ESAIM: M2AN 57 (2023) 785–815
https://doi.org/10.1051/m2an/2022089

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

NUMERICAL APPROXIMATION OF PROBABILISTICALLY WEAK AND STRONG SOLUTIONS OF THE STOCHASTIC TOTAL VARIATION FLOW

Ľubomír Baňas[1,*] and Martin Ondreját[2]

Abstract. We propose a fully practical numerical scheme for the simulation of the stochastic total variation flow (STVF). The approximation is based on a stable time-implicit finite element space-time approximation of a regularized STVF equation. The approximation also involves a finite dimensional discretization of the noise that makes the scheme fully implementable on physical hardware. We show that the proposed numerical scheme converges in law to a solution that is defined in the sense of stochastic variational inequalities (SVIs). Under strengthened assumptions the convergence can be show to holds even in probability. As a by product of our convergence analysis we provide a generalization of the concept of probabilistically weak solutions of stochastic partial differential equation (SPDEs) to the setting of SVIs. We also prove convergence of the numerical scheme to a probabilistically strong solution in probability if pathwise uniqueness holds. We perform numerical simulations to illustrate the behavior of the proposed numerical scheme as well as its non-conforming variant in the context of image denoising.

Mathematics Subject Classification. 60H17, 60H15, 60H35, 65C30, 65M60, 94A08, 60A10.

Received May 4, 2022. Accepted November 1, 2022.

1. Introduction

We study a numerical approximation of the stochastic total variation flow (STVF)

$$\begin{aligned} \mathrm{d}X &= \mathrm{div}\left(\frac{\nabla X}{|\nabla X|}\right)\mathrm{d}t - \lambda(X-g)\mathrm{d}t + B(X)\mathrm{d}W, && \text{in } (0,T)\times\mathcal{O}, \\ X &= 0 && \text{on } (0,T)\times\partial\mathcal{O}, \\ X(0) &= x^0 && \text{in } \mathcal{O}, \end{aligned} \tag{1}$$

where $\mathcal{O}\subset\mathbb{R}^d$, $d\geq 1$ is a bounded polyhedral domain, $\lambda\geq 0$, $T>0$ are fixed constants and x^0, $g\in\mathbb{L}^2$ are given functions, using the notation $\mathbb{L}^2 = L^2(\mathcal{O})$. We consider W to be a cylindrical Wiener process on ℓ_2 and a continuous mapping $B:\mathbb{L}^2\to\mathscr{L}_2(\ell_2;\mathbb{L}^2)$ where ℓ_2 denotes the Hilbert space of square summable sequences, $\mathscr{L}_2$ stands for the space of Hilbert–Schmidt operators such that

($\mathbf{B}_1$) $\|B(h)\|_{\mathscr{L}_2(\ell_2;\mathbb{L}^2)}\leq C(\|h\|+1)$ for every $h\in\mathbb{L}^2$,

Keywords and phrases. stochastic total variation flow, stochastic variational inequalities, image processing, finite element approximation, tightness in BV spaces.

[1] Department of Mathematics, Bielefeld University, 33501 Bielefeld, Germany.

[2] Institute of Information Theory and Automation, Pod Vodárenskou věží 4, CZ-182 00 Praha 8, Czech Republic.

*Corresponding author: banas@math.uni-bielefeld.de

$(\mathbf{B}_2)$ if $d \geq 2$, whenever $\{h_n\}$ is bounded in $\mathbb{L}^2$ and $h_n \to h$ a.e. in $\mathcal{O}$ then

$$\|B(h_n) - B(h)\|_{\mathscr{L}_2(\ell_2;\mathbb{L}^2)} \to 0.$$

For instance, an operator

$$[B(h)](\alpha) := \sum_j \alpha_j b_j(h), \qquad \alpha \in \ell_2,$$

where $b_j : \mathbb{R} \to \mathbb{R}$ are continuous and

$$|b_j(x)| \leq \lambda_j \varrho(x), \qquad x \in \mathbb{R},$$

holds for some $\lambda_j \geq 0$ and $\varrho : \mathbb{R} \to \mathbb{R}$ such that

$$\sum_j \lambda_j^2 < \infty \quad \text{and} \quad |\varrho(x)| \leq C(|x| + 1),$$

satisfies $(\mathbf{B}_1)$. If, moreover,

$$\lim_{|x|\to\infty} \frac{|\varrho(x)|}{|x|} = 0,$$

then also $(\mathbf{B}_2)$ holds for B.

We also consider the weakly lower semicontinuous energy functional $\mathcal{J} : \mathbb{L}^2 \to [0, \infty]$

$$\begin{aligned} \mathcal{J}(u) &:= \|\nabla u\|_{\mathrm{TV}(\mathcal{O})} + \int_{\partial\mathcal{O}} |u| \,\mathrm{d}x + \frac{\lambda}{2} \int_{\mathcal{O}} |u - g|^2 \,\mathrm{d}x \qquad && u \in \mathbb{L}^2 \cap BV(\mathcal{O}), \\ \mathcal{J}(u) &:= \infty && u \in \mathbb{L}^2 \setminus BV(\mathcal{O}), \end{aligned}$$

see Lemma B.1 for details.

Due to the singular character of total variation flow (1), it is convenient to perform numerical simulations using a regularized problem

$$\begin{aligned} \mathrm{d}X &= \mathrm{div}\left(\frac{\nabla X}{\sqrt{|\nabla X|^2 + \varepsilon^2}}\right)\mathrm{d}t - \lambda(X - g)\mathrm{d}t + B(X)\,\mathrm{d}W \qquad && \text{in } (0,T) \times \mathcal{O}, \\ X &= 0 && \text{on } (0,T) \times \partial\mathcal{O}, \\ X(0) &= x^0 && \text{in } \mathcal{O}, \end{aligned} \tag{2}$$

with a regularization parameter $\varepsilon > 0$. In the deterministic setting $(B(X) \equiv 0)$ the equation (2) corresponds to the gradient flow of the regularized energy functional

$$\mathcal{J}_\varepsilon(u) := \int_{\mathcal{O}} \sqrt{|\nabla u|^2 + \varepsilon^2} \,\mathrm{d}x + \frac{\lambda}{2} \int_{\mathcal{O}} |u - g|^2 \,\mathrm{d}x \qquad u \in \mathbb{H}^1_0.$$

Convergent finite element approximation of the deterministic total variation flow (*i.e.*, (1) and (2) with $B(X) \equiv 0$) has been proposed in [14]. In the stochastic setting, numerical approximation of probabilistically strong stochastic variational inequalities (SVI) solutions of (1) with $B(X) \equiv X$ has been analyzed recently in [5–7] by considering the regularized problem (2) within the framework of stochastic variational inequalities, *cf.*, [2]. In the present work we propose a fully implementable numerical approximation of (1) *via* the regularized problem (2): in addition to the discretization in space and time we also consider an implementable approximation of the noise term. We show that, in the limit, the numerical solutions satisfy a stochastic variational inequality. As a consequence, we obtain an extension of the concept of stochastic variational inequalities of [2].

Let us compare the present work with [5] where (probabilistically) strong solutions of (1) are constructed numerically in case the domain $\mathcal{O}$ is bounded, convex and with a piecewise C^2-smooth boundary, the equation is driven by a one-dimensional noise W, $B(X) = X$ and the interpolants $\overline{X}^{\varepsilon}_{\tau,h}$ of the numerical approximations converge to the unique solution X with paths continuous in $L^2(\mathcal{O})$ *via* the double limit

$$\lim_{\varepsilon\to 0}\lim_{(\tau,h)\to(0,0)}\left\|\overline{X}^{\varepsilon}_{\tau,h} - X\right\|_{L^2(\Omega\times(0,T);L^2(\mathcal{O}))} = 0.$$

In the present work $\mathcal{O}$ is an open convex polyhedral domain, B is a fairly general non-linearity (hence uniqueness is not expected to hold and we construct just (probabilistically weak) "martingale" solutions). Furthermore, the considered noise is an infinite dimensional random walk generated by a sequence of random variables (suitable for computer simulations) and $\overline{X}^{\varepsilon}_{\tau,h}$ converge to X in the joint limit as $(\varepsilon,\tau,h)\to(0,0,0)$. Our SVI solution concept is more general than the one in [5] but paths of the obtained solutions are only weakly continuous in $L^2(\mathcal{O})$ and we cover the case $B(X) = X$ only in $d = 1$. If, in addition, pathwise uniqueness holds then the approximations converge to a probabilistically strong solution in probability. We also note that the technique used for the construction of the probabilistically weak SVI solutions is straightforward, *i.e.*, we avoid the use of martingale and Skorokhod representation theorems as in [19]. We mention that the approach of [5] is restricted to spatial dimension $d = 1$, *cf.*, [7]. The higher dimensional case (for probabilistically strong solutions) is treated in [6] using a slightly different approach; so far the work [6] appears to be the only one to cover the case of linear multiplicative noise $B(X) = X$ for $d > 1$.

The paper is organized as follows. In Section 2 we introduce the notation and the numerical approximation of (2) and in Section 3 we state the main results of the paper (which are proven in Sects. 7 and 8). In Section 4 we show *a priori* estimate for the numerical solution. In Section 5 we present auxiliary results on compactness properties of locally convex spaces which are used to deduce tightness properties and convergence of the numerical approximation in Section 6. Numerical experiments for the conforming and non-conforming finite element approximation schemes are presented in Section 9. The proofs of auxiliary results are collected in the appendix.

2. Numerical approximation

We denote the standard Lebesgue and Sobolev functions spaces on $\mathcal{O}$ as $\mathbb{L}^p := L^p(\mathcal{O})$, $\mathbb{L}^p_w = (\mathbb{L}^p, \text{weak})$, $\mathbb{H}^1_0 := H^1_0(\mathcal{O})$, $\mathbb{W}^{1,1} := W^{1,1}(\mathcal{O})$; we use $\|\cdot\|_{\mathbb{L}^p}$ to denote the standard $\mathbb{L}^p$-norm and $\|\cdot\| := \|\cdot\|_{\mathbb{L}^2}$. The sets of rational and irrational numbers are denoted as $\mathbb{Q}$ and $\mathbb{Q}^{\complement}$, respectively. For time dependent random variables we often write $S_t(\cdot)$ instead of $S(\cdot,t)$ provided that it fits the context of presentation.

We consider the space of functions of bounded variation

$$BV(\mathcal{O}) = \left\{u\in\mathbb{L}^1;\ \nabla u \text{ is a vector measure}\right\},$$

equipped with the norm

$$\|u\|_{BV(\mathcal{O})} := \|u\|_{\mathbb{L}^1} + \|\nabla u\|_{\mathrm{TV}(\mathcal{O})},$$

see, for instance, Section 3.1 of [1]. We recall that for $u \in BV(\mathcal{O})$ the gradient ∇u is a vector measure whose total variation satisfies

$$\|\nabla u\|_{\mathrm{TV}(\mathcal{O})} = \sup\left\{-\int_{\mathcal{O}} u \operatorname{div}\mathbf{v}\,\mathrm{d}x;\ \mathbf{v}\in C_0^\infty(\mathcal{O},\mathbb{R}^d),\ \|\mathbf{v}\|_{\mathbb{L}^\infty}\le 1\right\}.$$

For $N\in\mathbb{N}$ we consider an equidistant partition $\{t_i = i\tau,\ i = 0,\dots,N\}$ of the time interval $[0,T]$ with a discrete time step $\tau = T/N$. Consequently, we consider a discrete filtration $\mathcal{F}_\tau := \{\mathcal{F}^i_\tau\}_{i=0}^N$ on a probability space $(\Omega_\tau,\mathscr{F}_\tau,\mathbb{P}_\tau)$ and sequence $\{\xi^{i,j}_\tau\}_{i,j=1}^N$ of independent random variables such that

a) $\mathbb{E}\left[\,\xi^{i,j}_\tau\right] = 0$,

b) $\mathbb{E}\big[|\xi_\tau^{i,j}|^2\big] = \tau$,

c) $\mathbb{E}\big[|\xi_\tau^{i,j}|^4\big] \leq C\tau^2$,

d) $\big(\xi_\tau^{i,1},\dots,\xi_\tau^{i,N}\big)$ is $\mathcal{F}_\tau^i$-measurable and independent of $\mathcal{F}_\tau^{i-1}$,

for every $i,j\in\{1,\dots,N\}$ and some fixed constant $C>0$ independent of $N\in\mathbb{N}$. A simple and easily implementable construction of the noise that satisfies the above properties is, for instance, $\xi_\tau^{i,j}=\sqrt{\tau}\chi^{i,j}$ where $\{\chi^{i,j}\}_{i,j=1}^N$ are independent with $\mathbb{P}[\chi^{i,j}=\pm 1]=\frac{1}{2}$; as another choice, one can consider Brownian increments $\xi_\tau^{i,j}=\Delta_i\beta^j := \beta^j(t_i)-\beta^j(t_{i-1})$ of independent Brownian motions β^j.

Let $\{\mathbb{V}_h\}_{h>0}$ be a family of standard $\mathbb{H}_0^1$-conforming finite element spaces of globally continuous functions which are piecewise linear over the quasi-uniform partitions $\{\mathcal{T}_h\}_{h>0}$ of $\mathcal{O}$ with corresponding mesh sizes $h>0$. Further, let $P_h:\mathbb{L}^2\to\mathbb{V}_h$ denote the $\mathbb{L}^2$-orthogonal projection on $\mathbb{V}_h$. We assume that the family $\{\mathbb{V}_h\}_{h>0}$ satisfies the following properties.

Assumption 2.1. (1) *$\mathbb{V}_h$ is a finite-dimensional subspace of $\mathbb{H}_0^1$,*
(2) *$\mathbb{V}_{h_2}\subseteq\mathbb{V}_{h_1}$ if $0<h_1<h_2$,*
(3) *$\|P_h v\|_{\mathbb{H}_0^1}\leq\kappa\|v\|_{\mathbb{H}_0^1}$ holds for every $v\in\mathbb{H}_0^1$ and $h>0$, for some $\kappa\in(0,\infty)$ (see [10]),*
(4) *$\bigcup_{h>0}\mathbb{V}_h$ is dense both in $\mathbb{H}_0^1$ and $\mathbb{L}^2$.*

It is well known that the above assumption is satisfied for $\mathbb{V}_h$, P_h see for instance [12]. We note that the stability of the $\mathbb{L}^2$-projection, Assumption $2.1_{(3)}$ and the density of $\{\mathbb{V}_h\}_{h>0}$ in $\mathbb{H}_0^1$ implies that $\|\nabla v-\nabla P_h v\|\to 0$ as $h\to 0$ for every $v\in\mathbb{H}_0^1$.

We consider the following fully-discrete approximation of (2): fix $N\in\mathbb{N}$, $h>0$ set $X^0=P_h x^0$ and determine $X^i\in\mathbb{V}_h$, $i=1,\dots,N$ as the solution of

$$\begin{aligned}\big(X^i-X^{i-1},v_h\big) = &-\tau\left(\frac{\nabla X^i}{\sqrt{|\nabla X^i|^2+\varepsilon^2}},\nabla v_h\right)\\ &-\tau\lambda\big(X^i-g,v_h\big)+\sum_{j=1}^N\big(B_j(X^{i-1}),v_h\big)\xi_\tau^{i,j} \qquad \forall v_h\in\mathbb{V}_h,\end{aligned}\tag{3}$$

where $B_j(X)$ is defined as $B(X)e_j$ and $\{e_j\}$ denotes the canonical orthonormal basis in ℓ_2. Existence of the unique $\mathcal{F}_\tau$-adapted $\mathbb{V}_h$-valued solution $\{X^i\}_{i=0}^N$ can be proved analogously to Lemma 3 of [5] therefore we omit the proof. The process $X^i\equiv X^i_{\varepsilon,h}$, $i=0,\dots,N$ depends on the parameters (τ,h,ε), to simplify the notation we suppress this dependence unless it matters.

Remark 2.1. All result of the paper remain valid for a modification of (3) where the discrete noise term is replaced by $\sum_{j=1}^R\big(B_j(X^{i-1}),v_h\big)\xi_\tau^{i,j}$ with an N-independent truncation parameter $R\in\mathbb{N}$. In this case the convergence of the scheme follows by considering the limit $R\to\infty$ along with $(\tau,h,\varepsilon)\to 0$, without further modifications of the arguments. We further note that the explicit evaluation of the terms $B_j(X^{i-1})$, $j=1,\dots,N$ at X^{i-1} (and not at X^i) is related to the fact that the multiplicative noise in (1) is considered to be of Itô type.

3. Summary of the main results

In this section, we summarize the main results of the paper. We start with the definition of the SVI solution of (1).

Definition 3.1. Let $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ be a stochastic basis with independent $(\mathcal{F}_t)$-Wiener processes $(W^k)_{k\in\mathbb{N}}$. An adapted process $X\in L^2([0,T]\times\Omega;\mathbb{L}^2)$ with weakly continuous paths in $\mathbb{L}^2$ is called an SVI solution of (1) provided that

$$\frac{1}{2}\mathbb{E}\Big[\|X(t)-I(t)\|^2\Big]+\mathbb{E}\left[\int_0^t\mathcal{J}(X(s))\,\mathrm{d}s\right]\leq\frac{1}{2}\big\|x^0-u^0\big\|^2+\mathbb{E}\left[\int_0^t\mathcal{J}(I(s))\,\mathrm{d}s\right]+\mathbb{E}\left[\int_0^t(G(s),X(s)-I(s))\,\mathrm{d}s\right]$$

$$+\frac{1}{2}\mathbb{E}\Big[\int_0^t \|B(X(s))-H(s)\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\Big], \tag{4}$$

holds for every $t\in[0,T]$ and for every test process

$$I(t)=u^0-\int_0^t G(s)\,\mathrm{d}s+\sum_{j=1}^{\infty}\int_0^t H_j(s)\,\mathrm{d}W^j, \qquad t\in[0,T], \tag{5}$$

that satisfies $\mathbb{P}\left[I(t)\in\mathbb{H}^1_0\right]=1$ for almost every $t\in[0,T]$ and

$$\mathbb{E}\left[\int_0^T \|I(t)\|_{\mathbb{H}^1_0}\,\mathrm{d}t\right]<\infty,$$

for some $u^0\in\mathbb{L}^2$ and $(\mathcal{F}_t)$-progressively measurable processes G and H in $L^2([0,T]\times\Omega;\mathbb{L}^2)$ and $L^2([0,T]\times\Omega;\mathscr{L}_2(\ell_2,\mathbb{L}^2))$ respectively.

