

ADAPTIVE CONCEPTS FOR STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS

ANDREAS PROHL AND CHRISTIAN SCHELLNEGGER

ABSTRACT. An adaptive time stepping method to numerically solve a general SPDE is proposed, where local step sizes are chosen in regard of the distance between empirical laws of subsequent time iterates and extrapolated data. The histogram-based estimator uses a data-driven partitioning of the high-dimensional state space, and efficient sampling by bootstrapping. Time adaptivity is then complemented by a local refinement/coarsening strategy of the spatial mesh of a stochastic version of the ZZ-estimator. Next to an improved accuracy, we observe a significantly reduced empirical variance of standard estimators, and therefore a reduced sampling effort. The performance of the adaptive strategies is studied for SPDEs with linear drift, including the convection-dominated case where the streamline diffusion method is adopted to attain a stable discretization, and the stochastic version of the non-linear harmonic map heat flow to the sphere \mathbb{S}^2 where approximate solutions exhibit discrete blow-up dynamics.

1. INTRODUCTION

We develop new space-time adaptive concepts which base on local changes of empirical laws of approximate solutions of the general SPDE

$$\begin{aligned} dX_t &= \mathcal{A}(X_t)dt + \iota\sigma(X_t)dW_t && \text{in } D_T := (0, T] \times D, \\ X_0 &= x_0 && \text{in } D, \end{aligned} \tag{1.1}$$

with proper boundary conditions. Here, $D \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, is a bounded polyhedral domain, $0 < T < \infty$ denotes the terminal time, and \mathcal{A} a (non-)linear second order differential operator; see (2.1) and (2.4). We denote by $W \equiv \{W_t; t \in [0, T]\}$ a Hilbert space-valued trace-class Wiener process; $\iota \in \mathbb{R}$ is later referred to as noise intensity, and σ is a diffusion operator; see Section 2.1 for details.

For deterministic PDEs, there is a rich literature on adaptive methods to automatically generate space-time meshes which capture the structure of the solution and thus significantly reduce the computational effort. For low-dimensional SDEs, different concepts of time adaptivity have been proposed e.g. in [24] to accurately approximate solutions

Date: March 8, 2019.

2000 Mathematics Subject Classification. Primary.

Key words and phrases. Space-time adaptivity of SPDEs, distance, probability measures, comparing distributions, partitioning of the high-dimensional state space, variance reduction.

Acknowledgment: The authors are grateful to interesting discussions with T. Dunst and E. Teufel (U Tübingen), and M. Tretyakov (U Nottingham).

pathwisely, or e.g. in [30] to approximate corresponding laws of iterates. The construction [25, 30] of an adaptive mesh is based on an (asymptotic) a posteriori estimate, whose computation involves global, coupled random dual backward equations; conceptually, the construction rests on using Kolmogorov's backward equation, which restricts the practical applicability to small SDE systems. The advantage, however, is a theoretical backup of this adaptive algorithm.

In order to develop adaptive concepts for large SDE systems, or even those which come from a Galerkin discretization of the SPDE (1.1), we rather *sample* empirical laws of related subsequent iterates in this work and estimate their distance. The algorithm then initiates refinement/coarsening of the temporal mesh at places where related empirical laws of subsequent iterates change rapidly/slowly. We motivate conceptual ideas: let $Y^j \in L^2(\Omega, \mathbb{V}_h^j)$ be an approximation of the solution X of (1.1) at time $t_j \in [0, T]$, which has already been obtained from the finite element-based discretization Scheme 3.1 of (1.1). We want to determine the new local time step size $\tau^{j+1} := t_{j+1} - t_j > 0$. For this purpose, we first check whether a refinement of the earlier time step τ^j is needed, by

- 1) computing the new random variable $Y^{1,j+1} \in L^2(\Omega, \mathbb{V}_h^j)$ with the help of the previous step size $\tau^{1,j+1} := \tau^j$ in Scheme 3.1, and then determining its empirical law $\widehat{\mu}_\tau^{1,j+1} := \widehat{\mathcal{L}}(Y^{1,j+1})$.
- 2) We then proceed correspondingly with the step size $\tau^{2,j+1} := \tau^j/2$, and use extrapolation to construct the empirical measure $\widehat{\mu}_\tau^{2,j+1} := \widehat{\mathcal{L}}(Y^{2,j+1})$.
- 3) Now we compute the (Hellinger-)distance $\mathfrak{d}(\widehat{\mu}_\tau^{1,j+1}, \widehat{\mu}_\tau^{2,j+1})$ of the both empirical probability measures $\{\widehat{\mu}_\tau^{\mathfrak{s},j+1}; \mathfrak{s} \in \{1, 2\}\}$ to decide whether to refine τ^j or not.

A corresponding strategy applies to steer coarsening; see Section 4.4 for further details.

A relevant step in this procedure is therefore to efficiently assemble and compare the involved empirical measures $\{\widehat{\mu}_\tau^{\mathfrak{s},j+1}; \mathfrak{s} \in \{1, 2\}\}$, both of which are defined on a σ -algebra over the state space \mathbb{R}^{L^j} , which is isomorphic to the L^j -dimensional finite element space \mathbb{V}_h^j . This σ -algebra is generated by a data-dependent partition $\widehat{\mathcal{P}}_{\tau; R_\tau}^{j+1} := \{\widehat{C}_{\tau; r}^{j+1}\}_{r=1}^{R_\tau}$ of the state space \mathbb{R}^{L^j} into $R_\tau \in \mathbb{N}$ many cells; see Figure 1(B) for an illustration in the case $L^j = 2$. While this strategy to approximate distributions on (low-dimensional) state spaces is well-known in the statistical literature [10, Chapter 21], it has only recently been used in [11] for the simulation of BSDEs resp. BSPDEs, which is again a different setting to the present one: in Section 4.1, we explain how a data-dependent partition of \mathbb{R}^{L^j} may efficiently be generated to assemble the two empirical measures from above. Sections 4.2 and 4.4 then detail the role of the distance of these empirical measures to attain the new time step τ^{j+1} .

These concepts lead to Algorithm 4.4, which may be used to steer time adaptivity for the following SDE from [25, Example 5.1],

$$dX_t = \frac{\mathbb{1}_{(1/3, 1]}(t)}{2\sqrt{t - 1/3 + 10^{-8}}} X_t dt + X_t dW_t \quad \forall t \in (0, 1], \quad X_0 = 1; \quad (1.2)$$

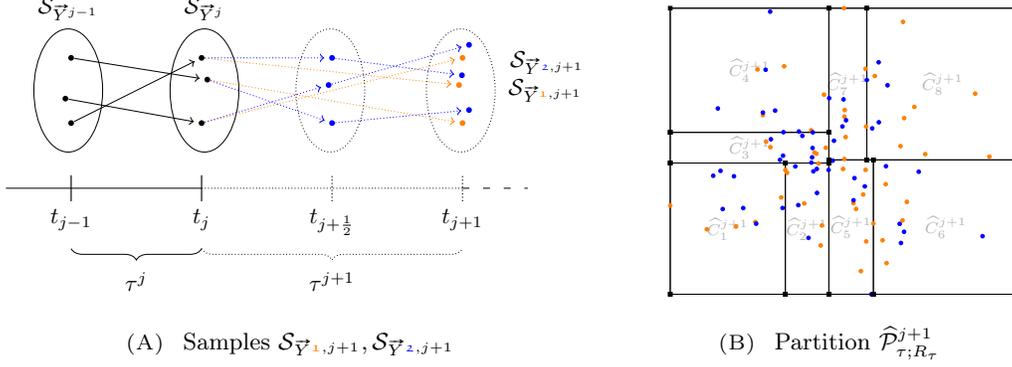


FIGURE 1. (Algorithm 4.4) Computation of the new local time step size $\tau^{j+1} := t_{j+1} - t_j > 0$ based on the previous step size τ^j and the sample $\mathcal{S}_{\bar{Y}^j}$: (A) The samples $\mathcal{S}_{\bar{Y}^{1,j+1}}$ and $\mathcal{S}_{\bar{Y}^{2,j+1}}$ are computed by the same discretization scheme, but with different time step sizes $\tau^{1,j+1}$ and $\tau^{2,j+1}$; see Section 4.4 for the used notation and further details. (B) Sketch to assemble a single realization of empirical laws $\hat{\mu}_\tau^{1,j+1}$ and $\hat{\mu}_\tau^{2,j+1}$ on the data-dependent partition $\hat{\mathcal{P}}_{\tau;R_\tau}^{j+1} \equiv \hat{\mathcal{P}}_{\tau;R_\tau}^{j+1}(\mathcal{S}_{\bar{Y}^{1,j+1}} \cup \mathcal{S}_{\bar{Y}^{2,j+1}})$ on the state space \mathbb{R}^2 with $R_\tau = 8$ cells (\blacksquare); see Section 4.2.

note that $t \mapsto \mathbb{E}[|X_t|]$ fastly changes in the vicinity of $t = 1/3$, and thus favors a higher resolution by a refined mesh in its neighborhood; see Figure 2(B).

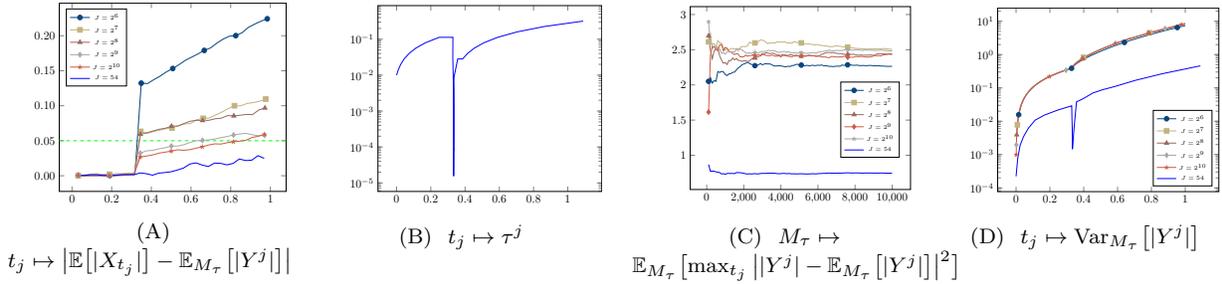


FIGURE 2. (Example 1.1 for $\mathbf{d} = \mathbf{d}_H$ (see Section 4.2), $\text{To1}_\tau = 0.05$ (---) and $R_\tau = 2^6$, $M_\tau = 10^5$) (A) Error for uniform ($\blacklozenge, \blacktriangle, \blackstar, \blackdiamond, \blacktriangleright$) vs. adaptive (—) time meshes via Algorithm 4.4, and (B) corresponding adaptive time step size. (C) Error $M_\tau \mapsto \mathbb{E}_{M_\tau}[\max_{t_j} ||Y^j| - \mathbb{E}_{M_\tau}[|Y^j|]|^2]$, and (D) empirical variance $t_j \mapsto \text{Var}_{M_\tau}[|Y^j|]$ for uniform vs. adaptive meshes.

Example 1.1. Figure 2(A) displays errors for a drift-implicit Euler discretization of (1.2) on uniform vs. adaptive time meshes ($M_\tau = 10^5$ MC simulations). A uniform mesh with $J = 2^{10}$ time steps is needed vs. an adaptively refined mesh with $J = 54$ time steps via Algorithm 4.4 (—) to stay below the given error threshold (---). The adapted local time step

size $\{\tau^j\}_j$ rapidly decays near $t = 1/3$, and afterwards coarsens again (see Figure 2(B)); compare also with [25, Figure 2]. Figure 2(C) shows a significantly reduced empirical variance on adaptive meshes ($\mathbb{E}_{M_\tau} [|\mathbf{Y}^j|] := \frac{1}{M_\tau} \sum_{k=1}^{M_\tau} |\mathbf{Y}^j(\omega_k)|$); Figure 2(D) displays the empirical variance $t_j \mapsto \text{Var}_{M_\tau} [|\mathbf{Y}^j|] := \mathbb{E}_{M_\tau} [|\mathbf{Y}^j| - \mathbb{E}_{M_\tau} [|\mathbf{Y}^j|]]^2$ to attribute this observation to a smaller empirical variance near $t = 1/3$ via Algorithm 4.4, as opposed to uniform grids.

An important feature of the adaptive Algorithm 4.4 is that the dimension of a system of SDEs does not affect its applicability — as opposed to [25, 30], where it is practically limited to small SDE systems.

Example 1.2. Consider

$$d\mathbf{X}_t = [-\mathbf{J} + \mathbf{D}(t)]\mathbf{X}_t dt + [1 + |\mathbf{X}_t|]d\mathbf{W}_t \quad \forall t \in (0, 1], \quad \mathbf{X}_0 = (1, 1, 1, 1)^\top, \quad (1.3)$$

with $\mathbf{J} = 16 \cdot \text{tridiag}[-1, 2, -1] \in \mathbb{R}^{4 \times 4}$, and $\mathbf{D}(t) = \text{diag}[\gamma_1(t), \dots, \gamma_4(t)] \in \mathbb{R}^{4 \times 4}$, where $\gamma_i(t) := \mathbb{1}_{(\beta_i, 1)}(t)/2\sqrt{t - \beta_i + 10^{-8}}$ and $\beta_i := i/5$. Here, $t \mapsto \mathbb{E}[|\mathbf{X}_t|]$ rapidly changes in the vicinity of times $t \in \{\beta_i\}_{i=1}^4$; see Figure 3(B). Computational studies for other choices $\mathbf{d} \in \{\mathbf{d}_{\text{KL}}, \mathbf{d}_{\text{TV}}\}$ (see Section 4.2) show comparable results.

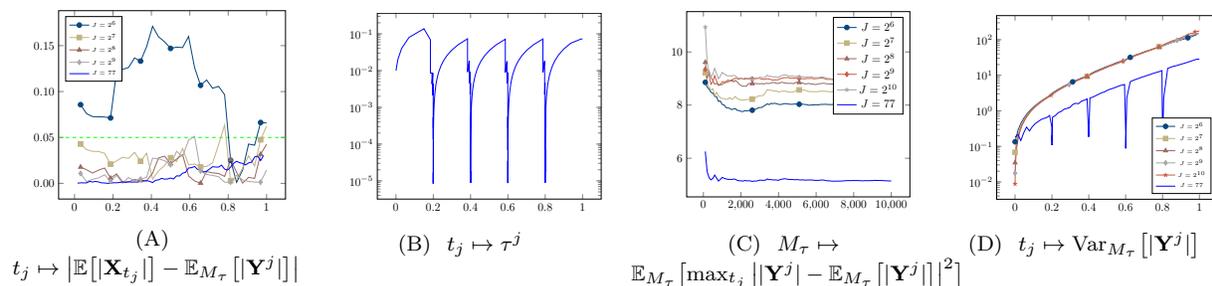


FIGURE 3. (Example 1.2 for $\mathbf{d} = \mathbf{d}_H$ (see Section 4.2), $\text{To1}_\tau = 0.05$ (---) and $R_\tau = 2^6$, $M_\tau = 10^5$) (A) Error for uniform ($\blacklozenge, \blacktriangle, \blackstar, \blackdiamond, \blackdagger$) vs. adaptive (—) time meshes via Algorithm 4.4, and (B) corresponding adaptive time step size. (C) $M_\tau \mapsto \mathbb{E}_{M_\tau} [\max_{t_j} ||\mathbf{Y}^j| - \mathbb{E}_{M_\tau} [|\mathbf{Y}^j|]|^2]$, and (D) empirical variance $t_j \mapsto \text{Var}_{M_\tau} [|\mathbf{Y}^j|]$ for uniform vs. adaptive meshes.

Large systems of SDEs arise after a spatial discretization of the SPDE (1.1), and thus finer partitions $\widehat{\mathcal{P}}_{\tau; R_\tau}^{j+1}$ are expected to properly resolve the high-dimensional state space \mathbb{R}^{L^j} . It is also plausible that the type of PDE matters when the number R_τ of used cells is chosen: for example, the simulations in Example 5.1 for SPDE (1.1) with $\mathcal{A}(X) = \varepsilon \Delta X - \beta \cdot \nabla X$ ($\varepsilon > 0$ small) suggest $R_\tau \approx 2^{12}$ for a stable selection of local time step sizes (see Figure 11): it is due to the convection-dominated drift that steep spatial gradients inside diffuse layers appear, which may easily be perturbed by the acting noise. In contrast, we found that only $R_\tau \approx 2^7$ cells are appropriate to reliably detect changes of subsequent empirical distributions in comparative studies for the drift operator $\mathcal{A}(X) = \Delta X$. Obviously, the choice of R_τ also depends on the specific noise, and its intensity ι in (1.1).

Next to time adaptivity, we use space adaptivity to discretize (1.1); see Figure 4. For this purpose, we adapt the idea of the ZZ-estimator [33] for deterministic PDEs to possibly refine/coarsen an element $K \in \mathcal{T}_h^j$, where \mathcal{T}_h^j denotes the triangulation of D at time t_j : in this vein, we compute the distance $\mathbf{d}(\widehat{\mu}_{h;K}^{1,j+1}, \widehat{\mu}_{h;K}^{2,j+1})$ for each K , where

- 1) $\widehat{\mu}_{h;K}^{1,j+1}$ is the empirical law of the \mathbb{R} -valued random variable $|\nabla Y^{j+1}|_K$, and
- 2) $\widehat{\mu}_{h;K}^{2,j+1}$ is the empirical law of the norm of the recovered gradient $G_h(\nabla Y^{j+1})$ (see equation (4.5)) which averages gradients locally.

The element $K \in \mathcal{T}_h^j$ is then refined/coarsened depending on the value of $\mathbf{d}(\widehat{\mu}_{h;K}^{1,j+1}, \widehat{\mu}_{h;K}^{2,j+1})$ — which is large for strong local variations of the solution. Constructing the empirical measures in 1)–2) is again with the help a data-dependent partitioning $\widehat{\mathcal{P}}_{h;R_h;K}^{j+1} := \{\widehat{C}_{h;r;K}^{j+1}\}_{r=1}^{R_h}$ of the state space, which here is \mathbb{R}_0^+ ; see Section 4.5 for further details. Adaptivity in space steers the re-distribution of spatial nodal points at each time step, leading e.g. to a better resolution of diffuse layers in Figure 4(A) in the case of the convection-dominated SPDE given in Example 5.1. The comparative studies in Figures 4(A) and 14(A) evidence that the choice of the distance $\mathbf{d} \in \{\mathbf{d}_H, \mathbf{d}_{KL}, \mathbf{d}_{TV}\}$ matters, and that \mathbf{d}_{TV} should be given preference; see also Table 2. However, spatial adaptivity not only enhances the resolution of the computation (see Figure 13), but also further reduces the empirical variance of estimators, which was already an outcome of adaptivity in time; see Figure 12. As a consequence, and if compared to uniform space-time meshes, space-time adaptivity will allow smaller values M_τ (resp. M_h) of needed samples — whose resourcing is the computationally most expensive part in the algorithm (see below) —, as well as smaller values of R_τ (resp. R_h).

