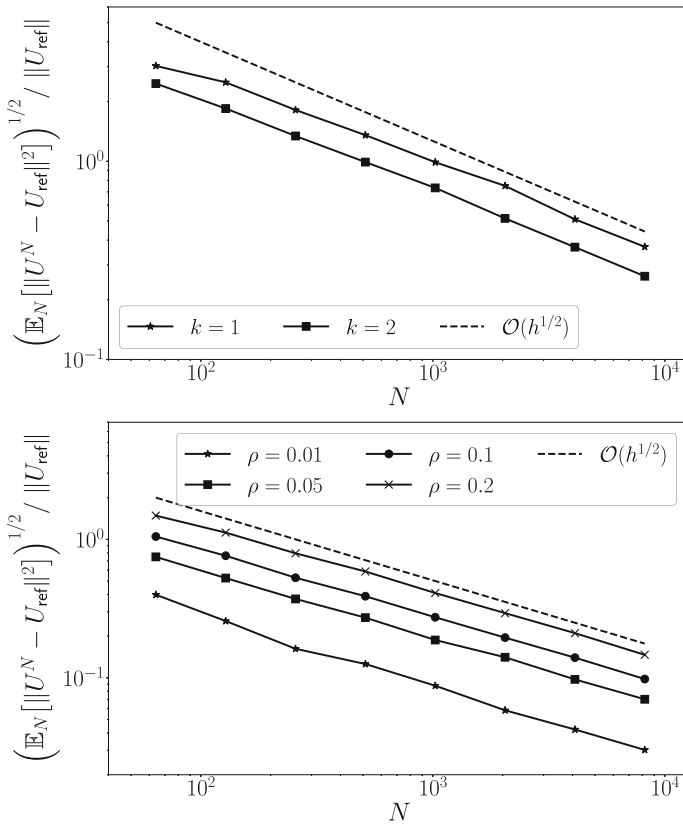
Fig. 1 The relative errors $\left(\mathbb{E}_N\left[\|U^N - U_{\text{ref}}\|^2\right]\right)^{1/2} / \|U_{\text{ref}}\|$ for the nonlinear setting described in Sect. 6.1. The upper plot uses the first randomized strategy and the lower plot uses the second strategy. We observe that the errors decay as $\mathcal{O}(h^{1/2})$, in line with Theorem 5.2, irrespective of the choice of ρ or k. A smaller ρ or larger k decreases the error, but of course also incurs a higher computational cost

possible due to the extra smoothness present in this linear problem. In the error bound of Theorem 5.2, the first term becomes small due to the used strategy, and because the solution is smooth the remaining terms are of size h^3 and h^2, respectively.

