

# EFFICIENT NUMERICAL APPROXIMATION OF A NON-REGULAR FOKKER–PLANCK EQUATION ASSOCIATED WITH FIRST-PASSAGE TIME DISTRIBUTIONS

UDO BOEHM, SONJA COX, GREGOR GANTNER, AND ROB STEVENSON

**ABSTRACT.** In neuroscience, the distribution of a decision time is modelled by means of a one-dimensional Fokker–Planck equation with time-dependent boundaries and space-time-dependent drift. Efficient approximation of the solution to this equation is required, e.g., for model evaluation and parameter fitting. However, the prescribed boundary conditions lead to a strong singularity and thus to slow convergence of numerical approximations. In this article we demonstrate that the solution can be related to the solution of a parabolic PDE on a rectangular space-time domain with homogeneous initial and boundary conditions by transformation and subtraction of a known function. We verify that the solution of the new PDE is indeed more regular than the solution of the original PDE and proceed to discretize the new PDE using a space-time minimal residual method. We also demonstrate that the solution depends analytically on the parameters determining the boundaries as well as the drift. This justifies the use of a sparse tensor product interpolation method to approximate the PDE solution for various parameter ranges. The predicted convergence rates of the minimal residual method and that of the interpolation method are supported by numerical simulations.

## 1. INTRODUCTION

In 1978 Ratcliff [Rat78] introduced a model for binary decision processes based on diffusion processes. This model turned out to agree well with experimental data; Gold and Shadlen [GS01] provides a neurophysiological explanation for its success. Indeed, the solution  $(X_t)_{t \geq 0}$  of a one-dimensional stochastic differential equation is assumed to describe the difference in activity of two competing neuron populations. At time  $t = 0$ , the value  $X_0 = x_0 \in \mathbb{R}$  represents the resting-state activity of the neuron populations. A decision is triggered when  $(X_t)_{t \geq 0}$  first reaches one of two (possibly time-dependent) critical values  $\alpha$  or  $\beta$ , each reflecting an outcome of the decision process.

In a typical decision experiment, scientists can only measure the decision time and outcome. Parameter fitting thus requires access to the decision time distributions, which are rarely known explicitly. Ad hoc numerical simulations are costly whence efficient simulation methods are much sought-after [HFW<sup>+</sup>15, FFGC20].

The goal of this article is to extend and improve a simulation method introduced in [VV08], which is based on the Fokker–Planck equation associated to the

---

*Date:* March 9, 2021.

*2010 Mathematics Subject Classification.* 30B40, 35A15, 35B65, 35K08, 65D05, 65M12.

*Key words and phrases.* Fokker–Planck equation, time-dependent spatial domain, space-time variational formulation, parameter dependent PDE, sparse tensor product interpolation.

decision time. To explain this approach consider the following stochastic differential equation:

$$(1.1) \quad dX_t^y = \mu(t, X_t^y) dt + \sigma dW_t \quad t \in [0, \infty), X_0^y = y.$$

Here  $(W_t)_{t \in [0, \infty)}$  is a Brownian motion,  $\sigma \in (0, \infty)$  is the diffusion parameter,  $\mu \in C([0, \infty) \times \mathbb{R})$  is the (time- and state-dependent) drift and  $y \in \mathbb{R}$  is the initial value. Let  $\alpha, \beta \in C^1([0, \infty))$  satisfy  $\alpha \leq \beta$ , and for all  $y \in [\alpha(0), \beta(0)]$  define the stopping times  $\hat{\alpha}_y, \hat{\beta}_y$  by

$$(1.2) \quad \begin{aligned} \hat{\alpha}_y &:= \inf\{t \in [0, \infty) : X_t^y \leq \alpha(t)\}, \\ \hat{\beta}_y &:= \inf\{t \in [0, \infty) : X_t^y \geq \beta(t)\}. \end{aligned}$$

The quantities of interest in neurophysiological decision models are the *first hitting time probabilities*:  $\mathbb{P}[\hat{\alpha}_y \leq \min(\tau, \hat{\beta}_y)]$ , where  $\tau \in (0, \infty)$  and  $y \in [\alpha(0), \beta(0)]$ . These probabilities can be linked to the solution of a parabolic PDE. Indeed, assume  $\alpha < \beta$  on  $[0, \tau]$  for some  $\tau \in (0, \infty)$ , set  $Q := \{(t, x) \in (0, \tau) \times \mathbb{R} : \alpha(\tau - t) < x < \beta(\tau - t)\}$ , and consider the following PDE:

$$(1.3) \quad \begin{cases} \partial_t F(t, x) = \frac{\sigma^2}{2} \partial_x^2 F(t, x) + \mu(\tau - t, x) \partial_x F(t, x) & (t, x) \in Q_\tau, \\ F(t, \alpha(\tau - t)) = 1, \quad F(t, \beta(\tau - t)) = 0 & t \in (0, \tau), \\ F(0, x) = 0 & x \in (\alpha(\tau), \beta(\tau)). \end{cases}$$

Under some additional regularity assumptions on  $\alpha$ ,  $\beta$ , and  $\mu$  it can be shown that a solution to (1.3) exists and satisfies

$$(1.4) \quad \mathbb{P}[\hat{\alpha}_y \leq \min(\tau, \hat{\beta}_y)] = F(\tau, y), \quad \alpha(0) \leq y \leq \beta(0).$$

(see [VV08, Appendix A] for the case that  $\alpha$  and  $\beta$  are constant and  $\mu$  does not depend on time or [Øks98, Chapter 7] for general Fokker–Planck equations, also known as Kolmogorov backwards equations).

In [VV08], a Crank–Nicolson method is used to approximate solutions to (1.3) in the case that  $\alpha$ ,  $\beta$ , and  $\mu$  are constant. One advantage of this setting is that one only needs to solve a single PDE of type (1.3) in order to obtain the first hitting time probabilities  $\mathbb{P}[\hat{\alpha}_y \leq \min(t, \hat{\beta}_y)]$  for all  $t \in [0, \tau]$ ,  $y \in [\alpha(0), \beta(0)]$ . However, due to the fact that  $F$  is discontinuous at  $(t, x) = (0, \alpha(\tau))$ , no proof of convergence of the Crank–Nicolson for decreasing step-sizes seems available. At best, reduced rates are to be expected. Moreover, various authors have argued that time-dependent boundaries  $\alpha$  and  $\beta$  and space-time-dependent drift  $\mu$  provide a more realistic model for decision processes, see [HFW<sup>+</sup>15, SK13] for an overview.

In this article we *extend* [VV08] to include diffusion models with time-dependent boundaries and non-constant drift. We *improve* the efficiency of the numerical simulation by not approximating the solution  $F$  to (1.3) directly, instead, we approximate the solution to a parabolic PDE on a rectangular domain with homogeneous initial and boundary conditions constructed such that its difference with  $F$  (transformed to the same rectangular domain) is a function for which a rapidly converging series expansion is known.

More specifically, in Section 2 we demonstrate that if  $\alpha, \beta$  are once continuously differentiable, then (1.3) can be transformed into a parabolic PDE on a rectangular domain with a space-time-dependent drift. Next, in Section 3 we demonstrate that by subtracting a known, discontinuous function, we obtain a parabolic PDE with homogeneous boundary conditions, see (3.1) below. We analyze the regularity of

the solution  $e$  to this equation and verify that it is indeed smoother than  $F$ , see Corollary 3.3 and Theorem 3.8.

In Section 4 we apply a minimal residual method [And13, SW20b, SW20a] to approximate the solution  $e$  to (3.1). This method is known to give quasi-best approximations from the selected trial space in the norm on a natural solution space being the intersection of two Bochner spaces. Taking as trial space the space of continuous piecewise bilinears with respect to a uniform partition of the space-time cylinder into rectangles with mesh width  $h$ , in Theorem 4.2 the optimal error bound of order  $h$  is shown for the solution  $e$  to (3.1).

In Section 5 we consider the situation that  $\mu$ ,  $\alpha$ , and  $\beta$  can be parametrized analytically and verify that in this case the corresponding solution  $e$  to (3.1) (transformed onto the unit square) depends analytically on these parameters as well as on the final time  $\tau$ , see Theorem 5.1. This justifies the use of a sparse tensor-product interpolation [NTW08] to determine the solution  $e$  to (3.1) efficiently for multiple end-time and parameter values. Finally, in Section 6 we provide numerical simulations for three different decision models taken from the neurophysiological literature.

In a parallel publication [BCGS21], aimed at the neuroscientific community, we provide further numerical experiments and code. In particular, in [BCGS21] we apply the Crank–Nicolson method to approximate the solution  $e$  to (3.1). In the examples we consider it appears that the Crank–Nicolson method leads to similar convergence as the minimal residual method. Although theoretical results for the minimal residual method are stronger, Crank–Nicolson may be preferred in practice as it is easier to implement. We refer to [BCGS21] for further details.

**1.1. Notation.** In this work, by  $C \lesssim D$  we mean that  $C$  can be bounded by a multiple of  $D$ , independently of parameters which  $C$  and  $D$  may depend on. Obviously,  $C \gtrsim D$  is defined as  $D \lesssim C$ , and  $C \approx D$  as  $C \lesssim D$  and  $C \gtrsim D$ .

For normed linear spaces  $E$  and  $F$ , by  $\mathcal{L}(E, F)$  we denote the normed linear space of bounded linear mappings  $E \rightarrow F$ , and by  $\mathcal{L}_{\text{iso}}(E, F)$  its subset of boundedly invertible linear mappings  $E \rightarrow F$ .

## 2. TRANSFORMING THE FOKKER–PLANCK EQUATION TO A RECTANGULAR SPACE-TIME DOMAIN

In this section we demonstrate that (1.3) can be transformed into a PDE on a rectangular space-time domain, see (2.3) below. The PDE in (2.3) below forms the starting point for the remainder of this article, which is why we use tildes in (2.1) below to distinguish the variables and coefficients of the non-transformed equation from those in (2.3). Indeed, let  $\tilde{T} \in (0, \infty]$ , assume  $a, b \in C^1([0, \tilde{T}])$  satisfy  $a(\tilde{t}) < b(\tilde{t})$  for all  $\tilde{t} \in [0, \tilde{T}]$ , set  $\tilde{Q} := \{(\tilde{t}, \tilde{x}) \in (0, \tilde{T}) \times \mathbb{R} : a(\tilde{t}) < \tilde{x} < b(\tilde{t})\}$ , let  $\tilde{v} \in L_\infty(\tilde{Q})$ , and consider the following parabolic initial- and boundary value problem:

$$(2.1) \quad \begin{cases} \partial_{\tilde{t}} \tilde{u}(\tilde{t}, \tilde{x}) = \partial_{\tilde{x}}^2 \tilde{u}(\tilde{t}, \tilde{x}) + \tilde{v}(\tilde{t}, \tilde{x}) \partial_{\tilde{x}} \tilde{u}(\tilde{t}, \tilde{x}) & (\tilde{t}, \tilde{x}) \in \tilde{Q}, \\ \tilde{u}(\tilde{t}, a(\tilde{t})) = 1, \quad \tilde{u}(\tilde{t}, b(\tilde{t})) = 0 & \tilde{t} \in (0, \tilde{T}), \\ \tilde{u}(0, \tilde{x}) = 0 & \tilde{x} \in (a(0), b(0)). \end{cases}$$

Note that this is (1.3) with  $\tilde{u}(\tilde{t}, \tilde{x}) = F(\frac{2\tilde{t}}{\sigma}, \tilde{x})$ ,  $\tilde{T} = \frac{\sigma\tau}{2}$ ,  $a(\tilde{t}) = \alpha(\frac{2}{\sigma}(\tilde{T} - \tilde{t}))$ ,  $b(\tilde{t}) = \beta(\frac{2}{\sigma}(\tilde{T} - \tilde{t}))$ ,  $\tilde{v}(\tilde{t}, \tilde{x}) = \mu(\frac{2}{\sigma}(\tilde{T} - \tilde{t}), \tilde{x})$ .

