On Wavelet-Galerkin Methods for Semilinear Parabolic Equations with Additive Noise

Mihály Kovács, Stig Larsson and Karsten Urban

Abstract We consider the semilinear stochastic heat equation perturbed by additive noise. After time-discretization by Euler's method the equation is split into a linear stochastic equation and a non-linear random evolution equation. The linear stochastic equation is discretized in space by a non-adaptive wavelet-Galerkin method. This equation is solved first and its solution is substituted into the nonlinear random evolution equation, which is solved by an adaptive wavelet method. We provide mean square estimates for the overall error.

1 Introduction

We consider the following semilinear parabolic problem with additive noise,

$$\begin{aligned} &\mathrm{d}u - \nabla\cdot(\kappa\nabla u)\,\mathrm{d}t = f(u)\,\mathrm{d}t + \mathrm{d}W, && x\in\mathscr{D},\ t\in(0,T),\\ &u=0, && x\in\partial\mathscr{D},\ t\in(0,T),\\ &u(\cdot,0)=u_0, && x\in\mathscr{D}. \end{aligned} \tag{1}$$

Here $T>0$, $\mathscr{D}\subset\mathbb{R}^d$, $d=1,2,3$, is a convex polygonal domain or a domain with smooth boundary $\partial\mathscr{D}$, and $\{W(t)\}_{t\ge0}$ is an $L_2(\mathscr{D})$-valued Q-Wiener process on a filtered probability space $(\Omega,\mathscr{F},\mathbb{P},\{\mathscr{F}_t\}_{t\ge0})$ with respect to the normal filtration

Mihály Kovács
University of Otago, Dept. of Mathematics and Statistics, P.O. Box 56, Dunedin, New Zealand
e-mail: mkovacs@maths.otago.ac.nz

Stig Larsson
Chalmers University of Technology, Mathematical Sciences, SE–412 96 Gothenburg, Sweden
e-mail: stig@chalmers.se

Karsten Urban
Ulm University, Inst. for Numerical Mathematics, Helmholtzstr. 18, DE–89069 Ulm, Germany
e-mail: karsten.urban@uni-ulm.de

$\{\mathscr{F}_t\}_{t\geq 0}$. We use the notation $H = L_2(\mathscr{D})$, $V = H_0^1(\mathscr{D})$ with $\|\cdot\| = \|\cdot\|_H$ and $(\cdot,\cdot) = (\cdot,\cdot)_H$. Moreover, $A\colon V \to V'$ denotes the linear elliptic operator $Au = -\nabla\cdot(\kappa\nabla u)$ for $u \in V$ where $\kappa(x) > \kappa_0 > 0$ is smooth. As usual we consider the bilinear form $a\colon V\times V \to \mathbb{R}$ defined by $a(u,v) = \langle Au, v\rangle$ for $u, v \in V$, and $\langle\cdot,\cdot\rangle$ denotes the duality pairing of V' and V. We denote by e^{-tA} the analytic semigroup in H generated by the realization of $-A$ in H with $D(A) = H^2(\mathscr{D}) \cap H_0^1(\mathscr{D})$. Finally, $f\colon H \to H$ is a nonlinear function, which is assumed to be globally Lipschitz continuous, i.e., there exists a constant L_f such that

$$\|f(u) - f(v)\| \leq L_f\|u - v\|, \quad u, v \in H. \tag{2}$$

It is well known that our assumptions on A and on the spatial domain $\mathscr{D}$ implies the existence of a sequence of nondecreasing positive real numbers $\{\lambda_k\}_{k\geq 1}$ and an orthonormal basis $\{e_k\}_{k\geq 1}$ of H such that

$$Ae_k = \lambda_k e_k, \quad \lim_{k\to+\infty} \lambda_k = +\infty.$$

Using the spectral functional calculus for A we introduce the fractional powers A^s, $s \in \mathbb{R}$, of A as

$$A^s v = \sum_{k=1}^{\infty} \lambda_k^s (v, e_k) e_k, \quad D(A^s) = \Big\{ v \in H : \|A^s v\|^2 = \sum_{k=1}^{\infty} \lambda_k^{2s} (v, e_k)^2 < \infty \Big\}$$

and spaces $\dot{H}^\beta = D(A^{\beta/2})$ with norms $\|v\|_\beta = \|A^{\beta/2} v\|$. It is classical that if $0 \leq \beta < 1/2$, then $\dot{H}^\beta = H^\beta$ and if $1/2 < \beta \leq 2$, then $\dot{H}^\beta = \{u \in H^\beta : u|_{\partial\mathscr{D}} = 0\}$, where H^β denotes the standard Sobolev space of order β. We also use the spaces $L_2(\Omega, \dot{H}^\beta)$ with the mean square norms $\|v\|_{L_2(\Omega,\dot{H}^\beta)} = \big(\mathbb{E}[\|v\|_\beta^2]\big)^{\frac{1}{2}}$.

We assume for some $\beta \geq 0$ that

$$\|A^{\frac{\beta-1}{2}} Q^{\frac{1}{2}}\|_{\mathrm{HS}} < \infty, \quad u_0 \in L_2(\Omega, \dot{H}^\beta). \tag{3}$$

Here Q is the covariance operator of W and $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm. The Hilbert-Schmidt condition in (3) can be viewed as a regularity assumption on the covariance operator Q. In particular, it holds with $\beta = 1$ if Q is a trace class operator and with $\beta < 1/2$ if $Q = I$ and $d = 1$. More generally, it holds if $\sum_{k=1}^{\infty} \lambda_k^{-\alpha} < \infty$ (thus $\alpha > d/2$) and $A^{\beta+\alpha-1}Q$ is a bounded linear operator on H (see, for example, [12, Theorem 2.1]).

It is known ([9], [10, Lemma 3.1]) that if (2) and (3) hold, then (1) has a unique mild solution, which is defined to be the solution of the fixed point equation

$$u(t) = \mathrm{e}^{-tA} u_0 + \int_0^t \mathrm{e}^{-(t-s)A} f(u(s))\,\mathrm{d}s + \int_0^t \mathrm{e}^{-(t-s)A}\,\mathrm{d}W(s). \tag{4}$$

This naturally splits the solution as $u = v + w$, where w is a stochastic convolution,

$$w(t) = \int_0^t \mathrm{e}^{-(t-s)A}\,\mathrm{d}W(s), \tag{5}$$

which is the solution of

$$\mathrm{d}w + Aw\,\mathrm{d}t = \mathrm{d}W,\ 0 < t \le T; \quad w(0) = 0, \tag{6}$$

and v is the solution of the random evolution equation

$$\dot{v} + Av = f(v+w),\ 0 < t \le T; \quad v(0) = u_0. \tag{7}$$

Our approach will be to first compute w and then to insert it into (7) which we then solve for v. Finally, $u = v + w$. For the numerical solution we use Rothe's method, where we first discretize with respect to time and then discretize the resulting elliptic problems with wavelet methods.

Thus, we fix a time step $\tau > 0$, set $t_n := n\tau$ with $t_N = T$, and consider a backward Euler discretization of (1). With $u^n \approx u(t_n)$ and increments $\Delta W^n = W(t_n) - W(t_{n-1})$ this reads

$$u^n + \tau A u^n = u^{n-1} + \tau f(u^n) + \Delta W^n,\ 1 \le n \le N; \quad u^0 = u_0. \tag{8}$$

Then we decompose $u^n = v^n + w^n$ to get time-discrete versions of (6) and (7):

$$w^n + \tau A w^n = w^{n-1} + \Delta W^n, \qquad 1 \le n \le N; \quad w^0 = 0, \tag{9a}$$

$$v^n + \tau A v^n = v^{n-1} + \tau f(v^n + w^n), \qquad 1 \le n \le N; \quad v^0 = u_0. \tag{9b}$$

This allows us to solve the linear problem (9a) first and use the result as an input for the nonlinear problem (9b). Moreover, the stochastic influence in (9b) is smoother than in (9a), which allows us to use fast nonlinear solvers.