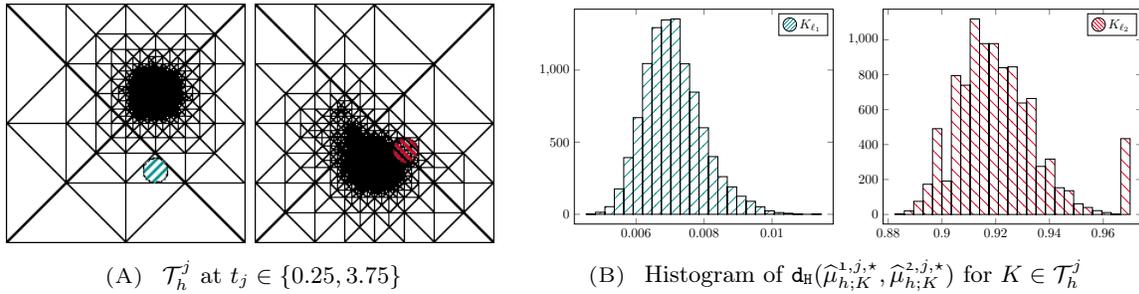


FIGURE 4. (Example 5.1 for $\iota = 0.3$, as well as $\text{To1}_h = 0.05$ and $R_h = 2^5$, $M_h = 10^3$ for each $K \in \mathcal{T}_h^j$): (A) Snapshots of different spatial meshes \mathcal{T}_h^j at evaluated times $t_j \in [0, 2\pi]$ with (B) corresponding histograms of $\mathbf{d}(\widehat{\mu}_{h;K}^{1,j,*}, \widehat{\mu}_{h;K}^{2,j,*})$ at selected elements K_{l_1} ($\textcircled{\text{blue}}$), K_{l_2} ($\textcircled{\text{red}}$) $\in \{\mathcal{T}_h^5, \mathcal{T}_h^{43}\}$ in (A) for $\mathbf{d} = \mathbf{d}_H$; see Section 4.5.

So far, the number of cells R_τ (resp. R_h) was fixed. However, the dimensionality and (hence) required resolution of state spaces may change in time (respectively in space), which, in turn, motivates to adjust the numbers R_τ (resp. R_h) of cells to build the above pairs of empirical laws in time. The computational studies in Section 5 (see Figure 15) favor

a heuristic strategy (5.5) which generates a non-constant sequence $\{\widehat{R}_\tau^{j+1}\}_j$ (resp. $\{\widehat{R}_h^{j+1}\}_j$) in time to vary the complexity of partitions, leading to further savings of computational resources while keeping the accuracy; the generation of corresponding partitions $\{\widehat{\mathcal{P}}_{\tau, \widehat{R}_\tau^{j+1}}^{j+1}\}$ (resp. $\{\widehat{\mathcal{P}}_{h, \widehat{R}_h^{j+1}}^{j+1}\}$) is then supplemented by parallel computation using OpenMP [7] for further speed-up; see Table 1. Another motivation for adaptivity of statistical parameters is due to the already mentioned role of space-time adaptivity to reduce the empirical variance of computed solutions (recall e.g. Figure 2(D)): as a consequence, sample sizes M_τ (resp. M_h) should be adaptively selected as well to avoid to compute large samples whenever possible: Figure 13 for Example 5.1 provides equally accurate, comparative simulations for uniformly vs. adaptively chosen statistical parameters M_τ, R_h which significantly reduce computational times in the latter case; see Section 5.1.3 for further details. To conclude, an intimate link of adaptive strategies for space-time discretization and statistics is needed to properly choose involved parameters.

A relevant issue which prevents an immediate, efficient realization of the given adaptive concepts so far is its huge computational complexity. The overall computational effort at time t_j is distributed across different tasks: it starts with the computation of two initial M_τ^j -samples of $\mathcal{O}(M_\tau^j(L^j)^{3/2})$ complexity [14], which is the most time consuming part. To then build the data-dependent partition $\widehat{\mathcal{P}}_{\tau, R_\tau^j}^{j+1}$ to initiate time adaptivity only causes low storage requirements $\mathcal{O}(R_\tau^j(L^j + \log(R_\tau^j)))$ with $R_\tau^j \ll M_\tau^j$, and is fastly accomplished since the partition is stored as a binary tree. Once $\widehat{\mathcal{P}}_{\tau, R_\tau^j}^{j+1}$ is available, we need another two fresh M_τ^j -samples to assemble a *single* realization of each $\{\widehat{\mu}_\tau^{s, j+1}; s \in \{1, 2\}\}$; see Figure 1(B) for an illustration. Obviously, it is now impractical to sample this tuple of empirical measures by a repeated computation of approximate solutions of the SPDE (1.1): instead, statistical inference about the distribution of $\mathbf{d}(\widehat{\mu}_\tau^{1, j+1}, \widehat{\mu}_\tau^{2, j+1})$ is then obtained via a bootstrap estimator, which draws with replacement from the existing sample, thus providing the needed fresh, independent sample copies; see Section 4.3. This part also requires no additional storage effort, since bootstrap samples are internally identified via associated index sets (see Section 4.3); for example, to provide a single sample of size $M_\tau = 10^5$ with $\mathcal{T}_h^j = \mathcal{T}_h^0$ for $h^0 = 2^{-6}$ (Example 5.1) requires approximatively 10 minutes of computational time, whereas its generation via bootstrap only takes approximatively 4 seconds. As a result, the combination of the above adaptive concepts discussed so far with the efficient sampling via bootstrap of the related empirical measures to steer space-time meshes locally leads to an efficient method. For example, we observe huge savings of more than 75% of computational time to obtain equally accurate simulations for Example 5.1 if compared to those with uniform discretization and statistical parameters; see Figure 13 and Section 5.1.3.

The proposed adaptive concepts are tested for two prototype problems: we start with a convection-dominated SPDE with linear drift in Section 2.1. In Section 3.1 we propose a stable discretization by a stochastic version of the SUPG method (see Section 3.1) which avoids spurious solutions that otherwise appear for the classical Galerkin scheme; see Figure 8. It turns out that the SUPG method not only yields a more accurate (see Figure 9),

stable approximation, see Lemma 3.2, but also reduces the empirical variance of related estimators. An SPDE with non-linear drift is the stochastic harmonic map flow to the sphere \mathbb{S}^2 , for which blow-up is known in the presence of super-critical initial data. For $D \subset \mathbb{R}^d$, $d \in \{2, 3\}$ this SPDE has a weak martingale solution (see e.g. [3]) rather than a probabilistically strong solution as for the former problem; however, our concept of space-time adaptivity is based on distributions, and therefore is applicable here as well. The simulations in Section 5 illustrate a refinement of space-time meshes close to singular behaviors of the solution, and a fast coarsening again beyond; concomitantly, the adaptively chosen number of cells R_τ^j (resp. R_h^j) grows close to the (discrete) blow-up time, and rapidly decays beyond it again. A corresponding dependence is observed for M_τ^j (resp. M_h^j), which again evidences a proper automatic adjustment of involved discretization and statistical parameters to efficiently resolve space-time varying behaviors of the solution for this nonlinear SPDE as well.

2. TWO PROTOTYPE SPDES

2.1. An SPDE with linear drift. Let $T > 0$, and $\mathcal{A} : \mathbb{W}_0^{1,2} \rightarrow \mathbb{W}^{-1,2}$ in (1.1) be of the form

$$\mathcal{A}(u) = \operatorname{div}(\mathbf{A}(t, \mathbf{x})\nabla u) - \boldsymbol{\beta}(t, \mathbf{x}) \cdot \nabla u, \quad (2.1)$$

with measurable $\mathbf{A} : D_T \rightarrow \mathbb{R}_{\operatorname{spd}}^{d \times d}$, and $\boldsymbol{\beta} \in L^\infty(D_T; [\mathbb{W}^{1,\infty}]^d)$. Let \mathbb{K} be a separable Hilbert space, and $\sigma : \mathbb{L}^2 \rightarrow \mathcal{L}(\mathbb{L}^2, \mathbb{K})$ be Lipschitz continuous, i.e., there exist constants $K_1, K_2 > 0$ such that for all $u, v \in \mathbb{L}^2$ holds

$$\|\sigma(u)\|_{\mathcal{L}(\mathbb{L}^2, \mathbb{K})} \leq K_1(1 + \|u\|_{\mathbb{L}^2}), \quad \|\sigma(u) - \sigma(v)\|_{\mathcal{L}(\mathbb{L}^2, \mathbb{K})}^2 \leq K_2\|u - v\|_{\mathbb{L}^2}^2, \quad (2.2)$$

where $\|\cdot\|_{\mathbb{L}^2}$ resp. $(\cdot, \cdot)_{\mathbb{L}^2}$ denotes the norm resp. the scalar product in $\mathbb{L}^2 := \mathbb{L}^2(D)$. The norm in $\mathbb{W}^{k,2} := \mathbb{W}^{k,2}(D)$ for $k \in \{0, 1\}$ is denoted by $\|\cdot\|_{\mathbb{W}^{k,2}}$; see e.g. [2].

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space, on which a \mathbb{K} -valued \mathbf{Q} -Wiener process $W \equiv \{W_t; t \in [0, T]\}$ with trace-class operator \mathbf{Q} is defined. Problem 1.1 with $x_0 \in \mathbb{L}^2$, and homogeneous Dirichlet boundary data may then be recast into the form: Find $X \equiv \{X_t; t \in [0, T]\}$ such that $X \in L^2_{\mathbb{F}}(\Omega; \mathcal{C}([0, T]; \mathbb{L}^2) \cap L^2(0, T; \mathbb{W}_0^{1,2}))$ satisfies \mathbb{P} -almost surely for every $t \in [0, T]$, and all $\Psi \in \mathbb{W}_0^{1,2}$,

$$(X_t, \Psi) - (x_0, \Psi) = - \int_0^t \left[(\mathbf{A}_s \nabla X_s, \nabla \Psi) + (\boldsymbol{\beta}_s \cdot \nabla X_s, \Psi) \right] ds + \iota \int_0^t (\sigma(X_s) dW_s, \Psi). \quad (2.3)$$

According to [27], problem (2.3) has a unique solution. Example 5.1 studies its numerical discretization via Scheme 3.1, and adaptivity in the ‘convection-dominated case’, i.e. $\mathbf{A} \equiv \epsilon \mathbf{I}$ for $\epsilon > 0$, and $\boldsymbol{\beta} \neq \mathbf{0}$. For this purpose, we approximate the \mathbb{K} -valued \mathbf{Q} -Wiener process \mathbf{W} by a \mathbb{K} -valued \mathbf{Q} -random walk. Let $\mathcal{I}_\tau := \{t_j\}_{j=0}^J$ with $\tau^{j+1} := t_{j+1} - t_j > 0$. A \mathbf{Q} -random walk on $(\Omega, \mathcal{F}, \mathbb{P})$ along \mathcal{I}_τ is a sequence $\{\xi_{j+1}\}_{j=0}^{J-1}$ of \mathbb{K} -valued independent identically distributed random variables such that for each $j \in \{0, 1, \dots, J-1\}$ the following conditions are satisfied:

- (1) $\mathbb{E}[\xi_{j+1}] = 0$, and $\mathbb{E}[(\xi_{j+1}, x)_{\mathbb{K}}(\xi_{j+1}, y)_{\mathbb{K}}] = \tau^{j+1}(\mathbf{Q}x, y)_{\mathbb{K}}$ for all $x, y \in \mathbb{K}$.
- (2) For every $p \in \mathbb{N}$, there exists $C_p > 0$ such that $\mathbb{E}[\|\xi_{j+1}\|_{\mathbb{K}}^{2p}] \leq C_p(\tau^{j+1})^p$.

2.2. An SPDE with non-linear drift — the stochastic harmonic map flow to \mathbb{S}^2 . Let $T > 0$. We look for an \mathbb{S}^2 -valued process $\mathbf{X} \equiv \{\mathbf{X}_t; t \in [0, T]\}$ that satisfies the non-linear SPDE

$$\begin{aligned} d\mathbf{X}_t + \mathbf{X}_t \times [\mathbf{X}_t \times \Delta \mathbf{X}_t] dt &= \iota \mathbf{X}_t \times \circ d\mathbf{W}_t && \text{in } D_T := (0, T] \times D, \\ \partial_{\mathbf{n}} \mathbf{X}_t &= 0 && \text{on } \partial D_T := (0, T] \times \partial D, \\ \mathbf{X}_0 &= \mathbf{x}_0 && \text{in } D, \end{aligned} \quad (2.4)$$

for $\mathbf{x}_0 \in \mathbb{W}^{1,2}(D; \mathbb{S}^2)$, and $\mathbf{W} = (W_1, W_2, W_3)^\top$ with three independent \mathbb{K} -valued \mathbf{Q} -Wiener processes $\{W_i\}_{i=1}^3$. Here, the Stratonovich integral is used which is indicated by ‘ \circ ’. Problem (2.4) is only known to have a weak martingale solution (see e.g. [3]), which relaxes the strong solution concept that applies for (2.3). Related computational experiments in [3] motivate possible blow-up dynamics of a solution of (2.4). A recent analytical study of this phenomenon is [19], which in fact characterizes a weak martingale solution as a regular, strong solution, apart from no more but finitely many space-time points (see [19, Theorem 3]). The computational studies in Section 5 illustrate a proper space-time resolution of singular behaviors of the solution of (2.4) by the adaptive method discussed in Section 4.

3. SPACE-TIME DISCRETIZATION OF THE SPDE (1.1)

In this section, we propose stable space-time discretizations for problems (2.3) and (2.4). Let $\{(\tau^j, \mathcal{T}_h^j)\}_{j=0}^J$ be a space-time mesh covering $[0, T] \times \bar{D}$. For $j \in \{0, 1, \dots, J\}$, we define sequences of $\mathbb{W}_0^{1,2}$ -conforming lowest order finite element spaces [4],

$$\mathbb{V}_h^j := \{\Psi \in \mathcal{C}(\bar{D}) : \Psi|_K \in \mathbb{P}_1(K) \forall K \in \mathcal{T}_h^j\},$$

with $L^j := \dim \mathbb{V}_h^j$ and elements $K \in \mathcal{T}_h^j$ of the regular mesh \mathcal{T}_h^j , with an associated set of (free) nodes $\{\mathbf{x}_\ell\}_{\ell=1}^{L^j}$. Moreover, let $\{\Psi_\ell\}_{\ell=1}^{L^j}$ denote the nodal basis of the finite element space \mathbb{V}_h^j , and $\vec{\Upsilon}_{\mathbb{V}_h^j} : \mathbb{V}_h^j \rightarrow \mathbb{R}^{L^j}$ the corresponding coordinate map.

A finite element discretization of problem (2.4) uses \mathbb{R}^3 -valued functions $\Psi \in [\mathbb{V}_h^j]^3$ at time t_j . We define a bilinear form $(\cdot, \cdot)_{h;j} : \mathcal{C}(\bar{D}; \mathbb{R}^3) \times \mathcal{C}(\bar{D}; \mathbb{R}^3) \rightarrow \mathbb{R}$ by [3, p. 107]

$$(\cdot, \cdot)_{h;j} := \int_D \mathcal{I}_h^j \left[\langle \Psi(\mathbf{x}), \Xi(\mathbf{x}) \rangle_{\mathbb{R}^3} \right] d\mathbf{x} = \sum_{\ell=1}^{L^j} \zeta_\ell \langle \Psi(\mathbf{x}_\ell), \Xi(\mathbf{x}_\ell) \rangle_{\mathbb{R}^3} \quad \forall \Psi, \Xi \in \mathcal{C}(\bar{D}; \mathbb{R}^3)$$

for certain weights $\zeta_\ell = \int_D \Psi_\ell d\mathbf{x} > 0$, $\ell \in \{1, \dots, L^j\}$. The (affine) nodal interpolation operator \mathcal{I}_h^j is a bounded map from $\mathcal{C}(\bar{D}; \mathbb{R})$ to \mathbb{V}_h^j . The discrete Laplace operator $\Delta_h : [\mathbb{V}_h^j]^3 \rightarrow [\mathbb{V}_h^j]^3$ is defined by $-(\Delta_h \Xi, \Psi)_{h;j} = (\nabla \Xi, \nabla \Psi)_{\mathbf{L}^2}$ for all $\Psi, \Xi \in [\mathbb{V}_h^j]^3$.

3.1. Discretization of the SPDE (2.3). A standard finite element discretization of (2.3) for $\iota = 0$ and with $\mathbf{A} \equiv \epsilon \mathbf{I}$ for some $\epsilon > 0$, and $\boldsymbol{\beta} \neq \mathbf{0}$ in (2.3) is known to possibly lead to approximates with spurious oscillations, which may e.g. be avoided by the stable streamline-upwind Petrov-Galerkin (SUPG) method; see e.g. [22]. The discretization in [23] for the related deterministic problem uses the continuous Galerkin method of order ≥ 1 , with (local) space-time test functions $\Psi + \delta_K(\partial_t \Psi + \boldsymbol{\beta} \cdot \nabla \Psi)$ on strips $S_j := \{(t, \mathbf{x}); t \in (t_j, t_{j+1}), \mathbf{x} \in D\}$, where $\{\delta_K\}_K \subset \mathbb{R}$ is a given family of mesh-dependent weights. For the convection-dominated SPDE (2.3), we rather suggest the SUPG method with spatial test functions $\Psi + \delta_K \boldsymbol{\beta} \cdot \nabla \Psi$, in combination with the implicit Euler-Maruyama scheme to control amplification of oscillatory numerical artefacts by the noise term, which is due to the limited time-regularity of solutions of the SPDE (1.1).