Now, consider the following ODE

$$(2.2) \quad \theta'(t) = (b(\theta(t)) - a(\theta(t)))^2, \quad \theta(0) = 0.$$

For some  $T \in (0, \infty]$ , it has a unique solution  $\theta \in C^1([0, T])$  that is strictly increasing with  $\theta(t) < \tilde{T}$  on  $[0, T)$ . This  $T$  can be taken such that  $\lim_{t \uparrow T} \theta(t) = \tilde{T}$ . Indeed, pick  $R \in (0, \tilde{T})$ , and let  $\varepsilon := \min_{\tilde{t} \in [0, R]} b(\tilde{t}) - a(\tilde{t}) > 0$ . Then as long as  $\theta(t) \in [0, R]$ , it holds that  $\theta'(t) \geq \varepsilon^2$ , and so  $\theta(t + \delta) \geq \theta(t) + \delta\varepsilon^2$ , which shows that there exists a  $\delta \in (0, \frac{R - \theta(t)}{\varepsilon^2}]$  with  $\theta(t + \delta) \geq R$ . In conclusion we have that  $\theta: [0, T) \rightarrow [0, \tilde{T})$  is a bijection.

With

$$\Omega := (0, 1),$$

and  $\zeta: [0, \tilde{T}) \times \overline{\Omega} \rightarrow \mathbb{R}$  defined by

$$\zeta(\tilde{t}, x) := (1 - x)a(\tilde{t}) + xb(\tilde{t}),$$

we have that

$$[0, T) \times \Omega \rightarrow \tilde{Q}: (t, x) \mapsto (\theta(t), \zeta(\theta(t), x))$$

is a bijection with inverse

$$(\tilde{t}, \tilde{x}) \mapsto \left( \theta^{-1}(\tilde{t}), \frac{\tilde{x} - a(\tilde{t})}{b(\tilde{t}) - a(\tilde{t})} \right).$$

Defining  $u, v: [0, T) \times \overline{\Omega} \rightarrow \mathbb{R}$  by

$$u(t, x) := \tilde{u}(\theta(t), \zeta(\theta(t), x)),$$

$$v(t, x) := (b(\theta(t)) - a(\theta(t))) [\tilde{v}(\theta(t), \zeta(\theta(t), x)) + (1 - x)a'(\theta(t)) + xb'(\theta(t))],$$

we have  $u(t, 0) = 1$ ,  $u(t, 1) = 0$  ( $t \in (0, T)$ ), and  $u(0, x) = 0$  ( $x \in \Omega$ ). Moreover, for  $(t, x) \in (0, T) \times \Omega$ , one has

$$\begin{aligned} \partial_x u(t, x) &= (b(\theta(t)) - a(\theta(t))) \partial_{\tilde{x}} \tilde{u}(\theta(t), \zeta(\theta(t), x)), \\ \partial_x^2 u(t, x) &= (b(\theta(t)) - a(\theta(t)))^2 \partial_{\tilde{x}}^2 \tilde{u}(\theta(t), \zeta(\theta(t), x)), \end{aligned}$$

and

$$\begin{aligned} \partial_t u(t, x) &= \theta'(t) \{ \partial_{\tilde{t}} \tilde{u}(\theta(t), \zeta(\theta(t), x)) + \partial_{\tilde{t}} \zeta(\theta(t), x) \partial_{\tilde{x}} \tilde{u}(\theta(t), \zeta(\theta(t), x)) \} \\ &= (b(\theta(t)) - a(\theta(t)))^2 \{ \partial_{\tilde{t}} \tilde{u}(\theta(t), \zeta(\theta(t), x)) + \\ &\quad [(1 - x)a'(\theta(t)) + xb'(\theta(t))] \partial_{\tilde{x}} \tilde{u}(\theta(t), \zeta(\theta(t), x)) \} \\ &= (b(\theta(t)) - a(\theta(t)))^2 \{ \partial_{\tilde{x}}^2 \tilde{u}(\theta(t), \zeta(\theta(t), x)) + \\ &\quad [\tilde{v}(\theta(t), \zeta(\theta(t), x)) + (1 - x)a'(\theta(t)) + xb'(\theta(t))] \partial_{\tilde{x}} \tilde{u}(\theta(t), \zeta(\theta(t), x)) \} \\ &= \partial_x^2 u(t, x) + v(t, x) \partial_x u(t, x). \end{aligned}$$

In other words, with

$$I := (0, T),$$

(2.1) is equivalent to finding  $u = u(v)$  that solves

$$(2.3) \quad \begin{cases} \partial_t u(t, x) = \partial_x^2 u(t, x) + v(t, x) \partial_x u(t, x) & (t, x) \in I \times \Omega, \\ u(t, 0) = 1, \quad u(t, 1) = 0 & t \in I, \\ u(0, x) = 0 & x \in \Omega. \end{cases}$$

For its numerical solution, we always consider system (2.3) for  $T < \infty$ .

*Example 2.1.* Bowman, Kording, and Gottfried [BKG12] suggested collapsing boundaries, i.e., in (1.3) they take  $\alpha(t) := \frac{\beta_0 t}{2T_0}$  and  $\beta(t) := \beta_0(1 - \frac{t}{2T_0})$  for some fixed parameters  $\beta_0, T_0 \in (0, \infty)$ . Translating this to the setting of (2.1), this leads to  $a(\tilde{t}) := \frac{\beta_0(\tilde{T}-\tilde{t})}{\sigma T_0}$  and  $b(\tilde{t}) := \beta_0(1 - \frac{\tilde{T}-\tilde{t}}{\sigma T_0})$  (note that it only makes sense to consider  $\tilde{T} \in (0, \frac{\sigma T_0}{2})$  in this setting). In this case the ODE (2.2) reads as

$$\theta'(t) = \beta_0^2 \left( \frac{\sigma T_0 - 2(\tilde{T} - \theta(t))}{\sigma T_0} \right)^2, \quad \theta(0) = 0,$$

which for  $\tilde{T} \in (0, \frac{\sigma T_0}{2})$  has the solution

$$\theta(t) = \frac{\beta_0^2 (\sigma T_0 - 2\tilde{T})^2 t}{\sigma^2 T_0^2 - 2\beta_0^2 (\sigma T_0 - 2\tilde{T}) t}, \quad t \in [0, \frac{\sigma^2 T_0^2}{2\beta_0^2 (\sigma T_0 - 2\tilde{T})}).$$

Note that  $\theta^{-1}(\tilde{t}) = \frac{\sigma^2 T_0^2 \tilde{t}}{\beta_0^2 (\sigma T_0 - 2\tilde{T}) (\sigma T_0 - 2\tilde{T} + 2\tilde{t})}$  and thus  $T = \theta^{-1}(\tilde{T}) = \frac{\sigma T_0 \tilde{T}}{\beta_0^2 (\sigma T_0 - 2\tilde{T})}$ .

Moreover,  $b(\theta(t)) - a(\theta(t)) = \frac{b_0(1 - 2\sigma T_0 \tilde{T})}{\sigma^2 T_0^2 - 2\beta_0^2 (\sigma T_0 - 2\tilde{T}) t}$ , and  $(1-x)a'(\theta(t)) + xb'(\theta(t)) = \frac{(2x-1)\beta_0}{\sigma T_0}$ , from which one derives  $v$  in terms of  $\tilde{v}$ .

### 3. WELL-POSEDNESS AND REGULARITY OF THE FOKKER-PLANCK EQUATION

Let  $u(v)$  denote the solution to (2.3) for some given drift function  $v$ . Due to the discontinuity between boundary and initial data, it is clear that  $u(v)$  is discontinuous at the corner  $(t, x) = (0, 0)$ . This reduces the rate of convergence of standard numerical methods and makes it difficult to provide a theoretical bound on the convergence rate. However, for *constant drift*  $v$ , a rapidly converging series expansion of  $u(v)$  is known ([GBK14]), which allows to efficiently approximate  $u(v)$  within any given positive tolerance. Knowing this, our approach to approximate  $u(v)$  for *variable*  $v \in C(\overline{I \times \Omega})$  is to approximate the difference  $e = e(v) = u(v) - u(v_0)$ , where  $v_0 := v(0, 0)$ . This function  $e(v)$  solves

$$(3.1) \quad \begin{cases} \partial_t e(t, x) = \partial_x^2 e(t, x) + v(t, x) \partial_x e(t, x) + (v(t, x) - v_0) \partial_x u(v_0) & (t, x) \in I \times \Omega, \\ e(t, 0) = 0, \quad e(t, 1) = 0 & t \in I, \\ e(0, x) = 0 & x \in \Omega, \end{cases}$$

which we solve approximately with a numerical method. To derive a priori bounds for the approximation error, we analyze the smoothness of  $e(v)$ , see Section 3.3. In particular, under additional smoothness conditions on  $v$ , and using that  $(v - v_0)(0, 0) = 0$ , we show that  $e(v)$  is more smooth than  $u(v_0)$ , and thus than  $u(v)$ , which shows the benefit of applying the numerical method to (3.1) instead of directly to (2.3).

It turns out that for any  $v$  the smoothness of  $u(v)$  is determined by that of the solution  $u_H$  of the heat equation on  $(0, \infty) \times \mathbb{R}$  that is 0 at  $t = 0$  and 1 at  $x = 0$ . Its smoothness is the topic of the next subsection.

**3.1. The heat kernel.** The function

$$H(t, x) := \frac{1}{2\sqrt{\pi t}} e^{-\frac{x^2}{4t}}$$

is the heat kernel. It satisfies

$$\begin{aligned} \partial_t H(t, x) &= \partial_x^2 H(t, x) \quad (t, x) \in (0, \infty) \times \mathbb{R}, \\ \lim_{t \downarrow 0} \int_{\mathbb{R}} H(t, x) \phi(x) dx &= \phi(0) \quad (\phi \in \mathcal{D}(\mathbb{R})). \end{aligned}$$

Following [Cos90, Ex. 2.14] and [FF03], for  $(t, x) \in (0, \infty) \times \mathbb{R}$  we define

$$u_H(t, x) := 2 \int_x^\infty H(t, y) dy = \frac{2}{\sqrt{\pi}} \int_{\frac{x}{2\sqrt{t}}}^\infty e^{-s^2} ds = \operatorname{Erfc}\left(\frac{x}{2\sqrt{t}}\right).$$

Knowing that  $\int_0^\infty \frac{1}{\sqrt{\pi t}} e^{-\frac{y^2}{4t}} dy = 1$ , and  $\lim_{t \downarrow 0} \int_x^\infty \frac{1}{\sqrt{\pi t}} e^{-\frac{y^2}{4t}} dy = 0$  for  $x > 0$ , we have

$$\begin{cases} \partial_t u_H(t, x) = \partial_x^2 u_H(t, x) & (t, x) \in (0, \infty) \times \mathbb{R}, \\ u_H(t, 0) = 1 & t > 0, \\ u_H(0, x) := \lim_{t \downarrow 0} u_H(t, x) = 0 & x > 0. \end{cases}$$

The following lemma turns out to be handy to analyze the smoothness of  $u_H$  restricted to  $I \times \Omega$ .

**Lemma 3.1.** *For  $p > 0$ ,  $\alpha, \beta \in \mathbb{R}$ , it holds that  $\int_0^T \int_0^1 |t^\alpha x^\beta e^{-\frac{x^2}{4t}}|^p dx dt < \infty$  if and only if  $p\beta > -1$  and  $p(2\alpha + \beta) > -3$ .*

*Remark 3.2.* The same result holds true when the domain of integration  $I \times \Omega$  reads as  $(0, \infty) \times \Omega$ .