We consider now the spatial discretization of (9). To this end, let S_J be a multiresolution space of order m (see (26) for the definition) and let $\{w_J^n\}_{n=0}^N \subset S_J$ be the corresponding Galerkin approximation of $\{w^n\}_{n=0}^N$, i.e.,

$$w_J^n + \tau A_J w_J^n = w_J^{n-1} + P_J \Delta W^n,\ 1 \le n \le N; \quad w_J^0 = 0. \tag{10}$$

We refer to Sect. 3 for further details. We enter this approximation instead of w^n into (9b). The corresponding equation reads

$$\bar{v}^n + \tau A \bar{v}^n = \bar{v}^{n-1} + \tau f(\bar{v}^n + w_J^n),\ 1 \le n \le N; \quad \bar{v}^0 = u_0. \tag{11}$$

For each $\omega \in \Omega$ and for each $n \ge 1$ the nonlinear equation in (11) is solved by an adaptive wavelet algorithm to yield an approximate solution v_ε^n with tolerance ε_n. Note that we use the same tolerance for each ω. More precisely, denoting $\bar{v}^n = E_n(\bar{v}^{n-1})$, where $E_n = (I + \tau A - \tau f(\cdot + w_J^n))^{-1}$ is the nonlinear one-step operator from (11), we assume that $v_\varepsilon^n = \tilde{E}_n(v_\varepsilon^{n-1})$, where $\tilde{E}_n$ is an approximation of E_n such that

$$\|E_n(v) - \tilde{E}_n(v)\| \leq \varepsilon_n, \quad 1 \leq n \leq N,\ v \in H. \tag{12}$$

The output of the computation will then be the sequence

$$u_\varepsilon^n = v_\varepsilon^n + w_J^n, \quad 0 \leq n \leq N. \tag{13}$$

The total error is $u_\varepsilon^n - u(t_n) = (v_\varepsilon^n - \bar{v}^n) + (\bar{v}^n - v^n) + (w_J^n - w^n) + (u^n - u(t_n))$. The contributions are bounded as follows, where the constants C depend on $\|u_0\|_{L_2(\Omega,\dot{H}^\beta)}$, $\|A^{\frac{\beta-1}{2}} Q^{\frac{1}{2}}\|_{\mathrm{HS}}$, and T, referring to assumption (3). We also assume $\tau L_f < \frac{1}{2}$.

First, in Sect. 2.1, an adaptive wavelet algorithm is described which realizes (12). In Theorem 2.4, we also analyze the computational effort of the algorithm applied to (11). We conclude the section by showing that

$$\max_{0 \leq t_n \leq T} \|v_\varepsilon^n - \bar{v}^n\|_{L_2(\Omega,H)} \leq C \sum_{n=1}^N \varepsilon_n. \tag{14}$$

The multiresolution approximation of the time-discrete stochastic convolution is studied in Sect. 3 and Theorem 3.3 shows that

$$\max_{0 \leq t_n \leq T} \|w_J^n - w^n\|_{L_2(\Omega,H)} \leq C 2^{-J \min(\beta,m)}. \tag{15}$$

In Sect. 4, Theorem 4.5, we study the time-discretization error and prove that

$$\max_{0 \leq t_n \leq T} \|u^n - u(t_n)\|_{L_2(\Omega,H)} \leq C \tau^{\frac{\beta}{2}}, \quad \text{if } 0 \leq \beta < 1. \tag{16}$$

Finally, in Sect. 5, we analyze the perturbation of the nonlinear term and obtain that

$$\max_{0 \leq t_n \leq T} \|\bar{v}^n - v^n\|_{L_2(\Omega,H)} \leq C \max_{0 \leq t_n \leq T} \|w_J^n - w^n\|_{L_2(\Omega,H)}. \tag{17}$$

Therefore, our main result is the following.

Theorem 1.1. *Assume* (3) *for some* $\beta \geq 0$. *Let* $\{w_J^n\}_{n=0}^N \subset S_J$ *be computed by a multiresolution Galerkin method of order m and* $\{v_\varepsilon^n\}_{n=0}^N$ *by an adaptive wavelet method with tolerances* ε_n. *Then for* $0 \leq \beta < 1$*, the total error in* (13) *is bounded by*

$$\max_{0 \leq t_n \leq T} \|u_\varepsilon^n - u(t_n)\|_{L_2(\Omega,H)} \leq C \tau^{\frac{\beta}{2}} + C 2^{-J \min(\beta,m)} + C \sum_{n=1}^N \varepsilon_n,$$

for $\tau L_f < \frac{1}{2}$, *where* $C = C(\|u_0\|_{L_2(\Omega,\dot{H}^\beta)}, \|A^{\frac{\beta-1}{2}} Q^{\frac{1}{2}}\|_{\mathrm{HS}}, T)$. *If* $\beta \geq 1$*, then first term is replaced by* $C_\delta \tau^{\frac{1}{2}-\delta}$*, for any* $\delta > 0$.

The literature on numerics for nonlinear stochastic parabolic problems is now rather large. We mention, for example, [15] on pure time-discretization and [13, 18] on complete discretization based on the method of lines, where the spatial discretization is first performed by finite elements and the resulting finite-dimensional evolution problem is then discretized. Wavelets have been used in [11] where the spatial approximation (without adaptivity) of stochastic convolutions were considered.

Our present paper is a first attempt towards spatial adaptivity by using Rothe's method together with known adaptive wavelet methods for solving the resulting nonlinear elliptic problems.

The spatial Besov regularity of solutions of stochastic PDEs is investigated in [2, 3]. The comparison of the Sobolev and Besov regularity is indicative of whether adaptivity is advantageous. For problems on domains with smooth or convex polygonal boundary with boundary adapted additive noise (that is, (3) holds for β high enough), where the solution can be split as $u = v + w$, we expect that the adaptivity is not needed for the stochastic convolution w, which then has sufficient Sobolev regularity. We therefore apply it only to the random evolution problem (7). Once the domain is not convex, or the boundary is not regular, or the noise is not boundary adapted, adaptivity might pay off also for the solution of the linear problem (9a).

The recent paper [1] is a first attempt for a more complete error analysis of Rothe's method for both deterministic and stochastic evolution problems. The overlap with our present work is not not too large, since we take advantage of special features of equations with additive noise.

2 Wavelet approximation

In this section, we collect the notation and the main properties of wavelets that will be needed in the sequel. We refer to [4, 8, 17] for more details on wavelet methods for PDEs. For the space discretization, let

$$\Psi = \{\psi_\lambda : \lambda \in \mathscr{J}^\Psi\}, \quad \tilde{\Psi} = \{\tilde{\psi}_\lambda : \lambda \in \mathscr{J}^\Psi\}$$

be a biorthogonal basis of H, i.e., in particular $(\psi_\lambda, \tilde{\psi}_\mu)_H = \delta_{\delta,\mu}$. Here, λ typically is an index vector $\lambda = (j,k)$ containing both the information on the level $j = |\lambda|$ and the location in space k (e.g., the center of the support of ψ_λ). Note that Ψ also contains the scaling functions on the coarsest level that are not wavelets. We will refer to $|\lambda| = 0$ as the level of the *scaling functions*.

In addition, we assume that $\psi_\lambda \in V$, which is an assumption on the regularity (and boundary conditions) of the primal wavelets. To be precise, we pose the following assumptions on the wavelet bases:

1. Regularity: $\psi_\lambda \in H^t(\mathscr{D})$, $\lambda \in \mathscr{J}^\Psi$ for all $0 \le t < s_\Psi$;
2. Vanishing moments: $((\cdot)^r, \psi_\lambda)_{0;\mathscr{D}} = 0$, $0 \le r < m_\Psi$, $|\lambda| > 0$.
3. Locality: $\mathrm{diam}(\mathrm{supp}\,\psi_\Lambda) \sim 2^{-|\lambda|}$.