Scheme 3.1. *Let $\epsilon > 0$, $Y^0 \in \mathbb{V}_h^0$, $\{(\tau^{j+1}, \mathcal{T}_h^{j+1}, \{\delta_K^{j+1}\}_K)\}_{j=0}^{J-1}$, and the \mathbf{Q} -random walk $\{\xi_{j+1}\}_{j=0}^{J-1}$ along \mathcal{I}_τ on $(\Omega, \mathcal{F}, \mathbb{P})$ be given. For all $j \in \{0, 1, \dots, J-1\}$, determine a \mathbb{V}_h^{j+1} -valued random variable Y^{j+1} such that \mathbb{P} -almost surely for all $\Psi \in \mathbb{V}_h^{j+1}$*

$$\begin{aligned} & (Y^{j+1} - Y^j, \Psi + \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla \Psi)_{\mathbb{L}^2} + \tau^{j+1} \epsilon (\nabla Y^{j+1}, \nabla \Psi)_{\mathbb{L}^2} \\ & \quad + \tau^{j+1} (\boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1}, \Psi + \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla \Psi)_{\mathbb{L}^2} \quad (3.1) \\ & = \iota (\sigma(Y^j) \xi_{j+1}, \Psi + \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla \Psi)_{\mathbb{L}^2}. \end{aligned}$$

The term scaled by $\epsilon > 0$ forgoes the further term $\tau^{j+1} \epsilon (\Delta Y^{j+1}, \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla \Psi)_{\mathbb{L}^2}$, due to the use of piecewise affine finite element functions. Computations for the deterministic counterpart in [20, 5] favor the uniform choice $\delta_K^{j+1} = \max_K h_K^{j+1} / 2 |\boldsymbol{\beta}_{t_{j+1}}|$ to achieve stable, accurate results on coarse meshes $\mathcal{T}_h^{j+1} \equiv \mathcal{T}_h^0$, since the small scales which require a stabilization are the spatial ones. Stability for this deterministic case and possible driving right-hand sides $f \in \text{BV}(0, T; \mathbb{L}^2)$ of the PDE has been shown in [5]; according to [5, Remark 5], a different choice for each δ_K^{j+1} of order $\mathcal{O}(\tau^{j+1})$ should however be preferred for a less regular f which, in particular, appears in the stochastic setting in Scheme 3.1; see also the simulations in Section 5 for computational evidence.

Lemma 3.2. *Let $\{Y^{j+1}\}_{j=0}^{J-1}$ be the solution of Scheme 3.1, and let $\max_K \delta_K^{j+1} \leq \tau^{j+1}/2$ be valid for each $j \in \{0, 1, \dots, J-1\}$. There exists $C \equiv C(T, \text{Tr}(\mathbf{Q}), \boldsymbol{\beta}) > 0$ such that*

$$\begin{aligned} & \max_{0 \leq j \leq J-1} \mathbb{E} \left[\|Y^{j+1}\|_{\mathbb{L}^2}^2 \right] + \sum_{j=0}^{J-1} \mathbb{E} \left[\|Y^{j+1} - Y^j\|_{\mathbb{L}^2}^2 \right] + \epsilon \sum_{j=0}^{J-1} \tau^{j+1} \mathbb{E} \left[\|\nabla Y^{j+1}\|_{\mathbb{L}^2}^2 \right] \\ & \quad + \sum_{j=0}^{J-1} \tau^{j+1} \mathbb{E} \left[\left\| \sqrt{\delta_K^{j+1}} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1} \right\|_{\mathbb{L}^2}^2 \right] \leq C \mathbb{E} \left[\|Y^0\|_{\mathbb{L}^2}^2 \right]. \end{aligned}$$

The Lax-Milgram lemma implies the existence and uniqueness of $Y^{j+1} \in \mathbb{V}_h^{j+1}$ for each j and \mathbb{P} -almost surely. We remark that the argument in [5] to prove stability in the deterministic case may not be applied in the present stochastic case, since (discrete) temporal

derivatives of the iterates $\{Y^{j+1}\}_j$ in Scheme 3.1 grow unboundedly for $\tau^{j+1} \downarrow 0$. For our numerical experiments in Section 5, we later choose $\delta_K^{j+1} := \min\{h_K^{j+1}, \tau^{j+1}/2\}$.

Proof. Consider (3.1) for a fixed $\omega \in \Omega$, and choose $\Psi = Y^{j+1}(\omega) \in \mathbb{V}_h^{j+1}$ as test function in equation (3.1). By binomial formula, we obtain

$$\begin{aligned} & \frac{1}{2} \left[\|Y^{j+1}\|_{\mathbb{L}^2}^2 - \|Y^j\|_{\mathbb{L}^2}^2 + \|Y^{j+1} - Y^j\|_{\mathbb{L}^2}^2 \right] + (Y^{j+1} - Y^j, \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1})_{\mathbb{L}^2} \\ & + \tau^{j+1} \epsilon \| \nabla Y^{j+1} \|_{\mathbb{L}^2}^2 - \tau^{j+1} (\boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1}, Y^{j+1})_{\mathbb{L}^2} \\ & + \tau^{j+1} \left\| \sqrt{\delta_K^{j+1}} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1} \right\|_{\mathbb{L}^2}^2 \\ & \leq \left(\sigma(Y^j) \xi_{j+1}, \{Y^{j+1} - Y^j\} + \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1} \right)_{\mathbb{L}^2} + \mathbb{I}_j, \end{aligned} \tag{3.2}$$

where $\mathbb{I}_j := (\sigma(Y^j) \xi_{j+1}, Y^j)_{\mathbb{L}^2}$. By Young's inequality, (2.2), and $\max_K \delta_K^{j+1} \leq \tau^{j+1}/2$ we resume

$$\begin{aligned} & \leq 2K_1^2 \left(1 + \|Y^j\|_{\mathbb{L}^2}\right)^2 \|\xi_{j+1}\|_{\mathbb{K}}^2 + \frac{1}{8} \|Y^{j+1} - Y^j\|_{\mathbb{L}^2}^2 \\ & + \frac{\tau^{j+1}}{8} \left\| \sqrt{\delta_K^{j+1}} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1} \right\|_{\mathbb{L}^2}^2 + \mathbb{I}_j. \end{aligned}$$

For the sixth term on the left-hand side of (3.2) we use integration by parts to get the bound

$$\tau^{j+1} |(\boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1}, Y^{j+1})_{\mathbb{L}^2}| = \frac{\tau^{j+1}}{2} |(\operatorname{div}(\boldsymbol{\beta}_{t_{j+1}}), |Y^{j+1}|^2)_{\mathbb{L}^2}| \leq C \frac{\tau^{j+1}}{2} \|Y^{j+1}\|_{\mathbb{L}^2}^2;$$

for the fourth term, we use again the assumption $\max_K \delta_K^{j+1} \leq \tau^{j+1}/2$ to conclude

$$|(Y^{j+1} - Y^j, \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1})_{\mathbb{L}^2}| \leq \frac{1}{4} \|Y^{j+1} - Y^j\|_{\mathbb{L}^2}^2 + \frac{\tau^{j+1}}{2} \left\| \sqrt{\delta_K^{j+1}} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1} \right\|_{\mathbb{L}^2}^2.$$

After absorbing terms, we sum over all iteration steps, take expectations, and use independence properties of increments $\{\xi_{j+1}\}_j$, the fact that $\mathbb{E}[\xi_{j+1}] = 0$, and the discrete Gronwall inequality to validate the assertion of the lemma. \square

3.2. Discretization of the non-linear SPDE (2.4). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. A stable discretization for (2.4) may be constructed as in [3, Section 2.2] for a related problem; see also [3, Chapter 2] for further details on the numerical analysis of the problem.

Scheme 3.3. Let $\mathbf{Y}^0 \in [\mathbb{V}_h^0]^3$ be given such that $|\mathbf{Y}^0(\mathbf{x}_\ell)| = 1$ for all $\ell \in \{1, \dots, L^0\}$, as well as $\{(\tau^{j+1}, \mathcal{T}_h^{j+1})\}_{j=0}^{J-1}$, and $\{\boldsymbol{\xi}_{j+1}\}_{j=0}^{J-1}$ be a \mathbf{Q} -random walk along \mathcal{I}_τ on $(\Omega, \mathcal{F}, \mathbb{P})$. For all $j \in \{0, 1, \dots, J-1\}$, determine a $[\mathbb{V}_h^{j+1}]^3$ -valued random variable \mathbf{Y}^{j+1} such that

\mathbb{P} -almost surely

$$\begin{aligned} & (\mathbf{Y}^{j+1} - \mathbf{Y}^j, \Psi)_{h;j+1} + \tau^{j+1} \left(\mathbf{Y}^{j+1/2} \times [\mathbf{Y}^{j+1/2} \times \Delta_h \mathbf{Y}^{j+1}], \Psi \right)_{h;j+1} \\ & = \iota(\mathbf{Y}^{j+1/2} \times \boldsymbol{\xi}_{j+1}, \Psi)_{h;j+1} \quad \forall \Psi \in [\mathbb{V}_h^{j+1}]^3. \end{aligned} \quad (3.3)$$

Here, $\mathbf{Y}^{j+1/2} := \frac{1}{2}(\mathbf{Y}^j + \mathbf{Y}^{j+1})$.

The existence of a solution \mathbf{Y}^{j+1} which satisfies (3.3) \mathbb{P} -almost surely may be shown with the help of Brouwer's fixed point theorem; see e.g. [16, p. 139]. The following stability properties for Scheme 3.3 may be adapted from [3, Theorem 2.11].

Theorem 3.4. *There exists a constant $C \equiv C(T, \text{Tr}(\mathbf{Q})) > 0$ such that*

- (1) $|\mathbf{Y}^{j+1}(\mathbf{x}_\ell)| = 1$ for all $\ell \in \{1, \dots, L^{j+1}\}$, and all $j \in \{0, 1, \dots, J-1\}$, \mathbb{P} -a.s.,
- (2) $\mathbb{E} \left[\sup_{0 \leq j \leq J-1} \|\nabla \mathbf{Y}^{j+1}\|_{\mathbf{L}^2}^2 + \sum_{j=0}^{J-1} \tau^{j+1} \|\mathbf{Y}^{j+1/2} \times \Delta_h \mathbf{Y}^{j+1}\|_{h;j+1}^2 \right] \leq C$.

A relevant property of iterates of (3.3) is that the length of initial profiles is preserved; cf. (1) in Theorem 3.4.

4. SPACE-TIME ADAPTIVITY BASED ON THE DISTANCE OF EMPIRICAL LAWS

We construct adaptive meshes in time ($\{\tau^{j+1}\}_j$) and space ($\{\mathcal{T}_h^{j+1}\}_j$) to compute iterates $\{\mathbf{Y}^{j+1}\}_j$ from Schemes 3.1 resp. 3.3. At time $t_{j+1} = \sum_{i=0}^j \tau^{i+1}$, we determine sequences of time step sizes $\{\tau_n^{j+1}\}_n$, of regular spatial meshes $\{\mathcal{T}_{h;n}^{j+1}\}_n$, and finite element spaces $\{\mathbb{V}_{h;n}^{j+1}\}_n$, indexed by $n \in \mathbb{N}_0$ until a threshold criterion is met. For this purpose, we propose

- the (BTC) strategy to build a data-dependent partition $\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1}$ (resp. $\widehat{\mathcal{P}}_{h;R_h;K;n}^{j+1}$) of the state space \mathbb{R}^{L^j} (resp. \mathbb{R}_0^+) into $R_\tau \equiv R_\tau(M_\tau)$ (resp. $R_h \equiv R_h(M_h)$) many cells, depending on the sample size M_τ (resp. M_h). This partition at time $t_j + \tau_n^{j+1}$ will allow a comparison of empirical probability measures $\{\widehat{\mu}_{\tau;n}^{s,j+1}\}_s$ (resp. $\{\widehat{\mu}_{h;K;n}^{s,j+1}\}_s$; $K \in \mathcal{T}_{h;n}^{j+1}$) to detect temporal (resp. spatial) changes.
- This temporal strategy uses different distances (Hellinger \mathbf{d}_H , Kullback-Leibler \mathbf{d}_{KL} , or total variation \mathbf{d}_{TV}) to quantify the change of related empirical measure pairs.
- a re-sampling strategy (bootstrap) is used for the estimator \mathbf{d}^* of this distance to prevent the ample computation of independent new samples.
- The basic sample size M_τ (resp. M_h), as well as the number of bootstrap replications B_τ (resp. B_h) are chosen, depending on the empirical variance of computed realizations.
- a space adaptive strategy is based on the ZZ-estimator, where the new regular spatial mesh $\mathcal{T}_{h;n+1}^{j+1}$ is obtained by refining/coarsening elements $K \in \mathcal{T}_{h;n}^{j+1}$ according to the distance of empirical probability measures of \mathbb{R}_0^+ -valued random variables $|\nabla Y_n^{j+1}|_K$ and $|G_h(\nabla Y_n^{j+1})|_K$ for a random variable $Y_n^{j+1} \in L^2(\Omega; \mathbb{V}_{h;n}^{j+1})$.

- Partitions $\widehat{\mathcal{P}}_{\tau; \widehat{R}_\tau^{j+1}; n}^{j+1}$ (resp. $\widehat{\mathcal{P}}_{h; \widehat{R}_{h;K}^{j+1}; K; n}^{j+1}$), where \widehat{R}_τ^{j+1} (resp. $\widehat{R}_{h;K}^{j+1}$) is chosen depending on the empirical variance of computed realizations are shown to perform equally well.

4.1. Partitioning. Let $Y \in L^2(\Omega; \mathbb{V}_h)$ be a random variable, with corresponding \mathbb{R}^L -valued coordinate map $\vec{Y} = \vec{Y}_{\mathbb{V}_h}(Y)$, where the finite element space \mathbb{V}_h is from Section 3, and $L := \dim \mathbb{V}_h$. We fix $R \equiv R(M)$, and choose $M \gg R$ for the sample $\mathcal{S}_{\vec{Y}} := \{\vec{Y}(\omega_k)\}_{k=1}^M$. We use the Binary Tree Cuboid (**BTC**) method as described in [11, 10] to generate a data-dependent partition of the state space \mathbb{R}^L :

$$\widehat{\mathcal{P}}_R \equiv \widehat{\mathcal{P}}_{M;R}(\mathcal{S}_{\vec{Y}}) = \{\widehat{C}_r\}_{r=1}^R, \quad \widehat{C}_r \cap \widehat{C}_s = \emptyset, \quad r \neq s, \quad \forall r, s \in \{1, \dots, R\},$$

such that all (closed) cells $\widehat{C}_r \subset \mathbb{R}^L$ are equally likely visited, i.e.,

$$\widehat{\mu}[\widehat{C}_r] \equiv \widehat{\mu}_{\mathcal{S}_{\vec{Y}}}[\widehat{C}_r] := \frac{\#\{k; \vec{Y}(\omega_k) \in \widehat{C}_r\}}{M} = \frac{\widehat{\nu}_r}{M} \quad \forall r \in \{1, \dots, R\},$$

where the frequency $\widehat{\nu}_r \in \mathbb{N}_0$ counts the number of realizations in $\mathcal{S}_{\vec{Y}}$ that lie in $\widehat{C}_r \in \widehat{\mathcal{P}}_R$, and $\sum_{r=1}^R \widehat{\nu}_r = M$. The partition $\widehat{\mathcal{P}}_R$ will be stored as a binary tree, see e.g. [10, Chapter 20]. Grouping the events uses geometry-based splittings of the index set of the spatial nodal points $\{\mathbf{x}_\ell\}_{\ell=1}^L$; see Figure 5 below.

Algorithm 4.1 (Binary Tree Cuboids (**BTC**)). *Choose $\kappa \in \mathbb{N}$. Let $\mathcal{S}_{0,1} := \mathcal{S}_{\vec{Y}}$.*

For $p = 1, \dots, \kappa + 1$ do:

For $q = 1, \dots, 2^{p-1}$ do:

- (I) *Define $\mathcal{S} := \mathcal{S}_{p-1,q}$, consisting of $\{\vec{Y}(\omega_k)\}_{k=1}^{2^{1-p} \cdot M}$.*
- (II) *Find the component $\ell \in \{1, \dots, L\}$ in the set of vectors \mathcal{S} which possesses the largest empirical standard deviation $\widehat{\sigma}_\ell \in \mathbb{R}$; denote this component by $\ell_{p,q} \in \{1, \dots, L\}$.*
- (III) *Compute the median $\mathbf{med}_{p,q} \in \mathbb{R}$ of $\{\vec{Y}_{\ell_{p,q}}(\omega_k)\}_{k=1}^{2^{1-p} \cdot M}$.*
- (IV) *Divide the (sub-)sample \mathcal{S} into two equal parts $\mathcal{S} = \mathcal{S}_{p,q} \cup \mathcal{S}_{p,2^{p-1}+q}$ according to the criterion $\vec{Y}_{\ell_{p,q}}(\omega_k) \leq \mathbf{med}_{p,q}$ ($k \in \{1, \dots, 2^{1-p}M\}$).*

Finding the entry $\ell_{p,q}$ is the most expensive part to create the binary tree; the computation of all empirical standard deviations at all spatial nodal points $\mathbf{x}_\ell \in D$, $\ell \in \{1, \dots, L\}$ may be done through $\mathcal{O}(LM)$ many operations in parallel to increase the efficiency; see Table 1. To determine the median $\mathbf{med}_{p,q} \in \mathbb{R}$ at entry $\ell_{p,q} \in \{1, \dots, L\}$ is accomplished by sorting all function values $\{\vec{Y}_{\ell_{p,q}}(\omega_k)\}_{k=1}^{2^{1-p} \cdot M}$, which may be done through $\mathcal{O}(M \log(M))$ many comparisons. Computational evidence supports that this division of (sub-)samples is stable against statistical outliers; cf. also [11]. Localizing the cell $\widehat{C}_r \in \widehat{\mathcal{P}}_R$ where a new realization of \vec{Y} lies in requires $\mathcal{O}(R)$ many checks. We remark that to uniformly partition a large subset of \mathbb{R}^L via hypercubes [17], or Voronoi meshes [11] is not favorable here due to the high dimensionality ($L \gg 1$), and the equi-probability of the cells is a relevant property of the (**BTC**) based partition.