*Proof.* The mapping

$$\Phi: \{(\lambda, x) \in (0, \infty) \times (0, 1) : x < 2\sqrt{\lambda T}\} \rightarrow (0, T) \times (0, 1) : (\lambda, x) \mapsto \left(\frac{x^2}{4\lambda}, x\right)$$

is a diffeomorphism, and  $|D\Phi(\lambda, x)| = \frac{x^2}{4\lambda^2}$ . One obtains

$$\begin{aligned} \int_0^T \int_0^1 |t^\alpha x^\beta e^{-\frac{x^2}{4t}}|^p dx dt &= \int_0^\infty \int_0^{\min(1, 2\sqrt{\lambda T})} \left(\frac{x^2}{4\lambda}\right)^{\alpha p} x^{\beta p} e^{-p\lambda \frac{x^2}{4\lambda^2}} \frac{x^2}{4\lambda^2} dx d\lambda \\ &= 4^{-\alpha p - 1} \int_0^\infty \lambda^{-\alpha p - 2} e^{-p\lambda} \int_0^{\min(1, 2\sqrt{\lambda T})} x^{2\alpha p + \beta p + 2} dx d\lambda. \end{aligned}$$

The integral over  $x$  is finite if and only if  $p(2\alpha + \beta) > -3$ , and if so, the expression is equal to

$$\frac{4^{-\alpha p - 1}}{2\alpha p + \beta p + 3} \left[ (2\sqrt{T})^{2\alpha p + \beta p + 3} \int_0^{\frac{1}{4T}} \lambda^{(\beta p - 1)/2} e^{-p\lambda} d\lambda + \int_{\frac{1}{4T}}^\infty \lambda^{-\alpha p - 2} e^{-p\lambda} d\lambda \right]$$

with the first integral being finite if and only if  $p\beta > -1$ .  $\square$

Following [Wlo87], we analyze the regularity of (2.3) and (3.1) in (intersections of) Bochner spaces. In particular, the space  $L_2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega))$  plays an important role in this and following sections. For the precise definition of this space and some properties we refer to [Wlo87, Chapter 25]. With  $H^1_{0,\{0\}}(I)$  denoting the closure in  $H^1(I)$  of the functions in  $C^\infty(I) \cap H^1(I)$  that vanish at 0, we have the following result concerning the smoothness of  $u_H$  restricted to  $I \times \Omega$ .

**Corollary 3.3.**  $u_H \in L_2(I; H^1(\Omega)) \cap H^1_{0,\{0\}}(I; H^{-1}(\Omega))$ , but  $u_H \notin H^1_{0,\{0\}}(I; L_2(\Omega))$  and  $u_H \notin L_2(I; H^2(\Omega))$ . Furthermore,  $t\partial_t\partial_x u_H, x\partial_x^2 u_H, t\partial_x^2 u_H \in L_2(I \times \Omega)$ , and  $x\partial_t\partial_x u_H \in L_2(I; H^{-1}(\Omega))$ .

*Proof.* By applications of Lemma 3.1, we infer that  $\partial_x u_H = -2H \in L_2(I \times \Omega)$ , and that  $\partial_t u_H(t, x) = \frac{1}{2\sqrt{\pi}} x t^{-\frac{3}{2}} e^{-\frac{x^2}{4t}} \notin L_2(I \times \Omega)$ . This yields  $u_H \in L_2(I; H^1(\Omega))$  and  $u_H \notin H^1_{0,\{0\}}(I; L_2(\Omega))$ .

If  $\partial_x F = f$ , then  $f \in L_2(I; H^{-1}(\Omega))$  if and only if  $F \in L_2(I \times \Omega)$ . We have  $\int_{-\infty}^x \partial_t u_H(t, y) dy = -\frac{t^{-\frac{1}{2}}}{\sqrt{\pi}} e^{-\frac{x^2}{4t}} \in L_2(I \times \Omega)$ , so indeed  $u_H \in H^1_{0,\{0\}}(I; H^{-1}(\Omega))$ .

It holds

$$\partial_x^2 u_H(t, x) = -2\partial_x H(t, x) = \frac{1}{2\sqrt{\pi}} x t^{-\frac{3}{2}} e^{-\frac{x^2}{4t}} \notin L_2(I \times \Omega),$$

or  $u_H \notin L_2(I; H^2(\Omega))$ , but  $x\partial_x^2 u_H, t\partial_x^2 u_H \in L_2(I \times \Omega)$ .

We have  $\partial_x \partial_t u_H = (t^{-\frac{3}{2}} - \frac{1}{2} x^2 t^{-\frac{5}{2}}) \frac{e^{-\frac{x^2}{4t}}}{2\sqrt{\pi}}$ , so  $t\partial_x \partial_t u_H \in L_2(I \times \Omega)$ . Proving that  $x\partial_x \partial_t u_H \in L_2(I; H^{-1}(\Omega))$  amounts to proving  $x t^{-\frac{3}{2}} e^{-\frac{x^2}{4t}}, t^{-\frac{5}{2}} x^3 e^{-\frac{x^2}{4t}} \in L_2(I; H^{-1}(\Omega))$ , i.e., proving that  $t^{-\frac{3}{2}} \int_{-\infty}^x y e^{-\frac{y^2}{4t}} dy, t^{-\frac{5}{2}} \int_{-\infty}^x y^3 e^{-\frac{y^2}{4t}} dy \in L_2(I \times \Omega)$ . The first function equals  $-2t^{-\frac{1}{2}} e^{-\frac{x^2}{4t}}$ , which is in  $L_2(I \times \Omega)$ , and the second function equals  $-8t^{-\frac{1}{2}} e^{-\frac{x^2}{4t}} - 2t^{-\frac{3}{2}} x^2 e^{-\frac{x^2}{4t}}$ , which is also in  $L_2(I \times \Omega)$ .  $\square$

Finally in this subsection, notice that from  $\frac{\partial}{\partial t} u_H(t, x) = \frac{1}{2\sqrt{\pi}} x t^{-\frac{3}{2}} e^{-\frac{x^2}{4t}}$ , it follows that for any  $x > 0$  and  $k \in \mathbb{N}_0$ ,

$$(3.2) \quad \lim_{t \downarrow 0} \frac{\partial^k}{\partial t^k} u_H(t, x) = 0.$$

**3.2. Regularity of the parabolic problem with homogeneous initial and boundary conditions.** Knowing that  $e(v)$  is the solution of the parabolic problem (3.1) that has homogeneous initial and boundary conditions, we study the regularity of such a problem.

Given functions  $v \in L_\infty(I \times \Omega)$  and  $f \in L_2(I; H^{-1}(\Omega))$ , let  $w$  solve

$$(3.3) \quad \begin{cases} \partial_t w(t, x) = \partial_x^2 w(t, x) + v(t, x) \partial_x w(t, x) + f(t, x) & (t, x) \in I \times \Omega, \\ w(t, 0) = 0, \quad w(t, 1) = 0 & t \in I, \\ w(0, x) = 0 & x \in \Omega, \end{cases}$$

where the spatial differential operators at the right-hand side should be interpreted in a weak sense, i.e.,  $((\partial_x^2 + v\partial_x)\eta)(\zeta) := \int_\Omega -\partial_x \eta \partial_x \zeta + v \partial_x \eta \zeta dx$ . It is

well-known that

$$(3.4) \quad L(v) := w \mapsto f \in \mathcal{L}_{\text{iso}}(L_2(I; H_0^1(\Omega)) \cap H_{0,\{0\}}^1(I; H^{-1}(\Omega)), L_2(I; H^{-1}(\Omega)))$$

(see, e.g., [Wlo87, Thm. 26.1]). Under additional smoothness conditions on the right-hand side  $f$  beyond being in  $L_2(I; H^{-1}(\Omega))$ , additional smoothness of the solution  $w$  can be demonstrated:

**Proposition 3.4.** *a) If  $v \in W_\infty^1(I \times \Omega)$ , then*

$$L(v)^{-1} \in \mathcal{L}\left(L_2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega)), H_{0,\{0\}}^1(I; H_0^1(\Omega)) \cap H^2(I; H^{-1}(\Omega)) \cap L_2(I; H^3(\Omega))\right).$$

*b) If  $v \in L_\infty(I \times \Omega)$ , then*

$$L(v)^{-1} \in \mathcal{L}\left(L_2(I \times \Omega), L_2(I; H^2(\Omega)) \cap H_{0,\{0\}}^1(I; L_2(\Omega))\right),$$

*Proof.* a) If  $f \in L_2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega))$ , then in particular  $f \in H^1(I; H^{-1}(\Omega))$ , and  $f(0, \cdot) \in L_2(\Omega)$  with its norm being bounded by some multiple of  $\|f\|_{L_2(I; H^1(\Omega))} + \|f\|_{H^1(I; H^{-1}(\Omega))}$  (see, e.g., [Wlo87, Thm. 25.5]). As shown in [Wlo87, Thm. 27.2 and its proof], from the last two properties of  $f$ , and  $v \in W_\infty^1(I; L_\infty(\Omega))$ , one has  $w = L(v)^{-1}f \in H_{0,\{0\}}^1(I; H_0^1(\Omega)) \cap H^2(I; H^{-1}(\Omega))$  with its norm being bounded by some multiple of  $\|f\|_{H^1(I; H^{-1}(\Omega))} + \|f(0, \cdot)\|_{L_2(\Omega)}$ .

To show the spatial regularity, i.e.,  $w \in L_2(I; H^3(\Omega))$ , given a constant  $\lambda$ , we define  $w_\lambda(t, \cdot) = w(t, \cdot)e^{-\lambda t}$ ,  $f_\lambda(t, \cdot) = f(t, \cdot)e^{-\lambda t}$ . One infers that

$$(3.5) \quad (-\partial_x^2 - v\partial_x + \lambda)w_\lambda = f_\lambda - \partial_t w_\lambda \quad \text{on } I \times \Omega, \quad w_\lambda(\cdot, 0) = 0 = w_\lambda(\cdot, 1) \quad \text{on } I,$$

where, as before, the spatial differential operators should be interpreted in a weak sense. Using that  $|\int_I \int_\Omega v(\partial_x w_\lambda) w_\lambda dx dt| \leq \|v\|_{L_\infty(I \times \Omega)} \|\partial_x w_\lambda\|_{L_2(I \times \Omega)} \|w_\lambda\|_{L_2(I \times \Omega)}$  and Young's inequality, one infers that for  $\lambda > \frac{1}{4} \|v\|_{L_\infty(I \times \Omega)}^2$  the bilinear form defined by the left-hand side of (3.5) is bounded and *coercive* on  $L_2(I; H_0^1(\Omega)) \times L_2(I; H_0^1(\Omega))$ . Thus for  $\lambda > \frac{1}{4} \|v\|_{L_\infty(I \times \Omega)}^2$  we have

$$A(v, \lambda) \in \mathcal{L}_{\text{iso}}(L_2(I; H_0^1(\Omega)), L_2(I; H^{-1}(\Omega))).$$