We assume the same properties for the dual wavelet basis with s_Ψ and m_Ψ replaced by $\tilde{s}_\Psi$ and $\tilde{m}_\Psi$. Note that the dual wavelet $\tilde{\psi}_\lambda$ does not need to be in V, typically one expects $\tilde{\psi}_\lambda \in V'$.

We will consider (often finite-dimensional) subspaces generated by (adaptively generated finite) sets of indices $\Lambda \subset \mathscr{J}^\Psi$ and

$$\Psi_\Lambda := \{\psi_\lambda : \lambda \in \Lambda\}, \quad S_\Lambda := \operatorname{clos}\operatorname{span}(\Psi_\Lambda),$$

where the closure is of course not needed if Λ is a finite set. If $\Lambda = \Lambda_J := \{\lambda \in \mathscr{J}^\Psi : |\lambda| \le J-1\}$, then $S_J := S_{\Lambda_J}$ contains all wavelets up to level $J-1$ so that S_J coincides with the multiresolution space generated by all scaling functions on level J, i.e.,

$$S_J = \operatorname{span}\Phi_J, \quad \Phi_J = \{\varphi_{J,k} : k \in \mathscr{I}_J\}, \tag{18}$$

where $\mathscr{I}_J$ is an appropriate index set.

2.1 Adaptive wavelet methods for nonlinear variational problems

In this section, we quote from [7] the main facts on adaptive wavelet methods for solving stationary nonlinear variational problems. Note, that all what is said in this section is taken from [7]. However, we abandon further reference for easier reading.

Let $F\colon V \to V'$ be a nonlinear map. We consider the problem of finding $u \in V$ such that

$$\langle v, R(u)\rangle := \langle v, F(u) - g\rangle = 0, \quad v \in V, \tag{19}$$

where $g \in V'$ is given. As an example, let F be given as $\langle v, F(u)\rangle := a(v,u) + \langle v, f(u)\rangle$ which covers (11). The main idea is to consider an equivalent formulation of (19) in terms of the wavelet coefficients $\mathbf{u}$ of the unknown solution $u = \mathbf{u}^T\Psi$. Setting

$$\mathbf{R}(\mathbf{v}) := (\langle \psi_\lambda, R(v)\rangle)_{\lambda \in \mathscr{J}^\Psi}, \quad v = \mathbf{v}^T\Psi,$$

the equivalent formulation amounts to finding $\mathbf{u} \in \ell_2(\mathscr{J}^\Psi)$ such that

$$\mathbf{R}(\mathbf{v}) = \mathbf{0}. \tag{20}$$

The next ingredient is a basic iteration in the (infinite-dimensional) space $\ell_2(\mathscr{J}^\Psi)$ and replacing the infinite operator applications in an adaptive way by finite approximations in order to obtain a computable version. Starting by some finite $\mathbf{u}^{(0)}$, the iteration reads

$$\mathbf{u}^{(n+1)} = \mathbf{u}^{(n)} - \Delta\mathbf{u}^{(n)}, \quad \Delta\mathbf{u}^{(n)} := \mathbf{B}^{(n)}\mathbf{R}(\mathbf{u}^{(n)}) \tag{21}$$

where the operator $\mathbf{B}^{(n)}$ is to be chosen and determines the nonlinear solution method (such as Richardson or Newton). The sequence $\Delta\mathbf{u}^{(n)} = \mathbf{B}^{(n)}\mathbf{R}(\mathbf{u}^{(n)})$ (possibly infinite even for finite input $\mathbf{u}^{(n)}$) is then replaced by some finite sequence $\mathbf{w}_\eta^{(n)} := \mathbf{RES}[\eta_n, \mathbf{B}^{(n)}, \mathbf{R}, \mathbf{u}^{(n)}]$ such that

$$\|\Delta\mathbf{u}^{(n)} - \mathbf{w}_\eta^{(n)}\| \le \eta_n.$$

Replacing $\Delta \mathbf{u}^{(n)}$ by $\mathbf{w}_\eta^{(n)}$ in (21) and choosing the sequence of tolerances $(\eta_n)_{n\in\mathbb{N}_0}$ appropriately results in a convergent algorithm such that any tolerance ε is reached after finitely many steps. We set $\bar{\mathbf{u}}(\varepsilon) := \mathbf{SOLVE}[\varepsilon, \mathbf{R}, \mathbf{B}^{(n)}, \mathbf{u}^{(0)}]$ such that we get $\|\mathbf{u} - \bar{\mathbf{u}}(\varepsilon)\| \leq \varepsilon$.

In terms of optimality, there are several issues to be considered:

- How many iterations $n(\varepsilon)$ are required in order to achieve ε-accuracy?
- How many "active" coefficients are needed to represent the numerical approximation and how does that compare with a "best" approximation?
- How many operations (arithmetic, storage) and how much storage is required?

In order to quantify that, one considers so-called *approximation classes*

$$\mathscr{A}^s := \{\mathbf{v} \in \ell_2(\mathscr{J}^\Psi) : \sigma_N(\mathbf{v}) \lesssim N^{-s}\}$$

of all those sequences whose *error of best N-term approximation*

$$\sigma_N(\mathbf{v}) := \min\{\|\mathbf{v} - \mathbf{w}\|_{\ell_2} : \#\operatorname{supp}\mathbf{w} \leq N\}$$

decays at a certain rate ($\operatorname{supp}\mathbf{v} := \{\lambda \in \mathscr{J}^\Psi : v_\lambda \neq 0\}$, $\mathbf{v} = (v_\lambda)_{\lambda\in\mathscr{J}^\Psi}$).

Let us first consider the case where $F = A$ is a linear elliptic partial differential operator, i.e., $Au = g \in V'$, where $A\colon V \to V'$, $g \in V'$ is given and $u \in V$ is to be determined. For the discretization we use a wavelet basis Ψ in H where rescaled versions admit Riesz bases in V and V', respectively. Then, the operator equation can equivalently be written as

$$\mathbf{A}\mathbf{u} = \mathbf{g} \in \ell_2(\mathscr{J}^\Psi),$$

where $\mathbf{A} := \mathbf{D}^{-1}a(\Psi,\Psi)\mathbf{D}^{-1}$, $\mathbf{g} := \mathbf{D}^{-1}(g,\Psi)$ and $\mathbf{u} := \mathbf{D}(u_\lambda)_{\lambda\in\mathscr{J}^\Psi}$, with u_λ being the wavelet coefficients of the unknown function $u \in V$, $\|u\|_V \sim \|\mathbf{u}\|_{\ell_2(\mathscr{J}^\Psi)}$. Wavelet preconditioning results in the fact that $\kappa_2(\mathbf{A}) < \infty$, [5].

The (biinfinite) matrix $\mathbf{A}$ is said to be *s^*-compressible*, $\mathbf{A} \in \mathscr{C}_{s^*}$, if for any $0 < s < s^*$ and every $j \in \mathbb{N}$ there exists a matrix $\mathbf{A}_j$ with the following properties: For some summable sequence $(\alpha_j)_{j\in\mathbb{N}}$, the matrix $\mathbf{A}_j$ is obtained by replacing all but the order of $\alpha_j 2^j$ entries per row and column in $\mathbf{A}$ by zero and satisfies

$$\|\mathbf{A} - \mathbf{A}_j\| \leq C\alpha_j 2^{-js}, \quad j \in \mathbb{N}.$$

Wavelet representations of differential (and certain integral) operators fall into this category. Typically, s^* depends on the regularity and the order of vanishing moments of the wavelets. Then, one can construct a linear counterpart $\mathbf{RES}_{\text{lin}}$ of $\mathbf{RES}$ such that $\mathbf{w}_\eta := \mathbf{RES}_{\text{lin}}[\eta, \mathbf{A}, \mathbf{g}, \mathbf{v}]$ for finite input $\mathbf{v}$ satisfies

$$\|\mathbf{w}_\eta - (\mathbf{A}\mathbf{v} - \mathbf{g})\|_{\ell_2} \leq \eta, \tag{22a}$$

$$\|\mathbf{w}_\eta\|_{\mathscr{A}^s} \lesssim (\|\mathbf{v}\|_{\mathscr{A}^s} + \|\mathbf{u}\|_{\mathscr{A}^s}), \tag{22b}$$

$$\#\operatorname{supp}\mathbf{w}_\eta \lesssim \eta^{-1/s}(\|\mathbf{v}\|_{\mathscr{A}^s}^{1/s} + \|\mathbf{u}\|_{\mathscr{A}^s}^{1/s}), \tag{22c}$$

where the constants in (22b), (22c) depend only on s. Here, we have used the quasi-norm

$$\|\mathbf{v}\|_{\mathscr{A}^s} := \sup_{N\in\mathbb{N}} N^s \sigma_N(\mathbf{v}).$$

This is the main ingredient for proving optimality of the scheme in the following sense.