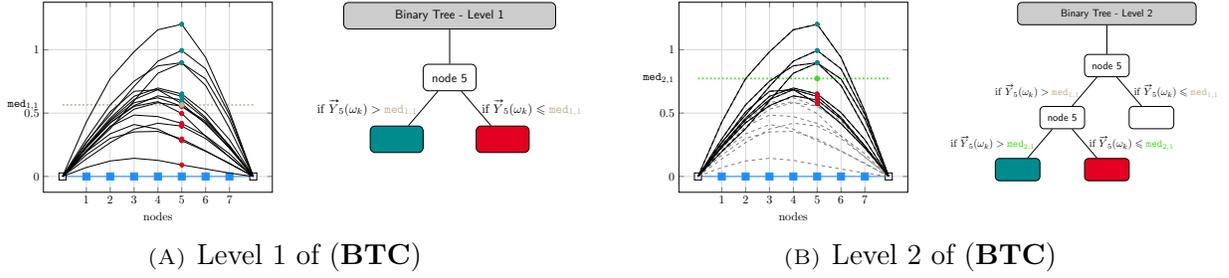


FIGURE 5. (A) First partition step to construct a **(BTC)** mesh: M many realizations of \vec{Y} (—) are divided according to their value at node \mathbf{x}_5 (■) and the median $\text{med}_{1,1}$ (⋯) into two subsets $\mathcal{S}_{1,1}$ (●) and $\mathcal{S}_{1,2}$ (●); see step (IV) in Algorithm 4.1 above. (B) Second partition step to construct a **(BTC)** mesh: the realizations in $\mathcal{S}_{1,1}$ (—) are divided according to their value at (here again) node \mathbf{x}_5 (■) and the median $\text{med}_{2,1}$ (⋯) into two subsets $\mathcal{S}_{2,1}$ (●) and $\mathcal{S}_{2,2}$ (●); from [11].

TABLE 1. Different number of cells R to build $\widehat{\mathcal{P}}_R$: the absolute simulation time (in seconds) in double precision arithmetic. Parallelized calculations are performed using OpenMP, cf. [7].

R	1024	2048	4096	8192
BTC	246s	468s	812s	1510s
(OpenMP) BTC	51s	61s	123s	201s

4.2. **The distance between two \mathbb{R}^L -valued samples.** Let $\mathcal{S}_{\vec{Y}^\mathfrak{s}}$ be an M -sample of realizations of $\vec{Y}^\mathfrak{s} \in L^2(\Omega; \mathbb{R}^L)$, where $\mathfrak{s} \in \{1, 2\}$. We want to compute the distance $d(\mathcal{L}(\vec{Y}^1), \mathcal{L}(\vec{Y}^2))$ of the related probability measures

$$\mu^\mathfrak{s} := \mathcal{L}(\vec{Y}^\mathfrak{s}) \quad (\mathfrak{s} \in \{1, 2\})$$

on $(\mathbb{R}^L, \mathcal{B}(\mathbb{R}^L))$, where $\mathcal{B}(\mathbb{R}^L)$ is the Borel σ -algebra on \mathbb{R}^L . For this purpose, we replace them by their empirical counterparts $\{\widehat{\mu}^\mathfrak{s}; \mathfrak{s} \in \{1, 2\}\}$ which are obtained from sampling on the underlying partition $\widehat{\mathcal{P}}_R$, and check whether

$$\mathbb{P}[d(\widehat{\mu}^1, \widehat{\mu}^2) \leq \text{To1}] > p, \quad (4.1)$$

for some given $\text{To1} > 0$, and $p \in [0, 1]$. The data-dependent partition $\widehat{\mathcal{P}}_R$ of \mathbb{R}^L uses the **(BTC)** method from Section 4.1,

$$\widehat{\mathcal{P}}_R \equiv \widehat{\mathcal{P}}_R(\mathcal{S}_{\vec{Y}^{1,2}}) = \{\widehat{C}_r\}_{r=1}^R \quad \text{where} \quad \mathcal{S}_{\vec{Y}^{1,2}} := \mathcal{S}_{\vec{Y}^1} \cup \mathcal{S}_{\vec{Y}^2}.$$

We then sample again — and refer below again to the new samples as $\mathcal{S}_{\vec{Y}^\mathfrak{s}}$ ($\mathfrak{s} \in \{1, 2\}$) —, to now compute frequency vectors $\widehat{\mathbf{v}}^\mathfrak{s} := (\widehat{v}_1^\mathfrak{s}, \dots, \widehat{v}_R^\mathfrak{s})^\top$, where $\widehat{v}_r^\mathfrak{s} := \#\widehat{A}_r^\mathfrak{s}$, $r \in \{1, \dots, R\}$,

with associated index set $\widehat{A}_r^{\mathfrak{s}} := \{k; \vec{Y}^{\mathfrak{s}}(\omega_k) \in \widehat{C}_r\}$, and thus obtain empirical probability measures $\widehat{\mu}^{\mathfrak{s}}$ via

$$\widehat{\mu}^{\mathfrak{s}} := \sum_{r=1}^R \widehat{q}_r^{\mathfrak{s}} \cdot \widehat{\xi}_r^{\mathfrak{s}}, \quad \text{where} \quad \widehat{q}_r^{\mathfrak{s}} := \frac{\widehat{\nu}_r^{\mathfrak{s}}}{M}, \quad \text{and} \quad \widehat{\xi}_r^{\mathfrak{s}} := \frac{1}{\widehat{\nu}_r^{\mathfrak{s}}} \sum_{k \in \widehat{A}_r^{\mathfrak{s}}} \delta_{\vec{Y}^{\mathfrak{s}}(\omega_k)} \quad (\mathfrak{s} \in \{\mathbf{1}, \mathbf{2}\}). \quad (4.2)$$

The Hellinger distance of the discrete measures $\{\widehat{\mu}^{\mathfrak{s}}; \mathfrak{s} \in \{\mathbf{1}, \mathbf{2}\}\}$ is

$$\mathbf{d}_H(\widehat{\mu}^{\mathbf{1}}, \widehat{\mu}^{\mathbf{2}}) = \left(\frac{1}{2} \sum_{r=1}^R \left(\sqrt{\widehat{q}_r^{\mathbf{1}}} - \sqrt{\widehat{q}_r^{\mathbf{2}}} \right)^2 \right)^{1/2}.$$

Other choices are the Kullback-Leibler distance resp. the total variation metric,

$$\mathbf{d}_{\text{KL}}(\widehat{\mu}^{\mathbf{1}}, \widehat{\mu}^{\mathbf{2}}) = \sum_{r=1}^R \widehat{q}_r^{\mathbf{1}} \log \left(\frac{\widehat{q}_r^{\mathbf{1}}}{\widehat{q}_r^{\mathbf{2}}} \right) \quad \text{resp.} \quad \mathbf{d}_{\text{TV}}(\widehat{\mu}^{\mathbf{1}}, \widehat{\mu}^{\mathbf{2}}) = \frac{1}{2} \sum_{r=1}^R |\widehat{q}_r^{\mathbf{1}} - \widehat{q}_r^{\mathbf{2}}|.$$

We refer to [15] for further details, and a comparison of these different distance functions. For comparison of the different distances, a normalization of \mathbf{d}_{KL} is achieved below via the non-linear transformation $1 - \exp(-\mathbf{d}_{\text{KL}})$ to take values in $[0, 1]$, and is denoted by $\widetilde{\mathbf{d}}_{\text{KL}}$.

Remark 4.2. 1) Likelihood-based inference for $\{\widehat{\mu}^{\mathfrak{s}}; \mathfrak{s} \in \{\mathbf{1}, \mathbf{2}\}\}$ via Fisher's non-parametric χ^2 -test of homogeneity uses the χ^2 -distance, see e.g. [26, Chapter 2],

$$\mathbf{d}_{\chi^2}^2(\widehat{\mu}^{\mathbf{1}}, \widehat{\mu}^{\mathbf{2}}) = \sum_{r=1}^R \frac{(\widehat{\nu}_r^{\mathbf{1}} - \widehat{\nu}_r^{\mathbf{2}})^2}{\widehat{\nu}_r^{\mathbf{1}} + \widehat{\nu}_r^{\mathbf{2}}} = M \sum_{r=1}^R \frac{(\widehat{q}_r^{\mathbf{1}} - \widehat{q}_r^{\mathbf{2}})^2}{\widehat{q}_r^{\mathbf{1}} + \widehat{q}_r^{\mathbf{2}}}. \quad (4.3)$$

Recall that if the null-hypothesis $H_0 : \mu^{\mathbf{1}} = \mu^{\mathbf{2}}$ is not rejected, there is no statistical evidence for $\mu^{\mathbf{1}} = \mu^{\mathbf{2}}$. As will be discussed in Remark 4.5, this fact limits the use of the χ^2 -test for time adaptivity.

2) Contrary to testing the null-hypothesis H_0 of homogeneity via $\mathbf{d}_{\chi^2}^2$ in (4.3) with the well-known asymptotic distribution χ_{R-1}^2 (cf. [18, p. 7]), required computations of probabilities $\mathbb{P}[\mathbf{d}(\widehat{\mu}^{\mathbf{1}}, \widehat{\mu}^{\mathbf{2}}) \geq \text{To1} \mid H_0]$ for other choices of \mathbf{d} must be approximated via MC simulations; here, the bootstrapping method (see Remark 4.5) is the relevant tool to drastically reduce the computational effort.

4.3. Bootstrapping. To sample empirical measures for (4.1) requires several independent M -samples $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ ($\mathfrak{s} \in \{\mathbf{1}, \mathbf{2}\}$), which is costly. An alternative way to obtain new independent M -samples is by a simple random draw with replacement from existent samples $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ via the bootstrap method [8]. A detailed algorithm is given in [28], where the 'Alias method' [32] is used as an improved alternative to inversion sampling. The effort to compute a sample via bootstrap is by far less expensive, and the computational studies in Section 5 confirm both, accuracy and speed-up. In the following, let quantities indexed by \star indicate bootstrap estimators for replications of $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$.

Let \mathbb{R}^L -valued M -samples $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}} := \{\vec{Y}^{\mathfrak{s}}(\omega_k)\}_{k=1}^M$ ($\mathfrak{s} \in \{\mathbf{1}, \mathbf{2}\}$) be given as in Section 4.2; we then generate $B \in \mathbb{N}$ many replications $\{\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}^{b,\star}\}_{b=1}^B$ ($\mathfrak{s} \in \{\mathbf{1}, \mathbf{2}\}$) of the existing sample $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$.

via [28, Algorithm 4.3.2], which are each of sample size M . Note that $\widehat{\mu}^{s,b,*} \equiv \widehat{\mu}_{\mathcal{S}_{\overline{Y}_n^s}}^{s,b,*}$ is a *single* realization of $\widehat{\mu}_{\mathcal{S}_{\overline{Y}_n^s}}^s$; according to [8, Theorem 29.1], we may now approximate the probability in (4.1) via

$$\mathbb{P}^*[\mathbf{d}(\widehat{\mu}^{1,*}, \widehat{\mu}^{2,*}) \leq \text{To1}] \approx \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{(-\infty, \text{To1}]}(\mathbf{d}(\widehat{\mu}^{1,b,*}, \widehat{\mu}^{2,b,*})). \quad (4.4)$$

4.4. Time adaptivity. Let t_j be fixed, $M_\tau \in \mathbb{N}$, $\mathbf{d} \in \{\mathbf{d}_H, \widetilde{\mathbf{d}}_{\text{KL}}, \mathbf{d}_{\text{TV}}\}$, and $\text{To1}_\tau > 0$. Below, we generate a finite sequence $\{\tau_n^{j+1}\}_{n \geq 0}$ with $\tau_0^{j+1} := \tau^j$ to (possibly) adaptively refine/coarsen τ_n^{j+1} . Therefore, for each $n \in \mathbb{N}_0$, we obtain empirical probability measures $\widehat{\mu}_{\tau;n}^{s,j+1}$ on \mathbb{R}^{L^j} , where $L^j := \dim \mathbb{V}_h^j$.

Let $\mathcal{S}_{\overline{Y}_n^{s,j+1}}$ be an M_τ -sample related to the random variable $Y_n^{s,j+1} \in L^2(\Omega; \mathbb{V}_h^j)$ with unknown law $\mu_{\tau;n}^{s,j+1}$ ($\mathbf{s} \in \{\mathbf{1}, \mathbf{2}\}$): the first random variable here is the solution of Scheme 3.1 with coarser time step size $\tau_n^{1,j+1} := \tau_n^{j+1}$, while the latter is obtained by extrapolation using the additional scale $\tau_n^{2,j+1} := \tau_n^{1,j+1}/2$; see e.g. [1], and Algorithm 4.3 below. We approximate the distance $\mathbf{d}(\mu_{\tau;n}^{1,j+1}, \mu_{\tau;n}^{2,j+1})$ with the help of the related empirical measures $\{\widehat{\mu}_{\tau;n}^{s,j+1}; \mathbf{s} \in \{\mathbf{1}, \mathbf{2}\}\}$ to then steer refinement or coarsening of $\tau_n^{j+1} > 0$. For this purpose, we sample on the underlying partition $\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1} := \bigcup_{r=1}^{R_\tau} \widehat{C}_{\tau;r;n}^{j+1}$ of \mathbb{R}^{L^j} which is obtained via Algorithm 4.1, i.e.,

$$\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1} \equiv \widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1}(\mathcal{S}_{\overline{Y}_n^{1,2;j+1}}) = \{\widehat{C}_{\tau;r;n}^{j+1}\}_{r=1}^{R_\tau},$$

where $\mathcal{S}_{\overline{Y}_n^{1,2;j+1}} := \mathcal{S}_{\overline{Y}_n^{1,j+1}} \cup \mathcal{S}_{\overline{Y}_n^{2,j+1}}$ and $\mathcal{S}_{\overline{Y}_n^{s,j+1}} := \{\overline{Y}_n^{s,j+1}(\omega_k)\}_{k=1}^{M_\tau}$. We then sample again — and refer below again to the new samples as $\mathcal{S}_{\overline{Y}_n^{s,j+1}}$ ($\mathbf{s} \in \{\mathbf{1}, \mathbf{2}\}$) —, to compute frequency vectors $\widehat{\mathbf{v}}_{\tau;n}^{s,j+1} := (\widehat{v}_{\tau;1;n}^{s,j+1}, \dots, \widehat{v}_{\tau;R_\tau;n}^{s,j+1})^\top$, where $\widehat{v}_{\tau;r;n}^{s,j+1} := \#\widehat{A}_{\tau;r;n}^{s,j+1}$, $r \in \{1, \dots, R_\tau\}$, with associated index set $\widehat{A}_{\tau;r;n}^{s,j+1} := \{k; \overline{Y}_n^{s,j+1}(\omega_k) \in \widehat{C}_{\tau;r;n}^{j+1}\}$, and thus obtain empirical probability measures $\widehat{\mu}_{\tau;n}^{s,j+1}$ on \mathbb{R}^{L^j} via

$$\widehat{\mu}_{\tau;n}^{s,j+1} = \sum_{r=1}^{R_\tau} \widehat{q}_{\tau;r;n}^{s,j+1} \cdot \widehat{\xi}_{\tau;r;n}^{s,j+1},$$

where

$$\widehat{q}_{\tau;r;n}^{s,j+1} := \frac{\widehat{v}_{\tau;r;n}^{s,j+1}}{M_\tau}, \quad \text{and} \quad \widehat{\xi}_{\tau;r;n}^{s,j+1} := \frac{1}{\widehat{v}_{\tau;r;n}^{s,j+1}} \sum_{k \in \widehat{A}_{\tau;r;n}^{s,j+1}} \delta_{\overline{Y}_n^{s,j+1}(\omega_k)} \quad (\mathbf{s} \in \{\mathbf{1}, \mathbf{2}\}).$$

We then generate $B_\tau \in \mathbb{N}$ many new bootstrap samples $\{\mathcal{S}_{\overline{Y}_n^{s,j+1}}^{b,*}\}_{b=1}^{B_\tau}$ ($\mathbf{s} \in \{\mathbf{1}, \mathbf{2}\}$) from $\mathcal{S}_{\overline{Y}_n^{s,j+1}}$ to eventually determine the distance of both via (4.4), and then choose the local time step size τ_n^{j+1} .

In order to reduce the empirical variance of these samples $\mathcal{S}_{\overline{Y}_n^{s,j+1}}$ ($\mathbf{s} \in \{\mathbf{1}, \mathbf{2}\}$), we interpolate the same Wiener process on $[t_j, t_j + \tau_n^{1,j+1}]$; see e.g. [12]. The following algorithm precises the generation of these two samples at time t_j .

Algorithm 4.3 (Richardson extrapolation). *Let $\tau_n^{j+1} > 0$, and $M_\tau \in \mathbb{N}$ be given. Initialize the samples $\mathcal{S}_{\overline{Y}_n^{1,j+1}} := \emptyset$, $\mathcal{S}_{\overline{Y}_n^{2,j+1}} := \emptyset$, as well as $\tau_n^{1,j+1} := \tau_n^{j+1}$ and $\tau_n^{2,j+1} := \tau_n^{1,j+1}/2$.*

For $k = 1, \dots, M_\tau$ do:

- (I) *Compute $\xi_{j+1}^{1,1;k} := \xi_{t_{j+1}}(\omega_k) - \xi_{t_j}(\omega_k)$ on $[t_j, t_j + \tau_n^{1,j+1}]$.*
- (II) *Compute the realization $Y_{n;k}^{1,j+1} \equiv Y_n^{1,j+1}(\omega_k)$ via one step of Scheme 3.1 with time step size $\tau_n^{1,j+1}$ and $\xi_{j+1}^{1,1;k}$. Set $\mathcal{S}_{\overline{Y}_n^{1,j+1}} := \mathcal{S}_{\overline{Y}_n^{1,j+1}} \cup \{\overline{Y}_{n;k}^{1,j+1}\}$.*
- (III) *Compute $\xi_{j+1}^{1,2;k} := \frac{1}{2}\xi_{j+1}^{1,1;k} + Z_{j,n}^{1,2;k}$, with $Z_{j,n}^{1,2;k} \equiv Z_{j,n}^{1,2}(\omega_k)$, for $Z_{j,n}^{1,2} \sim \mathcal{N}(0, t_j + \tau_n^{2,j+1})$.*
- (IV) *Compute $\xi_{j+1}^{2,2;k} := \frac{1}{2}\xi_{j+1}^{1,1;k} - Z_{j,n}^{2,2;k}$, with $Z_{j,n}^{2,2;k} \equiv Z_{j,n}^{2,2}(\omega_k)$, for $Z_{j,n}^{2,2} \sim \mathcal{N}(0, t_j + \tau_n^{2,j+1})$.*
- (V) *Compute the iterate $Y_{n;k}^{2,j+1} \equiv Y_n^{2,j+1}(\omega_k)$ via two steps of Scheme 3.1 with time step size $\tau_n^{2,j+1}$, and increments $\xi_{j+1}^{1,2;k}$, $\xi_{j+1}^{2,2;k}$. Define $Y_{n;k}^{2,j+1} := 2Y_{n;k}^{2,j+1} - Y_{n;k}^{1,j+1}$, and set $\mathcal{S}_{\overline{Y}_n^{2,j+1}} := \mathcal{S}_{\overline{Y}_n^{2,j+1}} \cup \{\overline{Y}_{n;k}^{2,j+1}\}$.*

We may now apply the tools from Section 4.2 to sample $d(\widehat{\mu}_{\tau;n}^{1,j+1}, \widehat{\mu}_{\tau;n}^{2,j+1})$ from (4.2). These steps are made precise in the following algorithm, and are sketched in Table 6.