Realizing that  $\|\cdot\|_{H^{k+2}(\Omega)}^2 = |\Delta(\cdot)|_{H^k(\Omega)}^2 + \|\cdot\|_{H^{k+1}(\Omega)}^2$ , an induction and tensor product argument shows  $A(0, 0)^{-1} \in \mathcal{L}(L_2(I; H^k(\Omega)), L_2(I; H^{k+2}(\Omega)))$  for any  $k \in \mathbb{N}_0$ . Writing

$$A(v, \lambda)^{-1} - A(0, 0)^{-1} = A(0, 0)^{-1}(v\partial_x - \lambda\text{Id})A(v, \lambda)^{-1},$$

and using that  $v\partial_x \in \mathcal{L}(L_2(I; H^1(\Omega)), L_2(I; L_2(\Omega)))$  by  $v \in L_\infty(I \times \Omega)$ , one verifies that  $A(v, \lambda)^{-1} \in \mathcal{L}(L_2(I \times \Omega), L_2(I; H^2(\Omega)))$ . Repeating the argument, now using that  $v\partial_x \in \mathcal{L}(L_2(I; H^2(\Omega)), L_2(I; H^1(\Omega)))$  by  $v \in L_\infty(I; W_\infty^1(\Omega))$ , one has  $A(v, \lambda)^{-1} \in \mathcal{L}(L_2(I; H^1(\Omega)), L_2(I; H^3(\Omega)))$ . Knowing that  $f_\lambda - \partial_t w_\lambda \in L_2(I; H^1(\Omega))$ , with its norm bounded by a multiple of  $\|f\|_{L_2(I; H^1(\Omega))} + \|f\|_{H^1(I; H^{-1}(\Omega))}$ , one infers that  $w_\lambda$  and thus  $w \in L_2(I; H^3(\Omega))$ , with its norm bounded by a multiple of the same quantity.

b) Similar to Part a), it suffices to show that

$$L(v, \lambda)^{-1} := f_\lambda \mapsto w_\lambda \in \mathcal{L}\left(L_2(I \times \Omega), L_2(I; H^2(\Omega)) \cap H_{0, \{0\}}^1(I; L_2(\Omega))\right).$$

Knowing that  $L(v, \lambda)^{-1} \in \mathcal{L}(L_2(I; H^{-1}(\Omega)), L_2(I; H_0^1(\Omega)) \cap H_{0, \{0\}}^1(I; H^{-1}(\Omega)))$ , and  $L(v, \lambda) - L(0, 0) = -v\partial_x + \lambda\text{Id} \in \mathcal{L}(L_2(I; H_0^1(\Omega)), L_2(I \times \Omega))$ , the proof is completed by  $L(v, \lambda)^{-1} - L(0, 0)^{-1} = L(0, 0)^{-1}(L(0, 0) - L(v, \lambda))L(v, \lambda)^{-1}$  and the *maximal regularity* result

$$L(0, 0)^{-1} \in \mathcal{L}\left(L_2(I \times \Omega), L_2(I; H^2(\Omega)) \cap H_{0, \{0\}}^1(I; L_2(\Omega))\right)$$

from, e.g., [dS64, DHP03].  $\square$

**3.3. The regularity of  $e(v) = u(v) - u(v_0)$ .** Recall that  $u_H$  denotes the solution of the heat equation studied in Section 3.1, that  $u(v)$  denotes the solution to (3.1) for given  $v \in C(\overline{I \times \Omega})$ , and  $v_0 := v(0, 0)$ . Since  $e(v)$  solves (3.1), i.e.,  $e(v)$  is the solution  $w$  of (3.3) for forcing function  $f$  given by

$$(3.6) \quad \begin{aligned} & (v - v_0)\partial_x u(v_0) \\ &= (v - v_0)\partial_x(u(v_0) - u(0)) + (v - v_0)\partial_x(u(0) - u_H) + (v - v_0)\partial_x u_H, \end{aligned}$$

in view of the regularity results proven in Proposition 3.4, we establish smoothness of  $e(v)$  by demonstrating smoothness of each of the three terms at the right-hand side of (3.6).

**Lemma 3.5.** *It holds that*

$$u(0) - u_H \in H_{0, \{0\}}^1(I; H_0^1(\Omega)) \cap H^2(I; H^{-1}(\Omega)) \cap L_2(I; H^3(\Omega)).$$

*Proof.* The function  $w(t, x) := u(0)(t, x) - (u_H(t, x) - xu_H(t, 1))$  satisfies the homogeneous initial and boundary conditions from (3.3), and  $\partial_t w(t, x) = \partial_x^2 w(t, x) + x\partial_t u_H(t, 1)$ . By (3.2) we have  $(t, x) \mapsto x\partial_t u_H(t, 1) \in L_2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega))$ , so that Proposition 3.4a) for  $v = 0$  and  $f(t, x) = x\partial_t u_H(t, 1)$  shows that

$$w \in H_{0, \{0\}}^1(I; H_0^1(\Omega)) \cap H^2(I; H^{-1}(\Omega)) \cap L_2(I; H^3(\Omega)).$$

Because, again by (3.2),  $(t, x) \mapsto xu_H(t, 1)$  is in the same space, the proof is completed.  $\square$

**Lemma 3.6.** *For any  $v_0 \in \mathbb{R}$ ,  $u(v_0) - u(0) \in L_2(I; H^2(\Omega)) \cap H_{0, \{0\}}^1(I; L_2(\Omega))$ .*

*Proof.* The function  $w := u(v_0) - u(0)$  satisfies the homogeneous initial- and boundary conditions from (3.3), and  $\partial_t w(t, x) = \partial_x^2 w(t, x) + v_0\partial_x w - v_0\partial_x u(0)$ . From  $\partial_x u(0) \in L_2(I \times \Omega)$  by Corollary 3.3 and Lemma 3.5, an application of Proposition 3.4b) for  $v = v_0$  and  $f = -v_0\partial_x u(0)$  completes the proof.  $\square$

**Lemma 3.7.** *If  $v \in W_\infty^1(I \times \Omega) \cap L_\infty(I; W_\infty^2(\Omega))$ , then*

$$(v - v_0)\partial_x u_H \in L_2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega)).$$

*Proof.* Abbreviate  $g := (v - v_0)\partial_x u_H$ . Throughout the proof, we use the estimates for  $u_H$  proven in Corollary 3.3.

We start with proving  $\partial_t g = (\partial_t v)\partial_x u_H + (v - v_0)\partial_t \partial_x u_H \in L_2(I; H^{-1}(\Omega))$ . Using  $v \in W_\infty^1(I; L_\infty(\Omega))$  and  $\partial_x u_H \in L_2(I \times \Omega)$ , the first term is even in  $L_2(I \times \Omega)$ . Writing the second term as

$$(v(t, x) - v_0)\partial_t \partial_x u_H(t, x) = \frac{v(t, x) - v(0, x)}{t} t \partial_t \partial_x u_H(t, x) + \frac{v(0, x) - v_0}{x} x \partial_t \partial_x u_H(t, x),$$

from  $t \partial_t \partial_x u_H \in L_2(I \times \Omega)$  and  $\frac{v(t, x) - v(0, x)}{t} \in L_\infty(I \times \Omega)$  by  $v \in W_\infty^1(I; L_\infty(\Omega))$ , we have  $\frac{v(t, x) - v(0, x)}{t} t \partial_t \partial_x u_H(t, x) \in L_2(I \times \Omega)$ . From  $x \partial_t \partial_x u_H(t, x) \in L_2(I; H^{-1}(\Omega))$  and  $\frac{v(0, x) - v_0}{x} \in L_\infty(I; W_\infty^1(\Omega))$  by  $v \in L_\infty(I; W_\infty^2(\Omega))$ , we have  $\frac{v(0, x) - v_0}{x} x \partial_t \partial_x u_H(t, x) \in L_2(I; H^{-1}(\Omega))$ , so that  $\partial_t g \in L_2(I; H^{-1}(\Omega))$ .

It remains to show that  $g \in L_2(I; H^1(\Omega))$ . It is clear that  $(v - v_0)\partial_x u_H \in L_2(I \times \Omega)$  and  $(\partial_x v)\partial_x u_H \in L_2(I \times \Omega)$  by  $v \in L_\infty(I; W_\infty^1(\Omega))$ . Writing

$$(v(t, x) - v_0)\partial_x^2 u_H(t, x) = \frac{v(t, x) - v(0, x)}{t} t \partial_x^2 u_H(t, x) + \frac{v(0, x) - v_0}{x} x \partial_x^2 u_H(t, x),$$

from  $\frac{v(t, x) - v(0, x)}{t}$ ,  $\frac{v(0, x) - v_0}{x} \in L_\infty(I \times \Omega)$  by  $v \in W_\infty^1(I \times \Omega)$ , and both  $t \partial_x^2 u_H(t, x)$  and  $x \partial_x^2 u_H(t, x) \in L_2(I \times \Omega)$ , we obtain  $g \in L_2(I; H^1(\Omega))$ , and the proof is completed.  $\square$

By combining the results of the preceding three propositions with the regularity result proven in Proposition 3.4 we obtain the following.

**Theorem 3.8.** *If  $v \in W_\infty^1(I \times \Omega) \cap L_\infty(I; W_\infty^2(\Omega))$ , then*

$$e(v) \in H_{0, \{0\}}^1(I; H_0^1(\Omega)) \cap H^2(I; H^{-1}(\Omega)) \cap L_2(I; H^3(\Omega)).$$

*Proof.* With Lemma 3.5 and 3.6, we see that  $(v - v_0)\partial_x(u(v_0) - u_H) \in L_2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega))$ , whilst Lemma 3.7 shows that  $(v - v_0)\partial_x u_H \in L_2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega))$ . We conclude that

$$(v - v_0)\partial_x u(v_0) \in L_2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega)),$$

so that an application of Proposition 3.4a) completes the proof.  $\square$

Notice that as a consequence of Corollary 3.3, Lemma 3.5 and 3.6,  $u(v_0) \notin H_{0, \{0\}}^1(I; L_2(\Omega)) \cup L_2(I, H^2(\Omega))$ . Comparing Corollary 3.3 with Theorem 3.8, we conclude that  $e = u(v) - u(v_0)$  is indeed more smooth than  $u(v_0)$ , and thus than  $u(v)$ , confirming the claim we made at the beginning of Section 3.

#### 4. MINIMAL RESIDUAL METHOD

For solving (3.3) (specifically for the forcing function  $f$  as in (3.6)), we write it in variational form, i.e., we multiply it by test functions  $z: I \times \Omega \rightarrow \mathbb{R}$  from a suitable collection, integrate it over  $I \times \Omega$ , and apply integration by parts with respect to  $x$ . We thus arrive at

$$\begin{aligned} (Bw)(z) &:= \int_I \int_\Omega \partial_t w(t, x) z(t, x) + \partial_x w(t, x) \partial_x z(t, x) - v(t, x) \partial_x w(t, x) z(t, x) \, dx \, dt \\ &= \int_I \int_\Omega f(t, x) z(t, x) \, dx \, dt =: f(z) \end{aligned}$$

for all those test functions. With

$$X := L_2(I; H_0^1(\Omega)) \cap H^1(I; H^{-1}(\Omega)), \quad Y := L_2(I; H_0^1(\Omega)),$$

it is known that  $(B, \gamma_0) \in \mathcal{L}_{\text{iso}}(X, Y' \times L_2(\Omega))$ , where  $\gamma_0 := w \mapsto w(0, \cdot)$  denotes the initial trace operator, see, e.g., [Wlo87, Chapter IV] or [SS09].

Already because  $X \neq Y \times L_2(\Omega)$ , the well-posed system  $Bw = f$  and  $\gamma_0 w = 0$  cannot be discretized by simple Galerkin discretizations. Given a family  $(X_h)_{h \in \Delta}$  of finite dimensional subspaces of  $X$ , as discrete approximations to  $w$  one may consider the minimizers  $\operatorname{argmin}_{\bar{w} \in X_h} \|B\bar{w} - f\|_{Y'}^2 + \|\gamma_0 \bar{w}\|_{L_2(\Omega)}^2$ . Since the dual norm  $\|\cdot\|_{Y'}$  cannot be evaluated, this approach is not immediately feasible either. Therefore, for  $(Y_h)_{h \in \Delta}$  being a second family of finite dimensional subspaces, now of  $Y$ , for  $h \in \Delta$  as a discrete approximation from  $X_h$  we consider

$$(4.1) \quad w_h := \operatorname{argmin}_{\bar{w} \in X_h} \|B\bar{w} - f\|_{Y'_h}^2 + \|\gamma_0 \bar{w}\|_{L_2(\Omega)}^2.$$

This minimal residual approach has been studied for general parabolic PDEs in, e.g., [And13, SW20b, SW20a], where  $\Omega$  can be a  $d$ -dimensional spatial domain for arbitrary  $d \geq 1$ .