Remark 3.1. Inequality (4) implies

$$\sup_{t\in[0,T]}\mathbb{E}\big[\|X(t)\|^2\big]+\mathbb{E}\left[\int_0^T \|X(t)\|_{BV(\mathcal{O})}\,\mathrm{d}t\right]<\infty.$$

Remark 3.2. The SVI solution in the sense of Definition 3.1 generalizes the definition of the SVI solution from [2] since the inequality (4) holds for a much larger class of test processes introduced in (5) than in [2]. Formally, inequality (4) can be derived by applying Itô's formula to the $\mathbb{L}^2$-norm of the difference of (1) and (5) and using the convexity of $\mathcal{J}$.

Definition 3.2. We say that pathwise uniqueness holds for SVI solutions of (1) satisfying (8) provided that any such two SVI solutions X^0 and X^1 on a stochastic basis $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ with respect to independent $(\mathcal{F}_t)$-Wiener processes $(W^k)_{k\in\mathbb{N}}$ satisfy $X^0(t)=X^1(t)$ a.s. for every $t\in[0,T]$.

Definition 3.3. We say that SVI solutions of (1) satisfying (8) are unique in law provided that any such two SVI solutions X^0 and X^1 on stochastic bases $(\Omega^0,\mathcal{F}^0,(\mathcal{F}^0_t),\mathbb{P}^0)$ and $(\Omega^1,\mathcal{F}^1,(\mathcal{F}^1_t),\mathbb{P}^1)$ with respect to independent $(\mathcal{F}^i_t)$-Wiener processes $(W^k_i)_{k\in\mathbb{N}}$, $i=0,1$ have the same finite-dimensional distributions, *i.e.*,

$$\mathbb{P}^0\big[X^0(r_0)\in A_0,\dots,X^0(r_n)\in A_n\big]=\mathbb{P}^1\big[X^1(r_0)\in A_0,\dots,X^1(r_n)\in A_n\big]$$

holds for every $0\le r_0\le\cdots\le r_n\le T$, every Borel sets $A_0,\dots,A_n$ in $\mathbb{L}^2$ and every $n\in\mathbb{N}$.

We define the piecewise linear interpolant of the solution of the scheme (3) as

$$X_\tau(t)=\frac{t-t_{i-1}}{\tau}X^i+\frac{t_i-t}{\tau}X^{i-1}\quad \text{for } t\in[t_{i-1},t_i], \tag{6}$$

as well as the piecewise constant interpolants

$$\overline{X}_\tau(t)=X^i \qquad \text{for } t\in(t_{i-1},t_i), \tag{7a}$$
$$\underline{X}_\tau(t)=X^{i-1} \qquad \text{for } t\in(t_{i-1},t_i), \tag{7b}$$

where the dependence on ε, h is not displayed.

Let $\mathcal{X}^{(1)}$ denote the space of weakly càglàd functions $f:[0,T]\to\mathbb{L}^2$ such that

$$\int_0^T \|f(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s<\infty,$$

let $\mathcal{X}^{(2)}$ denote the space of weakly càdlàg functions $f:[0,T]\to\mathbb{L}^2$, define $\mathcal{X}^{(3)}$ as $C([0,T];\mathbb{L}^2_w)$ and equip the spaces $\mathcal{X}^{(1)}$, $\mathcal{X}^{(2)}$ and $\mathcal{X}^{(3)}$ with the topology of uniform convergence in $\mathbb{L}^2_w$.

Theorem 3.1. *The random variables*

$$\left(\overline{X}_{\varepsilon,h,\tau}, \underline{X}_{\varepsilon,h,\tau}, X_{\varepsilon,h,\tau}\right) : (\Omega_\tau, \mathcal{F}_\tau, \mathbb{P}_\tau) \to \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)},$$

are Borel measurable and their laws under $\mathbb{P}_\tau \equiv \mathbb{P}_{\varepsilon,h,\tau}$ *are tight with respect to* ε, h, τ. *Moreover, every sequence* $(\varepsilon_n, h_n, \tau_n) \to (0,0,0)$ *has a subsequence* $(\varepsilon_{n_k}, h_{n_k}, \tau_{n_k})$ *such that laws of*

$$\left(\overline{X}_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}, \underline{X}_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}, X_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}\right),$$

under $\mathbb{P}_{\tau_{n_k}}$ *converge to a Radon probability measure* ν *on* $\mathscr{B}(\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)})$ *that satisfies*

$$\nu\Big((x_1, x_2, x_3) \in \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} : x_1 = x_2 = x_3\Big) = 1,$$

and there exists a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ *with independent* $(\mathcal{F}_t)$*-Wiener processes* $(W^k)_{k\in\mathbb{N}}$ *and a weakly continuous* $\mathbb{L}^2$*-valued SVI solution* X *of* (1) *in the sense of Definition 3.1 such that* $X(0) = x^0$, ν *is the law of* (X, X, X) *on* $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)}$ *and*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \|X(t)\|^4 + \left(\int_0^T \|X(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s\right)^2\right] \le C, \tag{8}$$

where $C = C(T, \mathcal{O}, \|x_0\|, \|g\|)$.

Proof. See Corollary 6.1 and Theorem 7.1 for the proof. □

Remark 3.3. Compared to the (probabilistically strong) SVI solutions in [2, 5] where the stochastic basis is given, the SVI solution obtained in this paper is probabilistically weak in the sense that the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and the Wiener processes $(W^k)_{k\in\mathbb{N}}$ are constructed as a part of the solution, *cf.*, Corollary 6.1 and Theorem 7.1.

Corollary 3.1. *If uniqueness in law holds for the SVI solution of* (1)*, cf., [2], then the laws of*

$$\left(\overline{X}_{\varepsilon,h,\tau}, \underline{X}_{\varepsilon,h,\tau}, X_{\varepsilon,h,\tau}\right),$$

under $\mathbb{P}_\tau$ *converge to* ν *on* $\mathscr{B}(\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)})$ *as* $(\varepsilon, h, \tau) \to (0,0,0)$. *In particular, there is no need to pass to a subsequence in Theorem 3.1.*

In case we work on a fixed stochastic basis independent of τ with a given Wiener process and pathwise uniqueness holds for (1) then we can construct probabilistically strong solutions.

Theorem 3.2. *Let* $(W^j)_{j\in\mathbb{N}}$ *be independent* $(\mathcal{F}_t)$*-Wiener processes on* $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ *and let*

$$\xi_\tau^{i,j} = W^j(t_i) - W^j(t_{i-1}), \qquad t_i = i\tau.$$

Furthermore, assume that pathwise uniqueness holds for SVI solutions of (1) *satisfying* (8)*. Then there exists an SVI solution* X *with respect to* $(W^j)_{j\in\mathbb{N}}$ *satisfying* (8) *such that*

$$\sup_{t\in[0,T]}\left|\left(\overline{X}_{\varepsilon,h,\tau}(t) - X(t), \varphi\right)\right|, \qquad \sup_{t\in[0,T]}\left|\left(\underline{X}_{\varepsilon,h,\tau}(t) - X(t), \varphi\right)\right|, \qquad \sup_{t\in[0,T]}\left|(X_{\varepsilon,h,\tau}(t) - X(t), \varphi)\right|$$

converge to 0 *in probability as* $(\varepsilon, h, \tau) \to (0,0,0)$ *for every* $\varphi \in \mathbb{L}^2$.

Proof. See Theorem 8.1. □

Remark 3.4. Theorems 3.1 and 3.2 can be strengthened considerably by Lemma 7.1. Assume that

$$K : \mathcal{X}^{(1)} \times \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \to [0,\infty],$$

satisfies the following property:

$$K\big(f^0, f^1, g, h\big) \leq \liminf_{n\to\infty} K\big(f^0_n, f^1_n, g_n, h_n\big), \tag{9}$$

for any sequence $(f^0_k, f^1_k, g_k, h_k) \in \mathcal{X}^{(1)}\times\mathcal{X}^{(1)}\times\mathcal{X}^{(2)}\times\mathcal{X}^{(3)}$ converging in $\mathcal{X}^{(1)}\times\mathcal{X}^{(1)}\times\mathcal{X}^{(2)}\times\mathcal{X}^{(3)}$ to (f^0, f^1, g, h) where in addition

$$\sup_k \int_0^T \Big[\big\|f^0_k(s)\big\|_{BV(\mathcal{O})} + \big\|f^1_k(s)\big\|_{BV(\mathcal{O})}\Big]\,\mathrm{d}s < \infty.$$

Then the variables $(\overline{X}_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}, \underline{X}_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}, X_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}})$ from Theorem 3.1 satisfy

$$\mathbb{E}[K(X,X,X,X)] \leq \liminf_{k\to\infty} \mathbb{E}_{\tau_k}\Big[K\Big(\overline{X}_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}, \overline{X}_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}, \underline{X}_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}, X_{\varepsilon_{n_k},h_{n_k},\tau_{n_k}}\Big)\Big], \tag{10}$$

and the random variables $(X, \overline{X}_{\varepsilon,h,\tau}, \underline{X}_{\varepsilon,h,\tau}, X_{\varepsilon,h,\tau})$ from Theorem 3.2 satisfy

$$\mathbb{E}[K(X,X,X,X)] \leq \liminf_{(\varepsilon,h,\tau)\to(0,0,0)} \mathbb{E}\big[K(X, \overline{X}_{\varepsilon,h,\tau}, \underline{X}_{\varepsilon,h,\tau}, X_{\varepsilon,h,\tau})\big]. \tag{11}$$

Obviously, if K is real bounded and (9) holds also for $-K$ then we get equalities and limits in (10) and (11). In particular, under the assumptions of Theorem 3.2,

$$K\big(X, \overline{X}_{\varepsilon,h,\tau}, \underline{X}_{\varepsilon,h,\tau}, X_{\varepsilon,h,\tau}\big) = \big\|\overline{X}_{\varepsilon,h,\tau} - X\big\|_{L^q((0,T);L^r(\mathcal{O}))} \to 0,$$

in probability as $(\varepsilon, h, \tau) \to (0,0,0)$ for every $r \in [1, \frac{d}{d-1})$ and every $q \in [1,\infty)$ such that $q(r-2) < r$.

4. *A priori* estimates

The numerical approximation (3) satisfies a discrete energy estimate.

Lemma 4.1. *Let $x^0, g \in \mathbb{L}^2$ and $T > 0$. Then there exists a constant $C > 0$ depending only on T and on the constants in $(\mathbf{B}_1)$ and in Section 2 such that the solutions of scheme (3) satisfy for any $\varepsilon, h \in (0,1]$, $N \in \mathbb{N}$*

$$\mathbb{E}\left[\left(\frac{1}{2}\sup_{i=1,\dots,N}\|X^i\|^2 + \sum_{i=1}^N\left(\frac{1}{4}\|X^i - X^{i-1}\|^2 + \tau\mathcal{J}_\varepsilon(X^i)\right)\right)^2\right] \leq C\left(\frac{1}{2} + \frac{1}{2}\|x^0\|^2 + |\mathcal{O}| + \frac{\lambda}{2}\|g\|^2\right)^2. \tag{12}$$

Proof. Analogously to Lemma 4.9 of [5], we set $v_h = X^i$ in (3) use the identity

$$\big(X^i - X^{i-1}, X^i\big) = \frac{1}{2}\|X^i\|^2 - \frac{1}{2}\|X^{i-1}\|^2 + \frac{1}{2}\|X^i - X^{i-1}\|^2,$$

and estimate using the Cauchy–Schwarz and Young's inequalities

$$\begin{aligned}\sum_{j=1}^N \big(B_j\big(X^{i-1}\big)\xi^{i,j}_\tau, X^i\big) &= \sum_{j=1}^N \big(B_j\big(X^{i-1}\big)\xi^{i,j}_\tau, X^{i-1}\big) + \sum_{j=1}^N \big(B_j\big(X^{i-1}\big)\xi^{i,j}_\tau, X^i - X^{i-1}\big)\\ &\leq \sum_{j=1}^N \big(B_j\big(X^{i-1}\big)\xi^{i,j}_\tau, X^{i-1}\big) + \frac{1}{4}\big\|X^i - X^{i-1}\big\|^2 + \left\|\sum_{j=1}^N B_j\big(X^{i-1}\big)\xi^{i,j}_\tau\right\|^2,\end{aligned}$$

and similarly

$$\big(X^i-g,X^i\big)=\big\|X^i-g\big\|^2+\big(X^i-g,g\big)\geq\frac{1}{2}\big\|X^i-g\big\|^2-\frac{1}{2}\|g\|^2.$$

After rearranging and absorbing the corresponding terms in the left hand side we obtain using the convexity of $f(s)=\sqrt{s^2+\varepsilon^2}$ that

$$\frac{1}{2}\|X^i\|^2-\frac{1}{2}\|X^{i-1}\|^2+\frac{1}{4}\|X^i-X^{i-1}\|^2+\tau\mathcal{J}_\varepsilon(X^i)\leq\tau\mathcal{J}_\varepsilon(0)+\sum_{j=1}^N\big(B_j(X^{i-1}),X^{i-1}\big)\xi_\tau^{i,j}+\left\|\sum_{j=1}^N B_j(X^{i-1})\xi_\tau^{i,j}\right\|^2,$$

for $1\leq i\leq N$.

We define

$$a^i=\frac{1}{2}+\frac{1}{2}\|X^i\|^2+\sum_{j=1}^i\left(\frac{1}{4}\big\|X^j-X^{j-1}\big\|^2+\tau\mathcal{J}_\varepsilon(X^j)\right)$$

$$b=\tau\mathcal{J}_\varepsilon(0),\quad c^i=\sum_{j=1}^N\big(B_j\big(X^{i-1}\big),X^{i-1}\big)\xi_\tau^{i,j},\quad d^i=\left\|\sum_{j=1}^N B_j\big(X^{i-1}\big)\xi_\tau^{i,j}\right\|^2,$$

and observe that

$$a^i-a^{i-1}\leq b+c^i+d^i,\qquad i=1,\ldots,N.$$

If $a^{i-1}\in L^2(\Omega)$ then $X^{i-1}\in L^4(\Omega;\mathbb{L}^2)$, $c^i\in L^2(\Omega)$, $d^i\in L^2(\Omega)$ and so, by induction, $a^i\in L^2(\Omega)$ for every $i=0,\ldots,N$. Next, observe that c^i is a square integrable martingale difference and that

$$\mathbb{E}\big[(c^i)^2\big]\leq C_B\tau\mathbb{E}\Big[\big(a^{i-1}\big)^2\Big],\qquad\mathbb{E}\big[(d^i)^2\big]\leq C_B\tau^2\mathbb{E}\Big[\big(a^{i-1}\big)^2\Big],$$

where C_B depends only on the growth constant in $(\mathbf{B}_1)$ and the constant in the assumption $\mathbb{E}\big[(\xi_\tau^{i,j})^4\big]\leq C\tau^2$. Let us define

$$a_*^i=\max_{j=0,\ldots,i}a^j\qquad i=0,\ldots,N.$$

Then

$$a_*^i\leq\big(a^0+Nb\big)+\max_{j=1,\ldots,i}\left|\sum_{\ell=1}^j c^\ell\right|+\sum_{j=1}^i d^j\qquad i=1,\ldots,N,$$

and

$$\big(a_*^i\big)^2\leq 3\big(a^0+Nb\big)^2+3\max_{j=1,\ldots,i}\left|\sum_{\ell=1}^j c^\ell\right|^2+3N\sum_{j=1}^i(d^j)^2\qquad i=1,\ldots,N.$$

Hence, by the discrete Burkholder–Davis–Gundy inequality, we obtain

$$\begin{aligned}\mathbb{E}\Big[\big(a_*^i\big)^2\Big]&\leq 3\big(a^0+Nb\big)^2+3C_2\sum_{j=1}^i\mathbb{E}\big[(c^j)^2\big]+3N\sum_{j=1}^i\mathbb{E}\big[(d^j)^2\big]\\&\leq 3\big(a^0+Nb\big)^2+3C_2\mathbf{c}\tau\sum_{j=1}^i\mathbb{E}\big[(a^{j-1})^2\big]+3\mathbf{c}\tau^2N\sum_{j=1}^i\mathbb{E}\big[(a^{j-1})^2\big]\\&\leq 3\big(a^0+Nb\big)^2+\frac{K_{C,T}}{N}\sum_{j=1}^i\mathbb{E}\big[(a_*^{j-1})^2\big]\qquad i=1,\ldots,N,\end{aligned}$$

and we get the result by the discrete Gronwall lemma. □

Next, we estimate the discrete time increments of the numerical solution.

Lemma 4.2. *For any* $0 \le n \le \ell + n \le N$ *it holds that*

$$\mathbb{E}\Big[\big\|X^{n+\ell} - X^n\big\|_{\mathbb{H}^{-1}}^4\Big] \le C t_\ell^2,$$

where C *does not depend on* ε, h, τ.