Algorithm 4.4 (Adaptivity in time). *Fix $j \geq 0$, $\text{ToI}_\tau > 0$ and $M_\tau \in \mathbb{N}$. Choose a distance $d \in \{d_H, \widehat{d}_{\text{KL}}, d_{\text{TV}}\}$ and $B_\tau \in \mathbb{N}$. Set $\tau_0^{j+1} := \tau^j$.*

For $n = 0, 1, 2, \dots$ do:

- (I) *Compute M_τ -samples $\mathcal{S}_{\overline{Y}_n^{s,j+1}}$ ($s \in \{1, 2\}$) for τ_n^{j+1} via Algorithm 4.3.*
- (II) *Create $\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1} \equiv \widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1}(\mathcal{S}_{\overline{Y}_n^{1,2,j+1}})$ via Algorithm 4.1.*
- (III) *Compute new M_τ -samples $\mathcal{S}_{\overline{Y}_n^{s,j+1}}$ ($s \in \{1, 2\}$) for τ_n^{j+1} via Algorithm 4.3.*
- (IV) *Generate B_τ independent bootstrap M_τ -samples $\{\mathcal{S}_{\overline{Y}_n^{s,j+1}}^{b,*}\}_{b=1}^{B_\tau}$ ($s \in \{1, 2\}$) from $\mathcal{S}_{\overline{Y}_n^{s,j+1}}$ to obtain B_τ realizations $\{\widehat{\mu}_{\tau;n}^{s,j+1,b,*}\}_{b=1}^{B_\tau}$. Then, approximate involved probabilities via (4.4), and decide:

 - (1) *If $\mathbb{P}^* [d(\widehat{\mu}_{\tau;n}^{1,j+1,*}, \widehat{\mu}_{\tau;n}^{2,j+1,*}) \geq 2\text{ToI}_\tau] > 95\%$, set $\tau_{n+1}^{j+1} := \tau_n^{j+1}/2$.*
 - (2) *If $\mathbb{P}^* [d(\widehat{\mu}_{\tau;n}^{1,j+1,*}, \widehat{\mu}_{\tau;n}^{2,j+1,*}) < \frac{1}{2}\text{ToI}_\tau] > 95\%$, set $\tau_{n+1}^{j+1} := 2\tau_n^{j+1}$ and stop.**

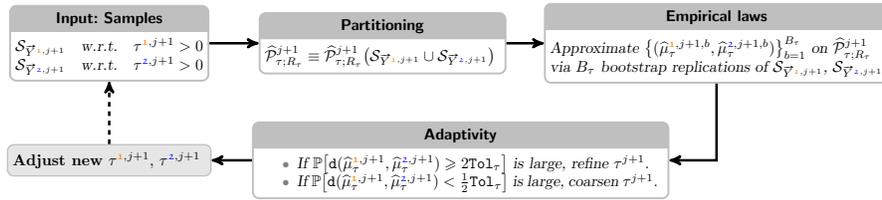


FIGURE 6. Construction of the local time step size τ^{j+1} .

Remark 4.5. The decision criteria in Step (IV) of Algorithm 4.4 involve the probability \mathbb{P}^* for the estimators $d(\widehat{\mu}_{\tau;n}^{1,j+1,*}, \widehat{\mu}_{\tau;n}^{2,j+1,*})$, for which the bootstrap method is used. Alternatively, a frequent sampling may be avoided by Fisher's χ^2 -test of homogeneity for which the asymptotic distribution χ_{R-1}^2 is well-known (see Remark 4.2, item 1)): for a fixed level

of significance $\alpha \in (0, 1)$, a local refinement of the time step size τ_n^{j+1} could be based on the rejection of the null-hypothesis $H_0 : \mathcal{L}(Y_{\tau;n}^{1,j+1}) = \mathcal{L}(Y_{\tau;n}^{2,j+1})$ by a *single* realization of $\mathbf{d}_{\chi^2}^2$ in (4.3), which is achieved by a modification of Step (IV) in Algorithm 4.4 to:

- Let $c_\alpha \approx \text{Inv-}\chi_{R-1}^2(1 - \alpha) \in \mathbb{R}_0^+$. If $\mathbf{d}_{\chi^2}^2(\widehat{\mu}_{\tau;n}^{1,j+1,*}, \widehat{\mu}_{\tau;n}^{2,j+1,*}) \geq c_\alpha$, set $\tau_n^{j+1} := \tau_n^{j+1}/2$.

Coarsening of the previously found τ_n^{j+1} (obtained by successive refinement) may be carried out by performing an additional *single* (i.e., $n = 0$) corresponding Fisher χ^2 -test of homogeneity with $\tau_0^{j+1} := 2\tau^{j+1}$. Unfortunately, the simulations in Figure 7(B) evidence an oscillatory behavior of $t_j \mapsto \tau^j$ for larger values of α in the case of Fisher's χ^2 -test. In fact, we see the following principle drawbacks that spontaneous refining/coarsening of τ_n^{j+1} via Fisher's χ^2 -test suffers from:

1) The computational studies show a dependence of the adaptive time step size $\{\tau^j\}_j$ on the sample size M_τ ; see Figure 7(A). In particular, Figure 7(A) illustrates smaller time step sizes for growing values of M_τ , which restricts the flexibility of the adaptive sampling algorithm, in particular for singular dynamics (see Examples 5.1, 5.2); see also Figure 13(A) as reference.

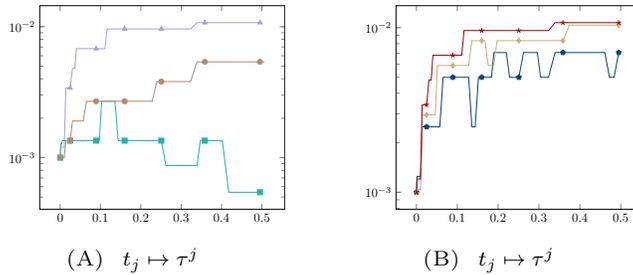


FIGURE 7. (Example 5.1 for the Algorithm in Remark 4.5, item 1) for fixed $\mathcal{T}_h^j \equiv \mathcal{T}_h^0$ with $h^0 = 2^{-6}$, $\tau^0 = 10^{-3}$, $n = 0$, and $R_\tau = 2^{12}$) (A) Behavior of $t_j \mapsto \tau^j$ for varying $M_\tau \in \{10^3$ (\blacktriangle), 10^4 (\bullet), 10^5 (\blacksquare) $\}$ and fixed $\alpha = 0.01$. (B) Adaptive time meshes for varying level of significance $\alpha \in \{0.10$ (\bullet), 0.05 (\blacklozenge), 0.01 (\star) $\}$ and fixed $M_\tau = 10^3$.

2) The error of second kind in this adaptive strategy is not controllable, which affects the stable selection of local time step sizes $\{\tau^j\}_j$. For example, in the case of refinement, this leads to unnecessarily small time step sizes.

As a consequence, we did not pursue this direction further.

4.5. Space adaptivity. Let $j \in \mathbb{N}_0$ be fixed, $M_h \in \mathbb{N}$ and $\tau^{j+1} > 0$ be chosen as detailed in Section 4.4. We generate a finite sequence $\{\mathcal{T}_{h;n}^{j+1}\}_{n \geq 0}$ resp. $\{\mathbb{V}_{h;n}^{j+1}\}_{n \geq 0}$, with $\mathcal{T}_{h;0}^{j+1} := \mathcal{T}_h^j$ resp. $\mathbb{V}_{h;0}^{j+1} := \mathbb{V}_h^j$ via the ZZ estimator by Zienkiewicz and Zhu [33] to (possibly) adaptively refine/coarsen each element $K \in \mathcal{T}_{h;n}^{j+1}$. Therefore, for every $n \in \mathbb{N}_0$, we obtain a family of parametrized empirical probability measures $\widehat{\mu}_{h;n}^{s,j+1} = \{\widehat{\mu}_{h;K;n}^{s,j+1}; K \in \mathcal{T}_{h;n}^{j+1}\}$ ($s \in \{1, 2\}$), each of which is supported on \mathbb{R}_0^+ , and whose construction is detailed next. For each

element $K \in \mathcal{T}_{h;n}^{j+1}$, we compare $|\nabla Y_n^{j+1}|_K$ with $|G_h(\nabla Y_n^{j+1})|_K$ to detect spatial changes in the solution, where the recovered gradient denoted by $G_h(\nabla Y_n^{j+1}) \in \mathbb{V}_{h;n}^{j+1}$ is computed via

$$G_h(\nabla Y_n^{j+1})(\mathbf{x}_{\ell;n}) := \frac{1}{|w_{\mathbf{x}_{\ell;n}}|} \int_{w_{\mathbf{x}_{\ell;n}}} \nabla Y_n^{j+1}(\mathbf{x}) \, d\mathbf{x}, \quad (4.5)$$

and $w_{\mathbf{x}_{\ell;n}} := \bigcup \{K \in \mathcal{T}_{h;n}^{j+1} : \mathbf{x}_{\ell;n} \in \overline{K}\}$ denotes the patch associated to $\mathbf{x}_{\ell;n} \in D$.

Let $\mathcal{S}_{Y_n^{j+1}}$ be an M_h -sample of realizations of $Y_n^{j+1} \in L^2(\Omega; \mathbb{V}_{h;n}^{j+1})$ computed with τ^{j+1} from Scheme 3.1. Our goal is now to measure the distance \mathbf{d} between the (unknown) probability measures $\mu_{h;K;n}^{1,j+1} := \mathcal{L}(|\nabla Y_n^{j+1}|_K)$ and $\mu_{h;K;n}^{2,j+1} := \mathcal{L}(|G_h(\nabla Y_n^{j+1})|_K)$ for each element $K \in \mathcal{T}_{h;n}^{j+1}$. For this purpose, we sample $\mathbf{d}(\hat{\mu}_{h;K;n}^{1,j+1}, \hat{\mu}_{h;K;n}^{2,j+1})$ of related empirical laws $\{\hat{\mu}_{h;K;n}^{\mathfrak{s},j+1}; \mathfrak{s} \in \{1, 2\}\}$ to steer refinement/coarsening of $K \in \mathcal{T}_{h;n}^{j+1}$. For its realization and a fixed $K \in \mathcal{T}_{h;n}^{j+1}$, we first provide independent M_h -samples

$$\mathcal{S}_{Y_n^{1,j+1}}^K := \{|\nabla Y_n^{j+1}(\omega_k)|_K\}_{k=1}^{M_h}, \quad \mathcal{S}_{Y_n^{2,j+1}}^K := \{|G_h(\nabla Y_n^{j+1}(\omega_k))|_K\}_{k=1}^{M_h} \quad (4.6)$$

to then define $\mathcal{S}_{Y_n^{1,2,j+1}}^K := \mathcal{S}_{Y_n^{1,j+1}}^K \cup \mathcal{S}_{Y_n^{2,j+1}}^K$; here, $\mathcal{S}_{Y_n^{1,j+1}}^K$ is easily obtained from the M_h -sample $\mathcal{S}_{Y_n^{j+1}}$ by restricting realizations to the given element $K \in \mathcal{T}_{h;n}^{j+1}$, while $\mathcal{S}_{Y_n^{2,j+1}}^K$ is obtained through a local averaging of associated realizations according to (4.5).

We may then generate the partition $\hat{\mathcal{P}}_{h;R_h;K;n}^{j+1} := \bigcup_{r=1}^{R_h} \hat{\mathcal{C}}_{h;r;K;n}^{j+1}$ of \mathbb{R}_0^+ via Algorithm 4.1 to prepare for the construction of $\{\hat{\mu}_{h;K;n}^{\mathfrak{s},j+1}; \mathfrak{s} \in \{1, 2\}\}$: we therefore sample again — and refer below again to the new samples as $\mathcal{S}_{Y_n^{\mathfrak{s},j+1}}^K$ ($\mathfrak{s} \in \{1, 2\}$) —, to compute frequency vectors $\hat{\mathcal{V}}_{h;K;n}^{\mathfrak{s},j+1} := (\hat{\mathcal{V}}_{h;1;K;n}^{\mathfrak{s},j+1}, \dots, \hat{\mathcal{V}}_{h;R_h;K;n}^{\mathfrak{s},j+1})^\top$ and associated index sets $\hat{A}_{h;r;K;n}^{\mathfrak{s},j+1}$ for $r \in \{1, \dots, R_h\}$, where

$$\begin{aligned} \hat{\mathcal{V}}_{h;r;K;n}^{1,j+1} &:= \#\hat{A}_{h;r;K;n}^{1,j+1}, & \hat{A}_{h;r;K;n}^{1,j+1} &:= \{k; |\nabla Y_n^{j+1}(\omega_k)|_K \in \hat{\mathcal{C}}_{h;r;K;n}^{j+1}\}, \\ \hat{\mathcal{V}}_{h;r;K;n}^{2,j+1} &:= \#\hat{A}_{h;r;K;n}^{2,j+1}, & \hat{A}_{h;r;K;n}^{2,j+1} &:= \{k; |G_h(\nabla Y_n^{j+1}(\omega_k))|_K \in \hat{\mathcal{C}}_{h;r;K;n}^{j+1}\}, \end{aligned}$$

and thus obtain empirical measures $\{\hat{\mu}_{h;K;n}^{\mathfrak{s},j+1}; \mathfrak{s} \in \{1, 2\}\}$ we were aiming for, where

$$\hat{\mu}_{h;K;n}^{\mathfrak{s},j+1} = \sum_{r=1}^{R_h} \hat{q}_{h;r;K;n}^{\mathfrak{s},j+1} \cdot \hat{\xi}_{h;r;K;n}^{\mathfrak{s},j+1}. \quad (4.7)$$

We refer to Figure 4(B) for an illustration, where $\hat{q}_{h;r;K;n}^{\mathfrak{s},j+1} := \hat{\mathcal{V}}_{h;r;K;n}^{\mathfrak{s},j+1}/M_h$, and

$$\hat{\xi}_{h;r;K;n}^{1,j+1} := \frac{1}{\hat{\mathcal{V}}_{h;r;K;n}^{1,j+1}} \sum_{k \in \hat{A}_{h;r;K;n}^{1,j+1}} \delta_{|\nabla Y_n^{j+1}(\omega_k)|_K}, \quad \hat{\xi}_{h;r;K;n}^{2,j+1} := \frac{1}{\hat{\mathcal{V}}_{h;r;K;n}^{2,j+1}} \sum_{k \in \hat{A}_{h;r;K;n}^{2,j+1}} \delta_{|G_h(\nabla Y_n^{j+1}(\omega_k))|_K}.$$

Hence, we arrive at two families of empirical measures $\{\hat{\mu}_{h;K;n}^{\mathfrak{s},j+1}; K \in \mathcal{T}_{h;n}^j\}$ ($\mathfrak{s} \in \{1, 2\}$) for every $n \in \mathbb{N}_0$. In the algorithm to follow, we involve the difference of first absolute moments of samples

$$\hat{\eta}_{h;K;n}^{1,j+1} := \mathbb{E}[|\nabla Y_n^{j+1}|_K], \quad \hat{\eta}_{h;K;n}^{2,j+1} := \mathbb{E}[|G_h(\nabla Y_n^{j+1})|_K] \quad (4.8)$$

to stabilize the refinement/coarsening procedure of a fixed element $K \in \mathcal{T}_{h;n}^j$ w.r.t. $n \in \mathbb{N}_0$. As for time adaptivity, statistical inferences about quantities are again obtained via bootstrapping.

Algorithm 4.6 (Adaptivity in space). *Fix $j \geq 0$, $\text{To1}_h > 0$, $M_h \in \mathbb{N}$. Choose a distance $\mathbf{d} \in \{\mathbf{d}_H, \tilde{\mathbf{d}}_{\text{KL}}, \mathbf{d}_{\text{TV}}\}$, $B_h \in \mathbb{N}$. Set $\mathcal{T}_{h;0}^{j+1} := \mathcal{T}_h^j$, as well as $\mathbb{V}_{h;0}^{j+1} := \mathbb{V}_h^j$. For $n = 0, 1, 2, \dots$ do:*

- (I) Compute a M_h -sample $\mathcal{S}_{Y_n^{j+1}} := \{Y_n^{j+1}(\omega_k)\}_{k=1}^{M_h}$ for τ^{j+1} with new $\{\xi_{j+1}(\omega_k)\}_{k=1}^{M_h}$.
- (II) For each $K \in \mathcal{T}_{h;n}^{j+1}$ do:
 - (1) Create $\widehat{\mathcal{P}}_{h;R_h;K;n}^{j+1}$ based on $\mathcal{S}_{Y_n^{1,2;j+1}}^K$ in (4.6) via Algorithm 4.1.
 - (2) Generate B_h new bootstrap samples $\{\mathcal{S}_{Y_n^{s,j+1}}^{K,b,*}\}_{b=1}^{B_h}$ ($\mathbf{s} \in \{1, 2\}$) from $\mathcal{S}_{Y_n^{s,j+1}}^K$ to approximate the first moments $\{\widehat{\eta}_{h;K;n}^{s,j+1,*}\}_{b=1}^{B_h}$ in (4.8).
 - (3) Generate B_h new bootstrap samples $\{\mathcal{S}_{Y_n^{s,j+1}}^{K,b,*}\}_{b=1}^{B_h}$ ($\mathbf{s} \in \{1, 2\}$) from $\mathcal{S}_{Y_n^{s,j+1}}^K$ to obtain B_h measures $\{\widehat{\mu}_{h;K;n}^{s,j+1,*}\}_{b=1}^{B_h}$. Then, approximate involved probabilities via (4.4), and decide:
 - (a) If $h_K^{j+1} |\widehat{\eta}_{h;K;n}^{1,j+1,*} - \widehat{\eta}_{h;K;n}^{2,j+1,*}| \mathbb{P}^*[\mathbf{d}(\widehat{\mu}_{h;K;n}^{1,j+1,*}, \widehat{\mu}_{h;K;n}^{2,j+1,*}) > 2\text{To1}_h] > 95\%$, mark K for refinement.
 - (b) If $h_K^{j+1} |\widehat{\eta}_{h;K;n}^{1,j+1,*} - \widehat{\eta}_{h;K;n}^{2,j+1,*}| \mathbb{P}^*[\mathbf{d}(\widehat{\mu}_{h;K;n}^{1,j+1,*}, \widehat{\mu}_{h;K;n}^{2,j+1,*}) < \frac{1}{2}\text{To1}_h] > 95\%$, mark K for coarsening.
- (III) If $\max_{K \in \mathcal{T}_{h;n}^{j+1}} h_K^{j+1} |\widehat{\eta}_{h;K;n}^{1,j+1,*} - \widehat{\eta}_{h;K;n}^{2,j+1,*}| \mathbb{P}^*[\mathbf{d}(\widehat{\mu}_{h;K;n}^{1,j+1,*}, \widehat{\mu}_{h;K;n}^{2,j+1,*}) \leq \text{To1}_h] > 95\%$, set $\mathcal{T}_h^{j+1} := \mathcal{T}_{h;n}^{j+1}$, $\mathbb{V}_h^{j+1} := \mathbb{V}_{h;n}^{j+1}$ and stop; otherwise continue.
- (IV) Obtain the new mesh $\mathcal{T}_{h;n+1}^{j+1}$ from $\mathcal{T}_{h;n}^{j+1}$ by local refinement resp. coarsening of the elements $K \in \mathcal{T}_{h;n}^{j+1}$ marked in the previous steps.