For parabolic differential operators with a possibly asymmetric spatial part, in our setting caused by a non-zero drift function  $v$ , in [SW20a, Thm. 3.1] it has been shown that if  $X_h \subset Y_h$  and

$$(4.2) \quad \varrho := \inf_{h \in \Delta} \inf_{0 \neq \bar{w} \in X_h} \frac{\|\partial_t \bar{w}\|_{Y'_h}}{\|\partial_t \bar{w}\|_{Y'}} > 0,$$

then

$$(4.3) \quad \|w - w_h\|_X \lesssim \min_{\bar{w} \in X_h} \|w - \bar{w}\|_X,$$

where the constant in (4.3) depends only on  $\varrho$  and an upper bound for  $\|v\|_{L_\infty(I \times \Omega)}$ , i.e.,  $w_h$  is a *quasi-best approximation* from  $X_h$  with respect to the norm on  $X$ .

*Remark 4.1.* This quasi-optimality result has been demonstrated under the condition that the spatial part of the parabolic differential operator is *coercive* on  $H_0^1(\Omega) \times H_0^1(\Omega)$  for a.e.  $t \in I$ , i.e.,

$$\int_{\Omega} \eta' \eta' - v(t, \cdot) \eta' \eta \, dx \gtrsim \|\eta\|_{H_0^1(\Omega)}^2 \quad (\eta \in H_0^1(\Omega)),$$

which holds true when  $\partial_x v \leq 0$  or  $\|v\|_{L_\infty(I \times \Omega)} \sup_{0 \neq \eta \in H_0^1(\Omega)} \frac{\|\eta\|_{L_2(\Omega)}}{\|\eta'\|_{L_2(\Omega)}} < 1$ , but which might be violated otherwise.

Although this coercivity condition might not be necessary, it can always be enforced by considering  $w_\lambda(t, \cdot) := w(t, \cdot)e^{-\lambda t}$ ,  $f_\lambda(t, \cdot) := f(t, \cdot)e^{-\lambda t}$  instead of  $w$  and  $f$  with  $\lambda$  sufficiently large, see also the proof of Proposition 3.4. By approximating  $w_\lambda$  by the minimal residual method, and by multiplying the obtained approximation by  $e^{\lambda t}$ , an approximation for  $w$  is obtained. Since qualitatively the transformations with  $e^{\pm \lambda t}$  do not affect the smoothness of solution or right-hand side, for convenience in the following we pretend that coercivity holds true for (3.3).

As in [SW20a, SW20b], we equip  $Y_h$  in (4.1) with the energy norm

$$\|z\|_Y^2 := (A_s z)(z) \quad (z \in Y_h),$$

where

$$(A_s z)(\bar{z}) := \int_I \int_{\Omega} \partial_x z(t, x) \partial_x \bar{z}(t, x) - \frac{v(t, x)}{2} (\partial_x z(t, x) \bar{z}(t, x) + z(t, x) \partial_x \bar{z}(t, x)) \, dx \, dt$$

denotes the symmetric part of the spatial differential operator. Equipping  $Y_h$  and  $X_h$  with bases  $\Phi^h = \{\phi_i^h\}$  and  $\Psi^h = \{\psi_j^h\}$ , and denoting by  $w^h$  the representation of the minimizer  $w_h$  with respect to  $\Psi_h$ ,  $w^h$  is found as the second component of the solution of

$$(4.4) \quad \begin{bmatrix} A_s^h & B^h \\ B^{h\top} & C^h \end{bmatrix} \begin{bmatrix} \mu^h \\ w^h \end{bmatrix} = \begin{bmatrix} f^h \\ 0 \end{bmatrix},$$

where  $(A_s^h)_{ij} := (A_s \phi_j^h)(\phi_i^h)$ ,  $B_{ij}^h := (B \psi_j^h)(\phi_i^h)$ ,  $C_{ij}^h := \int_{\Omega} \psi_j^h(0, x) \psi_i^h(0, x) dx$ , and  $f_i^h := f(\phi_i)$ . The operator  $A_s$  can be replaced by any other spectrally equivalent operator on  $Y_h$  without compromising the quasi-optimality result (4.3). We refer to [SW20a, SW20b] for details.

Let  $P_1$  be the set of polynomials of degree one. Taking for  $n := 1/h \in \mathbb{N}$ ,

$$(4.5) \quad \begin{aligned} V_{x,h} &:= \{\eta \in H_0^1(\Omega) : \eta|_{((i-1)h, ih)} \in P_1 \text{ for } i = 1, \dots, n\}, \\ V_{t,h} &:= \{\zeta \in H^1(I) : \zeta|_{((i-1)hT, ihT)} \in P_1 \text{ for } i = 1, \dots, n\}, \\ X_h &:= V_{t,h} \otimes V_{x,h}, \end{aligned}$$

it is known, cf. [SW20b, Sect. 4], that condition (4.2) is satisfied for

$$(4.6) \quad Y_h := \{\zeta \in L_2(I) : \zeta|_{((i-1)hT, ihT)} \in P_1 \text{ for } i = 1, \dots, n\} \otimes V_{x,h},$$

where obviously also  $X_h \subset Y_h$ .

Applying this approach for  $f = (v - v_0) \partial_x u(v_0)$ , in view of (4.3) the error of the obtained approximation for  $e(v)$  with respect to the  $X$ -norm can be bounded by the error of the best approximation from  $X_h$ . To bound the latter error we recall from Theorem 3.8 that for  $v \in W_{\infty}^1(I \times \Omega) \cap L_{\infty}(I; W_{\infty}^2(\Omega))$ , it holds that

$$e(v) \in (H_{0,\{0\}}^1(I) \otimes H_0^1(\Omega)) \cap (H^2(I) \otimes H^{-1}(\Omega)) \cap (L_2(I) \otimes H^3(\Omega)).$$

With  $Q_{x,h}$ ,  $Q_{t,h}$  denoting the  $L_2(\Omega)$ - or  $L_2(I)$ -orthogonal projectors onto  $V_{x,h}$  or  $V_{t,h}$ , respectively,  $Q_{t,h} \otimes Q_{x,h}$  is a projector onto  $X_h$ . Writing

$$\text{Id} - Q_{t,h} \otimes Q_{x,h} = (\text{Id} - Q_{t,h}) \otimes Q_{x,h} + \text{Id} \otimes (\text{Id} - Q_{x,h}),$$

and using that

$$\begin{aligned} \|\text{Id} - Q_{x,h}\|_{\mathcal{L}(H_0^1(\Omega) \cap H^2(\Omega), H_0^1(\Omega))} &\lesssim h, & \|Q_{x,h}\|_{\mathcal{L}(H_0^1(\Omega), H_0^1(\Omega))} &\lesssim 1, \\ \|\text{Id} - Q_{t,h}\|_{\mathcal{L}(H^1(I), L_2(I))} &\lesssim h, & \|\text{Id}\|_{\mathcal{L}(L_2(I), L_2(I))} &= 1 \end{aligned}$$

by standard interpolation estimates and uniform  $H^1$ -boundedness of these  $L_2$ -orthogonal projectors, see e.g. [BX91, §3], one infers that

$$\|\text{Id} - Q_{t,h} \otimes Q_{x,h}\|_{\mathcal{L}((L_2(I) \otimes (H_0^1(\Omega) \cap H^2(\Omega))) \cap (H^1(I) \otimes H_0^1(\Omega)), L_2(I) \otimes H_0^1(\Omega))} \lesssim h.$$

Similarly using that

$$\begin{aligned} \|\text{Id} - Q_{x,h}\|_{\mathcal{L}(L_2(\Omega), H^{-1}(\Omega))} &= \|\text{Id} - Q_{x,h}\|_{\mathcal{L}(H_0^1(\Omega), L_2(\Omega))} \lesssim h, \\ \|Q_{x,h}\|_{\mathcal{L}(H^{-1}(\Omega), H^{-1}(\Omega))} &= \|Q_{x,h}\|_{\mathcal{L}(H_0^1(\Omega), H_0^1(\Omega))} \lesssim 1, \\ \|\text{Id} - Q_{t,h}\|_{\mathcal{L}(H^2(I), H^1(I))} &\lesssim h, & \|\text{Id}\|_{\mathcal{L}(H^1(I), H^1(I))} &= 1, \end{aligned}$$

one infers that

$$\|\text{Id} - Q_{t,h} \otimes Q_{x,h}\|_{\mathcal{L}((H^1(I) \otimes L_2(\Omega)) \cap (H^2(I) \otimes H^{-1}(\Omega)), H^1(I) \otimes H^{-1}(\Omega))} \lesssim h.$$

Our findings are summarized in the following theorem.

**Theorem 4.2.** *For  $v \in W_\infty^1(I \times \Omega) \cap L_\infty(I; W_\infty^2(\Omega))$  and  $X_h, Y_h$  as defined in (4.5) and (4.6), the numerical approximation  $e_h = e_h(v) \in X_h$  to  $e = e(v)$  obtained by the application of the minimal residual method to (3.1)<sup>1</sup> satisfies*

$$\|e - e_h\|_X \lesssim h.$$

Notice that for this space  $X_h$  of continuous piecewise bilinears, this linear decay of the error  $\|e - e_h\|_X$  as function of  $h$  is generally the best that can be expected. In view of the order of the space  $X_h$ , one may hope that  $\|e - e_h\|_{L_2(I \times \Omega)}$  is  $\mathcal{O}(h^2)$ , but on the basis of the smoothness demonstrated for  $e$ , even for  $\inf_{\bar{e} \in X_h} \|e - \bar{e}\|_{L_2(I \times \Omega)}$  this cannot be shown.

## 5. INTERPOLATION FOR PARAMETRIZED DRIFT, BOUNDARIES, AND FINAL TIME

In this section we consider the case that  $v$  and  $T$  in (2.3) depend on a number of parameters  $(\rho_1, \dots, \rho_N) \in [-1, 1]^N$ , and that one is interested in the solution  $u(v)$  to (2.3) for multiple values of these parameters. As explained in Section 3, in order to find  $u(v)$  it suffices to obtain the solution  $e(v)$  to (3.1). Instead of simply solving  $e(v)$  for each of the desired parameter values, under the provision that  $v$  and  $T$  depend smoothly on the parameters, one may attempt to *interpolate*  $e(v)$  from its a priori computed approximations for a carefully selected set of parameters in  $[-1, 1]^N$ .

In order to be able to do so, first of all we need to get rid of the parameter dependence of the domain  $I \times \Omega = (0, T) \times (0, 1)$ . With  $\hat{I} := (0, 1)$ , the function  $\hat{u}$  on  $\hat{I} \times \Omega$  defined by  $\hat{u}(t, x) := u(tT, x)$  solves

$$(5.1) \quad \begin{cases} \partial_t \hat{u}(t, x) = T[\partial_x^2 \hat{u}(t, x) + \hat{v}(t, x) \partial_x \hat{u}(t, x)] & (t, x) \in \hat{I} \times \Omega, \\ \hat{u}(t, 0) = 1, \quad \hat{u}(t, 1) = 0 & t \in \hat{I}, \\ \hat{u}(0, x) = 0 & x \in \Omega, \end{cases}$$

where analogously  $\hat{v}(t, x) := v(tT, x)$ . Denoting this  $\hat{u}$  as  $\hat{u}(\hat{v}, T)$ , the difference

$$\hat{e} = \hat{e}(\hat{v}, T) := \hat{u}(\hat{v}, T) - \hat{u}(v_0, T) : (t, x) \mapsto e(tT, x)$$

solves

$$(5.2) \quad \begin{cases} \partial_t \hat{e}(t, x) = T[\partial_x^2 \hat{e}(t, x) + \hat{v}(t, x) \partial_x \hat{e}(t, x) + (\hat{v}(t, x) - v_0) \partial_x \hat{u}(v_0, T)] & (t, x) \in \hat{I} \times \Omega, \\ \hat{e}(t, 0) = 0, \quad \hat{e}(t, 1) = 0 & t \in \hat{I}, \\ \hat{e}(0, x) = 0 & x \in \Omega. \end{cases}$$

By simply replacing  $I = (0, T)$  by  $\hat{I} = (0, 1)$  and in particular  $X$  as well as  $Y$  by

$$\hat{X} := L_2(\hat{I}; H_0^1(\Omega)) \cap H^1(\hat{I}; H^{-1}(\Omega)), \quad \hat{Y} := L_2(\hat{I}; H_0^1(\Omega)),$$

in a number of places, it is clear that the results that we obtained about the smoothness of  $e$  and its numerical approximation  $e_h$  by the minimal residual method apply equally well to  $\hat{e}$  and its minimal residual approximation that we denote as  $\hat{e}_h$ .