Increasing the parameter ρ means that we disregard more of the information from the predictor, and as seen in Fig. 2 this causes the convergence order to decrease towards 1/2. On the other hand, setting $\rho = 0$ means that we always choose all the sub-domains and thereby do more computations than if we would simply solve the full problem directly. The parameter ρ is therefore a design parameter, and further research is required on how to choose it optimally for specific problem classes. Regardless of the choice, however, we still have $\mathcal{O}(h^{1/2})$-convergence.
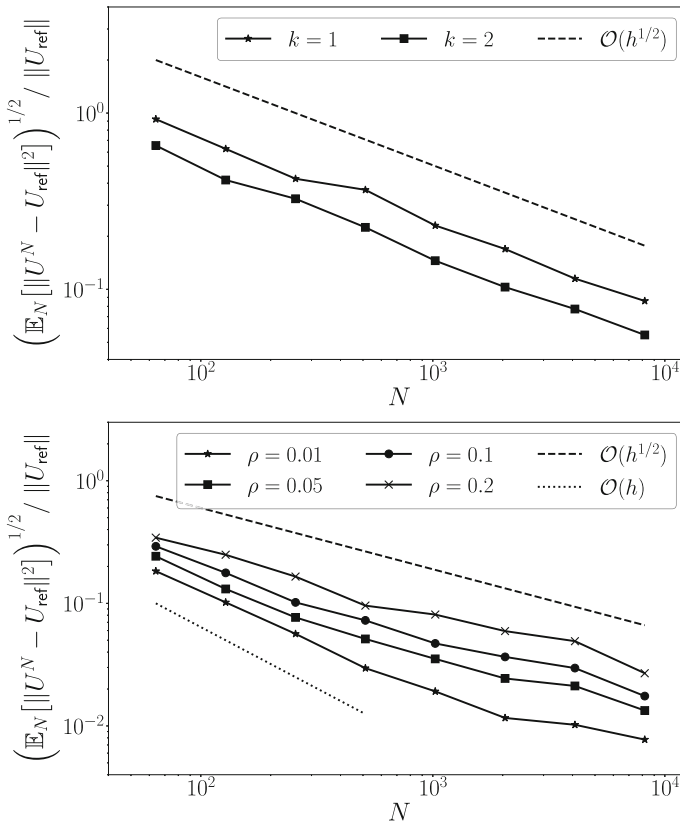
Fig. 2 The relative errors $\left(\mathbb{E}_N\left[\|U^N - U_{\text{ref}}\|^2\right]\right)^{1/2} / \|U_{\text{ref}}\|$ for the linear setting described in Sect. 6.2. The upper plot uses the first randomized strategy and the lower plot uses the second strategy. We observe that the errors for the first strategy decay as $\mathcal{O}(h^{1/2})$, similarly to the nonlinear case. For the second strategy, large ρ also leads to convergence of order 1/2, while sufficiently small ρ leads to faster convergence of order 1

Appendix A Auxiliary results

In this appendix, we collect a few useful inequalities and technical results that are needed in the paper.

Lemma A.1 (Discrete Grönwall inequality) *Let $(u_n)_{n \in N}$ and $(b_n)_{n \in N}$ be two nonnegative sequences that satisfy, for given $a \in [0, \infty)$ and $n \in \mathbb{N}$, that $u_n \leq a + \sum_{i=1}^{n} b_i u_i$. For $b_n \in [0, 1)$, it then follows that*

$$u_n \leq \frac{a}{1 - b_n} \exp \Big(\sum_{i=1}^{n-1} \frac{b_i}{1 - b_n} \Big).$$

Lemma A.2 (Weighted Young's inequality) *For $a, b \in [0, \infty)$, $\varepsilon \in (0, \infty)$, and $p, q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$, it follows that $ab \leq \varepsilon a^p + (\varepsilon p)^{-\frac{q}{p}} q^{-1} b^q$.*

A proof can be found in [12, Appendix B.2 d].

Lemma A.3 *Let Assumptions 2.1–2.5 be fulfilled. Let $Q \subseteq V$ be a countable, dense subset of V, V_ξ and H. Let the function $g \colon \Omega \times H \to V^*$ be given. Further, for $v \in H$ the mapping $\omega \mapsto \langle g(\omega, v), w \rangle_{V^* \times V}$ is measurable for $v \in H$ and $w \in Q$ and for almost every $\omega \in \Omega$ the mapping $v \mapsto \langle g(\omega, v), w \rangle_{V^* \times V}$ continuous for every $v, w \in V_{\xi(\omega)}$. For every $\omega \in \Omega$, the function g has a unique root which lies in $V_{\xi(\omega)}$. We denote this root by $r(\omega) \in V_{\xi(\omega)}$, i.e. $g(\omega, r(\omega)) = 0$. Then the function $r \colon \Omega \to H$ is measurable.*

A similar proof can be found in [5, Lemma 2.1.4] and [8, Lemma 4.3]. The main difference in this version is that the function g maps from $\Omega \times H$ instead of $\Omega \times V$ and therefore some small technical alterations have to be considered.