For each element $K \in \mathcal{T}_h^j$, the first moments in Algorithm 4.6 are approximated via (analogously for $\widehat{\eta}_{h;K;n}^{2,j+1,*}$ in (4.8))

$$\widehat{\eta}_{h;K;n}^{1,j+1,*} := \mathbb{E}^* [|\nabla Y_n^{j+1}|_K] \approx \frac{1}{B_h} \sum_{b=1}^{B_h} |\nabla Y_n^{j+1}(\omega_b)|_K. \quad (4.9)$$

To summarize, the overall space-time adaptivity strategy in the $(j+1)$ -th step consists of the following substeps:

- Time adaptivity according to Algorithm 4.4 to find τ^{j+1} .
- Space adaptivity according to Algorithm 4.6 to find \mathcal{T}_h^{j+1} resp. \mathbb{V}_h^{j+1} based on τ^{j+1} .

5. COMPUTATIONAL EXPERIMENTS

We computationally study stability and accuracy of the adaptive algorithms from Sections 4.4–4.5 with respect to the parameters (M_τ, B_τ, R_τ) (analogously for M_h, B_h, R_h in space), as well as the distances $\mathbf{d} \in \{\mathbf{d}_H, \tilde{\mathbf{d}}_{\text{KL}}, \mathbf{d}_{\text{TV}}\}$. For this purpose, we employ random

number generators from the GNU Scientific Library [13]. Local mesh refinement and coarsening of $\mathcal{T}_{h;n}^{j+1}$ in Algorithm 4.6 is performed using a bisection algorithm, and are based on the finite element code ALBERTA, cf. [29]. All computations are performed on an Intel Core i5-4670 3.40GHz processor with 16GB RAM in double precision arithmetic. Parallelized calculations are performed using OpenMP, cf. [7]. Arising linear algebraic systems are solved by Gaussian elimination, and the Software package [9].

The computational studies for Example 5.1 resp. Example 5.2 evidence that

- a robust adaptive time mesh may be obtained for all distances $\mathbf{d} \in \{\mathbf{d}_H, \tilde{\mathbf{d}}_{KL}, \mathbf{d}_{TV}\}$ in Algorithm 4.3 via partitions $\{\hat{\mathcal{P}}_{\tau;R_\tau^j}^j\}$ on which the empirical probability measures $\{\hat{\mu}_\tau^{\mathbf{s};j}; \mathbf{s} \in \{1, 2\}\}$ may be compared with the help of the bootstrap method.
- an efficient adaptive space mesh may be obtained for $\mathbf{d} = \mathbf{d}_{TV}$ in Algorithm 4.6, while $\mathbf{d} \in \{\mathbf{d}_H, \tilde{\mathbf{d}}_{KL}\}$ leads to unnecessarily fine meshes; cf. Table 2.
- fast computations of estimators as \mathbb{P}^* , and \mathbb{E}^* require bootstrapping.
- next to much smaller errors on adaptive space-time meshes, the empirical variance of the iterates $\{Y^j\}_j$ is significantly reduced, leading to reduced samples sizes M_τ^j (resp. M_h^j) and less cells R_τ^j (resp. R_h^j) for a coarser partition $\hat{\mathcal{P}}_{\tau;R_\tau^j}^j$ (resp. $\hat{\mathcal{P}}_{h;R_h^j;K}^j$).
- adaptivity of all involved discretization and statistical parameters is necessary to accurately resolve singular behaviors of the solution, as e.g. present in Example 5.2.

5.1. Computational experiments for an SPDE (2.3). We use Scheme 3.1 in combination with space-time adaptivity to approximate the solution of the convection-dominated linear SPDE (2.3).

Example 5.1. Let $D = (0, 1)^2$, $\boldsymbol{\beta} \equiv \boldsymbol{\beta}(x_1, x_2) = (\frac{1}{2} - x_2, x_1 - \frac{1}{2})^\top$, and $\iota \in \{0.1, 0.3, 0.5\}$. Consider ($t \in (0, 2\pi]$)

$$dX_t - (\varepsilon \Delta X_t - \theta \boldsymbol{\beta} \cdot \nabla X_t) dt = \iota(1 + |X_t|) \sum_{0 \leq |\mathbf{k}| \leq 3} \sigma_{\mathbf{k}} dW^{\mathbf{k}}(t), \quad X_0 = x_0, \quad (5.1)$$

with $X_t = 0$ on ∂D , $\varepsilon = 10^{-8}$, $\theta = 1.0$ (if not specified otherwise), and

$$\sigma_{\mathbf{k}}(\mathbf{x}) = \sqrt{2} \prod_{i=1}^2 \left(\frac{2}{(2k_i + 1)\pi} \right)^2 \sin(k_i \pi x_i)$$

for all $\mathbf{x} = (x_1, x_2)^\top \in D$, and multi-indices $\mathbf{k} = (k_1, k_2)^\top \in \mathbb{N}_0^2$ with $|\mathbf{k}| = k_1 + k_2$. The initial datum x_0 is given by the slotted cylinder; see [21] for an explicit formula of it.

We compute the reference solution \tilde{X} via Scheme 3.1 with mesh sizes $h^j \equiv h^0 = 2^{-10}$, $\tau^j \equiv \tau^0 = 10^{-5}$, and $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ according to Lemma 3.2 on uniform meshes, and $M_\tau = 10^5$.

Our first series of experiments for Example 5.1 serves to clarify the stability of computed empirical measures $\{\hat{\mu}_\tau^{\mathbf{s};j+1}\}_\mathbf{s}$ with respect to the used number of cells R_τ to resolve each state space $\{\mathbb{R}^{L^{j+1}}\}_j$ accurately. Next to this, we study reliability of adaptive space-time meshes with respect to the used distances \mathbf{d} (see Section 4.2) of empirical probability measures. In order to do this, we consider the following parameter constellations:

- Section 5.1.1: Adaptive $\{\tau^j\}_j$, uniform $\mathcal{T}_h^j \equiv \mathcal{T}_h^0$, and M_τ, R_τ, B_τ are fixed.
- Section 5.1.2: Adaptive $\{(\tau^j, \mathcal{T}_h^j)\}_j$, and M_τ, R_τ, B_τ are fixed.
- Section 5.1.3: Adaptive $\{(\tau^j, \mathcal{T}_h^j, M_\tau^j, R_\tau^j, B_\tau^j)\}_j$.

5.1.1. *Time adaptivity.* Figure 8 illustrates the stability properties of Scheme 3.1 for a uniform space-time mesh: the standard Galerkin method (i.e., $\delta_K^j \equiv 0$) yields highly oscillatory expectations of iterates to approximate (5.1); this (global Gibbs) phenomenon is well-known for convection-dominated PDEs, and here is amplified by the noise; see Figure 8(A). Corresponding simulations of higher moments show a significant reduction of spurious oscillations outside the diffuse layers. The simulations in Figure 8(B) via the SUPG based Scheme 3.1 show improved stability properties without oscillatory patterns attached to the cylindrical profile. The stabilization parameter $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ optimally balances stabilization with accuracy demands; see Lemma 3.2, and Figure 9(A)–9(B). In addition, this choice of δ_K^j leads to the smallest empirical variances of the computed

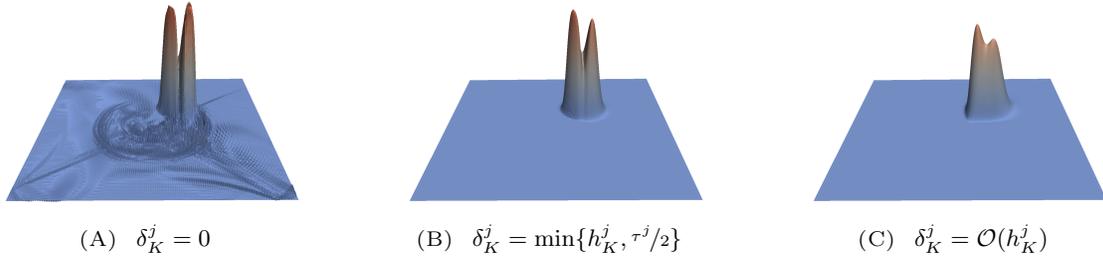


FIGURE 8. (Example 5.1 for $T = \pi/2$, and $\text{To1}_\tau = 0.05$, $M_\tau = 10^5$) $\mathbb{E}_{M_\tau} [|Y^j|]$ obtained with the Scheme 3.1 for uniform space-time meshes $\{(\tau^j, \mathcal{T}_h^j)\}_j$ with $h^j \equiv h^0 = 2^{-8}$, $\tau^j \equiv \tau^0 = 10^{-4}$, and $\iota = 0.1$: (A) Standard Galerkin FEM ($\delta_K^j = 0$), (B) $\delta_K^j = \min\{h_K^j, \tau^j/2\}$, and (C) $\delta_K^j = \mathcal{O}(h_K^j)$.

iterates $\{Y^j\}_j$; see Figure 9(C). Small-scale effects which are initiated by the driving Wiener process are not accurately recovered for $\delta_K^j = \mathcal{O}(h_K^j)$ which heavily diffuses the solution structure; see Figure 8(C). Thus, we choose the SUPG scheme with $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ for all simulations to follow.

Next, we study the dependence on the parameters (M_τ, R_τ) to obtain a stable, time-adaptive mesh for different distances $\mathbf{d} \in \{\mathbf{d}_H, \tilde{\mathbf{d}}_{\text{KL}}, \mathbf{d}_{\text{TV}}\}$ (see Section 4.2) of involved empirical probability measures. Consider

$$\mathcal{R}_\tau := \arg \min_{R_\tau} \left[\max_{t \in [0, T]} |\tau^{R_\tau}(t) - \tau^{2R_\tau}(t)| \leq \text{To1} \right] \quad (5.2)$$

with $\text{To1} := 10^{-2} \min_{t \in [0, T]} |\tau(t)|$, where $\tau(t) \equiv \tau^{R_\tau}(t)$ is the piecewise affine interpolation of the sequence $\{(t_j, \tau^j)\}_j$. Criterion (5.2) identifies the minimum value $\mathcal{R}_\tau \geq 1$ where the adaptive time mesh is not sensitive any more to refinement of the state space \mathbb{R}^{L^j} . The results in Figure 10(B) show that at least $\mathcal{R}_\tau \geq 2 \cdot 10^3$ cells are necessary to partition \mathbb{R}^{L^j}

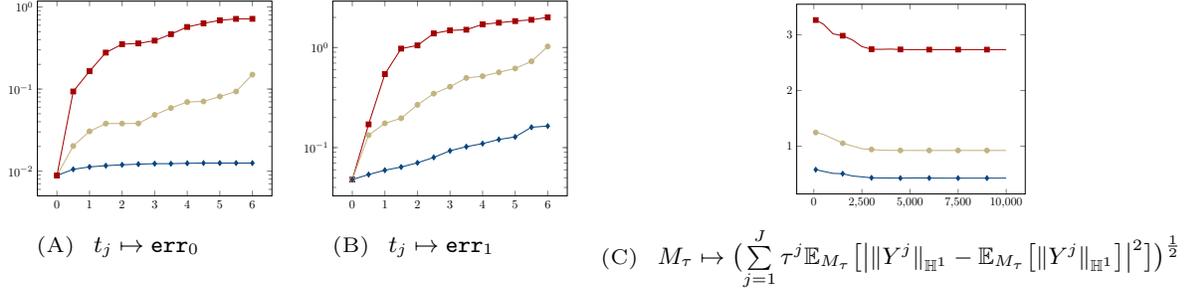


FIGURE 9. (Example 5.1 with reference space-time meshes $\{(\tau^j, \mathcal{T}_h^j)\}_j$ for $h^j \equiv h^0 = 2^{-10}$, $\tau^j \equiv \tau^0 = 10^{-5}$. Time adaptivity for $T = 2\pi$, $\iota = 0.3$, as well as $\mathbf{d} = \mathbf{d}_H$, $\text{ToI}_\tau = 0.05$, and $R_\tau = 2^{12}$, $M_\tau = 10^5$) Behavior of $t_j \mapsto \mathbf{err}_k := |\mathbb{E}_{M_\tau} [\|\tilde{X}_{t_j}\|_{\mathbb{W}^{k,\infty}} - \|Y^j\|_{\mathbb{W}^{k,\infty}}]|$ for $(k = 0)$ (A), $(k = 1)$ (B), and empirical variance (C) for the standard Galerkin FEM ($\delta_K^j = 0$) (\blacksquare), $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ (\blacklozenge), and $\delta_K^j = \mathcal{O}(h_K^j)$ (\bullet).

with constant $L^j \equiv L^0$ in the case of Example 5.1. This observation is common for all choices of distances \mathbf{d} ; see Figure 10(A).

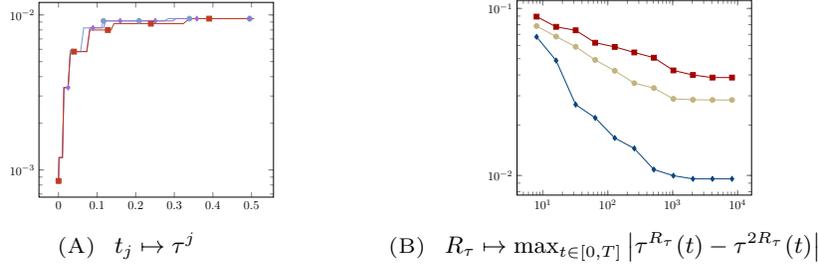


FIGURE 10. (Example 5.1 with uniform $\mathcal{T}_h^j \equiv \mathcal{T}_h^0$ for $h^j \equiv h^0 = 2^{-6}$. Time adaptivity for $T = 1/2$, $\iota = 0.3$, as well as $\tau^0 = 10^{-3}$, $\text{ToI}_\tau = 0.05$ and $R_\tau = 2^{12}$, $M_\tau = 10^5$) (A) Adaptive time meshes for $\mathbf{d} = \mathbf{d}_H$ (\bullet), $\mathbf{d} = \mathbf{d}_{TV}$ (\blacklozenge), $\mathbf{d} = \tilde{\mathbf{d}}_{KL}$ (\blacksquare), (B) and behavior of $R_\tau \mapsto \max_{t \in [0, T]} |\tau^{R_\tau}(t) - \tau^{2R_\tau}(t)|$ for $\iota = 0.1$ (\blacklozenge), $\iota = 0.3$ (\bullet), and $\iota = 0.5$ (\blacksquare) with respect to $\mathbf{d} = \mathbf{d}_H$.

5.1.2. *Space adaptivity.* So far, time adaptivity was performed for a fixed uniform spatial mesh. The results in Figure 11 display the required number of cells \mathcal{R}_τ to meet (5.2) for varying convection resp. noise intensity θ resp. ι in Example 5.1, both, for uniform and adaptive space-time meshes, and dimensions $d \in \{1, 2\}$. The studies indicate that the number of required cells \mathcal{R}_τ to partition each \mathbb{R}^{L^j} is significantly reduced in the case of space-time adaptivity: here, the meshes \mathcal{T}_h^j resolve regions in space where large gradients are likely to occur; see also Figure 4(A). Next to an increased spatial resolution, we observe

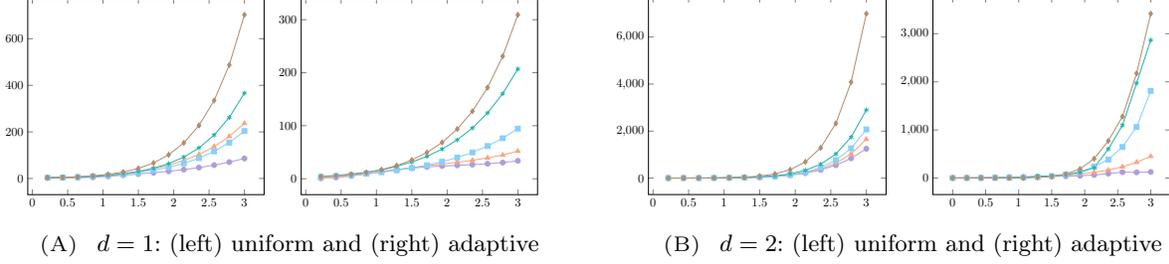


FIGURE 11. (Example 5.1 for $T = 1.0$, as well as $\mathbf{d} = \mathbf{d}_H$, $\text{ToI}_\tau = \text{ToI}_h = 0.05$, $h^0 = 2^{-5}$, $\tau^0 = 10^{-3}$, and $M_\tau = 10^5$) Number of required cells \mathcal{R}_τ to meet (5.2) for varying convection resp. noise intensity (θ, ι) with $\iota \in [0, 3]$ (abscissa) and $\theta = 1.0$ ($\text{---}\bullet\text{---}$), $\theta = 2.0$ ($\text{---}\blacktriangle\text{---}$), $\theta = 3.0$ ($\text{---}\blacksquare\text{---}$), $\theta = 4.0$ ($\text{---}\blacklozenge\text{---}$) and $\theta = 5.0$ ($\text{---}\blacklozenge\text{---}$) on uniform and space-time adaptive meshes for $d = 1$ (A) and $d = 2$ (B).

smaller empirical variances

$$\text{Var}_{M_\tau} [\|Y^j\|_{\mathbb{W}^{k,2}}] := \mathbb{E}_{M_\tau} \left[\left| \|Y^j\|_{\mathbb{W}^{k,2}} - \mathbb{E}_{M_\tau} [\|Y^j\|_{\mathbb{W}^{k,2}}] \right|^2 \right] \quad (k \in \{0, 1\}) \quad (5.3)$$

of computed realizations $\{Y^j\}_j$ (see Figure 12(A)–(C)) such that coupled space-time adaptivity can also be regarded as an importance sampling strategy, where corresponding solutions with multiple scales imprinted by the adaptive spatial mesh are chosen. As a consequence, smaller empirical variances of Y^j imply smaller confidence intervals and smaller image ranges $\vec{Y}^j[\Omega]$, such that smaller sample sizes M_τ (resp. M_h) and less cells R_τ (resp. R_h) are necessary to build an appropriate partition $\widehat{\mathcal{P}}_{\tau;R_\tau}^j$ (resp. $\widehat{\mathcal{P}}_{h;R_h;K}^j$).