Theorem 2.1 ([5, 7]). *If* $\mathbf{A} \in \mathscr{C}_{s^*}$ *and if the exact solution* $\mathbf{u}$ *of* $\mathbf{Au} = \mathbf{g}$ *satifies* $\mathbf{u} \in \mathscr{A}^s$, $s < s^*$, *then* $\bar{\mathbf{u}}(\varepsilon) = \mathbf{SOLVE}_{\mathrm{lin}}[\varepsilon]$ *satisfies*

$$\|\mathbf{u} - \bar{\mathbf{u}}(\varepsilon)\| \le \varepsilon, \tag{23a}$$

$$\#\operatorname{supp} \bar{\mathbf{u}}(\varepsilon) \lesssim \varepsilon^{-1/s}, \tag{23b}$$

$$\textit{computational complexity} \sim \#\operatorname{supp} \bar{\mathbf{u}}(\varepsilon). \tag{23c}$$

It turns out that most of what is said before also holds for the nonlinear case except that the analysis of the approximate evaluation of nonlinear expressions $\mathbf{R}(\mathbf{v})$ poses a constraint on the structure of the active coefficients, namely that it has *tree structure*. In order to define this, one uses the notation $\mu \prec \lambda$, $\lambda, \mu \in \mathscr{J}^\Psi$ to express that μ is a *descendant* of λ. We explain this in the univariate case with $\psi_\lambda = \psi_{j,k} = 2^{j/2}\psi(2^j\cdot - k)$. Then, the *children* of $\lambda = (j,k)$ are, as one would also intuitively define, $\mu = (j+1, 2k)$ and $\nu = (j+1, 2k+1)$. The descendants of λ are its children, the children of its children and so on. In higher dimensions and even on more complex domains this can also be defined – with some more technical effort, however.

Then, a set $\mathscr{T} \subset \mathscr{J}^\Psi$ is called a *tree* if $\lambda \in \mathscr{T}$ implies $\mu \in \mathscr{T}$ for all $\mu \in \mathscr{J}^\Psi$ with $\lambda \prec \mu$. Given this, the error of the *best N-term tree approximation* is given as

$$\sigma_N^{\mathrm{tree}}(\mathbf{v}) := \min\{\|\mathbf{v} - \mathbf{w}\|_{\ell_2} : \mathscr{T} := \#\operatorname{supp}\mathbf{w} \text{ is a tree and } \#\mathscr{T} \le N\}$$

and define the *tree approximation space* as

$$\mathscr{A}^s_{\mathrm{tree}} := \{\mathbf{v} \in \ell_2(\mathscr{J}^\Psi) : \sigma_N^{\mathrm{tree}}(\mathbf{v}) \lesssim N^{-s}\}$$

which is a quasi-normed space under the quasi-norm

$$\|\mathbf{v}\|_{\mathscr{A}^s_{\mathrm{tree}}} := \sup_{N\in\mathbb{N}} N^s \sigma_N^{\mathrm{tree}}(\mathbf{v}).$$

Remark 2.2. For the case $V = H^t$ (or, a closed subspace of H^t) it is known that the solution being in some Besov space $u \in B_q^{t+ds}(L_q)$, $q = (s + \frac{1}{2})^{-1}$, implies that $\mathbf{u} \in \mathscr{A}^r_{\mathrm{tree}}$, for $r < s$, see [6, Remark 2.3].

The extension of the s^*-compressibility $\mathscr{C}_{s^*}$ is the s^*-*sparsity* of the scheme **RES** which is defined by the following property: *If the exact solution* $\mathbf{u}$ *of* (20) *is in* $\mathscr{A}^s_{\mathrm{tree}}$ *for some* $s < s^*$, *then* $\mathbf{w}_\eta := \mathbf{RES}[\eta, \mathbf{B}, \mathbf{R}, \mathbf{v}]$ *for finite* $\mathbf{v}$ *satisfies*

$$\|\mathbf{w}_\eta\|_{\mathscr{A}^s_{\text{tree}}} \le C(\|\mathbf{v}\|_{\mathscr{A}^s_{\text{tree}}} + \|\mathbf{u}\|_{\mathscr{A}^s_{\text{tree}}} + 1),$$
$$\#\operatorname{supp}\mathbf{w}_\eta \le C\eta^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathscr{A}^s_{\text{tree}}} + \|\mathbf{u}\|^{1/s}_{\mathscr{A}^s_{\text{tree}}} + 1),$$
$$\textit{comp. complexity} \sim C(\eta^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathscr{A}^s_{\text{tree}}} + \|\mathbf{u}\|^{1/s}_{\mathscr{A}^s_{\text{tree}}} + 1) + \#\mathscr{T}(\operatorname{supp}\mathbf{v})),$$

where C *depends only on* s *when* $s \to s^*$ *and* $\mathscr{T}(\operatorname{supp}\mathbf{v})$ *denotes the smallest tree containing* $\operatorname{supp}\mathbf{v}$. Now, we are ready to collect the main result.

Theorem 2.3 ([7, Theorem 6.1]). *If* **RES** *is* s^**-sparse,* $s^* > 0$ *and if* $\mathbf{u} \in \mathscr{A}^s_{\text{tree}}$ *for some* $s < s^*$*, then the approximations* $\bar{\mathbf{u}}(\varepsilon)$ *satisfy* $\|\mathbf{u} - \mathbf{u}(\varepsilon)\| \le \varepsilon$ *with*

$$\#\operatorname{supp}\bar{\mathbf{u}}(\varepsilon) \le C\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathscr{A}^s_{\text{tree}}}, \quad \|\bar{\mathbf{u}}(\varepsilon)\|_{\mathscr{A}^s_{\text{tree}}} \le C\|\mathbf{u}\|_{\mathscr{A}^s_{\text{tree}}},$$

where C *depends only on* s *when* $s \to s^*$*. The number of operations is bounded by* $C\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathscr{A}^s_{\text{tree}}}$.

We remark that since the wavelet transform is of linear complexity the overall number of operations needed is the one mentioned in Theorem 2.3.

Next we show that the wavelet coefficients $\bar{\mathbf{v}}^n$ of the solution of (11) belong to a certain approximation class $\mathscr{A}^s_{\text{tree}}$ and hence, in view of Theorem 2.3, we obtain an estimate on the support of $\bar{\mathbf{v}}^n_\varepsilon$ and the number of operations required to compute it.

Theorem 2.4. *The wavelet coefficients* $\bar{\mathbf{v}}^n$ *of the solution of* (11) *belong to* $\mathscr{A}^s_{\text{tree}}$ *for all* $s < \frac{1}{2d-2}$*, where* $d \ge 2$ *is the spatial dimension of* $\mathscr{D}$*.*

Proof. It follows from [1, Lemma 5.15] that $r(\tau A) \in \mathscr{L}(L_2(\mathscr{D}), B^r_q(L_q))$ for $r = \frac{3d-2+4\varepsilon}{2d-2+4\varepsilon}$, where $1/q = (r-1)/d + 1/2$ and $\varepsilon > 0$. Thus, the statement follows from Remark 2.2 noting that $t = 1$ and hence $r = 1 + ds$.