Proof of Lemma A.3 To prove that r is measurable, we show that $r^{-1}(B) \in \mathcal{F}$ for every open set B in H. First, we notice that

$$
\begin{aligned}
r^{-1}(B) &= \{\omega \in \Omega \colon r(\omega) \in B\} \\
&= \{\omega \in \Omega \colon \text{ there exists } u \in B \text{ such that } g(\omega, u) = 0\} \\
&= \{\omega \in \Omega \colon \text{ there exists } u \in B \text{ such that } \langle g(\omega, u), v \rangle_{V^* \times V} = 0 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for all } v \in Q, \|v\|_V = 1\} \\
&= \bigcap_{v \in Q, \|v\|_V = 1} \{\omega \in \Omega \colon \text{ there exists } u \in B \text{ such that } \langle g(\omega, u), v \rangle_{V^* \times V} = 0\} \\
&= \bigcap_{v \in Q, \|v\|_V = 1} \bigcup_{u \in B} \{\omega \in \Omega \colon \langle g(\omega, u), v \rangle_{V^* \times V} = 0\}.
\end{aligned}
$$

Since $\omega \mapsto \langle g(\omega, v), w \rangle_{V^* \times V}$ is measurable for $v \in H$ and $w \in Q$, the set

$$\{\omega \in \Omega \colon \langle g(\omega, u), v \rangle_{V^* \times V} = 0\} = \big(\langle g(\cdot, u), v \rangle_{V^* \times V} \big)^{-1}(0)$$

is an element of $\mathcal{F}_\xi$ for $v \in Q$ and $u \in H$. If the set B only contains a countable amount of elements, it follows directly that $r^{-1}(B) \in \mathcal{F}_\xi$.

In the following, it remains to address the cases where B is not countable. For $\varepsilon \in (0, \infty)$ small enough and a fixed $v \in Q$, we introduce the multi-valued mapping

$$r_\varepsilon^v : \Omega \to 2^H, \quad r_\varepsilon^v(\omega) = \{u \in H : |\langle g(\omega, u), v \rangle_{V^* \times V}| < \varepsilon\}.$$

For $B \subseteq H$ open, it follows that

$$
\begin{aligned}
\left(r_\varepsilon^v\right)^{-1}(B) &= \{\omega \in \Omega : r_\varepsilon^v(\omega) \in B\} \\
&= \{\omega \in \Omega : \text{ there exists } u \in B \text{ such that } |\langle g(\omega, u), v \rangle_{V^* \times V}| < \varepsilon\} \\
&= \bigcup_{u \in B} \{\omega \in \Omega : |\langle g(\omega, u), v \rangle_{V^* \times V}| < \varepsilon\}.
\end{aligned}
$$

In the following, we will show that

$$\left(r_\varepsilon^v\right)^{-1}(B) = \left(r_\varepsilon^v\right)^{-1}(B \cap Q).$$

Since $B \cap Q \subseteq B$, it directly follows that $\left(r_\varepsilon^v\right)^{-1}(B \cap Q) \subseteq \left(r_\varepsilon^v\right)^{-1}(B)$. It remains to verify that $\left(r_\varepsilon^v\right)^{-1}(B) \subseteq \left(r_\varepsilon^v\right)^{-1}(B \cap Q)$. Let $\omega \in \left(r_\varepsilon^v\right)^{-1}(B)$, i.e. there exists $u \in B$ such that

$$u \in r_\varepsilon^v(\omega) = \{w \in H : |\langle g(\omega, w), v \rangle_{V^* \times V}| < \varepsilon\}.$$

Since $v \mapsto \langle g(\omega, v), w \rangle_{V^* \times V}$ is continuous for every $v, w \in V_{\xi(\omega)}$ and Q is dense in H, there exists $u_Q \in B \cap Q$ such that $|\langle g(\omega, u_Q), v \rangle_{V^* \times V}| < \varepsilon$. Thus, $u_Q \in r_\varepsilon^v(\omega)$ and in particular $\omega \in \left(r_\varepsilon^v\right)^{-1}(B \cap Q)$. This shows altogether that $\left(r_\varepsilon^v\right)^{-1}(B) = \left(r_\varepsilon^v\right)^{-1}(B \cap Q)$.