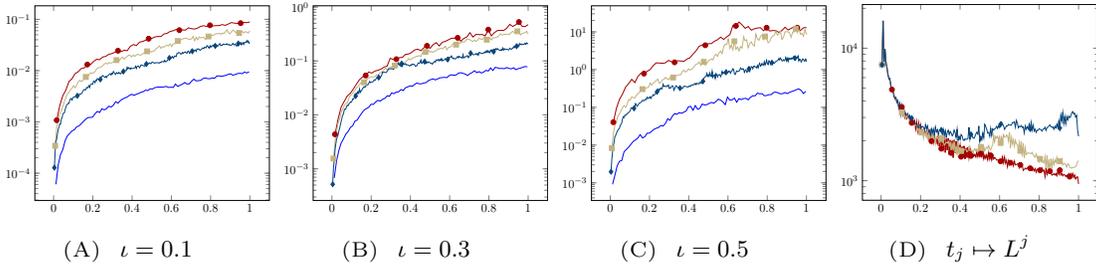


FIGURE 12. (Example 5.1 for $T = 1.0$, as well as $\mathbf{d} = \mathbf{d}_H$, $\text{ToI}_\tau = \text{ToI}_h = 0.05$, $h^0 = 2^{-5}$, $\tau^0 = 10^{-3}$, and $M_\tau = 10^5$, $R_\tau = 10^3$) (A)–(C) Behavior of the empirical variance in (5.3) with $k = 0$ on uniform meshes of sizes $(2^{-i}h^0, 2^{-i}\tau^0)$ with $i = 0$ ($\text{---}\bullet\text{---}$), $i = 1$ ($\text{---}\blacksquare\text{---}$) and $i = 2$ ($\text{---}\blacklozenge\text{---}$), and for adaptive (---) space-time meshes. (D) Behavior of $t_j \mapsto L^j$ for noise intensity $\iota = 0.1$ ($\text{---}\bullet\text{---}$), $\iota = 0.3$ ($\text{---}\blacksquare\text{---}$), and $\iota = 0.5$ ($\text{---}\blacklozenge\text{---}$).

The evolution of $t_j \mapsto L^j$ for different noise intensities ι is plotted in Figure 12(D). Related weak errors in Figure 13 are much smaller if compared to a uniform discretization,

or mere time adaptivity. Moreover, we found that a significantly smaller number of time steps ($J = 83$) is needed vs. the uniform grid to meet a given threshold criterion for the error $\max_j |\mathbb{E}_{M_\tau} [\|\tilde{X}_{t_j}\|_{\mathbb{W}^{k,2}} - \|Y^j\|_{\mathbb{W}^{k,2}}]|$ for $k \in \{0, 1\}$.

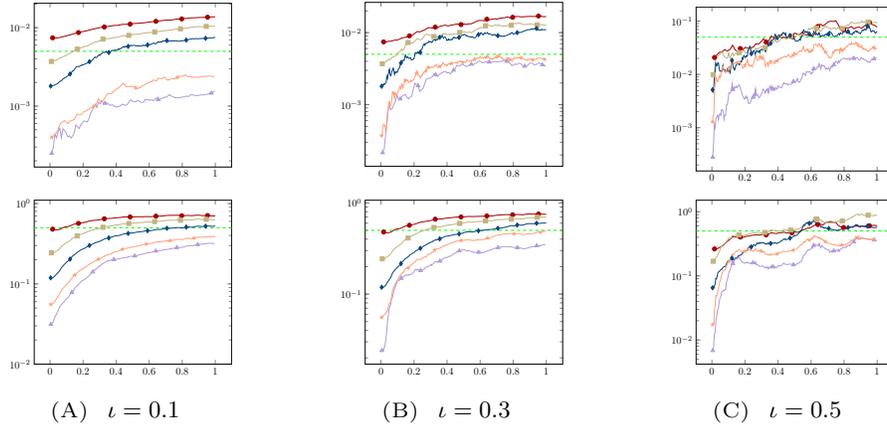


FIGURE 13. (Example 5.1 for $\mathbf{d} = \mathbf{d}_H$, $\text{Tol}_\tau = \text{Tol}_h = 0.05$, $h^0 = 2^{-5}$, $\tau^0 = 10^{-3}$, and $M_\tau = 10^5$, $M_h = 10^3$, $R_\tau = 10^3$, $R_h = 2^6$) Behavior of $t_j \mapsto |\mathbb{E}_{M_\tau} [\|\tilde{X}_{t_j}\|_{\mathbb{W}^{k,2}} - \|Y^j\|_{\mathbb{W}^{k,2}}]|$ for $t_j \in [0, 1]$ on uniform time meshes with $J = 2^8$ (—●—), $J = 2^9$ (—■—) and $J = 2^{10}$ (—◆—) required time steps for $k = 0$ (upper row) and $k = 1$ (lower row) for a given threshold (---). Corresponding space-time adaptive meshes for uniform (—▲—) vs. adaptive (—★—) statistical parameters (M_τ, B_τ, R_τ); see Section 5.1.3 for the latter.

So far, the distance $\mathbf{d} = \mathbf{d}_H$ was used for both, time and space adaptivity. Computational experiments give similar adaptive time meshes for all distances $\mathbf{d} \in \{\mathbf{d}_H, \tilde{\mathbf{d}}_{\text{KL}}, \mathbf{d}_{\text{TV}}\}$ in Algorithm 4.4. However, its choice is crucial for the efficiency of space adaptivity in Algorithm 4.6 (cf. Figure 4(A) and Figure 14(A)): we observe that the resolution of the arising diffuse layer of the solution in Example 5.1, and its width depend on the selected distance \mathbf{d} . According to [15], $\mathbf{d}_{\text{TV}} \leq \mathbf{d}_H \leq \sqrt{\mathbf{d}_{\text{KL}}}$, and the algorithm produces indeed more elements $K \in \mathcal{T}_h^j$ in the case of $\tilde{\mathbf{d}}_{\text{KL}}$ (see Section 4.2) to resolve diffusive layers, if e.g. compared to meshes obtained via \mathbf{d}_{TV} ; see Figure 14(A) and Table 2A. This ordering for used distances is also reflected by the related histograms, see Figure 14(B), since e.g. small values of the realization of \mathbf{d} favor coarsening of the corresponding $K \in \mathcal{T}_h^j$. These results motivate the choice \mathbf{d}_{TV} to properly balance costs and accuracy; cf. also Figures 4 and 14.

The simulations that we discussed so far use a fixed number of cells $R_\tau \equiv R_\tau^j$, as well as M_τ, B_τ . In the following, we discuss the relevancy to adaptively select the statistical parameters M_τ, B_τ , and R_τ (analogously in space) as well to further increase the efficiency of the numerical scheme.

5.1.3. *Adaptive choice of statistical parameters M_τ, B_τ, R_τ .* We study the time-dependent choice of M_τ, R_τ , and B_τ , with analogous results for M_h, R_h , and B_h for each $K \in \mathcal{T}_h^j$.

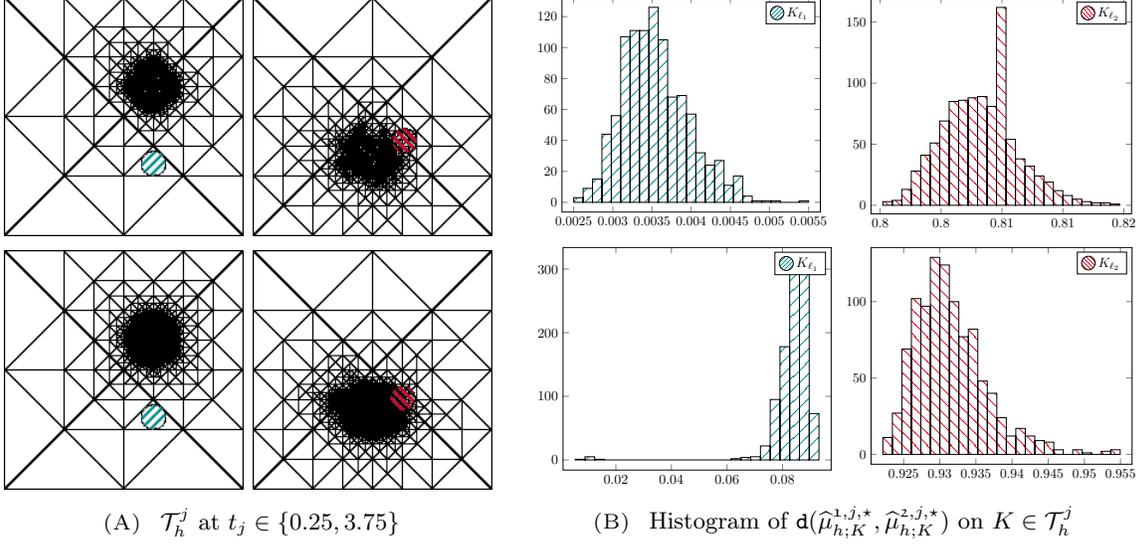


FIGURE 14. (Example 5.1 for $\iota = 0.3$, as well as $\text{To1}_h = 0.05$ and $R_h = 2^5$, $M_h = 10^3$ for each $K \in \mathcal{T}_h^j$): (A) Snapshots of different spatial meshes \mathcal{T}_h^j at evaluated times $t_j \in [0, 2\pi]$ with (B) corresponding histograms of $d(\widehat{\mu}_{h;K}^{1,j,*}, \widehat{\mu}_{h;K}^{2,j,*})$ at selected elements $K_{\ell_1} (\textcircled{\text{O}}), K_{\ell_2} (\textcircled{\text{O}}) \in \{\mathcal{T}_h^5, \mathcal{T}_h^{43}\}$ for d_{TV} (upper row) and $\widetilde{d}_{\text{KL}}$ (lower row).

TABLE 2. Error indicators $\mathbf{err}_k := \max_j |\mathbb{E}_{M_\tau} [\|\widetilde{X}_{t_j}\|_{\mathbb{W}^{k,2}} - \|Y^j\|_{\mathbb{W}^{k,2}}]|$ (Example 5.1) resp. $\mathbf{err}_k := \max_j |\mathbb{E}_{M_\tau} [\|\widetilde{\mathbf{X}}_{t_j}\|_{\mathbb{W}^{k,2}} - \|\mathbf{Y}^j\|_{\mathbb{W}^{k,2}}]|$ (Example 5.2) for $k \in \{0, 1\}$, and maximum number of degrees of freedom $L_{\max} := \max_j L^j$.

(A) Example 5.1 (Setup as in Figure 9)

d	d_{H}	d_{TV}	$\widetilde{d}_{\text{KL}}$
L_{\max}	29700	20600	41200
\mathbf{err}_0	0.0223	0.0231	0.0244
\mathbf{err}_1	0.0941	0.0987	0.1022

(B) Example 5.2 (Setup as in Figure 17)

d	d_{H}	d_{TV}	$\widetilde{d}_{\text{KL}}$
L_{\max}	81300	74600	85500
\mathbf{err}_0	0.0344	0.0337	0.0421
\mathbf{err}_1	0.1456	0.1417	0.1502

We recall the basic criterion for M_τ^j in terms of $\text{Var}[\|Y^j\|_{\mathbb{L}^2}]$ to achieve a certain accuracy:

$$\forall a > 0 : \lim_{M_\tau^j \uparrow \infty} \mathbb{P} \left[\left| \mathbb{E}[\|Y^j\|_{\mathbb{L}^2}] - \mathbb{E}_{M_\tau^j}[\|Y^j\|_{\mathbb{L}^2}] \right| \leq a \sqrt{\text{Var}[\|Y^j\|_{\mathbb{L}^2}]/M_\tau^j} \right] = 2\Phi(a) - 1, \quad (5.4)$$

where Φ is the standard normal distribution. Let $\text{To1} > 0$, and $a := \Phi^{-1}(1 - \frac{\alpha}{2}) > 0$ for some level of significance $\alpha \in (0, 1)$: to then control the statistical error in the Monte-Carlo estimation of $\mathbb{E}[\|Y^j\|_{\mathbb{L}^2}]$, we replace $\text{Var}[\|Y^j\|_{\mathbb{L}^2}]$ by its estimator $\text{Var}_{M_\tau^j}[\|Y^j\|_{\mathbb{L}^2}]$,

and deduce $M_\tau^j = \lfloor a^2 \text{Var}_{M_\tau^j}[\|Y^j\|_{\mathbb{L}^2}] / \tau \circ 1^2 \rfloor$ from (5.4). Here, $a > 0$ defines the confidence interval with probability $1 - \alpha = 95\%$.

The studies in Figure 15(B) suggest smaller sample sizes M_τ^j , leading to huge computational savings if compared to the uniformly chosen M_τ , while the accuracy is preserved; see Figure 13. For adaptivity in space, we proceed correspondingly to select M_h^j via (5.4), where $\|Y^j\|_{\mathbb{L}^2}$ is now replaced by $\|\nabla Y^j\|_{\mathbb{L}^2}$.

Accordingly, we base the selection of the number of required bootstrap replications B_τ in (4.4) resp. (4.9) on the empirical variance of the estimators \mathbf{d}^* , by replacing $(M_\tau^j, \|Y^j\|_{\mathbb{L}^2})$ with $(B_\tau^j, \mathbf{d}(\hat{\mu}_{\tau;n}^{1,j,*}, \hat{\mu}_{\tau;n}^{2,j,*}))$ in (5.4). For the simulations we choose at least $B_\tau^j \geq 10^3$ at each time t_j .

In order to adjust the number of cells $R_\tau = R_\tau^j$, we use a heuristic strategy from [6, Chapter 3.2.4], which in particular benefits from the statistically equivalent cell property of the partition via (BTC): assuming that the adaptive sample size M_τ^j is large enough, we set

$$R_\tau^j = 4\sqrt[5]{2}(M_\tau^j/a)^{2/5}, \quad (5.5)$$

which meets $R_\tau^j = \mathcal{O}(\log(M_\tau^j))$ from [10, Theorem 21.8], in particular. As a consequence, R_τ^j again depends on $\text{Var}_{M_\tau^j}[\|Y^j\|_{\mathbb{L}^2}]$ through the choice of M_τ^j via (5.4); see Figure 15.

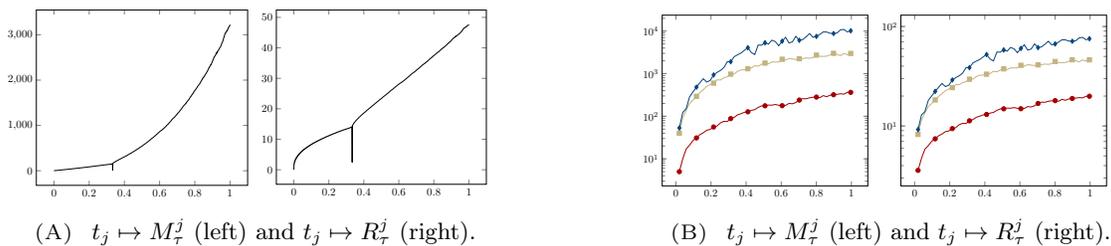


FIGURE 15. (Example 1.1 (A) and Example 5.1 (B) for $\|\cdot\|_{\mathbb{W}^{0,2}}$): Behavior of the required number of MC iterations M_τ^j resp. number of cells R_τ^j according to (5.4) resp. (5.5) for $\iota = 0.1$ (\bullet), $\iota = 0.3$ (\blacksquare), $\iota = 0.5$ (\blacklozenge).

We summarize our computational studies in Section 5.1 for Example 5.1. The combination of space-time adaptivity with adaptively chosen statistical parameters leads to significant savings in simulation times: it is the convection term in (5.1) that causes diffuse layers supporting rapidly changing values in space, and which need to be properly resolved. This resolution in space mainly triggers a variance reduction of computed realizations $\{Y^j\}_j$, and allows for significantly smaller statistical parameters M_τ^j : we recall that it is the most time consuming part of the algorithm to compute approximate solutions of the SPDE (1.1) to set up the required samples $\mathcal{S}_{\bar{\tau}^{\mathfrak{s},j+1}}$ ($\mathfrak{s} \in \{1, 2\}$); hence, smaller sample sizes M_τ^j are therefore very desirable to increase the overall efficiency of the adaptive strategy. We make the following observations:

- (adaptive discretization parameters) Let $M_\tau = 10^5$ be fixed. In order to meet the same error at $T = 1$ for space-time adaptivity (\blacktriangleleft in Figure 13), we need

$h^0 = 2^{-8}$, and $\tau^0 = 3 \cdot 10^{-4}$ for the *uniform* setting; the corresponding simulation time increases by more than 25% for the latter.

- (adaptive discretization and statistical parameters) We now choose statistical parameters adaptively as well. The studies in Figure 15(B) show a reduction of the sample sizes to 1/10-th in this setting: we observe overall computational time savings of up to 78% if compared to uniform discretization and statistical parameters.