We end this section by showing (14); that is, the overall error after n steps, when in every step (11) is solved approximately up to an error tolerance ε_n using the adaptive wavelet algorithm described above. Define

$$E^n_j = E_n \circ \cdots \circ E_{j+1},\ E^n_n = I; \quad 0 \le j < n \le N,$$

and similarly $\tilde{E}^n_j$. Then we have

$$\begin{aligned} v^n_\varepsilon - \bar{v}^n &= \tilde{E}^n_0(u_0) - E^n_0(u_0) \\ &= \sum_{j=0}^{n-1} \big(E^n_{j+1}(\tilde{E}^{j+1}_0(u_0)) - E^n_j(\tilde{E}^j_0(u_0))\big) \\ &= \sum_{j=0}^{n-1} \big(E^n_{j+1}(\tilde{E}^{j+1}_j(\tilde{E}^j_0(u_0))) - E^n_{j+1}(E^{j+1}_j(\tilde{E}^j_0(u_0)))\big) \\ &= \sum_{j=0}^{n-1} \big(E^n_{j+1}(\tilde{E}_{j+1}(v^j_\varepsilon)) - E^n_{j+1}(E_{j+1}(v^j_\varepsilon))\big). \end{aligned}$$

A simple argument shows that the Lipschitz constant of E_n is bounded by $(1-\tau L_f)^{-1} \le \mathrm{e}^{c\tau L_f}$ for some $c>0$, if $\tau L_f \le \frac{1}{2}$, cf. the proof of Lemma 5.1. Hence E^n_{j+1} has a Lipschitz constant bounded by $\mathrm{e}^{c(t_n-t_{j+1})} \le \mathrm{e}^{ct_N}$. Thus, using (12), we obtain

$$\|v^n_\varepsilon - \bar{v}^n\| \le \sum_{j=0}^{n-1} \mathrm{e}^{c(t_n-t_{j+1})} \|\tilde{E}_{j+1}(v^j_\varepsilon) - E_{j+1}(v^j_\varepsilon)\| \le \sum_{j=1}^{n} \mathrm{e}^{c(t_n-t_j)} \varepsilon_j \le \mathrm{e}^{ct_N} \sum_{j=1}^{n} \varepsilon_j.$$

After taking a mean square we obtain (14).

3 Error analysis for the stochastic convolution

Let $S_J = S_{\Lambda_J}$ be a multiresolution space (18). The multiresolution Galerkin approximation of the equation $Au = f$ in V' is to find $u_J \in S_J$ such that

$$a(u_J, v_J) = (f, v_J) \quad \forall v \in S_J. \tag{24}$$

Define the orthogonal projector $P_J \colon H \to S_J$ by

$$(P_J v, w_J) = (v, w_J), \quad v \in H,\ w_J \in S_J. \tag{25}$$

Note that P_J can be extended to V' by (25) since $S_J \subset V$. Next, we define the operator $A_J \colon S_J \to S_J$ by

$$a(A_J v_J, w_J) = a(v_J, w_J), \quad u_J, v_J \in S_J.$$

Then (24) reads $A_J u_J = P_J f$ in S_J. Alternatively we may write $u_J = R_J u$, where $R_J \colon V \to S_J$ is the *Ritz projector*, defined by

$$a(R_J v, w_J) = a(v, w_J), \quad v \in V,\ w_J \in S_J.$$

The multiresolution space is of order m if

$$\inf_{w_J \in S_J} \|v - w_J\| \lesssim 2^{-mJ} \|v\|_{m;\mathscr{D}}, \quad v \in H^m(\mathscr{D}) \cap V. \tag{26}$$

Standard arguments then show, using elliptic regularity thanks to our assumptions on $\mathscr{D}$, that $\|u_J - u\| \lesssim 2^{-mJ} \|u\|_{m;\mathscr{D}}$, or in other words

$$\|v - R_J v\| \lesssim 2^{-mJ} \|v\|_{m;\mathscr{D}}, \quad v \in H^m(\mathscr{D}) \cap V. \tag{27}$$

The next lemma is of independent interest and we state it in a general form.

Lemma 3.1. *Let $-A$ and $-B$ generate strongly continuous semigroups e^{-tA} and e^{-tB} on a Banach space X and let $r(s) = (1+s)^{-1}$. Then, for all $x, y \in X$, $N \in \mathbb{N}$, $\tau > 0$,*

$$\tau \sum_{n=1}^{N} \|r^n(\tau B)y - r^n(\tau A)x\|^p \le \int_0^\infty \|\mathrm{e}^{-tB}y - \mathrm{e}^{-tA}x\|^p \,\mathrm{d}t, \quad 1 \le p < \infty, \tag{28}$$

$$\|r^n(\tau B)y - r^n(\tau A)x\| \le \sup_{t\ge 0} \|\mathrm{e}^{-tB}y - \mathrm{e}^{-tA}x\|. \tag{29}$$

Proof. By the Hille-Phillips functional calculus, we have

$$r^n(\tau B)y - r^n(\tau A)x = \int_0^\infty (\mathrm{e}^{-t\tau B}y - \mathrm{e}^{-t\tau A}x) f_n(t)\,\mathrm{d}t, \tag{30}$$

where f_n denotes the nth convolution power of $f(t) = \mathrm{e}^{-t}$. Since $\|f_n\|_{L_1(\mathbb{R}_+)} = 1$ inequality (29) follows immediately by Hölder's inequality. To see (28) we note that f_n is a probability density and hence by Jensen's inequality and (30),

$$\begin{aligned} \tau \sum_{n=1}^{N} \|r^n(\tau B)y - r^n(\tau A)x\|^p &= \tau \sum_{n=1}^{N} \Big\| \int_0^\infty (\mathrm{e}^{-t\tau B}y - \mathrm{e}^{-t\tau A}x) f_n(t)\,\mathrm{d}t \Big\|^p \\ &\le \tau \sum_{n=1}^{N} \int_0^\infty \|\mathrm{e}^{-t\tau B}y - \mathrm{e}^{-t\tau A}x\|^p f_n(t)\,\mathrm{d}t \\ &= \int_0^\infty \|\mathrm{e}^{-tB}y - \mathrm{e}^{-tA}x\|^p \,\mathrm{d}t \sup_{t>0} \sum_{n=1}^{\infty} f_n(t). \end{aligned}$$

Finally, by monotone convergence, the Laplace transform of $\sum_{n=1}^\infty f_n$ is given by

$$\Big(\sum_{n=1}^{\infty} f_n\Big)^{\hat{}}(\lambda) = \sum_{n=1}^{\infty} \hat{f}_n(\lambda) = \sum_{n=1}^{\infty} \Big(\frac{1}{1+\lambda}\Big)^n = \frac{1}{\lambda}, \quad \lambda > 0.$$

Thus, $\sum_{n=1}^\infty f_n \equiv 1$ and the proof is complete.

Next we derive an error estimate for the multiresolution approximation of the semigroup e^{-tA} and its Euler approximation $r^n(\tau A)$.

Lemma 3.2. *Let S_J be a multiresolution space of order m and let A, A_J, and P_J be as above. Then, for $T \ge 0$, $N \ge 1$, τ, we have*

$$\Big(\int_0^T \|\mathrm{e}^{-tA_J}P_J v - \mathrm{e}^{-tA}v\|^2 \,\mathrm{d}t\Big)^{\frac{1}{2}} \le C2^{-J\beta}\|v\|_{\beta-1}, \quad 0 \le \beta \le m, \tag{31}$$

and

$$\Big(\tau \sum_{n=1}^{N} \|r^n(\tau A_J)P_J v - r^n(\tau A)v\|^2\Big)^{\frac{1}{2}} \le C2^{-J\beta}\|v\|_{\beta-1}, \quad 0 \le \beta \le m. \tag{32}$$

Proof. Estimate (31) is known in the finite element context, see for example [16, Theorem 2.5], and may be proved in a completely analogous fashion for using the approximation property (27) of the Ritz projection R_J, the parabolic smoothing (35), and interpolation. Finally, (32) follows from (31) by using Lemma 3.1 with $x = v$, $y = P_J v$, and $B = A_J$. (Note that C is independent of T.)