We can now finish the proof as

$$r^{-1}(B) = \bigcap_{v \in Q, \|v\|_V = 1} \bigcap_{i \in \mathbb{N}} \left(r_{\frac{1}{i}}^v\right)^{-1}(B) = \bigcap_{v \in Q, \|v\|_V = 1} \bigcap_{i \in \mathbb{N}} \left(r_{\frac{1}{i}}^v\right)^{-1}(B \cap Q) \in \mathcal{F}_\xi$$

is fulfilled. □

Lemma A.4 *Let Assumptions 2.1–2.5 be fulfilled. Further, let f_{ξ_n} be an element of $C([0, T]; V_{\xi_n}^*)$ almost surely for every $n \in \{1. \dots .N\}$. For $u' \in C^\gamma([0, T]; H)$, $\gamma \in (0, 1]$, and a maximal step size $h = \max_{i \in \{1, \dots, N\}} h_i$, it then follows that*

$$
\begin{aligned}
\sum_{i=1}^N h_i \mathbb{E}_{\xi_i} & \left[\left\| f_{\xi_i}(t_i) - A_{\xi_i}(t_i) u(t_i) - \frac{1}{h_i} \int_{t_{i-1}}^{t_i} (f(t) - A(t)u(t)) \, dt \right\|_H^2 \right] \\
& \leq 2 \sum_{i=1}^N h_i \mathbb{E}_{\xi_i} \left[\left\| f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - (f(t_i) - A(t_i)u(t_i)) \right\|_H^2 \right] \\
& \quad + 2h^{2\gamma} |u'|_{C^\gamma([0,T];H)}^2 T
\end{aligned}
$$

and

$$\sum_{i=1}^{N} h_i \left\| \mathbb{E}_{\xi_i} \left[f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - \frac{1}{h_i} \int_{t_{i-1}}^{t_i} \left(f(t) - A(t)u(t) \right) dt \right] \right\|_H^2$$
$$\leq h^{2\gamma} |u'|_{C^\gamma([0,T];H)}^2 T,$$

where $|u'|_{C^\gamma([0,T];H)}$ is the Hölder semi-norm with values in H of the function u'.

Proof To prove the first bound, we find that

$$\sum_{i=1}^{N} h_i \mathbb{E}_{\xi_i} \left[\left\| f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - \frac{1}{h_i} \int_{t_{i-1}}^{t_i} \left(f(t) - A(t)u(t) \right) dt \right\|_H^2 \right]$$

$$\leq 2 \sum_{i=1}^{N} h_i \mathbb{E}_i \left[\left\| f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - (f(t_i) - A(t_i)u(t_i)) \right\|_H^2 \right]$$

$$+ 2 \sum_{i=1}^{N} h_i \mathbb{E}_i \left[\left\| \frac{1}{h_i} \int_{t_{i-1}}^{t_i} \left(f(t_i) - A(t_i)u(t_i) - (f(t) - A(t)u(t)) \right) dt \right\|_H^2 \right]$$

$$\leq 2 \sum_{i=1}^{N} h_i \mathbb{E}_i \left[\left\| f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - (f(t_i) - A(t_i)u(t_i)) \right\|_H^2 \right]$$

$$+ 2 \sum_{i=1}^{N} \frac{1}{h_i} \left\| \int_{t_{i-1}}^{t_i} (u'(t_i) - u'(t)) dt \right\|_H^2.$$

To further bound the term in the last row, we apply Hölder's inequality and the regularity condition $u' \in C^\gamma([0, T]; H)$. We then find that

$$2 \sum_{i=1}^{N} \frac{1}{h_i} \left\| \int_{t_{i-1}}^{t_i} (u'(t_i) - u'(t)) dt \right\|_H^2 \leq 2 \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \|u'(t_i) - u'(t)\|_H^2 dt$$
$$\leq 2 h^{2\gamma} |u'|_{C^\gamma([0,T];H)}^2 T.$$