5.2. Computational experiments for the SPDE (2.4). We use the (\mathbb{P} -almost sure) length-preserving discretization Scheme 3.3, in combination with space-time and statistical adaptivity. The non-linear algebraic systems at each time step of Scheme 3.3 are solved by Newton's method.

In [31], the authors discuss (in-)stability of blow-up dynamics for (2.4) in the case $\iota = 0$. Below, we computationally study the role that noise exerts on the formation of this singular behavior, and 'discrete blow-up dynamics'. We start the evolution from a super-critical initial datum.

Example 5.2. Let $D = B_1(\mathbf{0}) := \{\mathbf{x} \in \mathbb{R}^2 \mid |\mathbf{x}| \leq 1\}$, and $\mathbf{x}_0 \in \mathbf{H}^1(D; \mathbb{S}^2)$ be given by

$$\mathbf{x}_0(\mathbf{x}) = \begin{cases} (0, 0, -1)^\top, & \text{for } |\mathbf{x}| > 1/2, \\ \left(\frac{2x_1 A}{A^2 + |\mathbf{x}|^2}, \frac{2x_2 A}{A^2 + |\mathbf{x}|^2}, \frac{A^2 - |\mathbf{x}|^2}{A^2 + |\mathbf{x}|^2} \right)^\top & \text{for } |\mathbf{x}| \leq 1/2, \end{cases} \quad \forall \mathbf{x} \in D, \quad (5.6)$$

and $A \equiv A(\mathbf{x}) := (1 - 2|\mathbf{x}|)^4$. The noise term in (3.3) is simulated by

$$\xi_{j+1}(\mathbf{x}) = \sum_{0 \leq |\mathbf{k}| \leq 3} \sum_{l=1}^3 \sqrt{2} \prod_{i=1}^2 \left(\frac{2}{(2k_i + 1)\pi} \right)^2 \sin(k_i \pi x_i) \mathbf{e}_l \xi_{j+1}^{\mathbf{k}, l} \quad \forall \mathbf{x} \in D,$$

where $\xi_{j+1}^{\mathbf{k}, l}$ are i.i.d. \mathbb{R} -valued Brownian increments, and \mathbf{e}_l for $l \in \{1, 2, 3\}$ are the canonical basis vectors of \mathbb{R}^3 .

For $\iota = 0$, a corresponding computational study on uniform space-time meshes for (5.6) in [3] supports a discrete blow-up at the origin $\mathbf{x}_{\ell^*} := (0, 0)^\top$ of the domain D at time $\tilde{t}_{j^*} := \min\{t_j \in [0, T]; \|\nabla \mathbf{Y}^j\|_{\mathbf{L}^\infty} \geq 1/h\}$. Once largest possible 'discrete gradients' are attained on a given spatial mesh, this setting is followed by a rapid switching of $\mathbf{Y}^{j^*}(\mathbf{x}_{\ell^*})$, and the solution becomes almost homogeneous. This dynamics favors Example 5.2 to test the adaptive concepts described in Section 4. Let $h_{\min} := \min_j \min_{K \in \mathcal{T}_h^j} h_K^j$. According to Figure 17(B) below, the computational studies for Example 5.2 via Scheme 3.3 and adaptivity indicate the occurrence of a 'discrete blow-up time' $t_{j^*} \approx 0.05$, where

$$t_{j^*} := \min \left\{ t_j \in [0, T]; \exists K \in \mathcal{T}_h^j : |\hat{\eta}_{h;K}^{1,j,*} - \hat{\eta}_{h;K}^{2,j,*}| \mathbb{P}^*[\mathbf{d}(\hat{\mu}_{h;K}^{1,j,*}, \hat{\mu}_{h;K}^{2,j,*}) \geq 95\%] \geq 1/h_{\min} \right\}, \quad (5.7)$$

where $\{\hat{\eta}_{h;K}^{\mathfrak{s}, j, *}; \mathfrak{s} \in \{1, 2\}\}$ are the first absolute moments of the empirical measures $\{\hat{\mu}_{h;K}^{\mathfrak{s}, j, *}; \mathfrak{s} \in \{1, 2\}\}$ defined in (4.7) resp. (4.8).

For time adaptivity, we use a modification of the (BTC) strategy in Section 4.1 to partition $(\mathbb{S}^2)^{L^j} = \bigcup_{r=1}^{R_r} \hat{C}_{\tau, r}^j$: grouping the events is based on finding the tuple (ℓ, i) , $\ell \in$

$\{1, \dots, L^j\}$, $i \in \{1, 2, 3\}$ in the M_τ^j -sample $\mathcal{S}_{\vec{\mathbf{Y}}^j} := \{\vec{\mathbf{Y}}^j(\omega_k)\}_{k=1}^{M_\tau^j}$ of the $(\mathbb{S}^2)^{L^j}$ -valued random variable $\vec{\mathbf{Y}}^j$ which possesses the largest empirical standard deviation. The snapshots in the top line in Figures 16(A)–(C) display partitions of (marginal) distributions for $\{\vec{\mathbf{Y}}^j(\mathbf{x}_{\ell^*})\}_j$ on \mathbb{S}^2 at \mathbf{x}_{ℓ^*} for different times near t_{j^*} ; the local discrete blow-up of $\mathbf{Y}^j(\mathbf{x}_{\ell^*})$ is well-detected and resolved by locally refined meshes which are generated by the adaptive algorithm (see bottom line in Figures 16(A)–(C)): more cells are created in areas where the random variable $\mathbf{Y}^j(\mathbf{x}_{\ell^*})$ is more likely to take values. The snapshot in Figure 16(D) displays both, the empirical expectation $\mathbb{E}_{M_\tau^j}[\mathbf{Y}^j]$ (top), and values of $|\mathbb{E}_{M_\tau^j}[\mathbf{Y}^j(\mathbf{x}_\ell)]|$ nearby \mathbf{x}_{ℓ^*} at time $t_j \approx t_{j^*}$, with $|\mathbb{E}_{M_\tau^j}[\mathbf{Y}^j(\mathbf{x}_{\ell^*})]| \approx 0.282$. For nodal points \mathbf{x}_ℓ satisfying $|\mathbf{x}_{\ell^*} - \mathbf{x}_\ell| > 0.1$, $|\mathbb{E}_{M_\tau^j}[\mathbf{Y}^j]|$ is almost 1.0. This shrinking of statistical averages of vectors in the neighborhood of $(t_{j^*}, \mathbf{x}_{\ell^*})$ is another indication of the discrete blow-up phenomenon; we also refer to [3] for corresponding studies of single trajectories in the case of uniform space-time meshes.

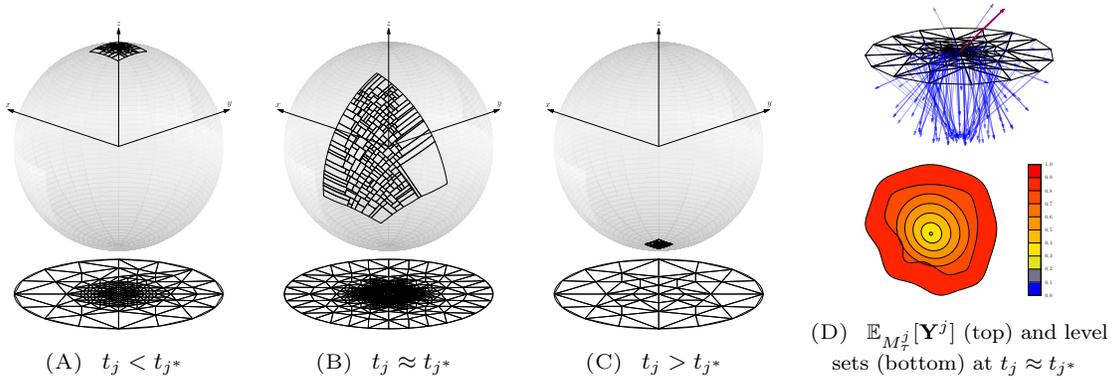


FIGURE 16. (Example 5.2 for $T = 0.1$, $\iota = 0.1$, as well as $\mathbf{d} = \mathbf{d}_{\text{TV}}$, $\text{ToI}_\tau = \text{ToI}_h = 0.05$, $M_\tau^0 = M_h^0 = 10^3$, $R_\tau^0 = 2^8$, $R_h = 2^5$, and $h^0 = 2^{-4}$, $\tau^0 = 10^{-4}$, $h_{\min} = 2^{-6}$) Snapshots of the partitioning of \mathbb{S}^2 supporting $\mathcal{L}(\vec{\mathbf{Y}}^j(\mathbf{x}_{\ell^*}))$ and corresponding spatial meshes \mathcal{T}_h^j for $t_j < t_{j^*}$ (A), $t_j \approx t_{j^*}$ (B), and $t_j > t_{j^*}$ (C). (D) Snapshots of $\mathbb{E}_{M_\tau^j}[\mathbf{Y}^j]$ (top) and corresponding level sets $\{\mathbf{x}_\ell : |\mathbf{x}_{\ell^*} - \mathbf{x}_\ell| \leq 0.1 \mid |\mathbb{E}_{M_\tau^j}[\mathbf{Y}^j(\mathbf{x}_\ell)]| = c\}$ (bottom) for values $c \in \{0.0, 0.1, \dots, 1.0\}$ at time $t_j \approx t_{j^*}$.

The discrete blow-up phenomenon is computationally detected by shrinking local step sizes $(\tau^j, \{h_K^j\}_K)$ to $\tau^j \approx h_{\min}^2$ near $j \approx j^*$, where $h_{\min} = 2^{-6}$ is chosen as the smallest admitted mesh-size to terminate repeated refinement; the Newton method requires the most iterations (up to 50 to meet a \mathbb{P} -almost sure threshold criterion) in the neighborhood of t_{j^*} , and its number decreases rapidly afterwards again. As for Example 5.1, and motivated by the results in Table 2B, the simulations again show superiority of \mathbf{d}_{TV} to perform adaptivity in space, in particular.

In a vicinity of t_{j^*} , the expected energy loss of iterates is approximately 4π ; see Figure 17(A). For times $t_j \uparrow t_{j^*}$, the simulations show a growing concentration of spatial nodal points at \mathbf{x}_{ℓ^*} of D (see Figure 16), and smaller time step sizes τ^j ; larger values are needed

here for M_τ^j , which is due to an increased empirical variance $\text{Var}_{M_\tau^j}[\|\nabla\mathbf{Y}^j\|_{\mathbf{L}^\infty}]$ of the iterates (see Figure 17(D)) near the discrete space-time blow-up. Beyond the time t_{j^*} , the time step size τ^j rapidly increases, and also \mathcal{T}_h^j coarsens again; accordingly, the adaptive selection of the statistical parameters M_h^j , B_h^j , and R_h^j recovers again to moderate values for $j \gg j^*$. For example, the chosen number $M_\tau^j \approx 3000$ at an initial time rapidly changes to $M_\tau^j \approx 180000$ close to t_{j^*} , where also $L^j \approx 75000$; cf. Table 2B. Hence, the adaptive concept from Section 5.1.3 concentrates computer resources to the discrete blow-up phenomenon: in fact, 76% of the overall required time steps ($J = 534$) are concentrated here. For comparison, we found that $J \approx 40000$ would be necessary for a uniform spatio-temporal discretization to obtain a similar error \mathbf{err}_1 as in Table 2B. As a conclusion, and according to our findings for Example 5.1, the direct comparison with uniform discretization and statistical parameters again shows a drastically reduced computational effort when the proposed adaptive concepts are used.

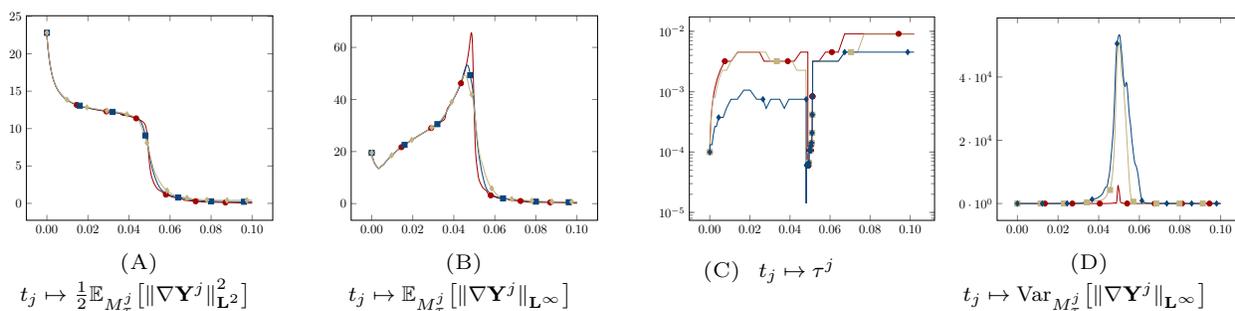


FIGURE 17. (Example 5.2 for the same setup as in Figure 16) (A) Behavior of $t_j \mapsto \frac{1}{2} \mathbb{E}_{M_\tau^j} [\|\nabla\mathbf{Y}^j\|_{\mathbf{L}^2}^2]$, (B) $t_j \mapsto \mathbb{E}_{M_\tau^j} [\|\nabla\mathbf{Y}^j\|_{\mathbf{L}^\infty}]$, (C) $t_j \mapsto \tau^j$, and (D) $t_j \mapsto \text{Var}_{M_\tau^j} [\|\nabla\mathbf{Y}^j\|_{\mathbf{L}^\infty}]$ with $\nu = 1.0$ (—●—), $\nu = 2.0$ (—■—), $\nu = 3.0$ (—◆—).

REFERENCES

- [1] O. Aboura. Weak error expansion of the implicit euler scheme. 2013.
- [2] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.
- [3] L. Bañas, Z. Brzeźniak, M. Neklyudov, and A. Prohl. *Stochastic ferromagnetism*, volume 58 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2014. Analysis and numerics.
- [4] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 2002.
- [5] E. Burman. Consistent SUPG-method for transient transport problems: stability and convergence. *Comput. Methods Appl. Mech. Engrg.*, 199(17-20):1114–1123, 2010.
- [6] R. B. D’Agostino and M. A. Stephens, editors. *Goodness-of-fit techniques*, volume 68 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York, 1986.
- [7] L. Dagum and R. Menon. Openmp: An industry-standard api for shared-memory programming. *IEEE Comput. Sci. Eng.*, 5(1):46–55, 1998.
- [8] A. DasGupta. *Asymptotic theory of statistics and probability*. Springer Texts in Statistics. Springer, New York, 2008.

- [9] T. A. Davis. Algorithm 832: UMFPACK V4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Software*, 30(2):196–199, 2004.
- [10] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*, volume 31 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1996.
- [11] T. Dunst and A. Prohl. The Forward-Backward Stochastic Heat Equation: Numerical Analysis and Simulation. *SIAM J. Sci. Comput.*, 38(5):A2725–A2755, 2016.
- [12] J. G. Gaines and T. J. Lyons. Variable step size control in the numerical solution of stochastic differential equations. *SIAM J. Appl. Math.*, 57(5):1455–1484, 1997.
- [13] M. Galassi. *GNU Scientific Library Reference Manual - Third Edition*, 2009.
- [14] A. George and E. Ng. On the complexity of sparse QR and LU factorization of finite-element matrices. *SIAM J. Sci. Statist. Comput.*, 9(5):849–861, 1988.
- [15] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *Internat. Stat. Rev.*, 70(3):419–435, 2002.
- [16] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations*, volume 5 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1986. Theory and algorithms.
- [17] E. Gobet, J.-P. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Probab.*, 15(3):2172–2202, 2005.
- [18] P. E. Greenwood and M. S. Nikulin. *A guide to chi-squared testing*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1996. A Wiley-Interscience Publication.
- [19] A. Hocquet. Struwe-like solutions for the stochastic harmonic map flow. *J. Evol. Equ.*, 18(3):1189–1228, 2018.
- [20] V. John and J. Novo. Error analysis of the SUPG finite element discretization of evolutionary convection-diffusion-reaction equations. *SIAM J. Numer. Anal.*, 49(3):1149–1176, 2011.
- [21] V. John and E. Schmeier. Finite element methods for time-dependent convection-diffusion-reaction equations with small diffusion. *Comput. Methods Appl. Mech. Engrg.*, 198(3-4):475–494, 2008.
- [22] C. Johnson. *Numerical solution of partial differential equations by the finite element method*. Cambridge University Press, Cambridge, 1987.
- [23] C. Johnson, U. Nävert, and J. Pitkäranta. Finite element methods for linear hyperbolic problems. *Comput. Methods Appl. Mech. Engrg.*, 45(1-3):285–312, 1984.
- [24] H. Lamba, J. C. Mattingly, and A. M. Stuart. An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA J. Numer. Anal.*, 27(3):479–506, 2007.
- [25] K.-S. Moon, A. Szepessy, R. Tempone, and G. E. Zouraris. Convergence rates for adaptive weak approximation of stochastic differential equations. *Stoch. Anal. Appl.*, 23(3):511–558, 2005.
- [26] E. J. G. Pitman. *Some basic theory for statistical inference*. Chapman and Hall, London; A Halsted Press Book, John Wiley & Sons, New York, 1979. Monographs on Applied Probability and Statistics.
- [27] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.
- [28] C. Schellnegger. *Adaptivity for Stochastic Magnetization Dynamics*. Phd thesis, Universität Tübingen, July 2018.
- [29] A. Schmidt and K. G. Siebert. *Design of adaptive finite element software*, volume 42 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2005.
- [30] A. Szepessy, R. Tempone, and G. E. Zouraris. Adaptive weak approximation of stochastic differential equations. *Comm. Pure Appl. Math.*, 54(10):1169–1214, 2001.
- [31] J. B. van den Berg and J. F. Williams. (In-)stability of singular equivariant solutions to the Landau-Lifshitz-Gilbert equation. *European J. Appl. Math.*, 24(6):921–948, 2013.
- [32] A. J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Trans. Math. Softw.*, 3(3):253–256, 1977.
- [33] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. I. The recovery technique. *Internat. J. Numer. Methods Engrg.*, 33(7):1331–1364, 1992.

MATHEMATISCHES INSTITUT, UNIVERSITÄT TÜBINGEN, AUF DER MORGENSTELLE 10, D-72076 TÜBINGEN, GERMANY

Email address: `prohl@na.uni-tuebingen.de`

MATHEMATISCHES INSTITUT, UNIVERSITÄT TÜBINGEN, AUF DER MORGENSTELLE 10, D-72076 TÜBINGEN, GERMANY

Email address: `schellnegger@na.uni-tuebingen.de`