Proof. Clearly, it suffices to treat the case $\ell \ge 1$. For any $v \in \mathbb{H}_0^1$ we set $v_h = P_h v$ in (3) and get after summing up for $i = n+1, \dots, n+\ell$ by the definition of projection P_h that

$$\begin{aligned}\big(X^{n+\ell} - X^n, v_h\big) = \big(X^{n+\ell} - X^n, P_h v\big) &\le \tau \sum_{i=n+1}^{n+\ell} \left\| \frac{\nabla X^i}{\sqrt{|\nabla X^i|^2 + \varepsilon^2}} \right\| \|\nabla P_h v\| \\ &+ \tau \sum_{i=n+1}^{n+\ell} \lambda\big(\|X^i\| + \|g\|\big)\|P_h v\| + \left\| \sum_{i=n+1}^{n+\ell} \sum_{j=1}^{N} B_j\big(X^{i-1}\big)\xi_\tau^{i,j} \right\| \|P_h v\| \qquad \forall v \in \mathbb{H}_0^1.\end{aligned}$$

On noting that that $\left|\frac{\nabla \cdot}{\sqrt{|\nabla \cdot|^2 + \varepsilon^2}}\right| \le 1$ we deduce by the stability of the $\mathbb{L}^2$ projection $\|P_h v\|_{\mathbb{H}_0^1} \le \kappa \|v\|_{\mathbb{H}_0^1}$ that

$$\big\|X^{n+\ell} - X^n\big\|_{\mathbb{H}^{-1}} \le C t_\ell \Big[1 + \max_{i=1,\dots,N}\big\{\big\|X^i\big\|\big\}\Big] + \left\| \sum_{i=n+1}^{n+\ell} \sum_{j=1}^{N} B_j(X^{i-1})\xi_\tau^{i,j} \right\|.$$

Hence, we obtain

$$\mathbb{E}\Big[\big\|X^{n+\ell} - X^n\big\|_{\mathbb{H}^{-1}}^4\Big] \le c t_\ell^4 + c t_\ell^2,$$

by the Burkholder–Rosenthal inequality (see *e.g.*, [20], Thm. 5.50), Lemma 4.1 and linear growth of $B : \mathbb{L}^2 \to \mathscr{L}_2(\ell_2; \mathbb{L}^2)$. Indeed, the martingale difference

$$\mathbf{d}_i = \sum_{j=1}^{N} B_j\big(X^{i-1}\big)\xi_\tau^{i,j},$$

satisfies for $p \in \{2, 4\}$

$$\begin{aligned}\sum_{i=n+1}^{n+\ell} \mathbb{E}\left[\|\mathbf{d}_i\|^p | \mathcal{F}_\tau^{i-1}\right] &\le c_\kappa \tau^{\frac{p}{2}} \sum_{i=n+1}^{n+\ell} \left[\sum_{j=1}^{N} \|B_j\big(X^{i-1}\big)\|^2\right]^{\frac{p}{2}} \le c_\kappa \tau^{\frac{p}{2}} \sum_{i=n+1}^{n+\ell} \big\|B\big(X^{i-1}\big)\big\|_{\mathscr{L}_2(\ell_2;\mathbb{L}^2)}^p \\ &\le c_\kappa \tau^{\frac{p}{2}} \sum_{i=n+1}^{n+\ell} \big[1 + \big\|X^{i-1}\big\|\big]^p \le c_\kappa t_\ell^{\frac{p}{2}} \Big[1 + \max_{i=1,\dots,N}\big\{\|X^i\|^p\big\}\Big].\end{aligned}$$

The Burkholder–Rosenthal inequality then yields

$$\mathbb{E}\left[\left\|\sum_{i=n+1}^{n+\ell} \mathbf{d}_i\right\|^4\right] \le \beta\, \mathbb{E}\left[\left(\sum_{i=n+1}^{n+\ell} \mathbb{E}\big[\|\mathbf{d}_i\|^2 | \mathcal{F}_\tau^{i-1}\big]\right)^2\right] + \beta \sum_{i=n+1}^{n+\ell} \mathbb{E}\big[\|\mathbf{d}_i\|^4\big] \le c t_\ell^2,$$

which completes the proof. ∎

Lemma 4.3. *Let $u^0 \in \mathbb{V}_h$, $G^1, \dots, G^N$ and $H^{0,j}, \dots, H^{N-1,j}$ be $\mathcal{F}_\tau$-adapted random variables in $L^2(\Omega; \mathbb{V}_h)$ for every $j \in \{1, \dots, N\}$ and define*

$$U^i - U^{i-1} = -\tau G^i + \sum_{j=1}^{N} H^{i-1,j} \xi_\tau^{i,j}, \qquad i = 1, \dots, N. \tag{13}$$

Then

$$\frac{1}{2}\mathbb{E}\Big[\big\|X^i - U^i\big\|^2\Big] + \tau \sum_{\ell=1}^{i} \mathbb{E}\big[\mathcal{J}_\varepsilon(X^\ell)\big] \le \frac{1}{2}\big\|x^0 - u^0\big\|^2 + \tau \sum_{\ell=1}^{i} \mathbb{E}\big[\mathcal{J}_\varepsilon(U^\ell) + \big(G^\ell, X^\ell - U^\ell\big)\big]$$
$$+ \frac{\tau}{2} \sum_{\ell=1}^{i} \sum_{j=1}^{N} \mathbb{E}\Big[\big\|P_h B_j\big(X^{\ell-1}\big) - H^{\ell-1,j}\big\|^2\Big], \quad 0 \le i \le N.$$

Proof. On noting (3), (13) we deduce that $D^i = X^i - U^i$ satisfies

$$\big(D^i - D^{i-1}, v_h\big) + \tau \left(\frac{\nabla X^i}{\sqrt{|\nabla X^i|^2 + \varepsilon^2}}, \nabla v_h \right) + \tau\lambda\big(X^i - g, v_h\big) = \tau\big(G^i, v_h\big) + \sum_{j=1}^{N} \big(B_j\big(X^{i-1}\big) - H^{i-1,j}, v_h\big)\xi_\tau^{i,j},$$

for all $v_h \in \mathbb{V}_h$.

We then set $v_h = D^i$ in the above expression, use that $(B_j(X^{i-1}), D^i) = (P_h B_j(X^{i-1}), D^i)$ and proceed analogously to the first part of the proof of Lemma 4.1 to obtain

$$\frac{1}{2}\|D^i\|^2 - \frac{1}{2}\|D^{i-1}\|^2 + \tau\mathcal{J}_\varepsilon(X^i) \le \tau\mathcal{J}_\varepsilon(U^i) + \tau\big(G^i, D^i\big) + \sum_{j=1}^{N} \big(B_j\big(X^{i-1}\big) - H^{i-1,j}, D^{i-1}\big)\xi_\tau^{i,j}$$
$$+ \frac{1}{2}\left\| \sum_{j=1}^{N} (P_h B_j(X^{i-1}) - H^{i-1,j})\xi_\tau^{i,j} \right\|^2,$$

for $1 \le i \le N$. The statement of the lemma then follows after summing up the above expression, taking the expectation and using the properties of $\xi_\tau^{i,j}$. ☐

5. Compactness in locally convex path spaces

In this section, Y stands for a Hausdorff locally convex space (typically a Hilbert space equipped with the strong or the weak topology), $Y^{[0,T]}$ denotes the space of functions from $[0,T]$ to Y on which we consider the topology of uniform convergence $\tau_{\mathbf{u}}$. We also define the subspaces $Q_n([0,T];Y)$, $n \in \mathbb{N}$ spanned by the functions $f \in Y^{[0,T]}$ that are constant on every interval (t_{i-1}^n, t_i^n) for $1 \le i \le n$ where $t_j^k = jT/k$, the Hausdorff locally convex path spaces

$$Q_\infty([0,T];Y) = \bigcup_{n=1}^{\infty} Q_n([0,T];Y), \qquad Q([0,T];Y) = \overline{Q_\infty([0,T];Y)},$$

and an important F_σ subset of $Q([0,T];Y)$

$$Q_c([0,T];Y) = Q_\infty([0,T];Y) \cup C([0,T];Y),$$

that contains both step-functions on equidistant partitions of $[0,T]$ and continuous functions, equipped with the uniform convergence topology, that is best suitable for our purposes in the sequel when piecewise constant processes will converge uniformly to a continuous process. The space of continuous Y-valued functions $C([0,T];Y)$ is also equipped with the topology of uniform convergence.

Further, we define the space

$$Q_{c,BV}([0,T];\mathbb{L}^2_w) = \left\{ f \in Q_c([0,T];\mathbb{L}^2_w) : \int_0^T \|f(s)\|_{BV(\mathcal{O})}\, \mathrm{d}s < \infty \right\},$$

as an F_σ subset of $Q([0,T];\mathbb{L}^2_w)$.

Finally, if M is a subset of $Q([0,T];Y)$, we define

$$M_n^{\uparrow} = M \setminus \bigcup_{m=1}^{n-1} Q_m([0,T];Y).$$

Remark 5.1. Every $f \in Q([0,T];Y)$, as a uniform limit of functions in $Q_\infty([0,T];Y)$, is bounded and continuous with an exception of an at most countable subset of rational numbers, and consequently f is Borel measurable.

Remark 5.2. If Y is sequentially complete then $Q([0,T];Y)$ coincides with the space of functions $f \in Y^{[0,T]}$ that are continuous at every $t \in (T\mathbb{Q}^{\complement}) \cap [0,T]$ and that have right and left limits at every $t \in [0,T]$.

Remark 5.3. The space $Q([0,T];Y)$ can be also equipped (alternatively) with the Skorokhod topology defined by neighborhoods

$$N_{O,\varepsilon}(f) = \{g : \exists \mu \text{ such that } \gamma(\mu) < \varepsilon \text{ and } g(\mu(t)) - f(t) \in O \text{ for every } t \in [0,T]\},$$

where O is an absolutely convex neighborhood of zero in Y, $\varepsilon > 0$, μ is an increasing bi-Lipschitz continuous homeomorphisms of $[0,T]$ onto $[0,T]$ and $\gamma(\mu) = \|\log \mu'\|_{L^\infty}$. But the Skorokhod topology is strictly weaker than the topology of uniform convergence. In other words, convergence in $Q([0,T];Y)$ implies convergence in the Skorokhod topology but not *vice versa*. Thus, for our purposes, the space $Q([0,T];Y)$ with the topology of uniform convergence is the better choice.

In the next theorem, we characterize compact sets in $Q_c([0,T];Y)$ which play an essential role in this paper. To this end, we present an Arzela–Ascoli theorem.

Theorem 5.1. *Let M be a non-empty subset in $Q([0,T];Y)$ and consider the following:*

(i) $\{f(t) : f \in M\}$ *is relatively compact in Y for every $t \in [0,T]$;*
(ii) *for every O being a neighbourhood of zero in Y, there exist $m \in \mathbb{N}$ and $\delta > 0$ such that*

$$\forall |t-s| \le \delta \quad \text{and} \quad \forall f \in M_m^{\uparrow} \quad \text{one has} \quad f(t) - f(s) \in O;$$

(iii) *the closure of M in $(Y^{[0,T]}, \tau_{\mathbf{u}})$ is a compact subset of $Q_c([0,T];Y)$;*
(iv) $\{f(t) : t \in [0,T],\ f \in M\}$ *is relatively compact in Y.*

Then

$$[\text{(i)} \quad \& \quad \text{(ii)}] \quad \Longleftrightarrow \quad \text{(iii)},$$
and
$$\text{(iii)} \quad \Longrightarrow \quad \text{(iv)}.$$

Proof. See Section A.1. □

Remark 5.4. If Y is sequentially complete and M is relatively compact in $Q([0,T];Y)$ then (iv) in Theorem 5.1 still holds with the same proof.

Corollary 5.1. *If compacts of Y are metrizable and M satisfies (iv) in Theorem 5.1 then M is also metrizable.*

Proof. See Section A.2. □

Now we provide an easy test for checking Borel measurability of $Q_c([0,T];Y)$-valued random variables. It turns out that pointwise measurability and Borel measurability coincide for mappings with a σ-compact range in $Q_c([0,T];Y)$ provided that compact sets in Y are metrizable.

Corollary 5.2. *Let compacts of Y be metrizable and let M be σ-compact in $Q_c([0,T];Y)$. Then*

$$V \in \mathscr{B}(Q([0,T];Y)) \quad \Longleftrightarrow \quad V \in \mathcal{Y}_T,$$

holds for every $V \subseteq M$ where

$$\mathcal{Y}_T = \sigma(\pi_s : \, s \in (T\mathbb{Q}) \cap [0,T]), \qquad \pi_s : Q([0,T];Y) \to Y : f \mapsto f(s).$$

Proof. See Section A.3. □

Remark 5.5. Let us recall that a compact K is metrizable if and only if there exists a countable family of real continuous functions on K separating points of K (see *e.g.*, [16]). In case of Hausdorff locally convex spaces Y, those functions can be chosen in such a way that they are linear and continuous on Y. Hence compacts are metrizable in all spaces where there exists a countable family of continuous functions separating points of that space. In particular, compact sets are metrizable *e.g.*, in analytic spaces (see *e.g.*, [9], Cor. 6.7.8) among which all separable Fréchet spaces equipped with any locally convex topology weaker than or equal to the metric one belong.

Example 5.1. If K is a set in Y then we denote by $C_n([0,T];K)$ the space of functions $f : [0,T] \to K$ that satisfy

$$f(t) = \frac{t - t_{i-1}}{\tau} f(t_i) + \frac{t_i - t}{\tau} f(t_{i-1}), \qquad t \in [t_{i-1}, t_i],$$

for every $i \in \{1, \ldots, n\}$ where $t_i = i\tau$ and $\tau = T/n$. If K is compact then $C_n([0,T];K)$ is compact in $C([0,T];Y)$.

Proof. Indeed, $C_n([0,T];K)$ is closed. Now, if O is an absolutely convex neighbourhood of zero then $K \subseteq \lambda O$ for some $\lambda > 0$, and so $f(t) - f(s) \in 2\lambda\tau^{-1}(t-s)O$ holds for every $s, t \in [0,T]$ and every $f \in C_n([0,T];K)$. Hence $C_n([0,T];K)$ is relatively compact by Theorem 5.1. □

We will need the following version of the Prokhorov theorem.

Theorem 5.2. *Let Z be a completely regular topological space, let $\{\mu_n\}$ be Borel probability measures such that there exist metrizable compacts K_j such that*

$$\sup_j \left[\inf_n \, \mu_n(K_j) \right] = 1.$$

Then there exists a subsequence $\{\mu_{n_k}\}$ that converges to a Radon probability measure μ on Z.

Proof. See Theorem 8.6.7. of [9]. □

Weak convergence of tight probability measures is actually more powerful than it might seem. Let us present a reinforcement of the Portmanteau theorem, *cf.*, Lemma 1.10 of [19].

Proposition 5.1. *Let Z be a completely regular topological space, let $\{\mu_n\}$ and μ be Radon probability measures on Z such that $\langle f, \mu_n\rangle \to \langle f, \mu\rangle$ for every $f \in C_b(Z)$ and, for every $r > 0$, there exist metrizable closed sets $K_{r,n} \searrow K_{r,\infty}$ such that*

$$\mu_n(K_{r,n}) \geq 1 - r, \qquad n \in \mathbb{N}.$$

Let $F_n, F : Z \to [-\infty, \infty]$ be such that $F_n|_{K_{r,n}}$, $F|_{K_{r,n}}$ are $\mathscr{B}(K_{r,n})$-measurable for every $r > 0$ and $n \in \mathbb{N}$, and denote by μ^* the outer measure associated with μ. Then F_n is μ_n-measurable for every $n \in \mathbb{N}$, F is μ-measurable and the following holds:

(1) *If F_n and F are non-negative and $\mu^*(D_r) = 0$ for every $r \in (0, 1)$ where*

$$D_r = \{x \in K_{r,\infty} : \exists x_n \in K_{r,n},\, x_n \to x,\, \liminf F_n(x_n) < F(x)\},$$

then

$$\int_Z F \,\mathrm{d}\mu \leq \liminf \int_Z F_n \,\mathrm{d}\mu_n.$$

(2) *If $\mu^*(D_r) = 0$ for every $r \in (0, 1)$ where*

$$D_r = \{x \in K_{r,\infty} : \exists x_n \in K_{r,n},\, x_n \to x,\, \limsup |F_n(x_n) - F(x)| > 0\},$$

and

$$\lim_{R\to\infty} \left[\sup_{n\in\mathbb{N}} \int_{[|F_n|>R]} |F_n| \,\mathrm{d}\mu_n \right] = 0,$$

then

$$\lim \int_Z F_n \,\mathrm{d}\mu_n = \int_Z F \,\mathrm{d}\mu.$$

Proof. See Section A.4. □

6. Tightness properties of the numerical approximation

We consider the interpolants X_τ, $\overline{X}_\tau$, $\underline{X}_\tau$ defined in (6), (7), respectively. As in the previous section, to simplify the notation, the dependence of X_τ, $\overline{X}_\tau$ and $\underline{X}_\tau$ on ε, h and τ will not be displayed for clarity reasons until it matters.

The next lemma is a direct consequence of the *a priori* estimates in Lemma 4.1.

Lemma 6.1. *The interpolants of the numerical solution of the scheme (3) satisfy the following bounds:*

$$\mathbb{E}\Big[\|\overline{X}_\tau\|^2_{L^1(0,T;\mathbb{W}^{1,1}_0)}\Big] \leq C,\ \mathbb{E}\Big[\|\underline{X}_\tau\|^2_{L^1(\tau,T;\mathbb{W}^{1,1}_0)}\Big] \leq C, \mathbb{E}\Big[\|X_\tau\|^2_{L^1(\tau,T;\mathbb{W}^{1,1}_0)}\Big] \leq C, \tag{14}$$

$$\mathbb{E}\Big[\|\overline{X}_\tau\|^4_{L^\infty(0,T;\mathbb{L}^2)}\Big] \leq C,\ \mathbb{E}\Big[\|\underline{X}_\tau\|^4_{L^\infty(0,T;\mathbb{L}^2)}\Big] \leq C,\ \mathbb{E}\Big[\|X_\tau\|^4_{C([0,T];\mathbb{L}^2)}\Big] \leq C, \tag{15}$$

$$\mathbb{E}\Big[\|\overline{X}_\tau - X_\tau\|^4_{L^q(0,T;\mathbb{L}^2)}\Big] \leq C\tau^{\frac{4}{q}},\ \mathbb{E}\Big[\|\overline{X}_\tau - \underline{X}_\tau\|^4_{L^q(0,T;\mathbb{L}^2)}\Big] \leq C\tau^{\frac{4}{q}}, \tag{16}$$

where C does not depend on ε, h, τ and $q \in [2, \infty]$.