<sup>1</sup>If necessary taking into account the transformations discussed in Remark 4.1.

Since the domain of  $\hat{e}$  is independent of parameters, we can apply the idea of interpolation. One option is to perform a ‘full’ tensor product interpolation. In this case, the number of interpolation points required for a fixed polynomial degree, i.e., the number of values of the parameters for which a numerical approximation for  $\hat{e} \in \hat{X}$  has to be computed, grows exponentially with the number  $N$  of parameters. As this is undesirable, we instead apply a sparse tensor product interpolation. More specifically, we choose the Smolyak construction, based on Clenshaw–Curtis abscissae in each parameter direction, see [NTW08]: For  $i \in \mathbb{N}$  let  $I_{i+1}$  denote the univariate interpolation operator with abscissae  $\cos j2^{-i}\pi$ ,  $j = 0, \dots, 2^i$ , onto the space of polynomials of degree  $2^i$ , let  $\hat{I}$  be the interpolation operator with abscissa 0 and let  $I_0 := 0$ . Then, for an integer  $q \geq N$ , we apply the sparse interpolator

$$\mathcal{I}_q := \sum_{\{\mathbf{i} \in \mathbb{N}_0^N : \sum_{n=1}^N i_n \leq q\}} \bigotimes_{n=1}^N (I_{i_n} - I_{i_n-1}).$$

It is known that the resulting interpolation error in  $C([-1, 1]^N; \hat{X})$  (for arbitrary Banach space  $\hat{X}$ ), equipped with  $\|\cdot\|_{L_\infty([-1, 1]^N; \hat{X})}$ , decays subexponentially in the number of interpolation points when  $\hat{e}$  as function of each of the parameters  $\rho_n$  has an extension to a differentiable mapping on a neighbourhood  $\Sigma$  of  $[-1, 1]$  in  $\mathbb{C}$ . For details about this statement we refer to [NTW08, Thm. 3.11]. [NTW08] also mentions that the result requires relatively large values of  $q$ . Thus, the authors additionally prove algebraic convergence [NTW08, Thm. 3.10] under the same assumptions but for arbitrary  $q$ .

Instead of  $\hat{e}$ , we interpolate a numerical approximation  $\hat{e}_h$ , specifically the one obtained by the minimal residual method described in Section 4. For the additional error we have

$$\|\mathcal{I}_q(\hat{e} - \hat{e}_h)\|_{L_\infty([-1, 1]^N; \hat{X})} \leq \|\mathcal{I}_q\|_{\mathcal{L}(C([-1, 1]^N), C([-1, 1]^N))} \|\hat{e} - \hat{e}_h\|_{L_\infty([-1, 1]^N; \hat{X})}.$$

In [Chk14, Sect. 5.3] it has been shown that the factor  $\|\mathcal{I}_q\|_{\mathcal{L}(C([-1, 1]^N), C([-1, 1]^N))}$ , known as the Lebesgue constant, is bounded by  $(\#\{\mathbf{i} \in \mathbb{N}_0^N : \sum_{n=1}^N i_n \leq q\})^2$ , which is only of polylogarithmic order as function of the number of interpolation points.

Concerning the factor  $\|\hat{e} - \hat{e}_h\|_{L_\infty([-1, 1]^N; \hat{X})}$ , in our derivation of Theorem 4.2 we have seen that for each parameter value  $(\rho_1, \dots, \rho_N) \in [-1, 1]^N$  the expression  $h^{-1}\|\hat{e} - \hat{e}_h\|_{\hat{X}}$  can be bounded by a constant multiple, only dependent on an upper bound for  $\|\hat{v}\|_{L_\infty(\hat{I} \times \Omega)}$  and for the norm of  $\hat{e}$  in  $H_{0, \{0\}}^1(\hat{I}; H_0^1(\Omega)) \cap H^2(\hat{I}; H^{-1}(\Omega)) \cap L_2(\hat{I}; H^2(\Omega))$ . For uniformly bounded  $T$  and  $T^{-1}$ , and  $\hat{v}$  that varies over a bounded set in  $W_\infty^1(\hat{I} \times \Omega) \cap L_\infty(\hat{I}; W_\infty^2(\Omega))$ , inspection of the estimates from Sect. 3 reveals that the latter norm of  $\hat{e}$  is uniformly bounded. So assuming that these conditions on  $T$ ,  $T^{-1}$  and  $v$  hold true for  $(\rho_1, \dots, \rho_N) \in [-1, 1]^N$ , we have that  $\|\hat{e} - \hat{e}_h\|_{L_\infty([-1, 1]^N; \hat{X})} \lesssim h$ .

What remains is to establish the differentiability of the solution  $\hat{e}$  as function of each of the parameters which is done in the following theorem.

**Theorem 5.1.** *For an open  $[-1, 1] \subset \Sigma \subset \mathbb{C}$ , let  $(\hat{v}, T) : \Sigma \rightarrow C(\bar{\hat{I}}; W_\infty^1(\Omega)) \times (0, \infty)$  be differentiable. For  $\rho \in \Sigma$  let  $\hat{e}(\hat{v}(\rho), T(\rho)) \in \hat{X}$  be the solution to (5.2). Then  $\rho \mapsto \hat{e} = \hat{e}(\hat{v}(\rho), T(\rho)) : \Sigma \rightarrow \hat{X}$  is differentiable.*

*Proof.* The proof is based on the fact that  $\hat{\ell}$  is the solution of a well-posed PDE with coefficients and a forcing term that are differentiable functions of  $\rho$ .

Analogously to (3.4), denoting by  $L(\hat{v}, T)$  the map  $w \mapsto f$  defined by  $\partial_t w = T(\partial_x^2 + \hat{v}\partial_x)w + f$  on  $\hat{I} \times \Omega$ ,  $w(t, 0) = 0 = w(t, 1)$  ( $t \in \hat{I}$ ), and  $w(0, x) = 0$  ( $x \in \Omega$ ), one has

$$(5.3) \quad \hat{\ell}(\hat{v}(\rho), T(\rho)) = L(\hat{v}(\rho), T(\rho))^{-1}T(\rho)(\hat{v}(\rho) - v_0(\rho))\partial_x \hat{u}(v_0(\rho), T(\rho)),$$

where  $v_0(\rho) := \hat{v}(\rho)(0, 0)$ . Below we demonstrate that

$$(5.4) \quad \rho \mapsto T(\rho)(\hat{v}(\rho) - v_0(\rho)): \Sigma \rightarrow L_\infty(\hat{I}; W_\infty^1(\Omega)) \text{ is differentiable,}$$

$$(5.5) \quad \rho \mapsto \hat{u}(v_0(\rho), T(\rho)): \Sigma \rightarrow L_2(\hat{I} \times \Omega) \text{ is differentiable,}$$

so that, from  $\partial_x \in \mathcal{L}(L_2(\hat{I} \times \Omega), \hat{Y}')$  and  $L_\infty(\hat{I}; W_\infty^1(\Omega))$ -functions being pointwise multipliers in  $\mathcal{L}(\hat{Y}', \hat{Y}')$ ,

$$(5.6) \quad \rho \mapsto T(\rho)(\hat{v}(\rho) - v_0(\rho))\partial_x \hat{u}(v_0(\rho), T(\rho)): \Sigma \rightarrow \hat{Y}' \text{ is differentiable.}$$

We proceed below to show that

$$(5.7) \quad \rho \mapsto L(\hat{v}(\rho), T(\rho))^{-1}: \Sigma \rightarrow \mathcal{L}(\hat{Y}', \hat{X}) \text{ is differentiable.}$$

Together, (5.6) and (5.7) complete the proof.

From  $\rho \mapsto \hat{v}(\rho): \Sigma \rightarrow C(\bar{I}_1; W_\infty^1(\Omega))$  being differentiable, it follows that  $\rho \mapsto v_0(\rho): \Sigma \rightarrow \mathbb{C}$  is differentiable, which together with  $T: \Sigma \rightarrow (0, \infty)$  being differentiable shows (5.4).

To show (5.7), we fix some arbitrary  $\rho_0 \in \Sigma$ , abbreviate  $L := L(\hat{v}(\rho), T(\rho))$  as well as  $L_0 := L(\hat{v}(\rho_0), T(\rho_0))$  and write

$$L^{-1} = L_0^{-1} + L_0^{-1}[L_0 - L]L_0^{-1} + L^{-1}\{[L_0 - L]L_0^{-1}\}^2.$$

This decomposition and the fact that  $L(\hat{v}(\rho), T(\rho))^{-1}$  is bounded in  $\mathcal{L}(\hat{Y}', \hat{X})$  for  $\rho$  in a neighbourhood of  $\rho_0$  ([SS09, Thm. 5.1]) imply that it suffices to show that for some  $K(\rho_0) \in \mathcal{L}(\mathbb{C}, \mathcal{L}(\hat{X}, \hat{Y}'))$ ,

$$(5.8) \quad L(\hat{v}(\rho_0), T(\rho_0)) - L(\hat{v}(\rho), T(\rho)) = K(\rho_0)(\rho - \rho_0) + o(\rho - \rho_0) \text{ in } \mathcal{L}(\hat{X}, \hat{Y}').$$

We have

$$\begin{aligned} & L(\hat{v}(\rho_0), T(\rho_0)) - L(\hat{v}(\rho), T(\rho)) \\ &= [T(\rho) - T(\rho_0)]\partial_x^2 + [(T(\rho) - T(\rho_0))\hat{v}(\rho) + T(\rho_0)(\hat{v}(\rho) - \hat{v}(\rho_0))]\partial_x. \end{aligned}$$

From  $T(\rho) - T(\rho_0) = T'(\rho_0)(\rho - \rho_0) + o(\rho - \rho_0)$ ,  $\hat{v}(\rho) - \hat{v}(\rho_0) = \hat{v}'(\rho_0)(\rho - \rho_0) + o(\rho - \rho_0)$  in  $C(\bar{I}_1, W_\infty^1(\Omega)) \hookrightarrow L_\infty(\hat{I} \times \Omega)$ ,  $\partial_x^2 \in \mathcal{L}(\hat{X}, \hat{Y}')$ ,  $\partial_x \in \mathcal{L}(\hat{X}, L_2(\hat{I} \times \Omega))$ ,  $L_\infty(\hat{I} \times \Omega)$ -functions being pointwise multipliers in  $\mathcal{L}(L_2(\hat{I} \times \Omega), L_2(\hat{I} \times \Omega))$ , and  $L_2(\hat{I} \times \Omega) \hookrightarrow \hat{Y}'$ , one concludes (5.8), and so (5.7).