Now we are ready to consider the multiresolution approximation of w^n in (9a).

Theorem 3.3. *Let S_J be a multiresolution space of order m and w and w_J^n the solutions of* (9a) *and* (10). *If* $\|A^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\|_{\mathrm{HS}} < \infty$ *for some* $0 \le \beta \le m$, *then*

$$\big(\mathbb{E}[\|w_J^n - w^n\|^2]\big)^{\frac{1}{2}} \le C2^{-J\beta}\|A^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\|_{\mathrm{HS}}.$$

Proof. Let $t_k = k\tau$, $k = 0,\dots,n$. By (10), (9a), and induction,

$$w_J^n - w^n = \sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\big[r^{n-k+1}(\tau A_J)P_J - r^{n-k+1}(\tau A)\big]\,\mathrm{d}W(s),$$

whence, by Itô's isometry, we get

$$\begin{aligned}\mathbb{E}[\|w_J^n - w^n\|^2] &= \sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\big\|\big[r^{n-k+1}(\tau A_J)P_J - r^{n-k+1}(\tau A)\big]Q^{\frac{1}{2}}\big\|_{\mathrm{HS}}^2\,\mathrm{d}s\\ &= \sum_{k=1}^{n}\tau\big\|\big[r^{k}(\tau A_J)P_J - r^{k}(\tau A)\big]Q^{\frac{1}{2}}\big\|_{\mathrm{HS}}^2.\end{aligned}$$

Let $\{e_l\}_{l=1}^{\infty}$ be an orthonormal basis of H. Then, using Lemma 3.2, we obtain

$$\begin{aligned}\mathbb{E}[\|w_J^n - w^n\|^2] &= \sum_{l=1}^{\infty}\sum_{k=1}^{n}\tau\|[r^k(\tau A_J)P_J - r^k(\tau A)]Q^{\frac{1}{2}}e_l\|^2\\ &\le C\sum_{l=1}^{\infty}2^{-2J\beta}\|Q^{\frac{1}{2}}e_l\|_{\beta-1}^2 = C2^{-2J\beta}\|A^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2.\end{aligned}$$

4 Pure time discretization

In the proofs below we will often make use of the following well-known facts about the analytic semigroup e^{-tA}, namely

$$\|A^{\alpha}\mathrm{e}^{-tA}\| \le Ct^{-\alpha}, \quad \alpha \ge 0,\ t > 0, \tag{33}$$

$$\|(\mathrm{e}^{-tA} - I)A^{-\alpha}\| \le Ct^{\alpha}, \quad 0 \le \alpha \le 1,\ t \ge 0, \tag{34}$$

for some $C = C(\alpha)$, see, for example, [14, Chapter II, Theorem 6.4]. Also, by a simple energy argument we may prove

$$\int_0^t \|A^{\frac{1}{2}}\mathrm{e}^{-sA}v\|^2\,\mathrm{d}s \le \frac{1}{2}\|v\|^2, \quad v \in H,\ t \ge 0. \tag{35}$$

We quote the following existence, uniqueness and stability result from [10, Lemma 3.1]. For the mild, and other solution concepts we refer to [9, Chapters 6 and 7].

Lemma 4.1. *If $\|A^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\|_{\mathrm{HS}} < \infty$ for some $\beta \geq 0$, $u_0 \in L_2(\Omega,H)$, and (2) holds, then there is a unique mild solution $\{u(t)\}_{t\geq 0}$ of (1) with $\sup_{t\in[0,T]}\mathbb{E}\|u(t)\|^2 \leq K$, where $K = K(u_0,T,L_f)$.*

Concerning the temporal regularity of the stochastic convolution we have the following theorem.

Theorem 4.2. *Let $\|A^{-\eta}Q^{\frac{1}{2}}\|_{\mathrm{HS}} < \infty$ for some $\eta \in [0,\frac{1}{2}]$. Then the stochastic convolution $w(t) := \int_0^t \mathrm{e}^{-(t-\sigma)A}\,\mathrm{d}W(\sigma)$ is mean square Hölder continuous on $[0,\infty)$ with Hölder constant $C = C(\eta)$ and Hölder exponent $\frac{1}{2}-\eta$, i.e.,*

$$\big(\mathbb{E}\|w(t)-w(s)\|^2\big)^{\frac{1}{2}} \leq C|t-s|^{\frac{1}{2}-\eta}, \quad t,s \geq 0.$$

Proof. For $\eta = \frac{1}{2}$ the result follows from Lemma 4.1. Let $\eta \in [0,\frac{1}{2})$ and, without loss of generality, let $s < t$. By independence of the increments of W,

$$\begin{aligned}\mathbb{E}\|w(t)-w(s)\|^2 = \mathbb{E}\Big\|\int_s^t \mathrm{e}^{-(t-\sigma)A}\,\mathrm{d}W(\sigma)\Big\|^2 \\ + \mathbb{E}\Big\|\int_0^s \mathrm{e}^{-(t-\sigma)A} - \mathrm{e}^{-(s-\sigma)A}\,\mathrm{d}W(\sigma)\Big\|^2 = I_1 + I_2.\end{aligned}$$

From Itô's isometry and (33) it follows that

$$\begin{aligned}I_1 &= \mathbb{E}\Big\|\int_s^t A^{\eta}\mathrm{e}^{-(t-\sigma)A}A^{-\eta}\,\mathrm{d}W(\sigma)\Big\|^2 = \int_s^t \|A^{\eta}\mathrm{e}^{-(t-\sigma)A}A^{-\eta}Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2\,\mathrm{d}\sigma \\ &\leq C\int_s^t (t-\sigma)^{-2\eta}\|A^{-\eta}Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2\,\mathrm{d}\sigma \leq \frac{C}{1-2\eta}(t-s)^{1-2\eta}\|A^{-\eta}Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2.\end{aligned}$$

Finally, let $\{e_k\}_{k=1}^{\infty}$ be an orthonormal basis of H. Then, by (34) and (35),

$$\begin{aligned}I_2 &= \int_0^s \|(\mathrm{e}^{-(t-\sigma)A} - \mathrm{e}^{-(s-\sigma)A})Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2\,\mathrm{d}\sigma \\ &= \sum_{k=1}^{\infty}\int_0^s \|(\mathrm{e}^{-(t-s)A} - I)A^{-(\frac{1}{2}-\eta)}A^{\frac{1}{2}-\eta}\mathrm{e}^{-(s-\sigma)A}Q^{\frac{1}{2}}e_k\|^2\,\mathrm{d}\sigma \\ &\leq C(t-s)^{1-2\eta}\sum_{k=1}^{\infty}\int_0^s \|A^{\frac{1}{2}}\mathrm{e}^{-(s-\sigma)A}A^{-\eta}Q^{\frac{1}{2}}e_k\|^2\,\mathrm{d}\sigma \\ &\leq C(t-s)^{1-2\eta}\sum_{k=1}^{\infty}\|A^{-\eta}Q^{\frac{1}{2}}e_k\| = C(t-s)^{1-2\eta}\|A^{-\eta}Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2.\end{aligned} \tag{36}$$

The next result shows that the time regularity of w transfers to the solution of the semilinear problem.