Furthermore, by Lemma 4.2 the following time-fractional bounds hold for the piecewise linear interpolant.

Lemma 6.2. *Let $\mathbf{m}$ denote the modulus of continuity of $\mathbb{H}^{-1}$-valued functions on $[0, T]$*

$$\mathbf{m}(f, \delta) := \sup\{\|f(t) - f(s)\|_{\mathbb{H}^{-1}} : s, t \in [0, T],\, |t - s| \leq \delta\}.$$

Then the following estimate holds for $\alpha \in (0, \frac{1}{2})$ and $s \in (0, \frac{1}{4})$

$$\mathbb{E}\Big[\|X_\tau\|^4_{W^{\alpha,4}(0,T;\mathbb{H}^{-1})}\Big] \le C, \qquad \mathbb{E}\Big[\sup_{\delta>0}\{\delta^{-s}\mathbf{m}(X_\tau,\delta)\}\Big] \le C,$$

$$\mathbb{E}\Big[\sup_{\delta>0}\{(\delta+\tau)^{-s}\mathbf{m}(\overline{X}_\tau,\delta)\}\Big] \le C, \qquad \mathbb{E}\Big[\sup_{\delta>0}\{(\delta+\tau)^{-s}\mathbf{m}(\underline{X}_\tau,\delta)\}\Big] \le C,$$

where C does not depend on ε, h, τ.

Proof. Use Lemmas 4.2, C.1 and the inequality

$$\max\{\mathbf{m}(\overline{X}_\tau,\delta),\mathbf{m}(\underline{X}_\tau,\delta)\} \le \mathbf{m}(X,\delta+\tau), \qquad \delta > 0.$$

□

With the notation and the parameters from Lemma 6.2, for $R > 0$ and $a \in [0,T]$, writing shortly Q_n for $Q_n([0,T];\mathbb{L}^2_w)$ and C for $C([0,T];\mathbb{L}^2_w)$, we consider the sets

$$V_{R,n,a} = \left\{ f \in Q_n : \sup_{t\in[0,T]} \|f(t)\| \le R, \quad \sup_{\delta>0} \frac{\mathbf{m}(f,\delta)}{(\delta+T/n)^s} \le R, \int_a^T \|f(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s \le R \right\},$$

$$V_{R,\infty,a} = \left\{ f \in C : \sup_{t\in[0,T]} \|f(t)\| \le R, \quad \sup_{\delta>0} \frac{\mathbf{m}(f,\delta)}{\delta^s} \le R, \int_a^T \|f(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s \le R \right\},$$

$$V^m_{R,b} = V_{R,\infty,b_*} \cup \bigcup_{n\in[m,\infty]} V_{R,n,b_n}, \qquad b_* := \limsup_{n\to\infty} b_n.$$

Proposition 6.1. *The random variables $\overline{X}_\tau$, $\underline{X}_\tau$, X_τ are Borel measurable as mappings from $(\Omega_\tau, \mathscr{F}_\tau, \mathbb{P}_\tau)$ to $Q_c([0,T];\mathbb{L}^2_w)$. The sets $V^m_{R,a}$ and V_{R,n,a_n} are compact in $Q_c([0,T];\mathbb{L}^2_w)$ and the sets V_{R,∞,a_∞} are compact in $C([0,T];\mathbb{L}^2_w)$ for every $m,n \in \mathbb{N}$ and $a \in [0,T]^{\mathbb{N}\cup\{\infty\}}$. Furthermore*

$$\mathbb{P}_\tau\left[\overline{X}_\tau \notin V_{R,N,0}\right] \le \frac{C}{R}, \qquad \mathbb{P}_\tau\left[\underline{X}_\tau \notin V_{R,N,T/N}\right] \le \frac{C}{R}, \qquad \mathbb{P}_\tau\left[X_\tau \notin V_{R,\infty,T/N}\right] \le \frac{C}{R},$$

holds for every $R > 0$ where C does not depend on ε, h, τ and R. In particular, the laws

$$\mathbb{P}_\tau\left[\overline{X}_{\varepsilon,h,\tau} \in \cdot\right], \qquad \mathbb{P}_\tau\left[\underline{X}_{\varepsilon,h,\tau} \in \cdot\right] \qquad \mathbb{P}_\tau\left[X_{\varepsilon,h,\tau} \in \cdot\right],$$

are tight on $Q_{c,BV}([0,T];\mathbb{L}^2_w)$, $Q_c([0,T];\mathbb{L}^2_w)$ and $C([0,T];\mathbb{L}^2_w)$ resp. with respect to ε, h, τ.

Proof. $\overline{X}_\tau$, $\underline{X}_\tau$ and X_τ are clearly $\mathcal{Y}_T$-measurable and $Q_N([0,T];\mathbb{L}^2_w)$ and $C_N([0,T];\mathbb{L}^2_w)$ are σ-compact in $C([0,T];\mathbb{L}^2_w)$ by Theorem 5.1 and Example 5.1. Hence $\overline{X}_\tau$, $\underline{X}_\tau$ and X_τ are Borel measurable by Corollary 5.2 as compact sets in $Q_c([0,T];\mathbb{L}^2_w)$ and $C([0,T];\mathbb{L}^2_w)$ are metrizable (Rem. 5.5). Now the sets V_{R,n,a_n} and $V_{R,a}$ are closed and relatively compact in $Q_c([0,T];\mathbb{L}^2_w)$ and the sets V_{R,∞,a_∞} are closed and relatively compact in $C([0,T];\mathbb{L}^2_w)$ by Theorem 5.1, as the weak topology and the $\mathbb{H}^{-1}$-topology coincide on bounded sets in $\mathbb{L}^2$. In the proof of closedness of the above sets, we use the fact that there exists a countable set $\mathcal{H}$ of smooth compactly supported functions such that

$$\|g\|_{BV(\mathcal{O})} = \sup\{(g,\varphi) : \varphi \in \mathcal{H}\}, \qquad \text{for } g \in L^1_{loc}(\mathcal{O}),$$

holds, *e.g.*, by Proposition 3.6 of [1] and by separability of $C^\infty_c(\mathcal{O})$, see *e.g.*, [13]. Hence

$$f \mapsto \int_0^T \|f(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s,$$

as a supremum of continuous functions, is lower semicontinuous on $Q([0,T];\mathbb{L}^2_w)$.

The tightness then follows directly from Lemmas 6.1 and 6.2. □

In the next lemma we obtain the convergence of the noise variables to a Wiener process.

Lemma 6.3. *Let W^j_τ, $1 \le j \le N$ be the piecewise linear processes on $[0,T]$ defined by*

$$W^j_\tau(t_i) = \sum_{\ell=1}^{i} \xi^{\ell,j}_\tau, \qquad 0 \le i \le N,$$

and W^j_τ is linear on $[t_{i-1}, t_i]$ for every $0 < i \le N$ where $\tau = T/N$ and $t_i = i\tau$. We also define $W^j_\tau = 0$ for $j > N$. Then the laws of W^j_τ converge to the Wiener measure on $C\,[0,T]$ as $\tau \to 0$, for every $j \in \mathbb{N}$.

Proof. Let $s \in (1/4, 1/2)$. Then,

$$\mathbb{E}\Big[\big|W^j_\tau(t_n) - W^j_\tau(t_{n-\ell})\big|^4\Big] \le C_\kappa t_\ell^2, \qquad 1 \le \ell \le n \le N,$$

hence, by Lemma C.1 we get

$$\mathbb{E}\Big[\,\|W^j_\tau\|^4_{B^s_{4,4}(0,T)}\Big] \le C_{\kappa,s,T}. \tag{17}$$

In particular, since $B^s_{4,4}(0,T)$ is embedded compactly in $C^\alpha([0,T])$ for every $0 < \alpha < s - \frac{1}{4}$ *e.g.*, by Corollary 26 of [21], the laws of $\{W^j_\tau\}$ are tight on $\mathscr{B}(C([0,T]))$. Since $(W^j_\tau(s_0), \dots, W^j_\tau(s_k))$ converge in law to the law of $(W(s_0), \dots, W(s_k))$ where W is a Wiener process, *e.g.*, by Theorem 18.2 in [8], we get the claim. □

Let us consider the completely regular space with metrizable compacts (see Rem. 5.5)

$$\mathbf{Z} = Q_{c,BV}\big([0,T];\mathbb{L}^2_w\big) \times Q_c\big([0,T];\mathbb{L}^2_w\big) \times C\big([0,T];\mathbb{L}^2_w\big) \times C([0,T]) \times C([0,T]) \times C([0,T]) \times \dots,$$

define the projections

$$\begin{aligned} &S^1 : \mathbf{Z} \to Q_{c,BV}([0,T];\mathbb{L}^2_w) && (f^1, f^2, f^3, w^1, w^2, w^3, \dots) \mapsto f^1,\\ &S^2 : \mathbf{Z} \to Q_c([0,T];\mathbb{L}^2_w) && (f^1, f^2, f^3, w^1, w^2, w^3, \dots) \mapsto f^2,\\ &S^3 : \mathbf{Z} \to C([0,T];\mathbb{L}^2_w) && (f^1, f^2, f^3, w^1, w^2, w^3, \dots) \mapsto f^3,\\ &W^j : \mathbf{Z} \to C([0,T]) && \big(f^1, f^2, f^3, w^1, w^2, w^3, \dots\big) \mapsto w^j, \end{aligned} \tag{18}$$

and the canonical filtration on $\mathbf{Z}$

$$\mathscr{Z}_t = \sigma\big(S^1_s, S^2_s, S^3_s, W^j_s : s \in [0,t],\, j \in \mathbb{N}\big), \qquad t \in [0,T].$$

If ν is a probability measure on $\mathscr{B}(\mathbf{Z})$ then $\mathscr{Z}^\nu_t$ stands for the augmentation of $\mathscr{Z}_t$ by ν-negligible Borel sets.

Corollary 6.1. *The random variables*

$$Z_{\varepsilon,h,\tau} = \big(\overline{X}_{\varepsilon,h,\tau}, \underline{X}_{\varepsilon,h,\tau}, X_{\varepsilon,h,\tau}, W^1_\tau, W^2_\tau, W^3_\tau, \dots\big),$$

are Borel measurable as mappings from $(\Omega_\tau, \mathscr{F}_\tau, \mathbb{P}_\tau)$ to $\mathbf{Z}$ and their laws under $\mathbb{P}_\tau$ are tight on $\mathscr{B}(\mathbf{Z})$ with respect to ε, h, τ. In particular, every sequence $(\varepsilon_n, h_n, \tau_n)$ has a subsequence $(\varepsilon_{n_k}, h_{n_k}, \tau_{n_k})$ such that laws of $Z_{\varepsilon_{n_k}, h_{n_k}, \tau_{n_k}}$ under $\mathbb{P}_{\tau_{n_k}}$ converge to a Radon probability measure μ on $\mathscr{B}(\mathbf{Z})$.

Proof. Since $\overline{X}_{\varepsilon,h,\tau}$, $\underline{X}_{\varepsilon,h,\tau}$ and $X_{\varepsilon,h,\tau}$ take values in σ-compact subsets of $Q_{c,BV}([0,T];\mathbb{L}^2_w)$, $Q_c([0,T];\mathbb{L}^2_w)$ and $C([0,T];\mathbb{L}^2_w)$ respectively by Theorem 5.1 and Example 5.1 and the fact that compact sets in all these spaces are metrizable (Rem. 5.5), we get that $Z_{\varepsilon,h,\tau}$ is Borel measurable *e.g.*, by Lemma 6.4.2/ii of [9]. Tightness follows from Proposition 6.1 and Lemma 6.3 and convergence of a subsequence by Theorem 5.2. □

7. Construction of a probabilistically weak SVI solution

Thanks to Corollary 6.1, in the sequel, we choose a subsequence $(\varepsilon_k, h_k, \tau_k) \to (0,0,0)$ such that the Borel laws of $Z_k = Z_{\varepsilon_k,h_k,\tau_k}$ under $\mathbb{P}_{\tau_k}$ converge to a Radon probability measure μ on $\mathscr{B}(\mathbf{Z})$.

Lemma 7.1. *Let $F_k, F : \mathbf{Z} \to [-\infty,\infty]$ be such that $F_k|_K$ and $F|_K$ are $\mathscr{B}(K)$-measurable for every compact K in $\mathbf{Z}$ (e.g., sequentially lower semicontinuous) and every $k \in \mathbb{N}$. Further, assume that one of the following*

(a) *F_k and F are non-negative and*

$$F\big(f,g,h,w^1,w^2,\dots\big) \le \liminf_{k\to\infty} F_k\big(f_k,g_k,h_k,w_k^1,w_k^2,\dots\big),$$

(b) *$\lim_{k\to\infty} F_k\big(f_k,g_k,h_k,w_k^1,w_k^2,\dots\big) = F\big(f,g,h,w^1,w^2,\dots\big)$ and*

$$\lim_{R\to\infty}\left[\sup_{k\in\mathbb{N}} \mathbb{E}_{\tau_k}\left[\mathbf{1}_{[|F_k(Z_k)|>R]}|F_k(Z_k)|\right]\right] = 0, \tag{19}$$

holds for every

(i) *$f_k \to f$ in $Q_{c,BV}([0,T];\mathbb{L}^2_w)$, $\sup_k \int_0^T \|f_k(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s < \infty$, $f \in C([0,T];\mathbb{L}^2_w)$,*
(ii) *$g_k \to g$ in $Q_c([0,T];\mathbb{L}^2_w)$, $\sup_k \int_{\tau_k^*}^T \|g_k(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s < \infty$, $g \in C([0,T];\mathbb{L}^2_w) \cap Q_{c,BV}([0,T];\mathbb{L}^2_w)$,*
(iii) *$h_k \to h$ in $C([0,T];\mathbb{L}^2_w)$, $\sup_k \int_{\tau_k^*}^T \|h_k(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s < \infty$, $h \in Q_{c,BV}([0,T];\mathbb{L}^2_w)$,*
(iv) *$w_k^j \to w^j$ in $C([0,T])$ for every $j \in \mathbb{N}$*

where $\tau_k^* = \max\{\tau_i : i \ge k\}$. If (a) holds then

$$\int_{\mathbf{Z}} F\,\mathrm{d}\mu \le \liminf_{k\to\infty} \mathbb{E}_{\tau_k}[F_k(Z_k)].$$

If (b) holds then

$$\int_{\mathbf{Z}} F\,\mathrm{d}\mu = \lim_{k\to\infty} \mathbb{E}_{\tau_k}[F_k(Z_k)].$$

Proof. The sets

$$\mathcal{K}_{R,n} = \left[\bigcup_{m\in[n,\infty]} V_{R,m,0}\right] \times \left[\bigcup_{m\in[n,\infty]} V_{R,m,T/m}\right] \times V_{R,\infty,T/n} \times C([0,T]) \times C([0,T]) \times \dots,$$

are closed, metrizable and decreasing in the second variable,

$$\mathcal{K}_{R,\infty} := \bigcap_{n=1}^{\infty} \mathcal{K}_{R,n} = V_{R,\infty,0} \times V_{R,\infty,0} \times V_{R,\infty,0} \times C([0,T]) \times C([0,T]) \times \dots,$$

and

$$\mathbb{P}_{\tau_k}\left[Z_k \notin \mathcal{K}_{R,T/\tau_k^*}\right] \le \mathbb{P}_{\tau_k}\left[Z_k \notin \mathcal{K}_{R,T/\tau_k}\right] \le \frac{C}{R},$$

by Proposition 6.1. The rest follows from Proposition 5.1. □

Remark 7.1. From the definition of the topological space $\mathbf{Z}$ we observe that $\overline{X}_{\varepsilon_k,h_k,\tau_k}$ converges in a significantly stronger (hence better) sense than $\underline{X}_{\varepsilon_k,h_k,\tau_k}$ and $X_{\varepsilon_k,h_k,\tau_k}$. This is due to the fact that we have $L^1(0,T;\mathbb{W}^{1,1}_0)$ apriori bounds for the processes $\overline{X}_{\varepsilon_k,h_k,\tau_k}$ in Lemma 6.1 that are not available on the full interval $[0,T]$ for $\underline{X}_{\varepsilon_k,h_k,\tau_k}$ and $X_{\varepsilon_k,h_k,\tau_k}$.

Corollary 7.1. *If $\alpha \in (0, \frac{1}{2})$ then the following holds:*

(I) *The $\mathbb{L}^2$-valued processes S^1, S^2, S^3 and the real-valued processes $(W^k)_{k\in\mathbb{N}}$ are $(\mathscr{Z}_t)$-progressively measurable.*
(II) $\mu\big(S^1 = S^2 = S^3\big) = 1$.
(III) *We have*

$$\int_{\mathbf{Z}} \left[\sup_{t\in[0,T]} \|S^3(t)\|^4 + \|S^3\|^4_{W^{\alpha,4}(0,T;\mathbb{H}^{-1})} + \left(\int_0^T \|S^1(t)\|_{BV(\mathcal{O})}\,\mathrm{d}t \right)^2 \right] \mathrm{d}\mu < \infty.$$

(IV) *The σ-algebras $\mathscr{Z}_t$ and $\sigma(W^j(b) - W^j(a) : t \le a \le b \le T,\ j \in \mathbb{N})$ are μ-independent.*
(V) *The processes $W^1, W^2, W^3, \ldots$ are μ-independent $(\mathscr{Z}_t)$-Brownian motions.*

Proof. (I) follows from Remark 5.1 as the processes S^1, S^2, S^3 are continuous with an exception of an at most countable set and they are $(\mathscr{Z}_t)$-adapted by definition, *cf.*, Proposition 1.13 of [17], and (II), (III) from Lemmas 6.1, 6.2 and 7.1.