To show (5.5), i.e., differentiability of  $\rho \mapsto \hat{u}(v_0(\rho), T(\rho))$ , we repeat the argument that led to (5.3) to obtain

$$\begin{aligned} \hat{u}(v_0(\rho), T(\rho)) &= \hat{u}(0, T(\rho)) + \hat{u}(v_0(\rho), T(\rho)) - \hat{u}(0, T(\rho)) \\ &= \hat{u}(0, T(\rho)) + T(\rho)v_0(\rho)L(v_0(\rho), T(\rho))^{-1}\partial_x \hat{u}(0, T(\rho)), \end{aligned}$$

and show that

$$(5.9) \quad \rho \mapsto \hat{u}(0, T(\rho)): \Sigma \hookrightarrow L_2(\hat{I} \times \Omega) \text{ is differentiable.}$$

Then  $\rho \mapsto \partial_x \hat{u}(0, T(\rho)) : \Sigma \mapsto \hat{Y}'$  is differentiable, and from both  $\rho \mapsto T(\rho)v_0(\rho) : \Sigma \rightarrow \mathbb{C}$  and  $\rho \mapsto L(v_0(\rho), T(\rho))^{-1} : \Sigma \rightarrow \mathcal{L}(\hat{Y}', \hat{X})$  being differentiable one infers (5.5).

To show (5.9), we apply our approach for the third time. Picking some  $\bar{\rho} \in \Sigma$ , we write

$$\hat{u}(0, T(\rho)) = \hat{u}(0, T(\bar{\rho})) + (T(\bar{\rho}) - T(\rho))L(0, T(\rho))^{-1}\partial_x^2 \hat{u}(0, T(\bar{\rho})).$$

Knowing that  $\partial_x^2 \hat{u}(0, T(\bar{\rho})) \in \hat{Y}'$ , and  $\rho \mapsto L(0, T(\rho))^{-1} : \Sigma \rightarrow \mathcal{L}(\hat{Y}', \hat{X})$  and  $\rho \mapsto T(\rho) : \Sigma \rightarrow \mathbb{C}$  are differentiable, the proof of (5.9) and thus of the theorem is completed.  $\square$

## 6. NUMERICAL RESULTS

We consider three relevant examples of the form (1.3) (or its equivalent reformulation (2.1)) with  $\sigma = 1$  from the literature. We transform the solution  $\tilde{u}$  of (2.1), which might live on a time-dependent spatial domain, to  $u$ , which satisfies (2.3) on the domain  $(0, T) \times (0, 1)$ . In each example the resulting drift function  $v$  as well as the end time point  $T$  depend on an up to  $N = 5$  dimensional parameter  $\rho \in [-1, 1]^N$ .

As  $u(v(\rho)(0, 0), T(\rho))$  can be computed efficiently as a truncated series, it suffices to consider the difference

$$e(v(\rho), T(\rho)) = u(v(\rho), T(\rho)) - u(v(\rho)(0, 0), T(\rho)),$$

which satisfies equation (3.1) and is provably smoother than  $u$  (Theorem 3.8).

Thinking of a multi-query setting, instead of approximating this difference for each individual parameter value of interest we want to use (sparse) interpolation in the parameter domain  $[-1, 1]^N$ . To that end, defining  $\hat{e}(t, x) := e(tT(\rho), x)$ , we get rid of the parameter-dependent domain  $(0, T(\rho)) \times \Omega$  on which  $e$  lives. This function  $\hat{e}(t, x)$  satisfies the parabolic problem equation (5.2) on the space-time domain  $\hat{I} \times \Omega = (0, 1)^2$  with forcing term

$$\begin{aligned} \bar{w} &\mapsto \int_0^1 \int_{\Omega} (\hat{v}(t, x) - v_0) \partial_x \hat{u}(v_0)(t, x) \bar{w}(t, x) dx dt \\ &= \int_0^1 \int_{\Omega} \hat{u}(v_0)(t, x) (-\partial_x \hat{v}(t, x) \bar{w}(t, x) - (\hat{v}(t, x) - v_0) \partial_x \bar{w}(t, x)) dx dt \end{aligned}$$

for all  $\bar{w} \in \hat{X} = L_2(\hat{I}; H_0^1(\Omega)) \cap H^1(\hat{I}; H^{-1}(\Omega))$ , and  $v_0 := v(\rho)(0, 0)$  and corresponding  $\hat{u}(v_0)$  solving (5.1) with  $\hat{v} = v_0$ .

For all sparse interpolation points, by applying the minimal residual method from Section 4 we approximate  $\hat{e}$  by the continuous piecewise affine function  $\hat{e}_h$  on a uniform tensor mesh with mesh-size  $h$ , where  $\hat{u}(v_0)$  inside the forcing term can be efficiently approximated at high accuracy as a truncated series.

Finally, for all parameter values  $\rho$  of interest, we apply the sparse tensor product interpolation analyzed in Section 5 giving rise to an overall error

$$\|\hat{e} - \mathcal{I}_q \hat{e}_h\|_{\hat{X}} \leq \|\hat{e} - \hat{e}_h\|_{\hat{X}} + \|\hat{e}_h - \mathcal{I}_q \hat{e}_h\|_{\hat{X}} \approx \|\hat{e}_{h/2} - \hat{e}_h\|_{\hat{X}} + \|\hat{e}_h - \mathcal{I}_q \hat{e}_h\|_{\hat{X}}$$

with  $q$  the parameter that steers the accuracy of the sparse interpolation. For each of the considered three examples, we compute the latter two errors for different  $h$  and  $q$  and parameter test set

$$(6.1) \quad \rho \in \{-1, -0.5, 0.5, 1\}^N.$$

By Theorem 4.2, we expect  $\|\hat{e}_{h/2} - \hat{e}_h\|_{\hat{X}} = \mathcal{O}(h)$  for the first term. Section 5 suggests subexponential convergence of the second term  $\|\hat{e}_h - \mathcal{I}_q \hat{e}_h\|_{\hat{X}}$  as function of the number of interpolation points (this was shown for  $\|\hat{e} - \mathcal{I}_q \hat{e}\|_{\hat{X}}$ ). However, we already mentioned there that subexponential convergence is only observed for very high  $q$  and in practice one should rather expect algebraic convergence.

Notice that  $\|\cdot\|_{\hat{X}}$  involves a negative order Sobolev norm. Thus, we compute an equivalent version of  $\|\cdot\|_{\hat{X}}$  for functions in the discrete trial space  $\bar{w} \in \hat{X}_h \subset \hat{X}$  (similarly for  $\bar{w} \in \hat{X}_{h/2}$ ) (see [SW20a, Proof of Thm. 3.1])

$$(6.2) \quad \|\bar{w}\|_{\hat{X}}^2 \approx (\bar{w}^h)^\top (\mathbf{B}^h)^\top (\mathbf{A}^h)^{-1} \mathbf{B}^h \bar{w}^h + (\bar{w}^h)^\top \mathbf{C}^h \bar{w}^h.$$

Here,  $\bar{w}^h$  is the coefficient vector of  $\bar{w}$  in the standard nodal basis  $\Psi^h = \{\psi_i^h\}$ ,  $\mathbf{B}^h$  and  $\mathbf{C}^h$  are defined as in (4.4) with the standard nodal basis  $\Phi^h = \{\phi_i^h\}$ , and  $A_{ij}^h := \int_I \int_\Omega \partial_x \phi_j^h(t, x) \partial_x \phi_i^h(t, x) dx dt$ .

**6.1. Time-dependent hyperbolic drift function.** As in [CKS08, HKS14], we consider

$$\mu(t, x) := \mu_0 + \mu_1 \frac{t}{t + t_0}$$

from Section 1 with parameters  $\mu_0, \mu_1 \in \mathbb{R}$  and  $t_0 > 0$ . The left and right boundary are given as

$$\alpha(t) := 0 \quad \text{and} \quad \beta(t) := \beta_0$$

with parameter  $\beta_0 > 0$ . Following [CKS08, HKS14], we particularly consider the following practical ranges:  $\mu_0 \in [-1.97, -1.64]$ ,  $\mu_1 \in [-2.31, -0.99]$ ,  $t_0 \in [0.13, 0.40]$ ,  $\beta_0 \in [1.38, 2.26]$ , and  $\tau \in [0.1, 2.5]$  for the end-time point. We have  $N = 5$  different parameters on which  $\bar{v}$  and thus  $v$  depend. After rescaling, the parameter space hence has the form  $[-1, 1]^5$ .

In Figure 1, we plot the maximal error  $\hat{e}_{h/2} - \hat{e}_h \approx \hat{e} - \hat{e}_h$  measured in the (equivalent)  $\hat{X}$ -norm (6.2) over the test set (6.1) for different values of  $h$ . Figure 2 depicts the maximal interpolation error  $\hat{e}_h - \mathcal{I}_q \hat{e}_h$  over the test set (6.1) for different values of  $h$  and  $q$ .

**6.2. Space-dependent linear drift function.** As in [Smi10], we consider

$$\mu(t, x) := \mu_0 + \mu_1(\beta_0 - x)$$

from Section 1 with parameters  $\beta_0 > 0$  and  $\mu_0, \mu_1 \in \mathbb{R}$ . The left and right boundary are again given as

$$\alpha(t) := 0 \quad \text{and} \quad \beta(t) := \beta_0.$$

Motivated by [MW09, Smi10], we particularly consider the following practical ranges:  $\mu_0 \in [-2, 2]$ ,  $\mu_1 \in [-4, 4]$ , and  $\beta_0 \in [0.5, 2]$ , and choose the end-time point as  $\tau := 2.5$ . We have  $N = 3$  different parameters on which  $\bar{v}$  and thus  $v$  depend. After rescaling, the parameter space hence has the form  $[-1, 1]^3$ .

In Figure 3, we plot the maximal error  $\hat{e}_{h/2} - \hat{e}_h \approx \hat{e} - \hat{e}_h$  measured in the (equivalent)  $\hat{X}$ -norm (6.2) over the test set (6.1). Figure 4 depicts the maximal interpolation error  $\hat{e}_h - \mathcal{I}_q \hat{e}_h$  over the test set (6.1) for different values of  $h$  and  $q$ .

**6.3. Constant drift function and time-dependent linear spatial domain.** We consider a constant drift function

$$\mu(t, x) := \mu_0$$

with parameter  $\mu_0 \in \mathbb{R}$ . As in [ETH20] (see also Example 2.1), we choose the left and right boundary as

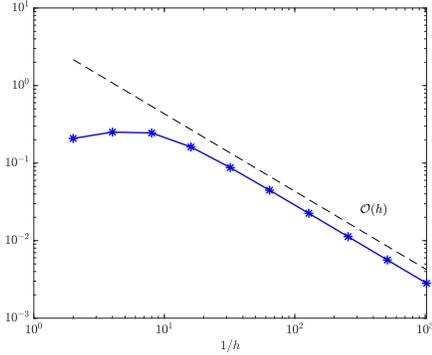
$$\alpha(t) := \beta_0 \frac{t}{2T_0} \quad \text{and} \quad \beta(t) := \beta_0 \left(1 - \frac{t}{2T_0}\right)$$

with parameters  $\beta_0, T_0 > 0$ . Recall from Example 2.1 the solution of the ODE (2.2)

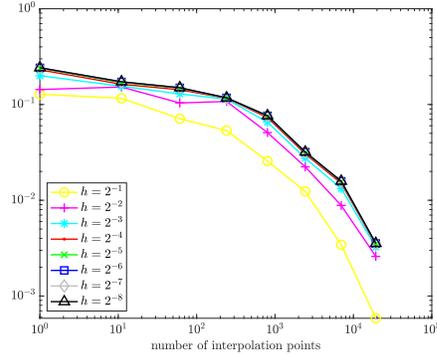
$$\theta(t) = \frac{\beta_0^2 (T_0 - 2\tilde{T})^2 t}{T_0^2 - 2\beta_0^2 (T_0 - 2\tilde{T})t}, \quad t \in \left[0, \frac{T_0^2}{2\beta_0^2 (T_0 - 2\tilde{T})}\right).$$

and that  $T = \theta^{-1}(\tilde{T}) = \frac{T_0 \tilde{T}}{\beta_0^2 (T_0 - 2\tilde{T})}$ . Following [ETH20], we particularly consider the following practical ranges:  $\mu_0 \in [-5.86, 0]$ ,  $\beta_0 \in [0.56, 3.93]$ ,  $T_0 \in [3, 20]$ , and  $\tau \in [0.1, 2.5]$  for the end-time point. We have  $N = 4$  different parameters on which  $\tilde{v}$  and thus  $v$  depend. After rescaling, the parameter space hence has the form  $[-1, 1]^4$ .