It remains to prove the second estimate of the lemma. Recall that $\mathbb{E}_{\xi_i}\left[f_{\xi_i}(t_i) \right] = f(t_i)$ and $\mathbb{E}_{\xi_i}\left[A_{\xi_i}(t_i)u(t_i) \right] = A(t_i)u(t_i)$ is fulfilled by Assumption 2.5. Using these equalities, it follows that

$$\sum_{i=1}^{N} h_i \left\| \mathbb{E}_{\xi_i} \left[f_{\xi_i}(t_i) - A_{\xi_i}(t_i)u(t_i) - \frac{1}{h_i} \int_{t_{i-1}}^{t_i} \left(f(t) - A(t)u(t) \right) dt \right] \right\|_H^2$$

$$= \sum_{i=1}^{N} h_i \left\| f(t_i) - A(t_i)u(t_i) - \frac{1}{h_i} \int_{t_{i-1}}^{t_i} (f(t) - A(t)u(t)) dt \right\|_H^2$$

$$\leq \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \|u'(t_i) - u'(t)\|_H^2 dt \leq h^{2\gamma} |u'|_{C^\gamma([0,T];H)}^2 T.$$

$\square$

References

1. Aronsson, G., Evans, L.C., Wu, Y.: Fast/slow diffusion and growing sandpiles. J. Differential Equations **131**(2), 304–335 (1996)
2. Bochacik, T., Goćwin, M., Morkisz, P.M., Przybyłowicz, P.: Randomized Runge-Kutta method-stability and convergence under inexact information. J. Complexity **65**, 101554 (2021)
3. Bottou, L., Curtis, F.E., Nocedal, J.: Optimization methods for large-scale machine learning. SIAM Rev. **60**(2), 223–311 (2018)
4. Daun, T.: On the randomized solution of initial value problems. J. Complexity **27**(3–4), 300–311 (2011)
5. Eisenmann, M.: Methods for the temporal approximation of nonlinear, nonautonomous evolution equations. PhD thesis, TU Berlin (2019)
6. Eisenmann, M., Hansen, E.: Convergence analysis of domain decomposition based time integrators for degenerate parabolic equations. Numer. Math. **140**(4), 913–938 (2018)
7. Eisenmann, M., Hansen, E.: A variational approach to the sum splitting scheme. IMA J. Numer. Anal. **42**(1), 923–950 (2022)
8. Eisenmann, M., Kovács, M., Kruse, R., Larsson, S.: On a randomized backward Euler method for nonlinear evolution equations with time-irregular coefficients. Found. Comput. Math. **19**(6), 1387–1430 (2019)
9. Eisenmann, M., Stillfjord, T., Williamson, M.: Sub-linear convergence of a stochastic proximal iteration method in Hilbert space. Comput. Optim. Appl. **83**(1), 181–210 (2022)
10. Emmrich, E.: Two-step BDF time discretisation of nonlinear evolution problems governed by monotone operators with strongly continuous perturbations. Comput. Methods Appl. Math. **9**(1), 37–62 (2009)
11. Emmrich, E., Thalhammer, M.: Stiffly accurate Runge-Kutta methods for nonlinear evolution problems governed by a monotone operator. Math. Comp. **79**(270), 785–806 (2010)
12. Evans, L.C.: Partial differential equations. American Mathematical Society, Providence, RI (1998)
13. Hansen, E., Henningsson, E.: Additive domain decomposition operator splittings–convergence analyses in a dissipative framework. IMA J. Numer. Anal. **37**(3), 1496–1519 (2017)
14. Hansen, E., Ostermann, A.: Dimension splitting for quasilinear parabolic equations. IMA J. Numer. Anal. **30**(3), 857–869 (2010)
15. Hansen, E., Stillfjord, T.: Convergence of the implicit-explicit Euler scheme applied to perturbed dissipative evolution equations. Math. Comp. **82**(284), 1975–1985 (2013)