As for (IV), it suffices to realize that

$$\mathscr{Z}_t = \sigma\big((\varphi, S^1_s), (\varphi, S^2_s), (\varphi, S^3_s), W^j_s : s \in [0,t],\ j \in \mathbb{N},\ \varphi \in \mathbb{L}^2\big).$$

If $u \ge t + \tau_k$ then $\sigma(W^j(b) - W^j(a) : u \le a \le b \le T,\ j \in \mathbb{N})$ and $\mathscr{Z}_t$ are $\mathbb{P}_{\tau_k}(Z_k \in \cdot\,)$-independent, hence also μ-independent by Lemma 7.1. Consequently, $\sigma(W^j(b) - W^j(a) : t < a \le b \le T,\ j \in \mathbb{N})$ and $\mathscr{Z}_t$ are μ-independent but the former coincides with $\sigma(W^j(b) - W^j(a) : t \le a \le b \le T,\ j \in \mathbb{N})$ since the processes W^j are continuous.

As for (V), the σ-algebras $\sigma(W^1), \sigma(W^2), \sigma(W^3), \ldots$ are $\mathbb{P}_{\tau_k}(Z_k \in \cdot\,)$-independent, hence also μ-independent by Lemma 7.1. And Lemma 6.3 yields that they are Brownian. ☐

Theorem 7.1. *The process S^3 defined in (18) is an SVI solution on $(\mathbf{Z}, \mathscr{B}(\mathbf{Z}), (\mathscr{Z}^\mu_t), \mu)$ with Wiener processes $(W^k)_{k\in\mathbb{N}}$ (also defined in defined in (18)) in the sense of Definition 3.1.*

Proof. The proof is divided into several steps. Recall that we consider the sub-sequence $(\varepsilon_k, h_k, \tau_k) \to (0,0,0)$ for $k \to 0$.

(i) First, we show that a discrete version (25) of (4) holds for simple step-processes G and H. For let $0 = s_0 < \cdots < s_m = T$, define $\mathbb{R}^{4M^2}$-valued continuous mappings on $\mathbf{Z}$ as

$$V^\alpha = \Big(\Big(\varphi_\beta, S^1_{r^\alpha_\gamma}\Big), \Big(\varphi_\beta, S^2_{r^\alpha_\gamma}\Big), \Big(\varphi_\beta, S^3_{r^\alpha_\gamma}\Big), W^j_{r^\alpha_\gamma} : \beta, \gamma, j \in \{1, \ldots, M\}\Big), \quad 0 \le \alpha \le m,$$

for some $r^\alpha_\gamma \in [0, s_\alpha]$ and $\varphi_\beta \in \mathbb{L}^2$ where we consider the product with φ_β to work with real-valued random variables, and let

$$g_\alpha, h_{\alpha,j} : \mathbb{R}^{4M^2} \to \mathbb{H}^1_0, \qquad \alpha \in \{0, \ldots, m\},\ j \in \mathbb{N},$$

be $\mathbb{H}^1_0$-bounded continuous functions such that $h_{\alpha,j} = 0$ for $j \ge j_0$ and some arbitrary $j_0 \in \mathbb{N}$, to simplify the argument. We define

$$G(t) = \sum_{\alpha=0}^{m-1} \mathbf{1}_{(s_\alpha, s_{\alpha+1}]}(t) g_\alpha(V^\alpha), \qquad H_j(t) = \sum_{\alpha=0}^{m-1} \mathbf{1}_{(s_\alpha, s_{\alpha+1}]}(t) h_{\alpha,j}(V^\alpha),$$

and

$$I(t) = u^0 - \int_0^t G(s)\,\mathrm{d}s + \sum_{j=1}^{j_0} \int_0^t H_j(s)\,\mathrm{d}W^j. \tag{20}$$

Setting $N_k = T/\tau_k$, $t_i := i\tau_k$ for $i \in \{0, \dots, N_k\}$ then $G_{t_i}(Z_k)$ and $H_{j,t_i}(Z_k)$ are $\mathcal{F}^i_{\tau_k}$-measurable, Lemma 4.3 yields

$$\begin{aligned}\frac{1}{2}\mathbb{E}_{\tau_k}\Big[\big\|S^1_{t_i}(Z_k) - P_{h_k}U^i(Z_k)\big\|^2\Big] + \mathbb{E}_{\tau_k}\Big[\int_0^{t_i} \mathcal{J}_{\varepsilon_k}\big(S^1_s(Z_k)\big)\,\mathrm{d}s\Big] &\le \frac{1}{2}\big\|x^0 - u^0\big\|^2\\ &+ \sum_{\ell=1}^{i}\mathbb{E}_{\tau_k}\Big[\int_{t_{\ell-1}}^{t_\ell}\big[\mathcal{J}_{\varepsilon_k}\big(P_{h_k}U^\ell(Z_k)\big) + \big(P_{h_k}G_{t_\ell}(Z_k), S^1_s(Z_k) - U^\ell(Z_k)\big)\big]\,\mathrm{d}s\Big]\\ &+ \frac{1}{2}\sum_{\ell=2}^{i}\mathbb{E}_{\tau_k}\Big[\int_{t_{\ell-2}}^{t_{\ell-1}}\big\|P_{h_k}B\big(S^1_s(Z_k)\big) - P_{h_k}H_{t_{\ell-1}}(Z_k)\big\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\Big]\\ &+ \frac{\tau_k}{2}\big\|P_hB(x^0)\big\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)},\end{aligned}$$

for $0 \le i \le N_k$ where

$$U^i = u^0 - \tau_k\sum_{\ell=1}^{i} G(t_\ell) + \sum_{\ell=1}^{i}\sum_{j=1}^{N_k}\big(W^j(t_\ell) - W^j(t_{\ell-1})\big)H_j(t_{\ell-1}) \qquad i \in \{0, \dots, N_k\}, \tag{21}$$

as $S^1_{t_i}(Z_k) = \overline{X}^i_{\varepsilon_k,h_k,\tau_k}$ by the definition of S^1 and Z_k. For $N_k \ge j_0$ we deduce that

$$\max_{1\le\ell\le N_k}\sup_{t\in[t_{\ell-1},t_\ell]}\|I(t) - U^\ell\|_{\mathbb{H}^1_0} \le C_G\tau_k + C_H\sum_{j=1}^{j_0}\mathbf{m}(W^j, \tau_k), \tag{22}$$

where $\mathbf{m}$ is the modulus of continuity of real-valued functions.
In the following, we replace U by I in the last but one inequality above, we proceed term by term. We note that

$$\mathbb{E}_{\tau_k}\Big[\max_{1\le\ell\le N_k}\big\|U^\ell(Z_k)\big\|^2_{\mathbb{H}^1_0} + \sup_{s\in[0,T]}\big\|S^1_s(Z_k)\big\|^2\Big] \le C, \tag{23}$$

and

$$\mathbb{E}_{\tau_k}\big[\mathbf{m}\big(W^j(Z_k), \tau_k\big)\big]^2 \le C\tau_k^{2\theta}, \tag{24}$$

hold for some $\theta \in (0, \frac{1}{4})$ by (17), the Doob maximal inequality for submartingales and Lemma 6.1. Next, we observe that

$$\begin{aligned}\Big|\mathbb{E}_{\tau_k}\Big[\big\|S^1_{t_i}(Z_k) - P_{h_k}U^i(Z_k)\big\|^2\Big] - \mathbb{E}_{\tau_k}\Big[\big\|S^1_{t_i}(Z_k) - P_{h_k}I_{t_i}(Z_k)\big\|^2\Big]\Big|&\\ \le \mathbb{E}_{\tau_k}\Big[\big\|P_{h_k}U^i(Z_k) - P_{h_k}I_{t_i}(Z_k)\big\|^2\Big]&\\ + 4\sqrt{C}\Big\{\mathbb{E}_{\tau_k}\Big[\big\|P_{h_k}U^i(Z_k) - P_{h_k}I_{t_i}(Z_k)\big\|^2\Big]\Big\}^{\frac{1}{2}}&\\ \le C\tau_k^\theta,&\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}_{\tau_k}\Big[\int_{t_{\ell-1}}^{t_\ell}\big|\mathcal{J}_{\varepsilon_k}\big(P_{h_k}U^\ell(Z_k)\big) - \mathcal{J}_{\varepsilon_k}(P_{h_k}I_s(Z_k))\big|\,\mathrm{d}s\Big]\\ &\qquad \le C\sum_{j=1}^{2}\int_{t_{\ell-1}}^{t_\ell}\mathbb{E}_{\tau_k}\Big[\big\|P_{h_k}U^\ell(Z_k) - P_{h_k}I_s(Z_k)\big\|^j_{\mathbb{H}^1_0}\Big]\,\mathrm{d}s\end{aligned}$$

$$+ C\sum_{j=1}^{2}\int_{t_{\ell-1}}^{t_\ell}\mathbb{E}_{\tau_k}\left[\left\|P_{h_k}U^\ell(Z_k) - P_{h_k}I_s(Z_k)\right\|_{\mathbb{H}_0^1}^{\frac{j}{2}}\left\|P_{h_k}U^\ell(Z_k)\right\|_{\mathbb{H}_0^1}^{\frac{\ell}{2}}\right]\mathrm{d}s$$
$$\leq C\tau_k^{1+\frac{\theta}{2}},$$

by the stability of the projections $\{P_h\}_{h>0}$ in $\mathbb{H}_0^1$ from Assumption 2.1$_{(3)}$.
Now denote by R_k the set of $\ell \in \{1,\dots,N_k\}$ such that the interval $(t_{\ell-1},t_\ell)$ is not fully contained in some of the intervals $(s_\alpha, s_{\alpha+1}]$ for $\alpha\in\{0,\dots,m-1\}$. If $\ell\in R_k$ then there exists unique α such that $s_\alpha < t_\ell \leq s_{\alpha+1}$. If $s_\alpha \leq t_{\ell-1}$ then this would contradict that $\ell\in R_k$ hence $s_\alpha < t_\ell < s_\alpha + \tau_k$. In particular, card $(R_k)\leq m$, and consequently

$$\sum_{\ell=1}^{N_k}\mathbb{E}_{\tau_k}\left[\int_{t_{\ell-1}}^{t_\ell}|(P_{h_k}G_{t_\ell}(Z_k) - P_{h_k}G_s(Z_k), S_s^1(Z_k) - U^\ell(Z_k))|\,\mathrm{d}s\right]$$
$$= \sum_{\ell\in R_k}\mathbb{E}_{\tau_k}\left[\int_{t_{\ell-1}}^{t_\ell}|(P_{h_k}G_{t_\ell}(Z_k) - P_{h_k}G_s(Z_k), S_s^1(Z_k) - U^\ell(Z_k))|\,\mathrm{d}s\right]$$
$$\leq Cm\tau_k.$$

Analogously, we estimate

$$\sum_{\ell=2}^{N_k}\mathbb{E}_{\tau_k}\left[\int_{t_{\ell-2}}^{t_{\ell-1}}\|B(S_s^1(Z_k)) - H_{t_{\ell-1}}(Z_k)\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\right]\leq Cm\tau_k,$$
$$\sum_{\ell=2}^{N_k}\mathbb{E}_{\tau_k}\left[\int_{t_{\ell-2}}^{t_{\ell-1}}\|B(S_s^1(Z_k)) - H_s(Z_k)\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\right]\leq Cm\tau_k,$$

by the linear growth of B assumed in $(\mathbf{B}_1)$. In the fourth step, we estimate

$$\mathbb{E}_{\tau_k}\left[\int_{t_{\ell-1}}^{t_\ell}|(P_{h_k}G_s(Z_k), U^\ell(Z_k)) - (P_{h_k}G_s(Z_k), I_s(Z_k))|\,\mathrm{d}s\right]\leq \mathbf{C}\tau_k^{1+\theta},$$

by boundedness of G. Hence, we conclude that

$$\begin{aligned}\frac{1}{2}\mathbb{E}_{\tau_k}\left[\left\|S_{t_i}^1(Z_k) - P_{h_k}I_{t_i}(Z_k)\right\|^2\right] + \mathbb{E}_{\tau_k}\left[\int_0^{t_i}\mathcal{J}(S_s^1(Z_k))\,\mathrm{d}s\right] &\leq \frac{1}{2}\mathbb{E}_{\tau_k}\left[\left\|S_{t_i}^1(Z_k) - P_{h_k}I_{t_i}(Z_k)\right\|^2\right]\\
&+ \mathbb{E}_{\tau_k}\left[\int_0^{t_i}\mathcal{J}_{\varepsilon_k}\big(S_s^1(Z_k)\big)\,\mathrm{d}s\right]\leq \frac{1}{2}\left\|x^0 - u^0\right\|^2\\
&+ \mathbb{E}_{\tau_k}\left[\int_0^{t_i}\left[\mathcal{J}_{\varepsilon_k}(P_{h_k}I_s(Z_k)) + \big(P_{h_k}G_s(Z_k), S_s^1(Z_k) - I_s(Z_k)\big)\right]\mathrm{d}s\right]\\
&+ \frac{1}{2}\mathbb{E}_{\tau_k}\left[\int_0^{t_i}\left\|B\big(S_s^1(Z_k)\big) - H_s(Z_k)\right\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\mathrm{d}s\right] + C\tau_k^{\frac{\theta}{2}},\end{aligned}\tag{25}$$

for $0\leq i\leq N_k$ and some C independent of i and k. Here we used $\mathcal{J}\leq\mathcal{J}_\varepsilon$ and the linear growth of B assumed in $(\mathbf{B}_1)$.

(ii) In the second step, we extend the discrete result from step (i) to the time-continuous case on the stochastic basis $(\mathbf{Z},\mathscr{B}(\mathbf{Z}),(\mathscr{Z}_t^\mu),\mu)$, yet still for the simple processes G and H defined in part (i).
We note that by construction the mapping $I:[0,T]\times\mathbf{Z}\to\mathbb{H}_0^1$ from (i) is continuous and the following properties hold for every k and $r\in[0,T]$:

(a) $\|S^1_r - P_{h_k} I_r\|^2$ is lower semicontinuous on $\mathbf{Z}$,
(b) $\int_0^r \mathcal{J}(S^1)\,\mathrm{d}s$ is lower semicontinuous on $\mathbf{Z}$ by Remark B.1,
(c) $\int_0^r [\mathcal{J}_{\varepsilon_k}(P_{h_k} I) + (P_{h_k} G, S^1 - I)]\,\mathrm{d}s$ is continuous on $\mathbf{Z}$ as $(P_{h_k} G, S^1) = (G, P_{h_k} S^1)$,
(d) $\int_0^r \|B(S^1) - H\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s$ is $\mathscr{B}(\mathbf{Z})$-measurable by Corollary 7.1 (I).

Furthermore, from the fact that $\mathcal{J}_\varepsilon \to \mathcal{J}$ for $\varepsilon \to 0$ and $\|\nabla v - \nabla P_h v\| \to 0$, $v \in \mathbb{H}^1$ for $h \to 0$ we deduce the convergence

$$\begin{aligned}
\left\|S^1_t(z) - I_t(z)\right\|^2 &\le \liminf_{k\to\infty} \left\|S^1_{t^k_{i_k}}(z_k) - P_{h_k} I_{t^k_{i_k}}(z_k)\right\|^2,\\
\int_0^r \mathcal{J}(I(z))\,\mathrm{d}s &= \lim_{k\to\infty} \int_0^r \mathcal{J}_{\varepsilon_k}(P_{h_k} I(z_k))\,\mathrm{d}s,\\
\int_0^r (G(z), S^1(z) - I(z))\,\mathrm{d}s &= \lim_{k\to\infty} \int_0^r \big(P_{h_k} G(z_k), S^1(z_k) - I(z_k)\big)\,\mathrm{d}s,\\
\int_0^r \left\|B(S^1(z)) - H(z)\right\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s &= \lim_{k\to\infty} \int_0^r \left\|B\big(S^1(z_k)\big) - H(z_k)\right\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s,
\end{aligned}$$

whenever $t^k_{i_k} \nearrow t$ and $z_k \to z$ in the sense of (i)–(iv) of Lemma 7.1 where, in the last step, we used the assumption $(\mathbf{B}_2)$ on continuity of B if $d \ge 2$ (if $d = 1$, continuity of B suffices). Indeed, assume that

$$\int_0^T \|B(f_k) - B(f)\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s \ge r > 0, \tag{26}$$

for some $f_k \to f$ in the sense of (i) of Lemma 7.1. Then $f_k \to f$ uniformly in $\mathbb{H}^{-d}$ and $\int_0^T \|f_k\|_{BV}\,\mathrm{d}s \le C$. Hence $\int_0^T \|f_k - f\|_{\mathbb{L}^1}\,\mathrm{d}s \to 0$ since $BV(\mathcal{O}) \hookrightarrow\hookrightarrow \mathbb{L}^1 \hookrightarrow \mathbb{H}^{-d}$. If $d = 1$ then even $\int_0^T \|f_k - f\|_{\mathbb{L}^2}\,\mathrm{d}s \to 0$ since $BV(\mathcal{O}) \hookrightarrow\hookrightarrow \mathbb{L}^2 \hookrightarrow \mathbb{H}^{-d}$. Thus there exists a subsequence k_l such that $f_{k_l}(s) \to f(s)$ a.e. on $\mathcal{O}$ (or in $\mathbb{L}^2$ if $d = 1$) for a.e. $s \in [0,T]$. In particular, $\|B(f_{k_l}(s)) - B(f(s))\|_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)} \to 0$ for a.s. $s \in [0,T]$, and the linear growth of B then yields that $\int_0^T \|B(f_{k_l}) - B(f)\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s \to 0$ which is a contradiction with (26). Finally,

$$\begin{aligned}
\|I_s(Z_k)\|_{\mathbb{H}^1_0} &\le c + c\sum_{j=1}^{j_0} \left\|W^j_{\tau_k}\right\|_{C([0,T])},\\
|\mathcal{J}_{\varepsilon_k}(P_{h_k}(I_s(Z_k)))| &\le c\Big[1 + \|I_s(Z_k)\|^2_{\mathbb{H}^1_0}\Big],
\end{aligned}$$

holds by stability of the projections $\{P_h\}_{h>0}$ in $\mathbb{H}^1_0$ so (19) is satisfied by (17), Lemma 6.1 and the linear growth of B. Hence on taking the limit $k \to \infty$ in (25) we conclude by Lemma 7.1 that (4) holds.