Theorem 4.3. *If $u_0 \in L_2(\Omega,\dot{H}^{\beta})$ and $\|A^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\|_{\mathrm{HS}} < \infty$ for some $0 \leq \beta < 1$, then there is $C = C(T,u_0,\beta)$ such that the mild solution u of (1) satisfies*

$$\big(\mathbb{E}\|u(t)-u(s)\|^2\big)^{\frac{1}{2}} \leq C|t-s|^{\frac{\beta}{2}}, \quad t,s \in [0,T].$$

Proof. Let $T > 0$ and $0 \le s < t \le T$. Then, by (4),

$$u(t) - u(s) = (\mathrm{e}^{-tA} - \mathrm{e}^{-sA})u_0 + \int_s^t \mathrm{e}^{-(t-r)A} f(u(r))\,\mathrm{d}r \\ + \int_0^s (\mathrm{e}^{-(t-r)A} - \mathrm{e}^{-(s-r)A}) f(u(r))\,\mathrm{d}r + w(t) - w(s).$$

In a standard way, for $0 \le \beta \le 2$, we have $\mathbb{E}\|(\mathrm{e}^{-tA} - \mathrm{e}^{-sA})u_0\|^2 \le C|t-s|^\beta \mathbb{E}\|u_0\|_\beta^2$. Using that f is Lipschitz and hence $\|f(u)\| \le C(1 + \|u\|)$, the norm boundedness of the semigroup e^{-tA}, and Lemma 4.1, we have that

$$\mathbb{E}\Big\| \int_s^t \mathrm{e}^{-(t-r)A} f(u(r))\,\mathrm{d}r \Big\|^2 \le C|t-s|^2 \Big(1 + \sup_{r\in[0,T]} \mathbb{E}\|u(r)\|^2\Big) \le C|t-s|^2.$$

For $0 \le \beta < 1$, by Lemma 4.1, (33) and (34), it follows that

$$\begin{aligned} &\mathbb{E}\Big\| \int_0^s (\mathrm{e}^{-(t-r)A} - \mathrm{e}^{-(s-r)A}) f(u(r))\,\mathrm{d}r \Big\|^2 \\ &\quad \le s\mathbb{E} \int_0^s \|(\mathrm{e}^{-(t-r)A} - \mathrm{e}^{-(s-r)A}) f(u(r))\|^2\,\mathrm{d}r \\ &\quad \le Cs\Big(1 + \sup_{r\in[0,T]} \mathbb{E}\|u(r)\|^2\Big) \int_0^s \|\mathrm{e}^{-(t-r)A} - \mathrm{e}^{-(s-r)A}\|^2\,\mathrm{d}r \\ &\quad \le Cs \int_0^s \|A^{\frac{\beta}{2}} \mathrm{e}^{-(s-r)A} (\mathrm{e}^{-(t-s)A} - I) A^{-\frac{\beta}{2}}\|^2\,\mathrm{d}r \le C|t-s|^\beta s^{2-\beta} \le C|t-s|^\beta. \end{aligned}$$

Finally, by Theorem 4.2 with $\eta = -\frac{\beta-1}{2}$, we have $\mathbb{E}\|w(t) - w(s)\|^2 \le C|t-s|^\beta$, which finishes the proof.

In order to analyze the order of the backward Euler time-stepping (8) we quote the following deterministic error estimates, where $r(\tau A) = (I + \tau A)^{-1}$.

Lemma 4.4. *The following error estimates hold for $t_n = n\tau > 0$.*

$$\|[\mathrm{e}^{-n\tau A} - r^n(\tau A)]v\| \le C\tau^{\frac{\beta}{2}} \|v\|_\beta, \quad 0 \le \beta \le 2, \tag{37}$$

$$\|[\mathrm{e}^{-n\tau A} - r^n(\tau A)]v\| \le C\tau t_n^{-1} \|v\|, \tag{38}$$

$$\sum_{k=1}^n \tau \Big\| [r^k(\tau A) - \mathrm{e}^{-k\tau A}]v \Big\|^2 \le C\tau^\beta \|v\|_{\beta-1}^2, \quad 0 \le \beta \le 2. \tag{39}$$

Proof. Estimates (37) and (38) are shown in, for example, [16, Chapter 7]. Estimate (39) can be proved in a similar way as (2.17) in [18, Lemma 2.8].

Theorem 4.5. *If $u_0 \in L_2(\Omega, \dot H^\beta)$ and $\|A^{\frac{\beta-1}{2}} Q^{\frac{1}{2}}\|_{\mathrm{HS}} < \infty$ for some $0 \le \beta < 1$, then there is $C = C(T, u_0, \beta)$ such that for $0 < \tau < \frac{1}{2L_f}$, the solutions u of* (4) *and u^n of* (8) *satisfy*

$$(\mathbb{E}\|u(t_n) - u^n\|^2)^{\frac{1}{2}} \le C\tau^{\beta/2}, \quad t_n = n\tau \in [0, T].$$

Proof. We have, with $e^n := u(t_n) - u^n$,

$$e^n = [\mathrm{e}^{-t_nA} - r^n(\tau A)]u_0 + \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} [\mathrm{e}^{-(t_n-s)A} - r^{n-k+1}(\tau A)]\,\mathrm{d}W(s)$$
$$+ \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \mathrm{e}^{-(t_n-s)A} f(u(s)) - r^{n-k+1}(\tau A) f(u_k)\,\mathrm{d}s = e_1 + e_2 + e_3.$$

The error e_1 is easily bounded, using (37), as

$$\mathbb{E}\|e_1\|^2 \le C\tau^\beta \mathbb{E}\|u_0\|_\beta^2, \quad 0 \le \beta \le 2.$$

The contribution of e_2 is the linear stochastic error. First, we decompose e_2 as

$$e_2 = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} [\mathrm{e}^{-t_{n-k+1}A} - r^{n-k+1}(\tau A)]\,\mathrm{d}W(s)$$
$$+ \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} [\mathrm{e}^{-(t_n-s)A} - \mathrm{e}^{-t_{n-k+1}A}]\,\mathrm{d}W(s) = e_{21} + e_{22}.$$

Let $\{f_l\}_{l=1}^\infty$ be an ONB of H. By Itô's isometry, the independence of the increments of W and (39),

$$\mathbb{E}\|e_{21}\|^2 = \sum_{k=1}^{n} \tau \|[r^k(\tau A) - \mathrm{e}^{-k\tau A}]Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2 \le = \sum_{l=1}^{\infty}\sum_{k=1}^{n} \tau \|[r^k(\tau A) - \mathrm{e}^{-k\tau A}]Q^{\frac{1}{2}} f_l\|^2$$
$$\le C \sum_{l=1}^{\infty} \tau^\beta \|Q^{\frac{1}{2}} f_l\|_{\beta-1}^2 = C\tau^\beta \|A^{\frac{\beta-1}{2}} Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2, \quad 0 \le \beta \le 2.$$

The term e_{22} can be bounded using a similar argument as in (36) by

$$\mathbb{E}\|e_{22}\|^2 \le C\tau^\beta \|A^{\frac{\beta-1}{2}} Q^{\frac{1}{2}}\|_{\mathrm{HS}}^2, \quad 0 \le \beta \le 2.$$

Next, we can further decompose e_3 as

$$e_3 = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} r^{n-k+1}(\tau A)[f(u(t_k)) - f(u_k)]\,\mathrm{d}s$$
$$+ \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} [\mathrm{e}^{-t_{n-k+1}A} - r^{n-k+1}(\tau A)] f(u(t_k))\,\mathrm{d}s$$
$$+ \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \mathrm{e}^{-t_{n-k+1}A}[f(u(s)) - f(u(t_k))]\,\mathrm{d}s$$
$$+ \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} [\mathrm{e}^{-(t_n-s)A} - \mathrm{e}^{-t_{n-k+1}A}] f(u(s))\,\mathrm{d}s = e_{31} + e_{32} + e_{33} + e_{34}.$$