16. Hundsdorfer, W., Verwer, J.: Numerical solution of time-dependent advection-diffusion-reaction equations. Springer, Berlin (2003)
17. Jakobsen, E.R., Karlsen, K.H.: Convergence rates for semi-discrete splitting approximations for degenerate parabolic equations with source terms. BIT **45**(1), 37–67 (2005)
18. Jentzen, A., Neuenkirch, A.: A random Euler scheme for Carathéodory differential equations. J. Comput. Appl. Math. **224**(1), 346–359 (2009)
19. Jin, S., Li, L., Liu, J.-G.: Random batch methods (RBM) for interacting particle systems. J. Comput. Phys. **400**, 108877 (2020)
20. Jin, S., Li, L., Liu, J.-G.: Convergence of the random batch method for interacting particles with disparate species and weights. SIAM J. Numer. Anal. **59**(2), 746–768 (2021)
21. Kamin, S., Vázquez, J.L.: Fundamental solutions and asymptotic behaviour for the p-Laplacian equation. Rev. Mat. Iberoam. **4**(2), 339–354 (1988)
22. Kruse, R., Wu, Y.: Error analysis of randomized Runge-Kutta methods for differential equations with time-irregular coefficients. Comput. Methods Appl. Math. **17**(3), 479–498 (2017)
23. Kruse, R., Wu, Y.: A randomized and fully discrete Galerkin finite element method for semilinear stochastic evolution equations. Math. Comp. **88**(320), 2793–2825 (2019)
24. Kruse, R., Wu, Y.: A randomized Milstein method for stochastic differential equations with non-differentiable drift coefficients. Discrete Contin. Dyn. Syst. Ser. B **24**(8), 3475–3502 (2019)
25. Kufner, A.: Weighted Sobolev Spaces. Teubner, Leipzig (1980)
26. Kuijper, A.: p-Laplacian driven image processing. In: 2007 IEEE International conference on image processing, Vol. 5, pp. V-257–V-260 (2007)
27. Leoni, G.: A First Course in Sobolev Spaces. American Mathematical Society, Providence, RI (2009)
28. Li, L., Xu, Z., Zhao, Y.: A random-batch Monte Carlo method for many-body systems with singular kernels. SIAM J. Sci. Comput. **42**(3), A1486–A1509 (2020)
29. Lions, P.-L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)

30. Mathew, T.: Domain decomposition methods for the numerical solution of partial differential equations. Springer, Berlin (2008)
31. Mathew, T.P., Polyakov, P.L., Russo, G., Wang, J.: Domain decomposition operator splittings for the solution of parabolic equations. SIAM J. Sci. Comput. **19**(3), 912–932 (1998)
32. Robbins, H., Monro, S.: A stochastic approximation method. Ann. Math. Statistics **22**, 400–407 (1951)
33. Roubíček, T.: Nonlinear partial differential equations with applications, 2nd edn. Birkhäuser, Basel (2013)
34. Stengle, G.: Numerical methods for systems with measurable coefficients. Appl. Math. Lett. **3**(4), 25–29 (1990)
35. Stengle, G.: Error analysis of a randomized numerical method. Numer. Math. **70**(1), 119–128 (1995)
36. Stone, D., Geiger, S., Lord, G.J.: Asynchronous discrete event schemes for PDEs. J. Comput. Phys. **342**, 161–176 (2017)
37. Vázquez, J.L.: The porous medium equation. Oxford mathematical monographs. The Clarendon Press, Oxford University Press, Oxford (2007)
38. Veldman, D.W.M., Zuazua, E.: A framework for randomized time-splitting in linear-quadratic optimal control. Numer. Math. **151**(2), 495–549 (2022)