(iii) In the last step, we prove the full result. The extension of (4) to $(\mathcal{Z}^\mu_t)$-progressively measurable processes in $L^2([0,T]\times\Omega;\mathbb{H}^1_0)$ and $L^2([0,T]\times\Omega;\mathscr{L}_2(\ell_2,\mathbb{H}^1_0))$ follows by a standard density argument (see *e.g.*, [17], Sect. 3.2, Lem. 2.4), and the general case can be obtained by considering $G_h = P_h G$ and $H_h = P_h H$, and then letting $h \to 0$.

□

8. Convergence to pathwise unique probabilistically strong solution

In this section we study convergence of the interpolants $\overline{X}_{\varepsilon,h,\tau}$, $\underline{X}_{\varepsilon,h,\tau}$ and $X_{\varepsilon,h,\tau}$ to a probabilistically strong SVI solution of (1) in probability. The proofs rely on a generalization of the the Gyongy–Krylov lemma to non-Polish spaces (see *e.g.*, the proof in [11], Thm. 2.10.3):

Lemma 8.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, Y a Hausdorff locally convex topological vector space such that there exists a continuous injection of Y into some Polish space P and let $U_n : \Omega \to Y$ be a tight sequence of Borel measurable random variables such that, for every two subsequences $\{n_k\}$, $\{m_k\}$, there exists a subsequence $\{k_j\}$ and a Borel probability measure θ on $Y \times Y$ supported in the diagonal of $Y \times Y$ such that $(U_{n_{k_j}}, U_{m_{k_j}})$ converges in law to θ. Then $\{U_n\}$ converges in Y in probability to some Borel measurable random variable $U : \Omega \to Y$.*

Theorem 8.1. *Let (W^j) be independent $(\mathcal{F}_t)$-Wiener processes on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and let*

$$\xi_\tau^{i,j} = W^j(t_i) - W^j(t_{i-1}), \qquad t_i = i\tau.$$

Assume also that pathwise uniqueness holds for the SVI solutions of (1) satisfying (8). Then $\overline{X}_{\varepsilon,h,\tau}$, $\underline{X}_{\varepsilon,h,\tau}$ and $X_{\varepsilon,h,\tau}$ converge to X in probability in $Q_{c,BV}([0,T];\mathbb{L}^2_w)$, $Q_c([0,T];\mathbb{L}^2_w)$ and $C([0,T];\mathbb{L}^2_w)$ respectively where X is a solution (1) with respect to $(W^j)_{j\in\mathbb{N}}$.

Proof. The proof is based on the Gyongy–Krylov Lemma 1.1 in [15]. Define

$$\mathbf{S} = Q_{c,BV}([0,T];\mathbb{L}^2_w) \times Q_{c,BV}([0,T];\mathbb{L}^2_w) \times C([0,T]) \times C([0,T]) \times C([0,T]) \times C([0,T]) \times \dots,$$

and the projections

$$\begin{aligned} &Y^1 : \mathbf{S} \to Q_{c,BV}\big([0,T];\mathbb{L}^2_w\big) && \big(f^1, f^2, w^1, w^2, w^3, \dots\big) \mapsto f^1,\\ &Y^2 : \mathbf{S} \to Q_{c,BV}\big([0,T];\mathbb{L}^2_w\big) && \big(f^1, f^2, w^1, w^2, w^3, \dots\big) \mapsto f^2,\\ &W^j : \mathbf{S} \to C([0,T]) && \big(f^1, f^2, w^1, w^2, w^3, \dots\big) \mapsto w^j, \end{aligned}$$

and the canonical filtration on $\mathbf{S}$

$$\mathcal{S}_t = \sigma\big(Y^1_s, Y^2_s, W^j_s :\, s \in [0,t],\, j \in \mathbb{N}\big), \qquad t \in [0,T].$$

We consider two different sequences of discretization parameters $(\varepsilon^i_k, h^i_k, \tau^i_k) \to (0,0,0)$ for $i = 1,2$, which are chosen as in Corollary 6.1, such that

$$Z_k := \Big(\overline{X}_{\varepsilon^1_k, h^1_k, \tau^1_k}, \overline{X}_{\varepsilon^2_k, h^2_k, \tau^2_k}, W^1, W^2, W^3, \dots\Big),$$

converge to a Radon probability measure θ on $\mathscr{B}(\mathbf{S})$. Analogously as in Corollary 7.1, the processes Y^1, Y^2 and $(W^k)_{k\in\mathbb{N}}$ are $(\mathcal{S}_t)$-progressively measurable, paths of Y^1 and Y^2 are continuous θ-a.s.,

$$\int_{\mathbf{S}} \left[\sup_{t\in[0,T]} \|Y^i(t)\|^4 + \left(\int_0^T \|Y^i(t)\|_{BV(\mathcal{O})}\,\mathrm{d}t\right)^2\right] \mathrm{d}\theta < \infty, \qquad i = 1,2,$$

the σ-algebras $\mathcal{S}_t$ and $\sigma(W^j(b) - W^j(a) :\, t \le a \le b \le T,\, j \in \mathbb{N})$ are θ-independent and $W^1, W^2, W^3, \dots$ are θ-independent $(\mathcal{S}_t)$-Brownian motions. The proof that Y^1 and Y^2 are SVI solutions with respect to $(W^k)_{k\in\mathbb{N}}$ and

$$\theta\left[Y^1(0) = Y^2(0) = x^0\right] = 1,$$

is analogous to the proof of Theorem 7.1, we point out the differences below.

In step (i) one modifies the definition of the $\mathbb{R}^{3M^2}$-valued random variables

$$V^\alpha = \Big(\Big(\varphi_\beta, Y^1_{r^\alpha_\gamma}\Big), \Big(\varphi_\beta, Y^2_{r^\alpha_\gamma}\Big), W^j_{r^\alpha_\gamma} : \beta, \gamma, j \in \{1, \dots, M\}\Big), \quad 0 \le \alpha \le m,$$

defined on $\mathbf{S}$, the functions $g_\alpha, h_{\alpha,j}$ map $\mathbb{R}^{3M^2}$ to $\mathbb{H}^1_0$ and have the same properties as in the proof of Theorem 7.1 and

$$G(t) = \sum_{\alpha=1}^{m-1} \mathbf{1}_{(s_\alpha, s_{\alpha+1}]}(t) g_\alpha(V^{\alpha-1}), \qquad H_j(t) = \sum_{\alpha=1}^{m-1} \mathbf{1}_{(s_\alpha, s_{\alpha+1}]}(t) h_{\alpha,j}(V^{\alpha-1}),$$

i.e., there is a backward time shift compared to the definition of G and H in the proof of Theorem 7.1. Once we set we set $N_k^i = T/\tau_k^i$, $t_\ell^i := \ell\tau_k^i$ for $\ell \in \{0,\dots,N_k^i\}$, $i=1,2$ the above modification ensures that $V^{\alpha-1}(Z_k)$ is $\mathcal{F}_{s_\alpha}$-measurable. Consequently, $G(t,Z_k)$ and $H_j(t,Z_k)$ are $(\mathcal{F}_t)$-adapted processes as long as τ_k^i, $i=1,2$ are smaller than the mesh of the partition $\{s_\alpha\}$.

Pathwise uniqueness of solutions of (1) yields that $Y^1 = Y^2$ holds $\mathbb{P}$-a.s. hence $\overline{X}_{\varepsilon,h,\tau}$ is convergent in $Q_{c,BV}([0,T];\mathbb{L}^2_w)$ in probability as $(\varepsilon,h,\tau)\to(0,0,0)$ by Lemma 8.1.

Now we apply the Gyongy–Krylov Lemma 8.1 once again. By Corollary 6.1 we deduce that the laws of the sequence

$$\left(\overline{X}_{\varepsilon_k^1,h_k^1,\tau_k^1}, \underline{X}_{\varepsilon_k^1,h_k^1,\tau_k^1}, X_{\varepsilon_k^1,h_k^1,\tau_k^1}, \overline{X}_{\varepsilon_k^2,h_k^2,\tau_k^2}, \underline{X}_{\varepsilon_k^2,h_k^2,\tau_k^2}, X_{\varepsilon_k^2,h_k^2,\tau_k^2}\right),$$

on

$$\mathscr{B}(Q_{c,BV}\times Q_c\times C\times Q_{c,BV}\times Q_c\times C),$$

where $Q_{c,BV} = Q_{c,BV}([0,T];\mathbb{L}^2_w)$, $Q_c = Q_c([0,T];\mathbb{L}^2_w))$ and $C = C([0,T];\mathbb{L}^2_w)$ converge to some probability measure ν. Consequently

$$\nu\{x_1=x_2=x_3,\ x_4=x_5=x_6,\ x_1=x_4\}=1,$$

by Corollary 7.1 (II) and the first part of the proof. Hence Theorem 2.10.3 of [11] yields that $\underline{X}_{\varepsilon,h,\tau}$ and $X_{\varepsilon,h,\tau}$ converge in probability in $Q_c([0,T];\mathbb{L}^2_w))$ and $C([0,T];\mathbb{L}^2_w)$ respectively as $(\varepsilon,h,\tau)\to(0,0,0)$. And the limit equals to X by (16).

Analogously as in the proof of Theorem 7.1 we set

$$G(t)=\sum_{\alpha=0}^{m-1}\mathbf{1}_{(s_\alpha,s_{\alpha+1}]}(t)g_\alpha \qquad H_j(t)=\sum_{\alpha=0}^{m-1}\mathbf{1}_{(s_\alpha,s_{\alpha+1}]}(t)h_{\alpha,j}, \tag{27}$$

for some $0=s_0<\cdots<s_m=T$ where g_α and $h_{\alpha,j}$ are simple $\mathbb{H}^1_0$-valued $\mathcal{F}_{s_\alpha}$-measurable random variables such that $h_{\alpha,j}=0$ for $j\geq j_0$ for some arbitrary $j_0\in\mathbb{N}$ and define the process I as in (20). Setting $N=T/\tau$, $t_i:=i\tau$ for $i\in\{0,\dots,N\}$ then, as in (25) in the proof of Theorem 7.1 we obtain that

$$\begin{aligned}\frac12\mathbb{E}\Big[\big\|\overline{X}_\tau(t_i)-P_hI(t_i)\big\|^2\Big]+\mathbb{E}\bigg[\int_0^{t_i}\mathcal{J}\big(\overline{X}_\tau(s)\big)\,\mathrm{d}s\bigg]&\leq\frac12\big\|x^0-u^0\big\|^2\\&\quad+\mathbb{E}\bigg[\int_0^{t_i}\big[\mathcal{J}_\varepsilon(P_hI(s))+\big(P_hG(s),\overline{X}_\tau(s)-I(s)\big)\big]\,\mathrm{d}s\bigg]\\&\quad+\frac12\mathbb{E}\bigg[\int_0^{t_i}\big\|B\big(\overline{X}_\tau(s)\big)-H(s)\big\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\bigg]+c\tau^{\frac{\theta}{2}},\end{aligned}$$

for $0\leq i\leq N$ if $N\geq j_0$. If $0\leq t\leq t_i<t+\tau$ then

$$\begin{aligned}\frac12\mathbb{E}\Big[\big\|\overline{X}_\tau(t_i)-P_hI(t_i)\big\|^2\Big]+\mathbb{E}\bigg[\int_0^{t}\mathcal{J}\big(\overline{X}_\tau(s)\big)\,\mathrm{d}s\bigg]&\leq\frac12\big\|x^0-u^0\big\|^2\\&\quad+\mathbb{E}\bigg[\int_0^{t}\big[\mathcal{J}_\varepsilon(P_hI(s))+\big(P_hG(s),\overline{X}_\tau(s)-I(s)\big)\big]\,\mathrm{d}s\bigg]\\&\quad+\frac12\mathbb{E}\bigg[\int_0^{t}\big\|B\big(\overline{X}_\tau(s)\big)-H(s)\big\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\bigg]+c\tau^{\frac{\theta}{2}}+c_1\tau.\end{aligned}$$

We deduce that the following holds for $(\varepsilon,h,\tau)\to(0,0,0)$:

– $\overline{X}_\tau(t_i)-P_hI(t_i)$ is tight in $\mathbb{L}^2_w$ and converges to $X(t)-I(t)$ in $\mathbb{L}^2_w$ in probability (hence also in law) thus

$$\mathbb{E}\Big[\big\|\overline{X}(t)-(t)\big\|^2\Big]\leq\liminf\mathbb{E}\Big[\big\|\overline{X}_\tau(t_i)-P_hI(t_i)\big\|^2\Big],$$

by Proposition 5.1,

– $\overline{X}_\tau$ converges to X in $Q_{c,BV}([0,T];\mathbb{L}^2_w)$ in probability (hence also in law) thus

$$\mathbb{E}\Big[\int_0^t \mathcal{J}(X(s))\,\mathrm{d}s\Big] \le \liminf \mathbb{E}\Big[\int_0^t \mathcal{J}(\overline{X}_\tau(s))\,\mathrm{d}s\Big],$$

as in the proof of Theorem 7.1,

–

$$\mathbb{E}\Big[\int_0^t \mathcal{J}(I(s))\,\mathrm{d}s\Big] = \lim \mathbb{E}\Big[\int_0^t \mathcal{J}_\varepsilon(P_h I(s))\,\mathrm{d}s\Big],$$

as in the proof of Theorem 7.1,

–

$$\mathbb{E}\Bigg[\int_0^T \big|\big(P_h G(s) - G(s), \overline{X}_\tau(s) - I(s)\big)\big|\,\mathrm{d}s\Bigg] \le C\mathbb{E}\Bigg[\int_0^T \|P_h G(s) - G(s)\|^2\,\mathrm{d}s\Bigg] \to 0,$$

–

$$\mathbb{E}\Bigg[\int_0^T \big|(G(s), \overline{X}_\tau(s) - X(s))\big|\,\mathrm{d}s\Bigg] = \sum_{\alpha=0}^{m-1} \mathbb{E}\Big[\int_{s_\alpha}^{s_{\alpha+1}} \big|\big(g_\alpha, \overline{X}_\tau(s) - X(s)\big)\big|\,\mathrm{d}s\Big]$$
$$\le T \sum_{\alpha=0}^{m-1} \mathbb{E}\Bigg[\sup_{s\in[0,T]} \big|\big(g_\alpha, \overline{X}_\tau(s) - X(s)\big)\big|\Bigg] \to 0,$$

since g_α are simple,

– we proved in the proof of Theorem 7.1 that if $f_n^j \to f^j$ in $Q_{c,BV([0,T];\mathbb{L}^2_w)}$ and

$$\int_0^T \|f_n^j(s)\|_{BV(\mathcal{O})}\,\mathrm{d}s \le C, \qquad j = 1,2,$$

then

$$\int_0^T \|B(f_n^1) - B(f_n^2)\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s \to \int_0^T \|B(f^1) - B(f^2)\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s.$$

Now $(\overline{X}_\tau, X)$ are tight in $Q_{c,BV([0,T];\mathbb{L}^2_w)} \times Q_{c,BV([0,T];\mathbb{L}^2_w)}$ and converge in probability (hence in law) to (X,X). By Proposition 5.1 we deduce

$$\lim \mathbb{E}\Bigg[\int_0^T \big\|B\big(\overline{X}_\tau(s)\big) - B(X(s))\big\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\Bigg] = \mathbb{E}\Bigg[\int_0^T \|B(X(s)) - B(X(s))\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\Bigg].$$

Hence, we obtain as in (ii) in the proof of Theorem 7.1 that

$$\frac{1}{2}\mathbb{E}[\|X(t) - I(t)\|^2] + \mathbb{E}\Bigg[\int_0^t \mathcal{J}(X(s))\,\mathrm{d}s\Bigg] \le \frac{1}{2}\|x^0 - u^0\|^2 + \mathbb{E}\Bigg[\int_0^t [\mathcal{J}(I(s)) + (G(s), X(s) - I(s))]\,\mathrm{d}s\Bigg]$$
$$+ \frac{1}{2}\mathbb{E}\Big[\int_0^t \|B(X(s)) - H(s)\|^2_{\mathscr{L}_2(\ell_2,\mathbb{L}^2)}\,\mathrm{d}s\Big].$$

The extension to general G, H and I is analogous to **(iii)** in the proof of Theorem 7.1.

□

9. Numerical experiments

We perform numerical experiments in $d = 2$ on the unit square $\mathcal{O} = (0,1)^2$. We construct a triangulation $\mathcal{T}_h$ of $\mathcal{O}$ with mesh size $h = 2^{-\ell}$ by partitioning the unit square into sub-squares of size h and subdividing each square into four equal right-angled triangles. Given the triangulation $\mathcal{T}_h$ we consider the finite element space $\mathbb{V}_h \equiv \mathbb{V}_h(\mathcal{T}_h) = \mathrm{span}\{\phi_j,\ j = 1,\dots,J\}$. For simplicity of implementation we consider a finite dimensional noise: given a scalar valued function σ (to be specified below) we set $B_j(X) = B_j(X,h) = \sigma(X)\phi_j$ and consider discrete increments $\Delta_i\beta_j := \beta_j(t_i) - \beta_j(t_{i-1})$ where $\beta_j,\ j = 1,\dots,J$ are independent scalar-valued Wiener processes.