By the stability of $r^n(\tau A)$ and the Lipschitz condition on f, we have

$$\mathbb{E}\|e_{31}\|^2 \le 2L_f^2\tau^2\mathbb{E}\|e^n\|^2 + 2L_f^2\tau^2 n\sum_{k=1}^{n-1}\mathbb{E}\|e^k\|^2 \le 2L_f^2\tau^2\mathbb{E}\|e^n\|^2 + C\tau\sum_{k=1}^{n-1}\mathbb{E}\|e^k\|^2.$$

By (38) and Lemma 4.1, with $\tau t_{n-k+1}^{-1} = (n-k+1)^{-1} = l^{-1}$,

$$\begin{aligned}\mathbb{E}\|e_{32}\|^2 &\le C\mathbb{E}\Big(\sum_{k=1}^{n}\tau\tau t_{n-k+1}^{-1}\|f(u(t_k))\|\Big)^2 \le C\tau^2\sum_{l=1}^{n}\frac{1}{l^2}\sum_{k=1}^{n}\mathbb{E}\|f(u(t_k))\|^2\\ &\le C\tau^2\sum_{k=1}^{n}(1+\mathbb{E}\|u(t_k)\|^2) \le C\tau t_n \le C\tau.\end{aligned}$$

Furthermore, by Theorem 4.3,

$$\mathbb{E}\|e_{33}\|^2 \le t_n\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\mathbb{E}\|f(u(s)) - f(u(t_k))\|^2\,\mathrm{d}s \le Ct_n^2\tau^\beta \le C\tau^\beta,\ 0\le\beta<1.$$

To estimate e_{34} we have, using again that $t_{n-k+1} = t_n - t_{k-1}$ and Lemma 4.1,

$$\begin{aligned}\mathbb{E}\|e_{34}\|^2 &= \mathbb{E}\Big(\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\|[A^{\frac{\beta}{2}}\mathrm{e}^{-(t_n-s)A}(I-\mathrm{e}^{-(s-t_{k-1})A})]A^{-\frac{\beta}{2}}f(u(s))\|\,\mathrm{d}s\Big)^2\\ &\le Ct_n\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}(t_n-s)^{-\beta}\tau^\beta\mathbb{E}\|f(u(s))\|^2\,\mathrm{d}s \le C\tau^\beta, \quad 0\le\beta<1.\end{aligned}$$

Putting the pieces together, we have

$$\mathbb{E}\|e^n\|^2 \le C\tau^\beta + 2L_f^2\tau^2\mathbb{E}\|e^n\|^2 + C\tau\sum_{k=1}^{n-1}\mathbb{E}\|e^k\|^2, \quad 0\le\beta<1.$$

Finally, if $\tau < \frac{1}{2L_f}$, then by the discrete Gronwall lemma,

$$\mathbb{E}\|e^n\|^2 \le C\tau^\beta\mathrm{e}^{Ct_n} \le C\tau^\beta,\ 0\le\beta<1,$$

and the theorem is established.

5 Error analysis for the nonlinear random problem

In this section we bound the term $\mathbb{E}\big[\|\bar{v}^n - v^n\|^2\big]$ in (17). We use the global Lipschitz condition (2).

Lemma 5.1. *Assume that $\tau L_f \le \frac{1}{2}$. Then, with $C = 2L_f T e^{2L_f T}$,*

$$\max_{1\le n\le N}\Big(\mathbb{E}\big[\|\bar{v}^n - v^n\|^2\big]\Big)^{\frac{1}{2}} \le C\max_{1\le n\le N}\Big(\mathbb{E}\big[\|w_J^n - w^n\|^2\big]\Big)^{\frac{1}{2}}.$$

Proof. Let $e^n := \bar{v}^n - v^n$. Then, we have by (9b) and (11)

$$e^n + \tau A e^n = \tau\big(f(\bar{v}^n + w_J^n) - f(v^n + w^n)\big) + e^{n-1}.$$

Since $e^0 = 0$, we get by induction

$$e^n = \tau \sum_{j=1}^{n} (I+\tau A)^{-(n+1-j)} \big(f(\bar{v}^j + w_J^j) - f(v^j + w^j)\big).$$

In view of the global Lipschitz condition (2), this results in the estimate

$$\begin{aligned}\|e^n\| &\le L_f \tau \sum_{j=1}^{n} \|(I+\tau A)^{-(n+1-j)}\| \, \|\bar{v}^j + w_J^j - v^j - w^j\| \\ &\le L_f \tau \sum_{j=1}^{n} \big(\|w_J^j - w^j\| + \|e^j\|\big),\end{aligned}$$

since $\|(I+\tau A)^{-1}\| \le 1$. Thus, we obtain

$$\|e^n\| \le (1 - L_f\tau)^{-1} L_f \tau \Big(\sum_{j=1}^{n} \|w_J^j - w^j\| + \sum_{j=1}^{n-1} \|e^j\|\Big).$$

With $L_f \tau \le \frac{1}{2}$ we complete the proof by the standard discrete Gronwall lemma.

References

1. Cioca, P.A., Dahlke, S., Döhring, N., Friedrich, U., Kinzel, S., Lindner, F., Raasch, T., Ritter, K., Schilling, R.: On the convergence analysis of Rothe's method. Preprint Nr. 124, DFG-Schwerpunktprogramm 1324 "Extraktion Quantifizierbarer Information aus Komplexen Systemen" (2012).
2. Cioca, P.A., Dahlke, S., Döhring, N., Kinzel, S., Lindner, F., Raasch, T., Ritter, K., Schilling, R.: Adaptive wavelet methods for the stochastic Poisson equation. BIT Numer. Math. **52**, 589–614 (2012).
3. Cioca, P.A., Dahlke, S., Kinzel, S., Lindner, F., Raasch, T., Ritter, K., Schilling, R.: Spatial Besov regularity for stochastic partial differential equations on Lipschitz domains. Stud. Math. **207**, 197–234 (2011).
4. Cohen, A.: Wavelet methods in numerical analysis. Handbook of Numerical Analysis **7**, 417–711 (2000).
5. Cohen, A., Dahmen, W., DeVore, R.A.: Adaptive wavelet schemes for elliptic operator equations — convergence rates. Math. Comp. **70**, 27–75 (2001).
6. Cohen, A., Dahmen, W., DeVore, R.A.: Sparse evaluation of compositions of functions using multiscale expansions. SIAM J. Math. Anal. **35**, 279–303 (2003).
7. Cohen, A., Dahmen, W., DeVore, R.A.: Adaptive wavelet schemes for nonlinear variational problems. SIAM J. Numer. Anal. **41**, 1785–1823 (2003).
8. Dahmen, W.: Wavelet and multiscale methods for operator equations. Acta Numerica **6**, 55–228 (1997).
9. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions. Cambridge University Press, Cambridge (1992).

10. Kovács, M., Lindgren, F., Larsson, S.: Strong convergence of the finite element method with truncated noise for semilinear parabolic stochastic equations with additive noise. Numer. Algorithms **53**, 309–320 (2010).
11. Kovács, M., Lindgren, F., Larsson, S.: Spatial approximation of stochastic convolutions. J. Comput. Appl. Math. **235**, 3554–3570 (2011).
12. M. Kovács, S. Larsson and F. Lindgren, Weak convergence of finite element approximations of stochastic evolution equations with additive noise, BIT, **52**, 85–108 (2012).
13. Kruse, R.: Optimal error estimates of Galerkin finite element methods for stochastic partial differential equations with multiplicative noise. arXiv:1103.4504v1 [math.NA] (2011).
14. Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations. Springer-Verlag, New York (1983).
15. Printems, J.: On the discretization in time of parabolic stochastic partial differential equations. Math. Model. Numer. Anal. **35**, 1055–1078 (2001).
16. Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems. Second ed. Springer-Verlag, Berlin (2006).
17. Urban, K.: Wavelet Methods for Elliptic Partial Differential Equations, Oxford University Press, Oxford (2009).
18. Yan, Y.: Galerkin finite element methods for stochastic parabolic partial differential equations. SIAM J. Numer. Anal. **43**, 1363–1384 (2005).