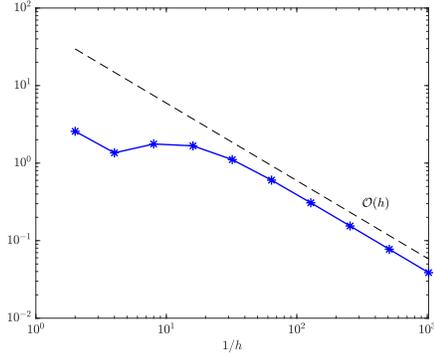
In Figure 5, we plot the maximal error  $\hat{e}_{h/2} - \hat{e}_h \approx \hat{e} - \hat{e}_h$  measured in the (equivalent)  $\hat{X}$ -norm (6.2) over the test set (6.1). Figure 6 depicts the maximal interpolation error  $\hat{e}_h - \mathcal{I}_q \hat{e}_h$  over the test set (6.1) for different values of  $h$  and  $q$ .



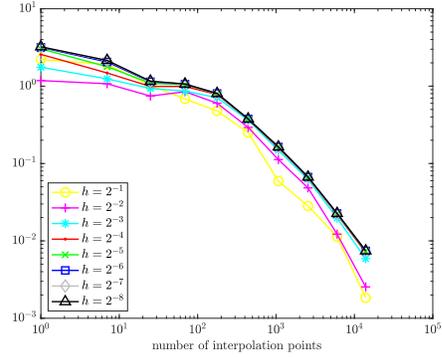
**Figure 1.** Maximal error  $\hat{e}_{h/2}(\hat{v}(\rho)) - \hat{e}_h(\hat{v}(\rho))$  measured in (equivalent)  $\hat{X}$ -norm over all  $\rho \in \{-1, -0.5, 0, 0.5, 1\}^5$  for time-dependent hyperbolic drift function from Section 6.1.



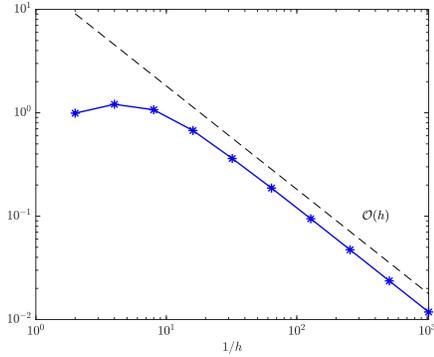
**Figure 2.** Maximal interpolation error  $\hat{e}_h(\hat{v}(\rho)) - (\mathcal{I}_q \hat{e}_h(\hat{v}(\cdot))) (\rho)$  for various choices of  $h$  measured in (equivalent)  $\hat{X}$ -norm over all  $\rho \in \{-1, -0.5, 0, 0.5, 1\}^5$  for time-dependent hyperbolic drift function from Section 6.1.



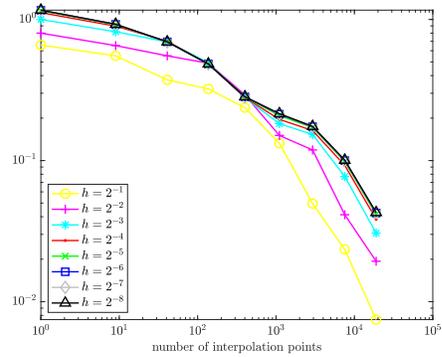
**Figure 3.** Maximal error  $\hat{e}_{h/2}(\hat{v}(\rho)) - \hat{e}_h(\hat{v}(\rho))$  measured in (equivalent)  $\hat{X}$ -norm over all  $\rho \in \{-1, -0.5, 0, 0.5, 1\}^3$  for space-dependent linear drift function from Section 6.2.



**Figure 4.** Maximal interpolation error  $\hat{e}_h(\hat{v}(\rho)) - (\mathcal{I}_q \hat{e}_h(\hat{v}(\cdot))) (\rho)$  with  $h = 2^{-1}, \dots, 2^{-8}$  measured in (equivalent)  $\hat{X}$ -norm over all  $\rho \in \{-1, -0.5, 0, 0.5, 1\}^3$  for space-dependent linear drift function from Section 6.2.



**Figure 5.** Maximal error  $\hat{e}_{h/2}(\hat{v}(\rho), T(\rho)) - \hat{e}_h(\hat{v}(\rho), T(\rho))$  measured in (equivalent)  $\hat{X}$ -norm over all  $\rho \in \{-1, -0.5, 0, 0.5, 1\}^4$  for constant drift function with time-dependent linear spatial domain from Section 6.3.



**Figure 6.** Maximal interpolation error  $\hat{e}_h(\hat{v}(\rho), T(\rho)) - (\mathcal{I}_q \hat{e}_h(\hat{v}(\cdot), T(\cdot))) (\rho)$  for various choices of  $h$  measured in (equivalent)  $\hat{X}$ -norm over all  $\rho \in \{-1, -0.5, 0, 0.5, 1\}^4$  for constant drift function with time-dependent linear spatial domain from Section 6.3.

## 7. CONCLUSION

We have developed a numerical solution method for solving the Fokker-Planck equation on a one-dimensional spatial domain and with a discontinuity between initial and boundary data and time-dependent boundaries. We first transformed the equation to an equation on a rectangular time-space domain. We then demonstrated that the solution of a corresponding equation with a suitable constant drift function, whose solution is explicitly available as a fast converging series expansion, captures the main singularity present in the solution for a variable drift function. The equation for the difference of both these solutions, which is thus more regular than both terms, is solved with a minimal residual method. This method is known to give a quasi-best approximation from the selected trial space.

Finally, in order to efficiently solve Fokker–Planck equations that depend on multiple parameters, we demonstrate that the solution is a holomorphic function of these parameters. Consequently, a sparse tensor product interpolation method can be shown to converge at a subexponential rate as function of the number of interpolation points. In one test example, this interpolation method works very satisfactory, but the results are less convincing in two other cases. We envisage that in those cases better results can be obtained by an adaptive sparse interpolation method as the one proposed in [CCS14].

## REFERENCES

- [And13] R. Andreev. Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. *IMA J. Numer. Anal.*, 33(1):242–260, 2013.
- [BCGS21] Udo Boehm, Sonja Cox, Gregor Gantner, and Rob Stevenson. Fast solutions for the first-passage distribution of diffusion models with space-time-dependent drift functions and time-dependent boundaries. *In preparation*, 2021.
- [BKG12] N.E. Bowman, K.P. Kording, and J.A. Gottfried. Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex. *Neuron*, 75:916–927, 2012.
- [BX91] J.H. Bramble and J. Xu. Some estimates for a weighted  $L^2$  projection. *Math. Comp.*, 56:463–476, 1991.
- [CCS14] A. Chkifa, A. Cohen, and Ch. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Found. Comput. Math.*, 14(4):601–633, 2014.
- [Chk14] A. Chkifa. *Sparse polynomial methods in high dimension: Application to parametric PDE*. PhD thesis, Université Pierre et Marie Curie - Paris VI, 2014.
- [CKS08] A.K. Churchland, R. Kiani, and M.N. Shadlen. Decision-making with multiple alternatives. *Nature Neuroscience*, 11(6):693–702, 2008.
- [Cos90] M. Costabel. Boundary integral operators for the heat equation. *Integral Equations Operator Theory*, 13(4):498–552, 1990.
- [DHP03] R. Denk, M. Hieber, and J. Prüss.  $\mathcal{R}$ -boundedness, Fourier multipliers and problems of elliptic and parabolic type. *Mem. Amer. Math. Soc.*, 166(788):viii+114, 2003.
- [dS64] L. de Simon. Un’applicazione della teoria degli integrali singolari allo studio delle equazioni differenziali lineari astratte del primo ordine. *Rend. Sem. Mat. Univ. Padova*, 34:205–223, 1964.
- [ETH20] N.J. Evans, J.S. Trueblood, and W.R. Holmes. A parameter recovery assessment of time-variant models of decision-making. *Behavior Research Methods*, 52:193–206, 2020.
- [FF03] N. Flyer and B. Fornberg. Accurate numerical resolution of transients in initial-boundary value problems for the heat equation. *J. Comput. Phys.*, 184(2):526–539, 2003.
- [FFGC20] Alexander Fengler, Michael Frank, Lakshmi Govindarajan, and Tony Chen. Likelihood approximation networks (lans) for fast inference of simulation models in cognitive neuroscience. *bioRxiv*, 2020.
- [GBK14] M. Gondan, S.P. Blurton, and M Kesselmeier. Even faster and even more accurate first-passage time densities and distributions for the Wiener diffusion model. *J. Math. Psych.*, 60:20–22, 2014.
- [GS01] J. I. Gold and M. N. Shadlen. Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, 5(1):10–16, 2001.
- [HFW<sup>+</sup>15] G.E. Hawkins, B.U. Forstmann, E.-J. Wagenmakers, R. Ratcliff, and S.D. Brown. Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, 35(6):2476–2484, 2015.
- [HKS14] T. Hanks, R. Kiani, and M.N. Shadlen. A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife*, 3:e02260, 2014.
- [MW09] D. Matzke and E.J. Wagenmakers. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, 16(5):798–817, 2009.
- [NTW08] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.
- [Øks98] B. Øksendal. *Stochastic Differential Equations*. Springer, Berlin, 5th ed. edition, 1998.

- [Rat78] R. Ratcliff. A theory of memory retrieval. *Psychological Review*, 85(2):59–108, 1978.
- [SK13] Michael N. Shadlen and Roozbeh Kiani. Decision making as a window on cognition. *Neuron*, 80(3):791–806, 2013.
- [Smi10] P.L. Smith. From poisson shot noise to the integrated ornstein-uhlenbeck process: Neurally principled models of information accumulation in decision-making and response time. *Journal of Mathematical Psychology*, 54:266–283, 2010.
- [SS09] Ch. Schwab and R.P. Stevenson. A space-time adaptive wavelet method for parabolic evolution problems. *Math. Comp.*, 78:1293–1318, 2009.
- [SW20a] R.P. Stevenson and J. Westerdiep. Minimal residual space-time discretizations of parabolic equations: Asymmetric spatial operators. Technical report, Korteweg-de Vries Institute, 2020. In preparation.
- [SW20b] R.P. Stevenson and J. Westerdiep. Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations. *IMA J. Numer. Anal.*, 2020.
- [VV08] A. Voss and J. Voss. A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology*, 52(52):1–9, 2008.
- [Wlo87] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge, 1987. Translated from the German by C. B. Thomas and M. J. Thomas.

KORTEWEG-DE VRIES (KDV) INSTITUTE FOR MATHEMATICS, UNIVERSITY OF AMSTERDAM, P.O. BOX 94248, 1090 GE AMSTERDAM, THE NETHERLANDS.

*Email address:* [u.bohm@uva.nl](mailto:u.bohm@uva.nl), [s.g.cox@uva.nl](mailto:s.g.cox@uva.nl), [g.gantner@uva.nl](mailto:g.gantner@uva.nl), [r.p.stevenson@uva.nl](mailto:r.p.stevenson@uva.nl)