The above choice of the noise yields a variant of scheme (3): for $i = 1,\dots,N$ we seek $X^i_{\varepsilon,h} \in \mathbb{V}_h$ that satisfies

$$\begin{aligned}\left(X^i_{\varepsilon,h}, v_h\right) &= \left(X^{i-1}_{\varepsilon,h}, v_h\right) - \tau\left(\frac{\nabla X^i_{\varepsilon,h}}{\sqrt{|\nabla X^i_{\varepsilon,h}|^2 + \varepsilon^2}}, \nabla v_h\right) \\ &\quad - \tau\lambda\left(X^i_{\varepsilon,h} - g_h, v_h\right) + \sum_{j=1}^{J}\left(\sigma(X^{i-1}_{\varepsilon,h})\phi_j, v_h\right)\Delta_i\beta_j \qquad \forall v_h \in \mathbb{V}_h, \qquad (28)\\ X^0_{\varepsilon,h} &= x^0_h.\end{aligned}$$

where $g_h,\ x^0_h \in \widetilde{\mathbb{V}}_h \subset \mathbb{V}_h$ are suitable approximations (see below) of the data g, x_0, respectively.

For comparison we also perform simulations using a non-conforming variant of (28) where the ($\mathbb{H}^1$-conforming) space $\mathbb{V}_h$ in (28) is replaced by a non-conforming finite element space $\mathbb{V}_{\mathrm{cr}} \not\subset \mathbb{H}^1_0$. Given a partition $\mathcal{T}_h$ of $\mathcal{O}$ we denote the set of all faces of elements $T \in \mathcal{T}_h$ as $\mathcal{S}_h = \cup_{T\in\mathcal{T}_h}\partial T$ and for a face $S \in \mathcal{S}_h$ we denote its barycenter by b_S. Then we define the non-conforming finite element space as

$$\begin{aligned}\mathbb{V}_{\mathrm{cr}} = \big\{&\varphi \in \mathbb{L}^2;\ \varphi|_T \in \mathcal{P}^1(T)\ \forall T \in \mathcal{T}_h,\ \varphi \text{ is continuous at } b_S\ \forall S \in \mathcal{S}_h \cap \mathcal{O} \\ &\text{and } \varphi(b_S) = 0 \text{ for } S \in \mathcal{S}_h \cap \partial\mathcal{O}\big\}.\end{aligned}$$

The above finite element space corresponds to the first order Crouzeix–Raviart finite element which is more suitable for the approximation of discontinuous solutions, *cf.*, [3] and the references therein, for its use in the context of image processing. We note that $\mathbb{V}_h(\mathcal{T}_h) \subset \mathbb{V}_{\mathrm{cr}}(\mathcal{T}_h)$ but since $\mathbb{V}_{\mathrm{cr}} \not\subset \mathbb{H}^1_0$ the elements of $\mathbb{V}_{\mathrm{cr}}$ have no (global) weak gradients in general. Hence for $w_h \in \mathbb{V}_{\mathrm{cr}}$ we define a discrete gradient $\nabla_h w_h$ *via* $\nabla_h w_h = \nabla(w_h|_T)$. Then the non-conforming counterpart of the scheme (3) is obtained by replacing $\mathbb{V}_h$ with $\mathbb{V}_{\mathrm{cr}}$ and the gradients ∇ in (3) by the discrete gradient ∇_h. The numerical solutions $X^i_{\varepsilon,h} \in \mathbb{V}_{\mathrm{cr}}$, $i = 1,\dots,N$ the exist and satisfy an energy law (counterpart of Lem. 4.1), however, the convergence of the non-conforming scheme is open so far.

To construct an approximation of the data g, x_0 we consider the space $\widetilde{\mathbb{V}}_h \equiv \mathbb{V}_h(\widetilde{\mathcal{T}}_h)$ with fixed mesh size $h = 2^{-6}$. We define the function $\tilde{g}_h \in \widetilde{\mathbb{V}}_h$, which represents an "exact image", as the composition of the characteristic function of a square with side $\frac{1}{2}$ at the center of $\mathcal{O}$ scaled by the factor $\frac{1}{2}$ and the characteristic function of a circle with radius $\frac{1}{4}$ shifted by 0.2 to the right of the center of $\mathcal{O}$ interpolated on the mesh $\widetilde{\mathcal{T}}_h$, see Figure 1 (left), *i.e.*, $\tilde{g}_h(x) = \sum_{j=1}^{\tilde{J}}\tilde{g}(x_j)\tilde{\phi}_j(x)$ where $\{\tilde{\phi}_j\}_{j=1}^{\tilde{J}}$ are the nodal basis functions associated with the nodes the $\{x_j\}_{j=1}^{\tilde{J}}$ of the mesh $\widetilde{\mathcal{T}}_h$. Hence, we set $g_h = \tilde{g}_h + \xi_h \in \widetilde{\mathbb{V}}_h$ with the "noise" $\xi_h(x) = 0.1\sum_{j=1}^{\tilde{J}}\tilde{\phi}_j(x)\xi_j$, $x \in \mathcal{O}$ where $\xi_j,\ j = 1,\dots,\tilde{J}$ are realizations of independent $\mathcal{U}(-1,1)$-distributed random variables. The corresponding realization of the noise ξ_h and the resulting "noisy image" g_h are displayed in Figure 1 (middle and right, respectively).

In all experiments we set $T = 0.1$, $\lambda = 200$, $\varepsilon = 10^{-4}$, $x^0_h = \tilde{g}_h$. The nonlinear algebraic system which corresponds to (28) is solved using a simple fixed-point iterative scheme with tolerance 10^{-4}, *cf.*, Section 5 of [4]. If not mentioned otherwise we use the time step $\tau = 10^{-3}$ and the mesh size $h = 2^{-6}$.

We consider the problem with additive noise $\sigma(X) \equiv \sigma = 1$ first. The time-evolution of the discrete energy functional $\mathcal{J}_\varepsilon$ for one realization of the space-time noise W_h is displayed in Figure 2 (left); P1 denotes the

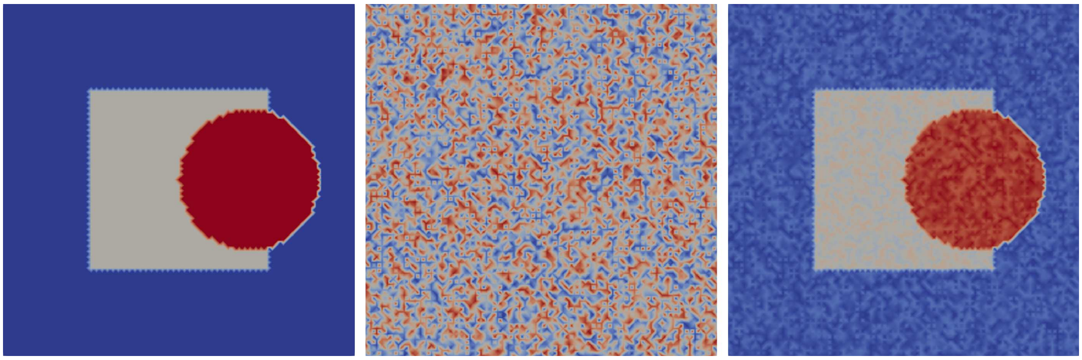

FIGURE 1. The original image $\tilde{g}_h$ (*left*), the noise ξ_h (*middle*) and the noisy image g_h (*right*).

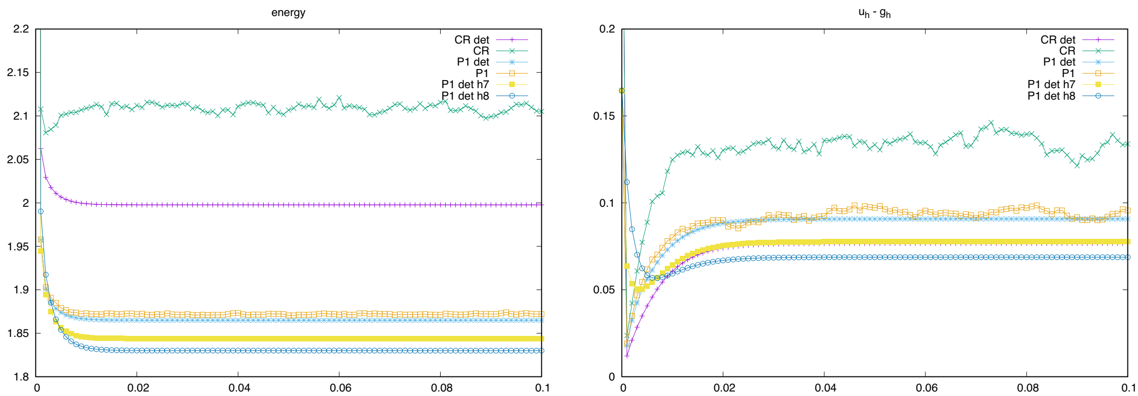

FIGURE 2. Evolution of the discrete total energy $t_i \to \mathcal{J}_\varepsilon(X^i_{\varepsilon,h})$ (*left*) and of the distance to the exact image $t_i \to \tilde{\mathcal{J}}_\lambda(X^i_{\varepsilon,h})$ (*right*).

solution with the conforming finite element approximation and CR denotes the non-conforming approximation, h7, h8 respectively denote the solution with mesh size $h = 2^{-7}$, 2^{-8} and *det* stands for the deterministic solution with $\sigma = 0$. The evolution of the fidelity term $\tilde{\mathcal{J}}_\lambda(X) := \frac{\lambda}{2}\|X - \tilde{g}_h\|^2$, which measures the quality of the approximation of the exact image $\tilde{g}_h$, is displayed in Figure 2 (right). We make the following observations for the conforming finite element method: in the deterministic case (*i.e.*, for $\sigma = 0$) the value of the term $\tilde{\mathcal{J}}_\lambda$ decreases with decreasing mesh size, in the stochastic case it oscillates around the values of its deterministic counterpart. For the non-conforming approximation we measure the quality of the approximation using a modified fidelity term $\tilde{\mathcal{J}}_\lambda(\Pi^0_h X)$, where Π^0_h is the projection onto piecewise constant functions on $\mathcal{T}_h$, see Figure 3 where we also display the solution of the conforming finite element scheme. As expected, *cf.*, [3], on the same mesh with $\sigma = 0$ the non-conforming finite element method yields a better approximation of the original image then the conforming method. The non-conforming approximation requires roughly 3 times more degrees of freedom than the conforming one but the approximation is still comparable to the conforming method with smaller mesh size $h = 2^{-7}$ (which involves 4 times more degrees of freedom than the approximation with $h = 2^{-6}$). Nevertheless, we also observe that the non-conforming approximation is more sensitive to the noise, *i.e.*, the corresponding fidelity term $\tilde{\mathcal{J}}_\lambda$ attains larger values for $\sigma = 1$ then its counterpart for the conforming approximation. For comparison in Figure 4 we also display the piecewise constant projections of the solutions computed with the conforming scheme with $h = 2^{-6}$ and $h = 2^{-8}$.

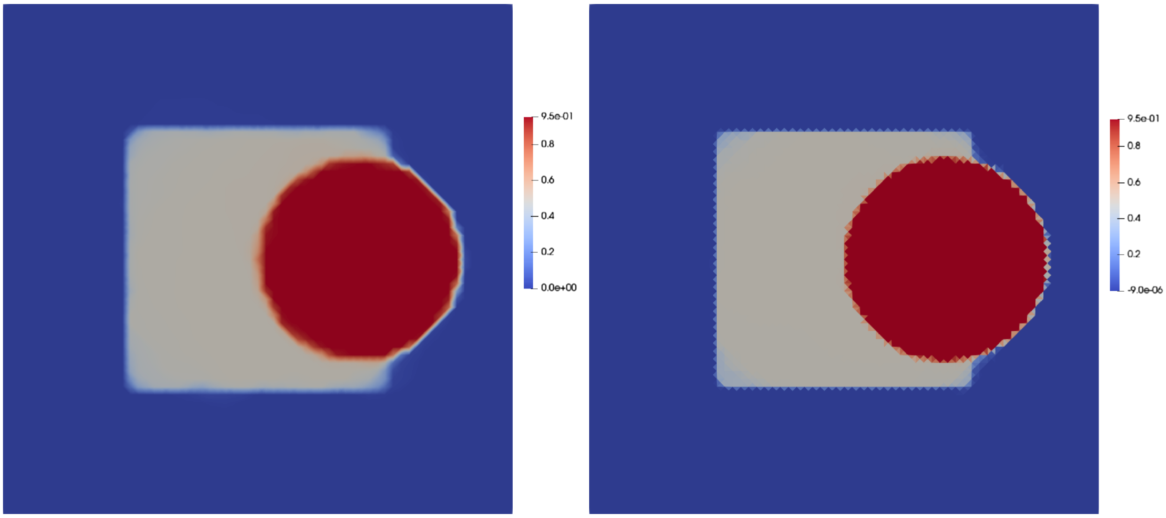

FIGURE 3. Solution computed with the conforming finite element scheme (*left*) and the projected solution of the non-conforming finite element scheme (*right*).

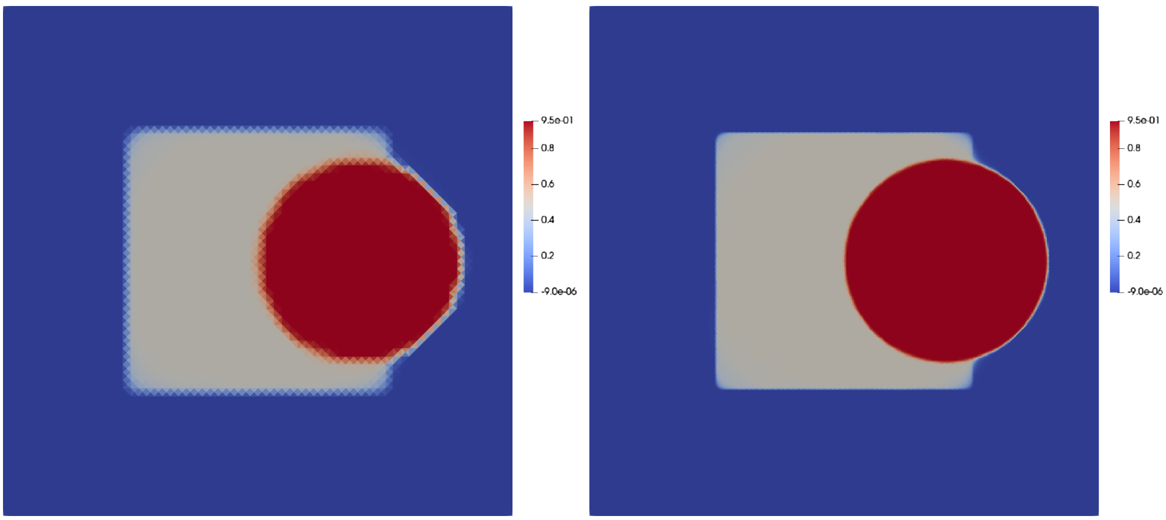

FIGURE 4. Projected solution of the conforming finite element scheme with $\sigma = 0$ for $h = 2^{-6}$ (*left*) and $h = 2^{-8}$ (*right*) at $T = 0.1$.

Next, we consider the scheme (28) (with the conforming spatial discretization) with multiplicative noise. We choose $\sigma = \{\sigma_1(X), \sigma_2(X)\}$ with $\sigma_1(X) = |X - g_h|$, $\sigma_2(X) = |X - g_h|^{1/2}$, note that σ_1 does not satisfy the conditions ($\mathbf{B}_1$), ($\mathbf{B}_2$) while σ_2 does. The remaining parameters were the same as in the previous experiments. In Figure 5 we display the evolution of the discrete energy functional $\mathcal{J}_\varepsilon$ and the evolution of the fidelity term $\tilde{\mathcal{J}}_\lambda$ for the two choices of the multiplicative noise as well as for the additive noise with $\sigma = 1$ and the deterministic case $\sigma = 0$. The energy in the multiplicative noise case is lower then in the case of additive noise. This can be attributed to the fact that the multiplicative noise with $\sigma = \{\sigma_1, \sigma_2\}$ has lower intensity when the numerical solution is close to the noisy image g_h. Analogously, the lower noise intensity in the multiplicative case results in

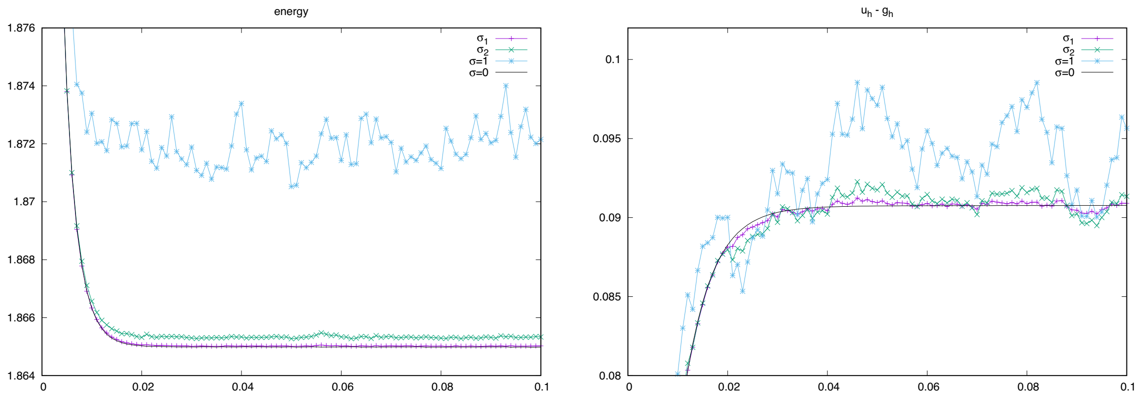

FIGURE 5. Evolution of the discrete total energy $t_i \to \mathcal{J}_\varepsilon(X^i_{\varepsilon,h})$ (*left*) and of the distance to the exact image $t_i \to \tilde{\mathcal{J}}_\lambda(X^i_{\varepsilon,h})$ (*right*) for multiplicative noise σ_1, σ_2 and for additive noise with $\sigma = 1$.

better approximation of the exact image (*i.e.*, lower values of $\tilde{\mathcal{J}}_\lambda$) then in the additive noise case. Furthermore, since $\sigma_1(s) < \sigma_2(s)$ for $|s| < 1$, the fidelity term $\tilde{\mathcal{J}}_\lambda$ attains slightly lower values for $\sigma = \sigma_1$ then for $\sigma = \sigma_2$. Graphically the numerical solutions with mutliplicative noise $\sigma = \{\sigma_1, \sigma_2\}$ were very similar to those shown in Figure 3 and are therefore not displayed.

APPENDIX A. PROOFS OF THE RESULTS FROM SECTION 5

A.1. Proof of Theorem 5.1

The proof is analogous to that of the generalized Arzela–Ascoli theorem *e.g.*, Theorem 7.6 of [18]. Denote by $\tau_{\mathbf{p}}$ the topology of pointwise convergence on $Y^{[0,T]}$. Apparently, $\tau_{\mathbf{p}} \subseteq \tau_{\mathbf{u}}$. Basically, (i) yields that $\overline{M}^{\tau_{\mathbf{p}}}$ is compact in $Y^{[0,T]}$ by the Tychonoff theorem, and the traces of $\tau_{\mathbf{p}}$ and $\tau_{\mathbf{u}}$ coincide on $\overline{M}^{\tau_{\mathbf{p}}}$ by (ii). To see the latter, fix an absolutely convex neighbourhood of zero O and get $\delta > 0$ and $m \in \mathbb{N}$ from (ii). Let D be a finite subset of $(T\mathbb{Q}) \cap [0,T]$ that contains all t^n_j for $0 \le j \le n \le m$, let D intersect each non-empty intersection $(t^k_{i-1}, t^k_i) \cap (t^l_{j-1}, t^l_j)$ whenever $1 \le i \le k \le m$, $1 \le j \le l \le m$, and let D be a δ-net in (t^k_{i-1}, t^k_i) for every $1 \le i \le k \le m$. With these preparations, if $f, g \in M$ are such that $f(r) - g(r) \in O$ for every $r \in D$ then $f(t) - g(t) \in 3O$ for every $t \in [0,T]$. Thus, if $f, g \in \overline{M}^{\tau_{\mathbf{p}}}$ are such that $f(r) - g(r) \in O$ for every $r \in D$ then $f(t) - g(t) \in 3\overline{O}$ for every $t \in [0,T]$. In particular, (ii) yields that $\tau_{\mathbf{p}}$ is stronger than $\tau_{\mathbf{u}}$ on $\overline{M}^{\tau_{\mathbf{p}}}$. But since $\tau_{\mathbf{p}}$ is weaker than $\tau_{\mathbf{u}}$, the topologies coincide on $\overline{M}^{\tau_{\mathbf{p}}}$. Now (ii) also yields

$$\overline{M}^{\tau_{\mathbf{p}}} = \bigcap_{n=1}^{\infty} \left\{ \overline{M_n^{\uparrow}}^{\tau_{\mathbf{p}}} \cup \bigcup_{m=1}^{n-1} \overline{M \cap Q_m}^{\tau_{\mathbf{p}}} \right\} \subseteq Q_\infty \cup \bigcap_{n=1}^{\infty} \overline{M_n^{\uparrow}}^{\tau_{\mathbf{p}}} \subseteq Q_\infty \cup C([0,T];Y) = Q_c. \tag{A.1}$$

The implication (iii) $\Rightarrow$ (i) is obvious and one gets (iii) $\Rightarrow$ (ii) by contradiction.

To prove (iii) $\Rightarrow$ (iv) and the assertion in Remark 5.4, we are going to use only the fact that $f(s+)$ and $f(t-)$ exist for every $0 \le s < t \le T$ and every f in $\overline{M}$. For let K be the closure of M and define

$$R = \{f(t-), f(t), f(t+) : t \in [0,T],\ f \in K\},$$

where $f(0-) := f(0)$ and $f(T_+) := f(T)$. The definition of R is correct since we know by (A.1) that $K \subseteq Q_\infty \cup C([0,T];Y)$ if (iii) holds, or we refer to Remark 5.2. Let us prove that R is compact in Y. For let $\mathcal{U}$ be an ultrafilter in R and define

$$S_U = \{(t,f) \in [0,T] \times C : \{f(t-), f(t), f(t+)\} \cap U \neq \emptyset\}.$$

Then $\{S_U : U \in \mathcal{U}\}$ is a basis of a filter in the compact space $[0,T] \times K$, and therefore it converges to some $(s,g) \in [0,T] \times K$. We conclude that

$$[(g(s-) + O) \cup (g(s) + O) \cup (g(s+) + O)] \cap U \neq \emptyset,$$

holds for every $U \in \mathcal{U}$ and every neighbourhood O of zero in Y. Since $\mathcal{U}$ is an ultrafilter,

$$[(g(s-) + O) \cup (g(s) + O) \cup (g(s+) + O)] \cap R \in \mathcal{U},$$

and so $\mathcal{U}$ converges to one of the elements in the set $\{g(s-), g(s), g(s+)\}$.

A.2. Proof of Corollary 5.1

Say that f takes values in some compact K for every $f \in M$, let $\{[|\cdot|_n < 1] : n \in \mathbb{N}\}$ be a basis of absolutely convex open neighbourhoods of zero in the compact set

$$C = \bigcup_{\max\{|a|,|b|\} \le 1} (aK + bK),$$

for some continuous pseudonorms $|\cdot|_n$ on Y and define

$$d(y_1, y_2) = \sum_{n=1}^{\infty} 2^{-n} \min\{1, |y_1 - y_2|_n\}, \qquad y_1, y_2 \in Y.$$

Then

$$D(f,g) = \sup\{d(f(t), g(t)) : t \in [0,T]\}, \qquad f, g \in Q([0,T];Y), \tag{A.2}$$

metrizes the topology on M.

A.3. Proof of Corollary 5.2

It suffices to prove the assertion for compact sets M in $Q_c([0,T];Y)$. The mapping $f \mapsto D(f,g)$ is $\mathcal{Y}_T$-measurable for every $g \in Q([0,T];Y)$ by Remark 5.1, hence the traces of $\mathscr{B}(Q([0,T];Y))$ and $\mathcal{Y}_T$ coincide on M as (M,D) is a separable metric space by Corollary 5.1. Now it suffices to prove that M itself belongs to $\mathcal{Y}_T$. According to Theorem 5.1, there exist $\{m_n : n \in \mathbb{N}\} \subseteq \mathbb{N}$ and $\{\delta_n : n \in \mathbb{N}\} \subseteq (0,\infty)$ such that $M \subseteq R$ where

$$R = \left[\bigcap_{t\in[0,T]} \pi_t^{-1}[K]\right] \cap \bigcap_{n=1}^{\infty} \left\{ \left[\bigcup_{j=1}^{m_n} Q_j\right] \cup \left[\bigcap_{|t-s|\le\delta_n} \{f : |f(t) - f(s)|_n \le 1\}\right]\right\},$$

and K and $\{|\cdot|_n\}$ are the same as in the proof of Corollary 5.1. But R is closed (as an intersection of closed sets), relatively compact in $Q_c([0,T];Y)$ by Theorem 5.1 (hence compact), and $\mathcal{Y}_T$-measurable as

$$R = \left[\bigcap_{t\in D_T} \pi_t^{-1}[K]\right] \cap \bigcap_{n=1}^{\infty} \left\{ \left[\bigcup_{j=1}^{m_n} Q_j\right] \cup \left[\bigcap_{t,s\in D_T,\, |t-s|\le\delta_n} \{f : |f(t) - f(s)|_n \le 1\}\right]\right\},$$

where $D_T = (T\mathbb{Q}) \cap [0,T]$. Thus the trace of $\mathscr{B}(Q([0,T];Y))$ on R is a subset of $\mathcal{Y}_T$ and, in particular, $M \in \mathcal{Y}_T$.

A.4. Proof of Proposition 5.1

It suffices to prove the first assertion for F_n and F real-valued (otherwise compose theses functions with $x \mapsto \min\{x, m\}$ and then let $m \to \infty$). If $t \in (0, \infty)$ then set $R = (-\infty, t]$, and we have, for every $r \in (0, 1)$,

$$\mu_n(F_n \in R) \leq r + \mu_n\left(\overline{\bigcup_{k=m}^{\infty} [F_k \in R] \cap K_{r,k}}\right), \qquad m \leq n,$$

so

$$\begin{aligned}\limsup \mu_n(F_n \in R) &\leq r + \mu\left(\bigcap_{m=1}^{\infty} \overline{\bigcup_{k=m}^{\infty} [F_k \in R] \cap K_{r,k}}\right)\\ &\leq r + \mu(F \in R) + \mu^*(D_r),\end{aligned}$$

by the classical Portmanteau theorem, *cf.*, Corollary 8.2.10 of [9], hence

$$\liminf \mu_n(F_n > t) \geq \mu(F > t)$$

and therefore

$$\int_X F \mathrm{d}\mu = \int_0^\infty \mu(F > t)\,\mathrm{d}t \leq \liminf \int_0^\infty \mu_n(F_n > t)\,\mathrm{d}t = \liminf \int_X F_n\,\mathrm{d}\mu,$$

by the Fatou lemma. The second part of the proof is analogous but we take any closed set R. In this way, we get

$$\limsup \mu_n(F_n \in R) \leq \mu(F \in R),$$

for every R closed, therefore $\limsup \mu_n(F_n \in \cdot\,) \Rightarrow \mu(F \in \cdot\,)$. The first part of the proof now yields that $|F|$ is integrable with respect to μ, and we get the claim by the assumption of uniform integrability of $|F_n|\,\mathrm{d}\mu_n$.

Appendix B. Bounded variation spaces

Lemma B.1. *The functional*

$$\mathcal{I}(u) = \sup\left\{\int_{\mathcal{O}} u \operatorname{div}\varphi\,\mathrm{d}x : \varphi \in C^\infty\big(\mathbb{R}^d; \mathbb{R}^d\big),\, |\varphi| \leq 1\right\}, \qquad u \in \mathbb{L}^1,$$

satisfies

$$\mathcal{I}(u) = \|\nabla u\|_{\mathrm{TV}(\mathcal{O})} + \int_{\partial\mathcal{O}} |u|\,\mathrm{d}x, \qquad \text{for } u \in BV(O),$$

and $\mathcal{I}(u) = \infty$ for $u \in \mathbb{L}^1 \setminus BV(O)$ where $\|\nu\|_{TV(\mathcal{O})}$ denotes the total variation of a vector measure on $\mathcal{O}$. In particular, $\mathcal{I}$ is lower semicontinuous on $(\mathbb{L}^1, \text{weak})$ and convex on $BV(\mathcal{O})$ and $\mathcal{J}$ is lower weakly semicontinuous on $\mathbb{L}^2$ and convex on $\mathbb{L}^2 \cap BV(\mathcal{O})$.

Proof. If $\mathcal{I}(u) < \infty$ then $u \in BV(O)$ *e.g.*, by Proposition 3.6 in [1]. If $u \in BV(O)$ then

$$\int_{\mathcal{O}} u \operatorname{div}\varphi\,\mathrm{d}x = \int_{\partial\mathcal{O}} u(\varphi, \nu)\,\mathrm{d}S - \int_{\mathcal{O}} \varphi \cdot \mathrm{d}\nabla u, \qquad \varphi \in C^\infty\big(\mathbb{R}^d; \mathbb{R}^d\big),$$

where ν is the outer normal vector field on ∂O by the integration by parts formula, see *e.g.*, (3.85) in [1], so

$$\mathcal{I}(u) = \sup\left\{\int_{\overline{\mathcal{O}}} \varphi \cdot \mathrm{d}\,\theta : \varphi : \mathbb{R}^d \to \mathbb{R}^d \text{ Borel measurable}, |\varphi| \leq 1\right\},$$

by a standard density argument (*cf.*, [1]) where $\theta = u\nu\mathcal{H}_{d-1}|_{\partial O} - \nabla u$. Hence

$$\mathcal{I}(u) = \|\theta\|_{\mathrm{TV}(\overline{\mathcal{O}})} = \|\nabla u\|_{\mathrm{TV}(\mathcal{O})} + \|u\nu\mathcal{H}_{d-1}\|_{\mathrm{TV}(\partial\mathcal{O})} = \|\nabla u\|_{\mathrm{TV}(\mathcal{O})} + \int_{\partial\mathcal{O}} |u|\,\mathrm{d}S.$$

□

Remark B.1. There exists a countable subset $\mathcal{H}$ of $C^\infty(\mathbb{R}^d)$ such that

$$\mathcal{I}(u) = \sup\left\{\int_{\mathcal{O}} u\phi \,\mathrm{d}x : \phi \in \mathcal{H}\right\}, \qquad u \in \mathbb{L}^1,$$

by separability of $\{\operatorname{div}\varphi : \varphi \in C^\infty(\mathbb{R}^d;\mathbb{R}^d),\ |\varphi| \le 1\}$ in $C^\infty(\mathbb{R}^d)$, see *e.g.*, [13].

Appendix C. Besov spaces

Lemma C.1. *Let Y be a Banach space and let $f : [0,T] \to Y$ be a continuous function linear on every $[t_i, t_{i+1}]$ for $i = 0, \dots, N-1$ and define*

$$f_{i,a} = \left[\tau \sum_{j=i}^{N} \|f(t_j) - f(t_{j-i})\|^a\right]^{\frac{1}{a}}, \qquad f_{i,\infty} = \max_{i \le j \le N} \|f(t_j) - f(t_{j-i})\|.$$

Then

$$\|f\|_{L^r(0,T)} \le \left[\sum_{i=0}^{N} \tau \|f(t_i)\|^r\right]^{\frac{1}{r}}, \qquad \|f\|_{L^\infty(0,T)} \le \max_{0 \le i \le N} \|f(t_i)\|,$$

$$[f]_{B^s_{p,q}} \le \frac{8}{s(1-s)} \left(\sum_{i=1}^{N-1} \tau \frac{f_{i,p}^q}{t_i^{1+sq}}\right)^{\frac{1}{q}}, \quad [f]_{B^s_{p,\infty}} \le 3 \max_{1 \le i < N} \frac{f_{i,p}}{t_i^s},$$

for every $s \in (0,1)$, $p \in [1,\infty]$ and $r, q \in [1,\infty)$.

Acknowledgements. We thank the referees for the careful reading of the manuscript and their valuable suggestions. The first named author was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1283/2 2021 – 317210226 while the second named author by the Czech Science Foundation grant no. 22-12790S.

References

[1] L. Ambrosio, N. Fusco and D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York (2000).

[2] V. Barbu and M. Röckner, Stochastic variational inequalities and applications to the total variation flow perturbed by linear multiplicative noise. *Arch. Ration. Mech. Anal.* **209** (2013) 797–834.

[3] S. Bartels, Nonconforming discretizations of convex minimization problems and precise relations to mixed methods. *Comput. Math. Appl.* **93** (2021) 214–229.

[4] Ľ. Baňas and A. Wilke, A posteriori estimates for the stochastic total variation flow. *SIAM J. Numer. Anal.* **60** (2022) 2657–2680.

[5] Ľ. Baňas, M. Röckner and A. Wilke, Convergent numerical approximation of the stochastic total variation flow. *Stoch. Part. Differ. Equ. Anal. Comput.* **9** (2021) 437–471.

[6] Ľ. Baňas, M. Röckner and A. Wilke, Convergent numerical approximation of the stochastic total variation flow with linear multiplicative noise: the higher dimensional case. arXiv:2211.04162 (2022).

[7] Ľ. Baňas, M. Röckner and A. Wilke, Correction to: Convergent numerical approximation of the stochastic total variation flow. *Stoch. Part. Differ. Equ. Anal. Comput.* (2022). DOI: 10.1007/s40072-022-00267-5.

[8] P. Billingsley, Convergence of probability measures. *Wiley Series in Probability and Statistics: Probability and Statistics*, 2nd edition. A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York (1999).

[9] V.I. Bogachev, Measure Theory. Vol. I, II. Springer-Verlag, Berlin (2007).

[10] J.H. Bramble, J.E. Pasciak and O. Steinbach, On the stability of the L^2 projection in $H^1(\Omega)$. *Math. Comp.* **71** (2002) 147–156.

[11] D. Breit, E. Feireisl and M. Hofmanová, Stochastically Forced Compressible Fluid Flows. Vol. 3 of *De Gruyter Series in Applied and Numerical Mathematics*. De Gruyter, Berlin (2018).

[12] S.C. Brenner and L.R. Scott, The Mathematical Theory of Finite Element Methods. Vol. 15 of *Texts in Applied Mathematics*, 3rd edition. Springer, New York (2008).

[13] J. Dieudonné, Sur les espaces de Montel métrisables. *C. R. Acad. Sci. Paris* **238** (1954) 194–195.

[14] X. Feng and A. Prohl, Analysis of total variation flow and its finite element approximations. *M2AN Math. Model. Numer. Anal.* **37** (2003) 533–556.

[15] I. Gyöngy and N. Krylov, Existence of strong solutions for Itô's stochastic equations via approximations. *Probab. Theory Relat. Fields* **105** (1996) 143–158.

[16] A. Jakubowski, The almost sure Skorokhod representation for subsequences in nonmetric spaces. *Teor. Veroyatnost. i Primenen.* **42** (1997) 209–216.

[17] I. Karatzas and S.E. Shreve, Brownian Motion and Stochastic Calculus. Vol. 113 of *Graduate Texts in Mathematics*, 2nd edition. Springer-Verlag, New York (1991).

[18] J.L. Kelley, General Topology. Graduate Texts in Mathematics, No. 27. Springer-Verlag, New York-Berlin (1975). Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.].

[19] M. Ondreját, A. Prohl and N. Walkington, Numerical approximation of nonlinear SPDE's. *Stoch. Part. Differ. Equ. Anal. Comput.* (2022). DOI: 10.1007/s40072-022-00271-9.

[20] G. Pisier, Martingales in Banach Spaces. Vol. 155 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge (2016).

[21] J. Simon, Sobolev, Besov and Nikolskiĭ fractional spaces: imbeddings and comparisons for vector valued spaces on an interval. *Ann. Mat. Pura Appl.* **157** (1990) 117–148.